



DIKLAT TEKNIS

MANAJEMEN PENDATAAN

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
PUSAT PENGEMBANGAN TENAGA KEPENDIDIKAN

Kompleks Kemdikbud Gd. D. Lt. 17, Jln. Jenderal Sudirman, Senayan, Jakarta Pusat, 10270, Telp/Faks (021) 57946110
Kampus: Jln Raya Cinangka Km. 19 Bojongsari, Depok 16517, Telp. (021) 7490411, Fax. (021) 7491174
Laman: <http://tendik.kemdiknas.go.id> Pos Elektronik: tendik@kemdiknas.go.id

.....



Modul 01

Pengumpulan Data

.....



DAFTAR ISI

DAFTAR ISI	i
Bab I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Deskripsi Singkat.....	2
1.3 Hasil Pembelajaran	2
1.4 Indikator Pembelajaran	2
1.5 Materi Pokok	2
1.6 Manfaat	3
Bab II Pengumpulan Data	5
2.1 Pengertian Pengumpulan Data	5
2.2 Perencanaan Pengumpulan Data	6
2.3 Metode Pengumpulan Data	8
2.4 Beberapa Cara Pengumpulan Data.....	11
2.5 Pelaksanaan Pengumpulan Data.....	13
2.6 Perancangan Kuesioner (Questioner Design).....	14
2.7 Rangkuman	22
2.8 Latihan Soal	23
Bab III Penutup	27
3.1 Kesimpulan	27
3.2 Tindak Lanjut	27
Daftar Pustaka	29
Lampiran	31

Bab I Pendahuluan

1.1 Latar Belakang

Dalam penelitian, teknik pengumpulan data merupakan faktor penting demi keberhasilan penelitian. Hal ini berkaitan dengan bagaimana cara mengumpulkan data, siapa sumbernya, dan apa alat yang digunakan.

Jenis sumber data adalah mengenai dari mana data tersebut diperoleh. Apakah data diperoleh dari sumber langsung (data primer) atau data diperoleh dari sumber tidak langsung (data sekunder).

Metode Pengumpulan Data merupakan teknik atau cara yang dilakukan untuk mengumpulkan data. Metode menunjuk suatu cara sehingga dapat diperlihatkan penggunaannya melalui angket, wawancara, pengamatan, tes, dokumentasi dan sebagainya.

Sedangkan Instrumen Pengumpul Data merupakan alat yang digunakan untuk mengumpulkan data. Karena berupa alat, maka instrumen dapat berupa lembar cek list, kuesioner (angket terbuka / tertutup), pedoman wawancara, *camera photo* dan lainnya.

Adapun tiga teknik pengumpulan data yang biasa digunakan adalah angket, observasi dan wawancara.

Data adalah catatan atas kumpulan fakta. Data merupakan bentuk jamak dari *datum*, berasal dari bahasa Latin yang berarti “sesuatu yang diberikan”. Dalam penggunaan sehari-hari data berarti suatu pernyataan yang diterima secara apa adanya. Pernyataan ini adalah hasil pengukuran atau pengamatan suatu variabel yang bentuknya dapat berupa angka, kata-kata, atau citra.

Dalam keilmuan (ilmiah), fakta dikumpulkan untuk menjadi data. Data kemudian diolah sehingga dapat diutarakan secara jelas dan tepat sehingga dapat dimengerti oleh orang lain yang tidak langsung mengalaminya sendiri, hal ini dinamakan deskripsi. Pemilahan banyak data sesuai dengan persamaan atau perbedaan yang dikandungnya dinamakan klasifikasi.

Menurut berbagai sumber lain, data dapat juga didefinisikan sebagai berikut:

- Menurut kamus bahasa Inggris-Indonesia, data berasal dari kata *datum* yang berarti fakta
- Dari sudut pandang bisnis, data bisnis adalah deskripsi organisasi tentang sesuatu (*resources*) dan kejadian (*transactions*) yang terjadi

Pengertian yang lain menyebutkan bahwa data adalah deskripsi dari suatu kejadian yang kita hadapi. Intinya data itu adalah suatu fakta-fakta tertentu sehingga menghasilkan suatu kesimpulan dalam menarik suatu keputusan.

1.2 Deskripsi Singkat

Mata Diklat Pengumpulan Data ini membahas pengertian pengumpulan data, perencanaan pengumpulan data, metode pengumpulan data, cara pengumpulan data, pelaksanaan pengumpulan data dan perancangan kuesioner (*questionnaire design*).

1.3 Hasil Pembelajaran

Peserta pelatihan diharapkan dapat mengerti dan memahami kegiatan pengumpulan data secara langsung di lapangan dan memperoleh pengetahuan mengenai merancang kuesioner sederhana yang sesuai dengan maksud dan tujuan penelitian.

1.4 Indikator Pembelajaran

Setelah mempelajari modul ini peserta pelatihan diharapkan dapat :

1. Menjelaskan aspek-aspek yang berhubungan dengan perencanaan pengumpulan data.
2. Membedakan antara metode pengumpulan data dengan sensus dan *sampling*.
3. Mempergunakan berbagai cara pengumpulan data.
4. Membuat kuesioner sederhana dengan memperhatikan jenis *item* pertanyaan, susunan *item* dalam daftar pertanyaan dan penyajian daftar pertanyaan.

1.5 Materi Pokok

Materi pokok yang dibahas pada modul Pengumpulan Data ini adalah :

1. Pengertian Pengumpulan data
2. Perencanaan Pengumpulan data
3. Metode Pengumpulan data
4. Beberapa cara Pengumpulan data
5. Pelaksanaan Pengumpulan data
6. Perancangan Kuesioner (*Questionnaire Design*)

1.6 Manfaat

Dengan mempelajari Modul Pengumpulan Data ini, peserta diharapkan mampu dan memiliki pemahaman tentang tata cara pengumpulan data primer. Dalam hal ini adalah data statistik yang baik, yang bisa dipercaya kebenarannya (*reliable*), dan tepat waktu (*up to date*).

Bab II Pengumpulan Data

2.1 Pengertian Pengumpulan Data

Dalam era kemajuan di bidang ilmu pengetahuan dan teknologi dewasa ini, data statistik sebagai sumber informasi mempunyai peranan guna membantu dalam menjelaskan dan menjawab permasalahan yang dihadapi oleh pemerintah maupun pihak swasta. Akibatnya, berbagai data statistik, baik sebagai sumber informasi utama (primer) maupun sumber informasi pendukung (sekunder) dikumpulkan untuk dianalisis secara objektif bagi keperluan pengambilan kebijaksanaan.

Sebagai teladan dalam kehidupan sehari-hari misalnya, seorang ibu yang ingin membeli suatu barang, akan mengumpulkan informasi seperti, di mana barang tersebut bisa diperoleh, berapa harganya, bagaimana kualitasnya, dan sebagainya. Dari informasi yang diperoleh akan diproses menurut cara berpikir ibu tersebut dengan dasar pertimbangan dan asumsi-asumsi yang ada padanya, sehingga akhirnya ia mengambil keputusan apakah akan membeli barang tersebut atau tidak. Beberapa teladan lainnya seperti usaha seorang pelajar lulusan SLTA yang akan memilih perguruan tinggi yang mana sebagai pendidikan lanjutannya, atau seorang wartawan suatu surat kabar yang ingin menulis berita mengenai keberhasilan-keberhasilan dan kegagalan-kegagalan suatu daerah dalam melaksanakan pembangunan daerahnya. Wartawan tersebut akan mencari berbagai 'ukuran yang objektif', sehingga ia dapat menggambarkan kepada pembaca, apakah daerah itu telah berhasil atau cenderung gagal.

Dari beberapa teladan di atas ternyata tanpa disadari kita telah menggunakan cara berpikir dengan prinsip statistika. Apabila cara berpikir tersebut dilakukan dengan metode ilmiah secara efisien dan sistematis maka aspek pengumpulan informasi, memprosesnya, sampai pada pengambilan keputusan merupakan tahapan dari kegiatan statistika.

Statistika adalah pengetahuan yang berhubungan dengan pengumpulan, pengolahan, penyajian data, analisis data, penarikan kesimpulan, serta pembuatan keputusan yang cukup beralasan berdasarkan data dan analisis yang dilakukan. Dari uraian ini jelas, bahwa pengumpulan data merupakan bagian yang paling awal dari suatu kegiatan statistika, sehingga dengan demikian akan merupakan bagian yang sangat penting.

Pengumpulan data yang akan dibahas dalam modul ini dititikberatkan pada suatu proses pengadaan data primer. Dalam hal ini adalah untuk menghasilkan data statistik yang **'baik'**, yaitu data yang bisa dipercaya kebenarannya (*reliable*) dan tepat waktu (*up to date*). Agar tujuan ini tercapai dengan sebaik-baiknya, maka proses pengumpulan datanya harus terencana dengan baik, mencakup semua unit yang menjadi obyek penelitian serta informasi yang dikumpulkan itu menurut keadaan yang sebenarnya (fakta), dengan metode dan cara yang ditetapkan.

Mengenai metode dan cara pengumpulan data, masing-masing akan diuraikan lebih lanjut dalam bagian tersendiri, dan diakhiri dengan kegiatan pelaksanaan pengumpulan data. Di samping itu, karena salah satu perangkat pengumpulan data primer yang umum digunakan adalah daftar pertanyaan (kuesioner), maka langkah-langkah penyusunannya juga akan diuraikan dalam bagian tersendiri.

2.2 Perencanaan Pengumpulan Data

Pada bagian ini akan diuraikan berbagai hal yang perlu dipersiapkan dalam Kegiatan Pengumpulan Data Primer di lapangan.

1. Persiapan Administrasi

Yang meliputi persiapan administrasi dalam menyusun rancangan pengumpulan data adalah:

a. Mempersiapkan jadwal waktu.

Jadwal waktu mencakup seluruh kegiatan pengumpulan data merupakan hal yang sangat penting, sekaligus dapat dipergunakan untuk pemantauan pelaksanaan tahap demi tahap.

b. Izin resmi.

Sebelum data dikumpulkan di lapangan, perlu diusahakan izin resmi dari wilayah yang menjadi sasaran penelitian agar pelaksanaan pengumpulan data berjalan lancar.

c. Mempersiapkan peneranga/Sosialisasi.

Penerangan dapat dilakukan setelah izin resmi pengumpulan data diperoleh. Peranan penerangan sangat penting agar masyarakat (responden) membantu pelaksanaan pengumpulan data dengan memberikan jawaban yang benar kepada para petugas pengumpul data (pencacah).

Disamping itu, responden juga akan mengetahui tujuan dari pengumpulan data, sehingga petugas pencacah tidak banyak mendapatkan kesulitan dalam menjelaskan tujuan kedatangannya. Dalam pelaksanaannya penerangan harus diatur sebaik mungkin agar tujuan tercapai, dan media penerangannya dapat melalui koran, majalah, radio, televisi, spanduk, brosur, dan sebagainya.

d. Organisasi, pengawasan dan koordinasi.

Agar pengumpulan data berjalan dengan lancar, maka perlu dipersiapkan dan diatur organisasi pengumpulan datanya, supaya masing-masing mempunyai tanggung jawab dalam tugas. Selain itu perlu pula direncanakan sistem pengawasannya secara terpadu dari pimpinan sampai pelaksana di lapangan. Koordinasi tiap tahapan pekerjaan harus dirancang, dari tahap perencanaan sampai pada pelaksanaan menurut jadwal waktu kegiatan.

e. Rancangan pembiayaan.

Dalam merencanakan anggaran biaya pengumpulan data, biaya yang direncanakan meliputi upah, perlengkapan, biaya latihan dan lain-lain.

f. Tenaga kerja dan barang.

Tenaga kerja yang diperlukan antara lain adalah petugas pencacahan staf pelaksana yang sesuai dengan kebutuhan. Barang atau peralatan yang amat penting untuk mengumpulkan data harus pula direncanakan.

2. Persiapan Teknis

Persiapan yang bersifat teknis dalam pengumpulan data meliputi:

a. Mempersiapkan 'konsep dan definisi'

Konsep dan definisi didiskusikan antara pengumpul data dan pemakai data, atau dengan badan yang berkepentingan atau yang memesan data yang akan dikumpulkan. Konsep dan definisi harus disepakati bersama terlebih dahulu.

b. Menentukan bentuk kegiatan

Dalam perencanaan pengumpulan data ditentukan bentuk pengumpulan data apa saja yang akan dilakukan. Bentuk kegiatan tentu saja harus disesuaikan dengan metode dan cara yang dipergunakan dalam pengumpulan data. Misalnya saja pengumpulan data menggunakan metode pencacahan lengkap seluruh unit populasi (sensus), bentuk kegiatannya tentu saja disesuaikan dengan metodenya, yaitu anggota populasi didaftar secara lengkap dan dikumpulkan sifat-sifat yang diinginkan dari seluruh anggota populasi. Demikian pula, apabila cara yang dikehendaki dalam pengumpulan data adalah melalui pos (*mailing system*), maka bentuk kegiatan adalah mempersiapkan segala sesuatunya mengikuti cara yang telah dipilih.

c. Menentukan ruang lingkup

Dalam persiapan teknis, ruang lingkup penelitian lebih ditegaskan lagi sesuai dengan sasaran yang akan diteliti. Misalnya: dalam melaksanakan pengumpulan data industri dengan sensus, maka semua usaha industri termasuk para pengrajin merupakan anggota populasi yang akan menjadi sasaran penelitian.

d. Perencanaan daftar pertanyaan

Dalam pengumpulan data statistik dengan cara wawancara, kita menggunakan daftar pertanyaan untuk mencatat keterangan yang diperoleh dari responden. Agar keterangan yang diperoleh sesuai dengan maksud dan tujuan penelitian, maka daftar pertanyaan perlu dirancang sebaik mungkin.

e. Mempersiapkan latihan petugas

Dalam kegiatan pengumpulan data yang memerlukan petugas pengumpul data, maka petugas yang akan melaksanakan kegiatan tersebut perlu terlebih dahulu dilatih, sehingga ada kesatuan bahasa serta memahami maksud dan tujuan penelitian di dalamnya. Perencanaan latihan petugas yang perlu disiapkan antara lain:

- 1) Bahan serta jadwal latihan
- 2) Jumlah petugas lapangan yang diperlukan serta klasifikasi pendidikan petugas
- 3) Cara-cara penyelenggaraan latihan

f. Rencana pengolahan dan publikasi

Dalam perencanaan pengumpulan data, juga telah dipikirkan pengolahannya, misalkan apakah akan diolah secara manual atau komputer. Demikian juga dalam hal rencana publikasi sudah harus dipikirkan bentuk publikasi dan penyajian datanya. Hal ini berhubungan dengan rancangan daftar pertanyaan yang akan disusun dalam kegiatan pengumpulan datanya.

2.3 Metode Pengumpulan Data

Dalam pengumpulan data statistik, dapat digunakan dua metode, yaitu metode sensus dan metode *sampling*. **Metode sensus** adalah suatu metode pengumpulan data yang setiap unit dalam populasi diteliti, dalam arti kata

setiap unit dicacah. Sedangkan **metode *sampling***, adalah suatu metode pengumpulan data, dimana hanya sebagian unit saja dari populasi yang diteliti. Pelaksanaan pengumpulan data melalui survei dengan metode *sampling* ini disebut survei contoh (*sample survey*).

Untuk menentukan pilihan, metode mana yang akan di pakai dalam pengumpulan data, berikut ini akan diuraikan keuntungan dan kelemahan dari masing-masing metode.

I. Pengumpulan Data dengan Metode Sensus

Keuntungan metode sensus antara lain :

1. Dari hasil sensus akan dapat diketahui gambaran sebenarnya suatu populasi.
2. Data hasil sensus sebagai kerangka dalam melakukan survei dengan metode *sampling*.
3. Tidak mempunyai *sampling error* (kesalahan dalam penggunaan metode pengumpulan datanya).

Disamping itu, dalam metode ini ditemukan juga kelemahan-kelemahan, yaitu:

1. Biaya dan fasilitas sangat besar.
2. Tenaga yang dibutuhkan sangat banyak
3. Waktu pelaksanaan lama.
4. Kesalahan sulit diperkirakan.
5. Jenis data yang diperoleh terbatas dan sifat data sederhana.

Walaupun pengumpulan data dengan metode sensus akan memakan biaya yang besar dan waktu yang lama, namun suatu saat harus dilakukan, sensus diperlukan untuk mendapatkan gambaran secara menyeluruh dari suatu populasi dan dasar untuk membuat kerangka contoh bagi penelitian-penelitian selanjutnya.

Badan Pusat Statistik, dalam beberapa kegiatan pengumpulan datanya menggunakan metode sensus. Sensus dilaksanakan sepuluh tahun sekali berdasarkan Undang-Undang yang telah ditetapkan tentang sensus. Pengumpulan data dengan metode sensus yang dilakukan Badan Pusat Statistik mencakup:

1. Sensus Penduduk, diadakan tahun 1961, 1971 dan 1980, 1990, 2010 dan selanjutnya pada tahun-tahun yang berakhiran 0
2. Sensus Pertanian, diadakan tahun 1963, 1973, 1983,1993,2013 dan selanjutnya tahun-tahun yang berakhiran 3.
3. Sensus Industri diadakan tahun 1964, 1974, kemudian diperluas dan disebut dengan Sensus Ekonomi, dilaksanakan pada tahun 1986, selanjutnya tahun-tahun yang berakhiran 6.

II. Pengumpulan Data dengan Metode *Sampling*

Karena tidak seluruh unit di dalam populasi diteliti, maka jelas bahwa metode *sampling* akan mempunyai keuntungan antara lain adalah:

1. Biaya jauh lebih sedikit
2. Tenaga yang diperlukan tidak banyak
3. Waktu dapat lebih cepat
4. Kegiatan-kegiatan mudah dikuasai dan diawasi
5. Kesalahan dapat diperkirakan atau diukur
6. Jenis data yang dicakup lebih banyak dan terinci
7. Meminimalkan kerusakan dari kegiatan bersifat merusak, misalnya: mungkin semua kaleng dibuka.

Di samping itu, kekurangannya antara lain adalah:

1. Gambaran populasi hanya merupakan taksiran.
2. Dalam survei dengan metode *sampling* perlu kerangka contoh yang berasal dari sensus.
3. Pengambilan sampel tidak selalu tepat, untuk memperkirakan nilai sesungguhnya. Kesalahan ini dinamakan *sampling error*.

Walaupun metode sensus mencakup semua unit dalam populasi, dalam banyak hal, pengumpulan data dilakukan dengan metode *sampling*.

Dasar pertimbangannya antara lain :

1. Populasi sulit dan tidak mungkin dikuasai sehingga dilakukan dengan metode *sampling*. Misalnya, populasi ikan di suatu danau, populasi tambang mineral, dan sebagainya.
2. Karena biaya terbatas, sehingga penelitian tidak mungkin dilakukan pada semua unit dalam populasi.
3. Waktu yang mendesak, sehingga tidak mungkin dilakukannya pengumpulan secara lengkap.
4. Penelitian yang bersifat merusak, sehingga tidak mungkin dilakukan pada semua unit. Misalnya mengetes kualitas makanan dalam kaleng, tidak mungkin semua kaleng dibuka
5. Karena pertimbangan akurasi data maka perlu direkrut tenaga yang berkualitas tinggi dengan latihan yang intensif. Dengan *sampling* dapat diperoleh tenaga yang jauh lebih sedikit tetapi berkualitas tinggi.

Berdasarkan pertimbangan-pertimbangan di atas, kita dapat memilih metode mana yang dianggap tepat dan sesuai dengan kondisi maupun sumber daya yang tersedia. Di dalam pengumpulan data dengan menggunakan metoda *sampling* ini, suatu hal yang paling penting adalah metode penarikan contohnya, karena contoh (*sample*) ini harus bisa dianalisis secara statistik. Ada empat syarat yang harus dipenuhi agar suatu contoh dikatakan sebagai contoh yang baik dan memenuhi syarat statistika :

1. Mewakili populasi (*representatif*)

Karena nilai contoh digunakan untuk memperkirakan nilai populasi, maka contoh yang terpilih harus mewakili atau mencerminkan keadaan populasinya. Untuk itu semua unit yang ada dalam populasi harus diikutsertakan (mempunyai kesempatan/peluang yang sama) dalam pemilihan contoh.

2. Mempunyai tingkat kebenaran yang bisa diukur

Bahwa nilai contoh tidak sama dengan nilai populasi bisa dipahami, tetapi yang penting bagaimana kita mengukur seberapa jauh perbedaan yang terjadi, atau seberapa besar kesalahan yang dihasilkan contoh.

3. Mempunyai nilai perkiraan yang tidak bias

Dalam pengambilan contoh, diharapkan nilai perkiraannya sama dengan nilai yang sebenarnya. Apakah harapan ini terpenuhi tergantung dari pemilihan contohnya. Yang dimaksud dengan nilai perkiraan tidak bias (*unbiased estimator*) apabila harapan nilai perkiraan (*statistic*) dari seluruh contoh yang mungkin sama dengan nilai sebenarnya (*parameter*).

4. Efisien

Karena hanya sebagian kecil saja yang diselidiki, pelaksanaan pengumpulan datanya haruslah lebih mudah. Juga ditinjau dari biaya, tenaga dan waktu *sampling* harus lebih efisien.

Untuk memenuhi syarat-syarat tersebut di atas harus dipakai metode *sampling* yang sesuai baik dalam prosedur penarikan contoh maupun penghitungan nilai-nilai perkiraan. Dari hasil pengujian secara empiris maupun secara matematis maka hanya *sampling* peluang (*probability sampling*) yang memenuhi syarat-syarat itu.

Pengumpulan data dengan metode sensus dan *sampling* ini ditujukan untuk memperoleh informasi utama (*primer*) yang memang sudah terjadi di lapangan. Selain kedua metode tersebut, pengadaan data *primer* dapat pula diperoleh melalui perancangan dan pengawasan dari suatu percobaan (*eksperimen*), yang disebut data hasil percobaan. Misalnya saja, untuk mengetahui pertambahan berat badan balita dengan adanya pemberian dosis makanan tambahan.

2.4 Beberapa Cara Pengumpulan Data

Apabila metode pengumpulan datanya telah ditentukan, maka agar informasi yang diperlukan dapat terkumpul ada beberapa cara yang dapat dilakukan, antara lain:

1. Wawancara (*interview*)
2. Diisi sendiri oleh responden

3. Pengamatan (*observation*)

4. Pengukuran (*measurement*)

Untuk menentukan cara mana yang sebaiknya dipakai, marilah kita kaji masing-masing cara tersebut.

1. Wawancara

Pencacah menghubungi responden (unit penelitian) dan mengajukan pertanyaan secara langsung atau dengan menggunakan daftar pertanyaan yang telah disiapkan (terstruktur). Pilihan wawancara dapat dilaksanakan melalui telepon, wawancara perorangan secara pribadi atau kelompok.

Keunggulan dalam menggunakan wawancara adalah :

- a. Kembalinya daftar pertanyaan (kalau menggunakan daftar pertanyaan) lebih banyak, karena pencacah mengisi langsung. Tidak adanya jawaban dari responden (*non response*) dapat ditekan sekecil mungkin.
- b. Data lebih terjamin kebenarannya, karena pencacah mendapatkan data langsung dari responden
- c. Kelengkapan jawaban terhadap semua pertanyaan dalam kuesioner dapat dijamin.
- d. Pencacah dapat lebih memperjelas pertanyaannya, kalau responden masih ragu akan maksud pertanyaan.
- e. Pencacah dapat bertanya dengan baik.
- f. Pencacah dapat menggunakan bahasa yang dapat dimengerti oleh responden.

Kelemahan dalam menggunakan wawancara adalah, yaitu :

- a. Sifat pribadi pencacah dapat mempengaruhi jawaban
- b. Responden seringkali memberikan jawaban yang bersifat pribadi
- c. Responden kadangkala sulit dijumpai
- d. Biaya juga akan bertambah karena harus melatih pencacah
- e. Perlu organisasi lapangan yang baik

2. Diisi sendiri oleh responden

Dalam hal ini, daftar pertanyaan dikirim kepada responden, kemudian setelah diisi oleh responden, daftar pertanyaan dikirim kembali (*mailing system*).

Keunggulan dari cara ini :

- a. Biaya lebih kecil dan ruang lingkup daerah bisa lebih luas
- b. Tidak memerlukan petugas pencacah, sehingga tidak ada biaya untuk latihan pencacah
- c. Responden dapat memberikan jawaban secara terbuka.

Kelemahannya yaitu:

- a. Banyak daftar isian yang tidak dikembalikan oleh responden dan juga banyak daftar yang tidak sampai kepada responden, *non*

- response* menjadi sangat besar.
- b. Banyak kesalahan pengisian daftar, karena responden tidak menangkap maksud pertanyaan yang sebenarnya.
 - c. Pertanyaan yang diajukan mungkin tidak dapat dijawab atau sulit dijawab secara tertulis.

3. Pengamatan

Cara ini dilakukan tanpa mengajukan pertanyaan-pertanyaan, sering dipergunakan dalam penelitian di dalam bidang sosial dan ekonomi serta terutama sekali dapat digunakan untuk penelitian di mana obyeknya tidak bisa diajak wawancara (komunikasi secara verbal). Misalnya pengamatan untuk mengetahui perilaku bayi, penelitian upacara adat, melihat kesibukan para karyawan yang bekerja di suatu kantor atau program TVRI apa yang digemari pirsawan dan sebagainya. Dalam pelaksanaan pengumpulan datanya dapat dilakukan oleh orang atau oleh suatu alat secara mekanis.

Kebaikan dalam cara pengamatan ini adalah, bahwa pencatatan bisa dilakukan pada waktu kejadian-kejadian itu terjadi, tanpa tergantung pada jawaban dari obyek yang diselidiki.

Sedangkan kelemahannya di antaranya adalah tidak bisa dipergunakan untuk memperoleh data tentang sikap dan motivasi responden, apa yang diperbuat responden pada waktu lalu dan rencana yang akan datang.

4. Pengukuran

Dalam hal ini pencacah langsung mengadakan pengukuran terhadap obyek yang akan diteliti.

Keuntungan dari cara ini adalah:

- a. Cara ini paling teliti, karena obyek dikuasai pencacah.
- b. Data hasil pengukuran dapat diisi seutuhnya dalam daftar pertanyaan.
- c. Data tidak dipengaruhi pendapat pencacah.

2.5 Pelaksanaan Pengumpulan Data

Apabila cara-cara pengumpulan data di atas kita kelompokkan, maka ada dua kelompok cara pengumpulan data, yaitu:

1. Pelaksanaan pengumpulan secara langsung. Dalam hal ini pencacah langsung mengumpulkan data di lapangan, seperti pengumpulan data yang dilakukan dengan wawancara, pengamatan dan pengukuran.
2. Pelaksanaan pengumpulan yang dilakukan secara tidak langsung. Dalam hal ini pencacah tidak langsung mengumpulkan data di lapangan, tapi hanya dengan mengirimkan daftar pertanyaan pada responden.

Dalam bagian ini pembahasan dibatasi hanya pada pelaksanaan secara langsung. Sedangkan pengumpulan yang dilakukan secara tidak langsung tidak dibicarakan karena dalam hal ini tidak diperlukan tenaga

lapangan, di samping pelaksanaannya sudah langsung dari pusat pengumpulan data.

Pengumpulan data yang dilakukan secara langsung di lapangan, diperlukan petugas pencacah. Petugas pencacah ini dilatih terlebih dahulu guna menjelaskan maksud tujuan penelitian, konsep dan definisi, cara pengisian daftar pertanyaan, tata cara mencacah dan lain-lain. Petugas pencacah selain dilatih, juga dilengkapi dengan buku pedoman yang memuat segala sesuatu mengenai petunjuk pelaksanaan pengumpulan data. Buku pedoman ini berguna untuk membimbing petugas apabila telah sampai di lapangan, memecahkan berbagai kesulitan yang mungkin timbul. Petugas pencacah juga haruslah orang yang tajam pikirannya, kreatif, teliti, bebas dari kecenderungan pribadi dan prasangka, bijaksana, ramah, menyenangkan, sehat, waspada dan lincah, di samping memiliki kemampuan menulis yang rapi dan dapat dibaca.

2.6 Perancangan Kuesioner (*Questioner Design*)

1. Perencanaan Kuesioner

Pada umumnya suatu kuesioner berisi sekumpulan pertanyaan yang tersusun secara *sistematis* dan *standard* sehingga pertanyaan yang sama dapat diajukan terhadap setiap responden. Sistematis di sini dimaksudkan *item-item* pertanyaan disusun menurut logika (*logical sequence*) yang sesuai dengan maksud dan tujuan penelitian. Sedangkan *standard* diartikan setiap *item* pertanyaan mempunyai pengertian yang sama.

Daftar pertanyaan ini perlu dikembangkan dengan teliti dan dicoba sebelum benar-benar diterapkan secara meluas. Oleh karenanya dalam mempersiapkan suatu kuesioner, perlu diperhatikan jenis pertanyaan, bentuk pertanyaan, pemilihan kata-kata dan juga urutannya.

Kesalahan yang sering terjadi adalah pada jenis pertanyaan yang tidak bisa dijawab, atau tidak perlu dijawab, dan malahan tidak mencantumkan pertanyaan yang seharusnya dijawab. Hendaknya setiap pertanyaan diperiksa sehingga bisa dilihat apakah menunjang tujuan penelitian.

Usaha untuk membuat kuesioner suatu penelitian yang baik diarahkan pada dua tujuan utama yaitu:

1. Memperoleh data yang berhubungan dengan maksud dan tujuan penelitian.
2. Mengumpulkan informasi dengan kecermatan dan ketelitian yang dapat dipertanggungjawabkan.

Untuk memenuhi tujuan pertama rancangan kuesioner harus betul-betul sesuai dengan situasi dengan lingkup topik yang diselidiki. Informasi yang dikumpulkan harus berupa fakta dan bersifat obyektif, sesuai dengan tujuan survei. Selain itu pertanyaan-pertanyaan tersebut hanya ditujukan kepada responden yang diketahui berhak dan mampu menjawabnya. Sedangkan tujuan kedua, untuk tingkat ketelitian informasi yang dikumpulkan dapat diperoleh apabila kuesioner disusun secara sederhana, mudah dimengerti serta adanya keseragaman peristiwa dan petunjuk pengisiannya.

Meskipun bukan merupakan satu-satunya alat pengumpulan data, tetapi kuesioner adalah alat yang efektif untuk berbagai macam cara pengumpulan data seperti wawancara, pengamatan maupun *mailing system*.

Keuntungan penggunaan kuesioner dalam suatu survei dibandingkan dengan alat yang lain terutama dapat diperolehnya data *standard* yang dapat dipertanggung jawabkan untuk keperluan analisis menyeluruh tentang karakteristik populasi yang diselidiki.

2. Bentuk Pertanyaan

Bentuk pertanyaan dibedakan atas pertanyaan terbuka dan pertanyaan tertutup. Perbedaan antara pertanyaan terbuka dan pertanyaan tertutup terletak pada tingkat kebebasan responden menjawab pertanyaan-pertanyaan dalam kuesioner. Pertanyaan terbuka memungkinkan responden memberikan jawaban yang dikehendaknya dengan kata-kata yang dipilihnya sendiri. Sedangkan pertanyaan tertutup membatasi jawaban responden dengan keharusan diantara jawaban-jawaban yang sudah tercantum dalam kuesioner.

a. Pertanyaan Tertutup

Ada dua macam bentuk sederhana dari pertanyaan tertutup, yaitu pilihan dwicabang dan pilihan ganda.

1). Pertanyaan Pilihan Dwicabang (*Dichotomy*)

Di dalam pertanyaan pilihan dwicabang, suatu pertanyaan telah diberikan dua alternatif jawaban misalnya :

Teladan 1 :

Apakah dalam satu tahun terakhir ini suami Ibu bekerja ?

1. Ya 2. Tidak

Dalam teladan tersebut hanya diberikan dua alternatif jawaban ya dan tidak. Atau seperti pada teladan lainnya di bawah ini.

Teladan 2 :

Di manakah tempat pekerjaan ibu ?

1. Di desa yang sama 2. Di luar desa

2). Pertanyaan pilihan Ganda

Pada pertanyaan pilihan ganda, dalam satu pertanyaan diberikan lebih dari dua alternatif jawaban, seperti pada teladan berikut

Teladan 3 :

Apakah lapangan pekerjaan suami Ibu ?

- | | |
|-----------------------------------------|-------------------------------------------------------|
| <input type="checkbox"/> 1. Pertanian | <input type="checkbox"/> 5. Kerajinan |
| <input type="checkbox"/> 2. Perkebunan | <input type="checkbox"/> 6. Perindustrian |
| <input type="checkbox"/> 3. Perikanan | <input type="checkbox"/> 7. Jasa-jasa/Pemerintah/ABRI |
| <input type="checkbox"/> 4. Perdagangan | <input type="checkbox"/> 8. Lainnya |

Teladan 4 :

Status kepemilikan rumah yang Saudara tempati adalah :

- | | |
|-------------------------------------------|----------------------------------------|
| <input type="checkbox"/> 1. Milik sendiri | <input type="checkbox"/> 4. Bebas Sewa |
| <input type="checkbox"/> 2. Kontrak | <input type="checkbox"/> 5. Lainnya |
| <input type="checkbox"/> 3. Sewa | |

Syarat yang harus dipenuhi pada pertanyaan pilihan dwicabang dan ganda:

Apabila dikehendaki pembedaan yang tegas antara dua atau lebih alternatif jawaban, maka jelas harus dipakai pertanyaan pilihan dwicabang atau pilihan ganda. Sebagai dasar analisis dari kedua cara ini adalah penghitungan frekuensi untuk setiap jawaban.

Para peneliti seringkali memulai penyelidikannya dengan dugaan bahwa suatu pilihan jawaban tertentu lebih disukai responden daripada pilihan jawaban yang lain. Dari penghitungan frekuensi dapat pula diuji apakah proporsi dari pilihan-pilihan jawaban sejalan dengan harapan atau dugaan peneliti sebelumnya.

Anggapan-anggapan yang dipakai bagi syarat sahnya data yang diperoleh dari pertanyaan pilihan dwicabang atau pilihan ganda adalah bahwa pilihan-pilihan jawaban yang diberikan dalam daftar pertanyaan harus saling asing (*mutually exclusive*) dan lengkap.

Saling asing artinya responden harus memilih satu dan hanya satu dari pilihan yang diberikan, sedangkan **lengkap** artinya pilihan-pilihan jawaban yang diberikan harus lengkap meliputi semua kemungkinan jawaban yang ada.

Perhatikan teladan berikut ini :

Teladan 5 :

Termasuk kategori yang mana umur Saudara (dalam Tahun) ?

- | | |
|-------------------------------------|-------------------------------------|
| <input type="checkbox"/> 1. 20 - 30 | <input type="checkbox"/> 3. 40 - 50 |
| <input type="checkbox"/> 2. 30 - 40 | <input type="checkbox"/> 5. 50 - 60 |

Kategori umur pada teladan di atas selain tidak saling asing (adanya batas kelas yang berulang) tetapi juga tidak lengkap tidak ada

perincian bagi mereka yang berumur di bawah 20 tahun). Penggolongan umur di atas dapat dirubah menjadi :

- | | |
|--------------------------------------------|---------------------------------------|
| <input type="checkbox"/> 1. kurang dari 20 | <input type="checkbox"/> 4. 40 - 49 |
| <input type="checkbox"/> 2. 20 - 29 | <input type="checkbox"/> 5. ≥ 50 |
| <input type="checkbox"/> 3. 30 - 39 | |

b. Pertanyaan Terbuka

Bentuk pertanyaan terbuka umumnya dapat dibagi menjadi empat yaitu:

1). Pertanyaan Pendahuluan

Teladan 6 : Bagaimana kesan Saudara terhadap pelayanan penerbangan pada perusahaan Garuda ?

2). Permintaan Saran

Teladan 7: Menurut Saudara apakah yang seharusnya dilakukan oleh pemakai jalan untuk mengurangi kemacetan lalu lintas ?

3). Petunjuk untuk membuka daya ingat responden

Teladan 8: Apakah yang Saudara ketahui tentang Keluarga Berencana?

4). Pertanyaan yang bersifat menyelidik

Teladan 9: Selain gaji tetap perbulan, adakah penghasilan lain di luar gaji ?

Sebutkan !

Pertanyaan terbuka mempunyai **kelebihan** sebagai berikut :

- 1) Memungkinkan diperolehnya jawaban yang spontan dan bebas. Dalam menjawab pertanyaan ini responden dapat mengikuti jalan pikirannya sendiri dengan menggunakan kata-katanya sendiri.
- 2) Bermanfaat untuk menjajaki jawaban responden terutama yang menyangkut segi kualitatif (pendapat, sikap dan aspirasi responden).
- 3) Untuk segi kuantitatif jenis pertanyaan ini juga bermanfaat untuk memperoleh rentang nilai yang diperlukan seperti: biaya sewa rumah, pendapatan dan konsumsi rumah tangga perbulan.
- 4) Memberikan pemanasan (*warming up*) terutama apabila topik penelitian merupakan hal yang baru bagi responden dan mereka hampir tidak memiliki pengetahuan dan pengalaman tentang bahan penelitian. Dalam hal yang demikian pertanyaan tertutup akan memberikan hasil yang menyesatkan.

Di samping kelebihan-kelebihan tersebut, jenis pertanyaan terbuka juga mempunyai **kekurangan** antara lain:

- 1) Bentuk pertanyaan terbuka biasanya memakan waktu lebih lama pada saat pencacahannya/pengisiannya.
- 2) Untuk keperluan analisis perlu dilakukan pengelompokan terhadap jawaban-jawaban responden. Hal ini kadang-kadang sulit dilakukan, apabila variasi jawaban responden sangat besar.
- 3) Cenderung menimbulkan bias yang bersumber dari responden dan pencacah akibat salah tafsir dari salah satu pihak. Atau bias dalam pengolahan data yang disebabkan oleh pemberian bobot yang tidak sesuai untuk jawaban pada waktu pengelompokan.

3. Kalimat Pertanyaan

Ada beberapa hal yang harus dipertimbangkan dalam menyusun kalimat pertanyaan antara lain:

- a) Pertanyaan mudah dimengerti baik oleh responden maupun pencacah.
- b) Setiap pertanyaan mempunyai satu arti yang tepat/unik. Tidak ada arti lain bagi pertanyaan tersebut yang mengakibatkan pengertiannya menjadi kabur.
- c) Susunan pertanyaan secara keseluruhan mengikuti jalan pikiran yang urut dan lengkap.
- d) Mudah dalam pengelolaan datanya.
- e) Setiap pertanyaan harus bermanfaat sesuai dengan maksud dan tujuan penelitian.

a. Pemilihan Kata

Bahasa yang dipakai dalam daftar pertanyaan harus sederhana, mudah dimengerti dan kalimatnya tidak terlalu panjang. Dalam hal ini harus diperhatikan kata-kata yang mempunyai pengertian sehari-hari di masyarakat. Berikut ini diberikan beberapa teladan pertanyaan yang mempunyai arti kurang jelas.

Teladan 10 : Berapa pendapatan Saudara ?

Kata-kata saudara disini ditunjukkan kepada responden, pribadi atau keluarganya? Kapan saat yang dimaksud? Seminggu yang lalu, sebulan yang lalu atau setahun sebelum penelitian. Kemudian apa yang termasuk dalam "pendapatan" disamping gaji atau upah; apakah penghasilan dari sumber lain seperti bunga bank, sewa rumah atau laba jual mobil termasuk dalam pengertian pendapatan ini?

Teladan 11: Berapa banyak beras yang biasanya saudara butuhkan? Dalam pertanyaan ini apakah responden menjawab dalam kg, kwintal atau liter? Kapan, tidak dijelaskan saatnya. Penggunaan kata biasanya juga mempunyai pengertian yang kabur, apakah setiap hari, setiap minggu atau sebulan tergantung tafsiran yang dipakai terhadap kata "biasanya".

Teladan 12: Berapa umur isteri Saudara ?

Pertanyaan sebaiknya dapat diajukan kepada semua responden yang mampu menjawabnya. Dalam teladan 12, pertanyaan tersebut seharusnya didahului oleh pertanyaan lain untuk menentukan apakah responden mempunyai isteri. Apabila tidak didahului oleh anggapan di atas, maka pertanyaan pada Teladan 12, tidak dapat digunakan terhadap responden terutama bagi responden wanita atau laki-laki yang belum kawin. Selain janggal pertanyaan demikian bisa menyinggung perasaan responden.

b. Pertanyaan Ganda

Dalam menyusun kalimat pertanyaan pada daftar pertanyaan harus dihindari dua pertanyaan atau lebih dalam satu kalimat. Pertanyaan semacam ini akan membingungkan baik pencacah maupun responden. Responden menjadi bingung untuk menjawab jenis pertanyaan tersebut, khususnya apabila dia mempunyai jawaban yang berbeda untuk setiap bagian pertanyaan. Pencacah juga akan mengalami kesulitan mengartikannya apabila bagian pertanyaan yang satu saling bertentangan dengan bagian yang lain. Perhatikan teladan pertanyaan ganda berikut ini:

Teladan 13: Apakah seminggu lalu tidak bekerja dan berusaha mencari pekerjaan?

Kemungkinan jawaban responden adalah sebagai berikut:

- a. tidak bekerja dan tidak mencari pekerjaan
- b. tidak bekerja dan mencari pekerjaan
- c. bekerja dan tidak mencari pekerjaan
- d. bekerja dan mencari pekerjaan

Pertanyaan ganda ditandai dengan munculnya kata penghubung dan atau atau. Pertanyaan ini bisa dipecah menjadi dua, untuk Teladan 13 pertanyaan menjadi :

13. Apakah seminggu yang lalu saudara bekerja ?

1. Ya 2. Tidak

14. Jika tidak apakah saudara berusaha mencari pekerjaan?

1. Ya 2. Tidak

4. Susunan *Item* Dalam Daftar Pertanyaan

Pentingnya Urutan *Item* Pertanyaan

Tujuan penelitian yang diwujudkan dalam bentuk pertanyaan harus disusun menurut urutan yang logis/masuk akal agar tidak akan berpengaruh buruk bagi mutu data yang dihasilkan.

Pertanyaan hendaknya dirancang untuk membangkitkan minat responden menjawab pertanyaan yang diajukan. Pertanyaan yang loncat dari hal yang satu ke hal yang lain yang tidak bersangkutan membuat bingung dan bosan.

Bagi pencacah urutan pertanyaan harus jelas dan mudah pengisiannya. Selain itu instruksinya pun harus jelas, kapan dan bagaimana mengalihkan pertanyaan yang satu ke pertanyaan yang lain. Kepentingan utama peneliti adalah diperolehnya data yang bermutu tinggi, oleh karenanya urutan pertanyaan seharusnya tidak menimbulkan bias.

Beberapa Segi Urutan Pertanyaan :

- a. Responden berkewajiban menjawab semua pertanyaan-pertanyaan yang diajukan sesuai undang-undang yang berlaku. Meskipun demikian masih dijumpai kesulitan-kesulitan yang disebabkan responden tidak bersedia diwawancarai. Cara paksaan tentu saja kurang baik, karena data yang diperoleh mungkin tidak benar. Dalam hal ini perlu diberikan keterangan-keterangan yang dapat mendorong minat responden untuk bekerjasama dalam penelitian. Selain itu jaminan kerahasiaan keterangan yang diberikan juga perlu ditekankan untuk menambah keyakinan responden.
- b. Keterangan Pengenal
Keterangan pengenal yang berkenaan dengan nama organisasi yang melaksanakan penelitian, judul penelitian, keterangan tempat, keterangan mengenai diri responden dan pencacah ditempatkan pada bagian awal daftar pertanyaan. Hal ini memudahkan pengaturan administrasi, mengendalikan prosedur pemilihan contoh dan juga untuk penelitian mengendalikan *non sampling error*.
- c. Pertanyaan Pembuka
Pertanyaan pembuka yang biasanya berupa pertanyaan yang bersifat pengantar harus mendorong minat responden untuk tidak hanya menjawab pertanyaan itu saja tetapi juga seluruh pertanyaan. Pertanyaan yang berhubungan dengan tujuan penelitian harus diselesaikan dulu. Dalam hal ini kecakapan pencacah untuk membawa pembicaraan mengarah langsung pada tugas-tugas yang harus dijalaninya.
Dalam beberapa penelitian pertanyaan pembuka berguna untuk mendaftar anggota populasi yang menjadi sasaran penelitian. Misalnya untuk penelitian pertanian akan didaftar responden yang bekerja di bidang pertanian (rumah tangga tani). Selain itu pertanyaan pembuka berguna sebagai latihan, untuk melatih responden mengenal tujuan penelitian dan keterangan apa saja yang dikehendaki dari dirinya.
- d. Untaian *Item* Pertanyaan
Satu hal yang perlu diingat dalam menyusun *item* pertanyaan adalah peneliti harus dapat mengikuti jalan pikiran responden. Aliran pertanyaan satu ke pertanyaan lain harus lancar, sehingga responden dapat memahami hubungan pertanyaan tersebut dengan tujuan penelitian. Salah satu yang sering digunakan adalah dengan memasukkan pertanyaan yang bersifat transisi untuk mengalihkan perhatian responden dari masalah yang satu dengan masalah yang lain. Misalnya (lihat lampiran) pada Blok II dalam kuesioner dimulai dengan kalimat "Kami ingin menanyakan sedikit tentang riwayat ibu". Kemudian untuk kelompok pertanyaan selanjutnya diawali dengan "Sekarang kami ingin menanyakan tentang suami ibu". Tanpa pertanyaan transisi perubahan periode mungkin akan membingungkan responden.

e. Letak Pertanyaan yang Bersifat Peka

Pertanyaan yang bersifat peka bisa menghambat jalannya wawancara. Responden yang merasa rahasia pribadi atau perusahaannya tidak mau diketahui orang lain tentu saja tidak mau mengungkapkannya. Supaya pertanyaan ini tidak mengganggu jalannya wawancara sebaiknya diletakkan pada bagian akhir daftar pertanyaan. Juga diperlukan kepandaian pencacah untuk meyakinkan responden bahwa data akan dijamin kerahasiaannya.

f. Penggolongan Data

Penggolongan data biasanya dipakai untuk keperluan tabulasi. Penggolongan dibuat menurut golongan ekonomi, sosial, dan ciri lainnya. Dalam beberapa penelitian ada yang menempatkan data ini pada bagian akhir daftar pertanyaan dengan alasan ada hubungan langsung dengan tujuan penelitian, meskipun ada juga yang menempatkannya di bagian depan. Termasuk dalam data ini misalnya umur, jenis kelamin, pendidikan, status perkawinan dan jabatan.

5. Penyajian Kuesioner

a. Tata Letak

Hal yang dijadikan pertimbangan dalam tata letak daftar pertanyaan antara lain:

1. Penampilan

Penampilan yang ditunjukkan pada bagian luar daftar pertanyaan mempengaruhi sikap responden terhadap minat untuk berwawancara. Kesan pertama yang diakibatkan oleh bentuk daftar pertanyaan yang tidak rapi dan kotor menyebabkan responden enggan untuk diwawancarai.

2. Memudahkan Pencatatan Data

Tata letak daftar pertanyaan harus memberikan kemudahan bagi pencacah untuk mencatat data yang diperoleh. Pencatatan yang lama dalam wawancara menyebabkan responden harus menunggu dari pertanyaan yang satu ke pertanyaan yang lain, bisa menyebabkan responden jemu.

3. Memudahkan Pengolahan Data

Tata letak daftar pertanyaan harus memudahkan proses pengolahan data. Adalah sangat melelahkan untuk mencari isian daftar pertanyaan yang tidak teratur letaknya untuk dipindahkan ke lembar kerja dalam pengolahan manual.

4. Ukuran Daftar Pertanyaan

Ukuran daftar pertanyaan harus cukup besar untuk memungkinkan jarak yang cukup antara *item-item* pertanyaan, tetapi juga tidak terlalu besar sehingga sulit dilipat dan disimpan. Ukuran kuarto adalah sangat baik dan mudah disimpan dalam map atau tas.

b. Petunjuk Pengisian

Untuk memudahkan pengisian, setiap *item* pertanyaan diberi nomorurut. Petunjuk diberikan langsung pada *item* pertanyaan yang bersangkutan atau di halaman depan. Bisa juga digunakan tanda panah. Petunjuk berisikan 'konsep dan definisi' serta cara pengisian,

baik untuk pencacah maupun pengawas, bisa dipisahkan dalam buku pedoman tersendiri.

c. Kertas dan Pencetakan Daftar Pertanyaan

Kertas yang dipakai sebaiknya yang bermutu baik, tidak terlalu tipis atau terlalu tebal. Dapat pula digunakan warna kertas yang berlainan untuk masing-masing kelompok pertanyaan. Nomor dan huruf cetakan harus rapi dan mudah dibaca.

6. Uji Coba

Satu hal yang penting adalah uji coba kuesioner, yaitu mencoba *item-item* pertanyaan yang telah dirancang dalam keadaan yang sebenarnya. Misalnya apakah bahasanya bisa dimengerti, tidak terlalu panjang, dan susunan *item* sudah efisien. Hasil uji coba ini dapat dipakai untuk memperbaiki istilah, kalimat maupun susunan yang tidak sesuai dengan keadaan yang sebenarnya. Dalam uji coba diambil contoh secukupnya untuk mendapatkan responden.

2.7 Rangkuman

Statistika adalah pengetahuan yang berhubungan dengan pengumpulan, pengolahan, penyajian data, analisis data, penarikan kesimpulan, serta pembuatan keputusan yang cukup beralasan berdasarkan data dan analisis yang dilakukan. Bahwa pengumpulan data merupakan bagian yang paling awal dari suatu kegiatan statistika, sehingga dengan demikian akan merupakan bagian yang sangat penting.

Pengumpulan data yang akan dibahas dalam **modul** ini dititik beratkan pada suatu proses pengadaan data primer. Dalam hal ini adalah untuk menghasilkan data statistik yang '**baik**', yaitu data yang bisa dipercaya kebenarannya (*reliable*) dan tepat waktu (*up to date*).

Dalam pengumpulan data, ada dua metode yaitu metode sensus dan metode *sampling*. Kedua metode tersebut masing-masing ada kelebihan dan kelemahannya.

Beberapa cara dalam pengumpulan data, antara lain ; Wawancara (*interview*), Diisi sendiri oleh responden, Pengamatan (*observasion*), dan Pengukuran (*measurement*).

Pelaksanaan pengumpulan data dapat dikelompokkan menjadi dua kelompok, yaitu pengumpulan data secara langsung dan tidak langsung.

Sedangkan perancangan kuesioner (*Questionnaire Design*) meliputi perencanaan kuesioner, bentuk pertanyaan, kalimat pertanyaan, susunan *item* dalam daftar pertanyaan, penyajian kuesioner dan uji coba.

2.8 Latihan Soal

Soal – soal Latihan Modul Pengumpulan Data

<u>A. Pilihan Berganda</u>

Petunjuk :

Di bawah ini ada 20 buah pernyataan. Masing-masing pernyataan mempunyai 4 buah kemungkinan jawaban. Pilih satu yang paling tepat.

1. Tahap awal dari kegiatan statistik adalah :
 - a. Pengolahan Data
 - b. Penyajian Data
 - b. Pengumpulan Data
 - c. Analisis Data
 2. Titik berat pengumpulan data dalam modul ini adalah pengumpulan data :
 - a. Primer
 - b. Sekunder
 - d. Tersier
 - e. Primer dan sekunder
 3. Salah satu perencanaan pengumpulan data adalah persiapan administrasi, hal-hal dibawah ini yang tidak termasuk didalam persiapan administrasi adalah :
 - a. Mempersiapkan jadwal waktu
 - b. Izin resmi
 - c. Mempersiapkan penerangan
 - d. Mempersiapkan latihan petugas
 4. Persiapan teknis perencanaan pengumpulan data adalah sebagai berikut, kecuali :
 - a. Mempersiapkan konsep dan definisi
 - b. Menentukan bentuk kegiatan
 - b. Menentukan ruang lingkup
 - c. Organisasi, pengawasan dan koordinasi
 5. Keuntungan pengumpulan data dengan metode sensus adalah :
 - a. Tidak memiliki *sampling* error (kesalahan dalam penggunaan metode pengumpulan data)
 - b. Biaya jauh lebih sedikit
-

- d. Tenaga yang dibutuhkan sedikit
 - e. Kegiatan mudah dikuasai dan diawasi
6. Kelemahan pengumpulan data dengan metode sensus adalah :
- a. Tidak mempunyai *sampling error*
 - b. Jenis data yang diperoleh terbatas dan sifat data sederhana
 - c. Data hasil sensus sebagai kerangka dalam melakukan survei dengan metode *sampling*
 - d. Dari hasil sensus akan dapat diketahui gambaran sebenarnya suatu populasi
7. Kegiatan sensus yang dilakukan oleh Badan Pusat Statistik pada tahun-tahun yang berakhir dengan angka 3 (tiga) adalah :
- a. Sensus Penduduk
 - b. Sensus Pertanian
 - d. Sensus Ekonomi
 - e. Sensus Kemiskinan
8. Keuntungan pengumpulan data dengan metode *sampling* adalah :
- a. Kegiatan-kegiatannya mudah dikuasai atau diawasi
 - b. Tenaga yang dibutuhkan banyak
 - c. Biaya Banyak
 - d. Butuh waktu lama atau banyak
9. Kelemahan atau kekurangan pengumpulan data dengan metode *sampling* adalah :
- a. Gambaran populasi hanya merupakan taksiran
 - b. Dalam survei dengan metode *sampling* tidak perlu kerangka contoh yang berasal dr sensus
 - c. Pengambilan sampel selalu tepat atau tidak ada *sampling error*
 - d. jawaban b dan c benar
10. Berikut adalah beberapa syarat yang harus dipenuhi agar suatu sample/ contoh dikatakan sebagai sampel yang baik dan memenuhi syarat statistika, kecuali :
- a. Mewakili populasi (representatif)
 - b. Mempunyai tingkat kebenaran yang bisa diukur
 - c. efektif
 - d. Mempunyai nilai perkiraan yang tidak bias
11. Berikut adalah beberapa cara pengumpulan data kecuali :
- a. Wawancara (*interview*)
 - b. Pengamatan (*observasion*)
 - c. Pengukuran (*measurement*)
 - d. Penafsiran

12. Kebaikan pengumpulan data dengan wawancara adalah :
 - a. Biaya besar karena harus melatih pencacah
 - b. Kelengkapan jawaban terhadap semua pertanyaan dalam kuesioner dapat dijamin
 - c. Terkadang petugas atau pencacah sulit ketemu dengan responden
 - d. Perlu organisasi lapangan yang baik

 13. Kelemahan pengumpulan data dengan wawancara adalah :
 - a. Responden seringkali memberikan jawaban yang bersifat pribadi
 - b. Kembalinya daftar pertanyaan lebih banyak, karena pencacah mengisi langsung
 - c. Pencacah dapat bertanya dengan baik
 - d. Pencacah Dapat menggunakan bahasa yang dapat dimengerti oleh responden

 14. Kelemahan pengumpulan data dengan mengisi daftar pertanyaan (*mailing system*) adalah :
 - a. Responden dapat memberikan jawaban secara terbuka
 - b. Pertanyaan yang diajukan mungkin tidak dapat dijawab atau sulit dijawab secara tertulis
 - c. Biaya lebih kecil dan lingkup daerah lebih luas
 - d. Tidak memerlukan pelatihan petugas pencacah

 15. Untuk mengetahui perilaku bayi dan melihat kesibukan para karyawan yang bekerja di suatu kantor, pengumpulan datanya dilakukan dengan cara :
 - a. Wawancara (*interview*)
 - b. Diisi sendiri oleh responden
 - c. Pengamatan (*observation*)
 - d. Pengukuran (*measurement*)

 16. Keuntungan pengumpulan data dengan pengukuran (*measurement*) adalah sebagai berikut, kecuali :
 - a. Teliti, karena obyek dikuasai pencacah
 - b. data hasil pengukuran dapat diisi seutuhnya dalam daftar pertanyaan
 - c. Responden dapat memberikan jawaban secara terbuka
 - d. Data tidak dipengaruhi pendapat pencacah

 17. Pengumpulan data yang dilakukan secara tidak langsung adalah.:
 - a. Wawancara (*interview*)
 - b. Diisi sendiri oleh responden
 - c. Pengamatan (*observation*)
 - d. Pengukuran (*measurement*)
-

18. Dalam mempersiapkan suatu kuesioner, yang tidak perlu diperhatikan adalah :
- Jenis pertanyaan
 - Bentuk pertanyaan
 - Pemilihan kata-kata dan urutannya
 - Sifat responden

19. Apakah dalam satu tahun terakhir ini suami ibu bekerja ?

1. Ya Tidak

Pertanyaan diatas merupakan contoh pertanyaan :

- Pertanyaan tertutup pilihan dwicabang (dichotomy)
 - Pertanyaan tertutup pilihan ganda
 - Pertanyaan terbuka pendahuluan
 - Pertanyaan terbuka yang bersifat menyelidik
20. Beberapa hal yang harus dipertimbangkan dalam menyusun kalimat pertanyaan adalah sebagai berikut, kecuali :
- Pertanyaan mudah dimengerti baik oleh responden maupun pencacah
 - Susunan pertanyaan secara keseluruhan mengikuti jalan pikiran yangurut dan lengkap
 - Mudah dalam pengelolaan datanya
 - Setiap pertanyaan harus sesuai dengan keinginan responden

B. Soal Uraian

- Dalam kegiatan pengumpulan data primer di lapangan, ada beberapa persiapan yang harus dilakukan. Coba sebutkan dan jelaskan.
- Sebutkan beberapa cara pengumpulan data dan jelaskan.
- Ada berapa macam bentuk pertanyaan, **sebutkan dan jelaskan**, serta **berilah contoh** masing-masing 2 (dua) pertanyaan.
- Kepentingan utama penelitian adalah diperolehnya data yang berkualitas tinggi, oleh karena itu urutan pertanyaan seharusnya tidak menimbulkan bias, dan untuk itu ada beberapa segi urutan pertanyaan. **Sebutkan dan jelaskan** beberapa segi urutan pertanyaan tersebut.
- Hal apa saja yang dijadikan pertimbangan dalam menyusun tata letak daftar pertanyaan, **sebutkan dan jelaskan**.

Bab III Penutup

3.1 Kesimpulan

1. Pengumpulan data merupakan bagian yang paling awal dari kegiatan statistika.
2. Pengumpulan data yang dibahas dalam modul ini dititik beratkan pada proses pengadaan data primer
3. Pengumpulan data diawali dengan perencanaan pengumpulan data, yang meliputi persiapan administrasi dan persiapan teknis
4. Dalam pengumpulan data statistik, dapat digunakan metode sensus dan metode *sampling*
5. Ada empat cara yang dapat dilakukan didalam pengumpulan data, yaitu ; wawancara (*interview*), diisi sendiri oleh responden, pengamatan (*observasion*) dan pengukuran (*measurement*)
6. Pelaksanaan pengumpulan data dapat dilakukan secara langsung dan secara tidak langsung
7. Kegiatan perancangan kuesioner (*questionnaire design*) diawali dari perencanaan kuesioner, bentuk pertanyaan, kalimat pertanyaan, susunan *item* dalam daftar pertanyaan, penyajian kuesioner dan uji coba

3.2 Tindak Lanjut

Setelah membaca modul ini peserta diklat mampu dan memahami konsep-konsep pengumpulan data sehingga dapat mengikuti pembelajaran selanjutnya yaitu pengolahan data dan penyajian data. Selain itu peserta juga dapat mengembangkan pengetahuan tentang teknik pengambilan sampel.

Daftar Pustaka

- Dayan, Anto. 1990. *Pengantar Metode Statistik Jilid I*. Jakarta: LP3S
- Supranto, J. 1992. *Statistik, Teori & Aplikasi Jilid I*. Jakarta: Erlangga.
- Nazir, M. 1985. *Metode Penelitian*. Jakarta: Galia Indonesia.
- Sudjana. 2001. *Metode Statistika Deskriptif*. Bandung: Tarsito.
- Wiroidikromo, Sartono. 2006. *Matematika IPA Kelas XI*. Jakarta: Erlangga.
- Ismail. 2003. *Pelatihan Terintegrasi Berbasis Kompetensi Mata Pelajaran Matematika: Statistika*. Jakarta: Direktorat Lanjutan Pertama.
- Heriyanto dan HM. Akib Hamid. 1993-1994. *Statistika Dasar*. Jakarta: Digdamen.
- Winarno dan Ganung Anggraeni. 2001. *Pengantar Statistika*. Yogyakarta: PPPG Matematika.
- Agus Irianto. 2004. *Statistik Konsep Dasar Dan Aplikasinya*. Jakarta: Prenada Media.
- Budi Santosa, Purbayu dan Ashari. 2006. *Analisis Statistik dengan Menggunakan MS Excel dan SPSS*. Yogyakarta: Andi.
- Kuswadi dan Erna Mutiara. 2004. *Statistik Berbasis Kompute untuk orang-orang Non – Statistik*. Jakarta: PT Elex Media Komputindo.
- Santosa, Singgih. 2000. *Buku latihan SPSS Statistik Parametrik*. Jakarta: PT Elex Media Komputindo.
-

Lampiran

Jawaban Soal-Soal Latihan Modul Pengumpulan Data

1. C
 2. A
 3. D
 4. D
 5. A
 6. B
 7. B
 8. A
 9. B
 10. C
 11. D
 12. B
 13. A
 14. B
 15. C
 16. C
 17. B
 18. D
 19. A
 20. D
-



Modul 02

Metode Penarikan Sampel



DAFTAR ISI

KATA PENGANTAR.....	ii
DAFTAR TABEL	v
DAFTAR GAMBAR.....	v
Bab I Pendahuluan.....	1
1.1 Latar Belakang	1
1.2 Deskripsi Singkat	1
1.3 Hasil Belajar (Tujuan Pembelajaran Umum).....	1
1.4 Indikator Hasil Belajar (Tujuan Pembelajaran Khusus)	1
1.5 Materi Pokok.....	2
1.6 Manfaat	2
Bab II Pengantar Metode Pengambilan Sampel.....	3
2.1 Pendahuluan	3
2.2 Jenis Data	3
2.3 Cara Pengumpulan Data	4
2.4 Konsep dan Definisi.....	5
2.5 Arti dan Pentingnya Sampling	6
Tabel 1. Perbandingan Survei Sampel dan Sensus	6
2.6 Perbandingan Survei dan Sensus	7
2.7 Proses Sampling.....	7
2.8 Latihan	7
2.9 Rangkuman	7
Bab III Dasar-Dasar Teori Penarikan Sampel.....	9
3.1 Teknik Pengambilan Sampel	9
3.2 Kemungkinan Sampel	10
3.3 Penyimpangan Nilai Dugaan Dari Nilai Populasi.....	11
3.4 Kerangka Sampel	13
3.5 Pemilihan Sampel Secara Acak (Menggunakan TAR).....	13
3.6 Latihan	16
3.7 Rangkuman	16
Bab IV. Sampel Acak Sederhana.....	19
4.1 Definisi Sampel Acak Sederhana.....	19
4.2 Metode Penduga Rata-Rata dan Varians	19
4.3 Penentuan Ukuran Sampel	21
4.4 Latihan	23
4.5 Rangkuman	24

Bab V. Sampel Sistematis (<i>Systematic Sampling</i>)	27
5.1 Definisi Sampel Sistematis	27
5.2 Metode Penduga Rata-Rata dan Varians	27
5.3 Latihan	29
5.4 Rangkuman	30
Bab VI. Penutup	29
6.1 Simpulan	29
6.2 Tindak Lanjut	29
Daftar Pustaka	31

DAFTAR TABEL

Judul Tabel	Halaman
Tabel 1. Perbandingan Survei Sampel dan Sensus	6
Tabel 2. Tabel Angka Random	15

DAFTAR GAMBAR

Judul Gambar	Halaman
Gambar 1. Hubungan Sampling Error dan Non Sampling Error	5

Bab I Pendahuluan

1.1 Latar Belakang

Modul Metode Penarikan Sampel merupakan salah satu media pembelajaran yang disediakan khusus untuk Diklat Fungsional Statistisi Tingkat Terampil. Modul ini telah disesuaikan dengan butir-butir penilaian dari tugas/pekerjaan seorang pejabat fungsional statistisi terampil khususnya yang berkaitan dengan pengambilan/ penarikan sampel. Kompetensi yang ingin dicapai setelah mempelajari modul ini adalah peserta dapat memahami tentang cara-cara penarikan sampel yang terbaik sesuai dengan kaidah ilmu statistik, sehingga dapat menunjang tugasnya sebagai pejabat fungsional statistisi tingkat terampil.

Dalam sebuah penelitian dimana pengumpulan datanya menggunakan teknik survei maka unit analisisnya adalah unit sampel. Cara penentuan jumlah sampel dan unit-unit sampel dari suatu populasi disebut sebagai metode penarikan sampel. Untuk mendapatkan sampel yang dapat mewakili dengan baik keadaan populasinya, diperlukan pemilihan metode pengambilan sampel yang tepat sesuai dengan karakteristik populasi tersebut. Salah satu contoh survei yang menggunakan data sampel dalam penelitiannya adalah Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan oleh BPS untuk memotret keadaan sosial (pendidikan, kesehatan, ketenagakerjaan dan lain-lain) dan keadaan ekonomi masyarakat di Indonesia, seperti pola pengeluaran untuk konsumsi makanan dan non makanan.

Modul ini mengantarkan para peserta untuk memahami pemilihan metode penarikan sampel yang tepat, penentuan jumlah sampel dan langkah-langkah dalam penarikan sampel serta penghitungan *standard error*-nya. Disamping itu, modul ini juga sebagai *guidance* bagi fasilitator dalam mendesain pembelajaran mata diklat Metode Penarikan Sampel.

1.2 Deskripsi Singkat

Modul ini membahas tentang pengantar metode penarikan sampel, dasar-dasar teori pengambilan sampel, sampel acak sederhana, sampel acak berlapis dan sampel acak berkelompok.

1.3 Hasil Belajar (Tujuan Pembelajaran Umum)

Setelah membaca modul ini, peserta dapat memahami dan menerapkan beberapa metode penarikan sampel sesuai karakteristik populasinya dari suatu penelitian dengan teknik survei.

1.4 Indikator Hasil Belajar (Tujuan Pembelajaran Khusus)

Setelah mempelajari modul ini secara tuntas, peserta diharapkan dapat:

1. Menjelaskan jenis-jenis data, pengertian sensus dan survei.
-

2. Mampu menggunakan metode pengambilan sampel secara acak sederhana dan sistematis.
3. Mampu menentukan ukuran sampel sesuai dengan tujuan penelitian.
4. Mampu melakukan pengambilan sampel acak berlapis.
5. Mampu menentukan sampling error

1.5 Materi Pokok

Materi pokok yang dibahas dalam modul ini adalah:

1. Pengantar metode penarikan sampel;
2. Dasar-dasar teori pengambilan sampel;
3. Sampel acak sederhana;
4. Sampel Sistematis (*systematic sampling*).

1.6 Manfaat

Berbekal hasil belajar pada modul Metode Penarikan Sampel ini, peserta diharapkan mampu menerapkan beberapa metode penarikan sampel dalam kegiatan survei yang dilakukan di unit kerjanya masing-masing dengan baik guna meningkatkan kinerja instansinya.

Bab II Pengantar Metode Pengambilan Sampel

2.1 Pendahuluan

Di dalam era globalisasi yang sudah mulai kita rasakan, banyak terjadi perubahan-perubahan dratis di segala sektor. Jenis produksi industri dan jasa yang beraneka ragam makin menyebar dengan tujuan untuk memperoleh nilai tambah setinggi-tingginya. Begitu pula pola pendapatan penduduk yang berubah, akan merubah pula keadaan sosial ekonomi masyarakat secara umum.

Untuk mengendalikan perubahan-perubahan tersebut, bagi pemerintah tidak ada jalan lain kecuali harus ikut campur mengendalikannya. Pemerintah sebagai pengendali banyak mempergunakan informasi (data) statistik untuk melihat perubahan-perubahan, menganalisa dan akhirnya sebagai dasar menentukan kebijaksanaan-kebijaksanaan. Di kalangan swasta, informasi statistik diperlukan sebagai bahan acuan untuk mengelola perusahaan secara lebih efisien dalam mencari peluang-peluang baru yang menguntungkan. Sedangkan di masyarakat umum, data statistik diperlukan untuk bahan penelitian dan kajian. Dengan demikian kebutuhan informasi statistik dilihat dari ragam dan kualitasnya yang baik akan semakin meningkat.

Dari penjelasan di atas dapat disimpulkan bahwa data statistik diperlukan baik oleh pemerintah, swasta, maupun masyarakat umum sebagai dasar perencanaan, pemantauan dan evaluasi penentuan kebijakan.

2.2 Jenis Data

Jenis data dapat dibedakan menurut sifat, sumber, cara memperoleh dan waktu pengumpulannya.

Data menurut sifat:

- a. Kualitatif, yaitu data yang disajikan dalam bentuk kata-kata yang mengandung makna. Contohnya: Nilai ujian yang diperoleh peserta pelatihan cukup baik.
- b. Kuantitatif, yaitu data yang disajikan dalam bentuk angka-angka. Contohnya: Nilai ujian yang diperoleh peserta pelatihan memiliki rata-rata 80.

Data menurut sumber:

- a. Internal, yaitu data yang menggambarkan situasi dan kondisi pada suatu organisasi secara internal. Contohnya: data pegawai, data keuangan, data penjualan, data produksi, dan lain-lain.
- b. Eksternal, yaitu data yang menggambarkan situasi serta kondisi yang ada di luar organisasi. Contohnya adalah data jumlah penjualan tiket kereta, tingkat kepuasan konsumen, persebaran penduduk, dan lain-lain.

Data menurut cara memperolehnya:

- a. Primer, yaitu data yang secara langsung diambil dari obyek penelitian oleh peneliti perorangan maupun organisasi. Contohnya: data hasil
-

wawancara langsung dari konsumen mengenai kepuasan terhadap suatu produk.

- b. Sekunder, yaitu data yang didapat tidak secara langsung dari obyek penelitian. Biasanya didapatkan dari data yang sudah jadi yang dikumpulkan oleh pihak lain. Contohnya adalah data yang diperoleh dari publikasi BPS, data hasil riset surat kabar atau majalah, dan lain-lain.

Data menurut waktu pengumpulannya:

- a. *Cross Section*, yaitu data yang menunjukkan titik waktu tertentu. Contohnya: data penduduk Indonesia tahun 2010.
- b. *Time Series*, yaitu data yang menggambarkan sesuatu dari waktu ke waktu. Contohnya: data penduduk Indonesia tahun 2000-2025.

2.3 Cara Pengumpulan Data

Secara umum ada berbagai cara yang dapat dilakukan untuk mengumpulkan data. Cara tersebut antara lain adalah:

- a. Registrasi atau catatan administrasi.
Registrasi adalah pencatatan secara individu yang dilaksanakan oleh setiap institusi. Misalnya adalah pencatatan penduduk di desa yang meliputi pencatatan kelahiran, kematian perpindahan dan sebagainya. Apabila pencatatan ini dilakukan dengan baik dan teratur maka akan dapat dijadikan sebagai sumber data statistik yang cukup baik.
- b. Sensus.
Sensus adalah pencacahan lengkap terhadap seluruh unit populasi yang menjadi obyek pengamatan pada suatu wilayah. Misalnya adalah sensus penduduk, sensus pertanian, sensus ekonomi, dan lain-lain.
- c. Survei.
Survei adalah pencacahan yang dilakukan hanya pada sebagian dari unit populasi yang menjadi obyek pengamatan pada suatu wilayah. Misalnya adalah Susenas, Sakernas, SDKI, dan lain-lain.
- d. Eksperimen atau percobaan.
Eksperimen adalah pengumpulan data melalui unit-unit yang telah ditentukan secara spesifik untuk tujuan-tujuan khusus. Biasanya eksperimen dilakukan di laboratorium dimana setiap unit diberikan perlakuan yang berbeda kemudian diamati pengaruh dari perlakuan tersebut. Misalnya adalah tanaman padi yang diberikan jenis pupuk yang berbeda dan diamati pertumbuhannya.

2.4 Konsep dan Definisi

Sampling adalah suatu proses yang dilakukan untuk memilih dan mengambil sampel secara benar dari suatu populasi sehingga sampel tersebut dapat mewakili populasinya.

Elemen adalah unit yang digunakan untuk mendapatkan informasi.

Populasi adalah kumpulan dari seluruh elemen sejenis namun dapat dibedakan satu sama lain dimana perbedaan yang ada disebabkan oleh adanya nilai karakteristik yang berlainan.

Populasi target adalah keseluruhan unit dalam areal/wilayah/lokasi/kurun waktu yang sesuai dengan tujuan penelitian.

Sampel adalah bagian dari populasi yang ingin diteliti, yang ciri-ciri dan keberadaannya diharapkan dapat mewakili atau menggambarkan cirri-ciri dan keberadaan populasi yang sebenarnya.

Parameter adalah nilai yang dihitung berdasarkan seluruh elemen populasi.

Statistik adalah nilai yang dihitung berdasarkan sampel atau bagian dari populasi.

Unit observasi adalah unit dimana informasi diperoleh baik secara langsung maupun melalui responden tertentu.

Unit sampling adalah unit yang dijadikan sebagai dasar penarikan sampel baik berupa elemen ataupun kumpulan elemen (klaster).

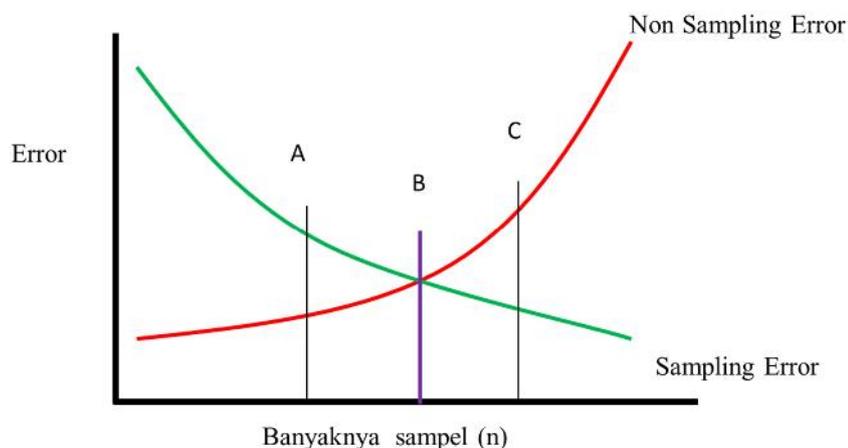
Kerangka sampel merupakan seluruh unit dalam populasi yang akan dijadikan dasar penarikan sampel.

Sampling Error adalah kesalahan yang disebabkan oleh teknik pengambilan sampel.

Non sampling Error adalah kesalahan yang disebabkan bukan dari teknik pengambilan sampel. Sumber-sumber nonsampling error:

1. Kegagalan mengukur beberapa unit dalam sampel terpilih.
2. Kesalahan pengamatan karena teknik pengukuran yang tidak sempurna.
3. kesalahan-kesalahan pada waktu mengedit, memberi kode, mengentri data, dan mentabelkan hasil-hasil survei.

Hubungan antara *Sampling Error* dan *Non sampling Error* adalah seperti Gambar 2.1 berikut ini.



Gambar 1 Hubungan Sampling Error dan Non Sampling Error

Standard Error adalah satuan pengukuran untuk “rata-rata” dari kesalahan-kesalahan dari seluruh distribusi sample.

Relative Standard Error adalah ratio antara standar error terhadap ekspektasi dari penduganya. Dengan RSE dapat diketahui sejauh mana tingkat ketelitian suatu teknik sampling.

2.5 Arti dan Pentingnya Sampling

Di dalam praktik sehari-hari, untuk mengetahui suatu keadaan, kita sering menggunakan sampel untuk bisa mengambil suatu kesimpulan. Ibu rumah tangga yang sedang memasak, akan mencicipi sebagian kecil dari masakannya. Apabila masakan yang dicicipi rasanya kurang asin, maka ia menyimpulkan bahwa masakan tersebut secara keseluruhan kurang asin, sehingga perlu ditambah garam. Contoh lain suatu bank memberikan kuesioner pada 500 nasabah, untuk mengetahui respon nasabah terhadap sistem layanan baru apakah mendapat respon yang baik dari seluruh nasabah atau sebaliknya.

Dari uraian di atas jelas bahwa di dalam metoda survei, kita hanya mengambil sebagian kecil dari unit-unit di dalam populasi untuk diteliti. Selanjutnya dari penelitian sampel tersebut kita gunakan untuk menduga (estimasi) nilai karakteristik populasi yang diteliti. Akibat hanya sebagian unit dalam populasi yang diteliti, maka jelas bahwa survei akan lebih menghemat tenaga, waktu dan biaya dibandingkan dengan sensus.

Beberapa hal yang menyebabkan dilakukan pengambilan sampel di dalam penelitian (proses pengumpulan data) adalah:

- a. Pertimbangan praktis, karena terbatasnya biaya, tenaga dan waktu.
- b. Seringkali tidak mungkin mengamati seluruh anggota populasi, karena akan merusak atau bahkan tidak akurat.
- c. Manajemen proyek lebih gampang, karena pengawasan dan perbaikan akan menjadi lebih mudah.

Tabel 1. Perbandingan Survei Sampel dan Sensus

Segi	Survei Sampel	Sensus
Tenaga	<ul style="list-style-type: none"> ▪ Jumlah relatif sedikit ▪ Dapat dipilih yang berkualitas 	<ul style="list-style-type: none"> ▪ Jumlah sangat besar ▪ Lebih sulit untuk memilih yang berkualitas seluruhnya
Waktu	<ul style="list-style-type: none"> ▪ Lebih cepat 	<ul style="list-style-type: none"> ▪ Lebih lama
Biaya	<ul style="list-style-type: none"> ▪ Lebih murah 	<ul style="list-style-type: none"> ▪ Lebih mahal
Kedalaman dan kualitas data	<ul style="list-style-type: none"> ▪ Biasanya kualitas data lebih baik ▪ Pertanyaan yang lebih sulit bisa dipergunakan 	<ul style="list-style-type: none"> ▪ Kualitas data kurang baik, hal ini akibat dari kualitas tenaga pengumpul ▪ Pertanyaan sederhana
Penyajian Data	<ul style="list-style-type: none"> ▪ Data tidak bisa disajikan sampai ke tingkat yang paling rendah 	<ul style="list-style-type: none"> ▪ Data bisa disajikan sampai ke tingkat yang paling rendah, karena semua unit dalam populasi dikumpulkan
Kesalahan (Error)	<ul style="list-style-type: none"> ▪ Adanya kesalahan sampel ▪ Adanya kesalahan bukan dari sampel, namun relatif kecil 	<ul style="list-style-type: none"> ▪ Tidak ada kesalahan sampel ▪ Adanya kesalahan bukan dari sampel yang besar

2.6 Perbandingan Survei dan Sensus

Perbandingan survei sampel dengan sensus dapat dilihat dari beberapa segi yaitu antara lain segi tenaga yang dipergunakan, waktu, biaya, kedalaman dan kualitas data yang dikumpulkan serta penyajian datanya. Lebih Jelasnya dapat dilihat pada Tabel 1 di atas.

2.7 Proses Sampling

Dalam melakukan proses sampling perlu dilakukan tahapan-tahapan sebagai berikut:

a. *Defining population*

Mendefinisikan populasi merupakan proses awal yang sangat penting dari suatu pengumpulan data. Dalam tahapan ini, perlu dipahami apa yang menjadi tujuan penelitian sehingga dapat menentukan apa atau siapa yang menjadi target populasi.

b. *Developing sampling frame*

Ketersediaan kerangka sampel merupakan hal yang harus dipenuhi ketika menggunakan *probability sampling*, baik diperoleh melalui hasil *listing* sendiri maupun dari informasi yang telah ada. Apabila kerangka sampel sulit diperoleh maka dapat menggunakan *non-probability sampling*.

c. *Specifying Sampling Method*

Menentukan Metode Penarikan Sampel yang akan digunakan harus didasari oleh pengetahuan peneliti mengenai karakteristik dari populasi yang ingin ditelitinya.

d. *Determining Sample Size*

Setelah mengetahui Metode Penarikan Sampel yang akan digunakan, maka tahapan selanjutnya adalah menentukan berapa besar sampel yang harus diambil. Yang harus diingat adalah dalam menentukan ukuran sampel perlu diperhatikan faktor karakteristik, waktu, tenaga dan biaya serta keakuratan yang diinginkan.

e. *Selecting Sample*

Yang terakhir adalah bagaimana memilih sampelnya. Apakah menggunakan *with replacement* atau *without replacement*.

2.8 Latihan

1. Sebutkan jenis-jenis data dan jelaskan tehnik pengumpulan data apa saja yang bisa diterapkan pada jenis data tersebut.
2. Jelaskan secara singkat mengenai teknik pengumpulan data sensus dan survei serta sebutkan keuntungan dari kedua tehnik tersebut
3. Sebutkan dan jelaskan tahapan dalam proses sampling.

2.9 Rangkuman

Data dan informasi statistik yang didapatkan dari hasil survei berguna untuk melihat perubahan-perubahan, menganalisa dan akhirnya sebagai dasar menentukan

kebijakan-kebijakan. Penelitian dengan menggunakan metode survei yang menghasilkan informasi dari data sampel memberikan beberapa keuntungan antara lain yaitu efisiensi waktu, biaya dan tenaga kerja dibandingkan dengan pengumpulan data dengan metode sensus.

Bab III Dasar-Dasar Teori Penarikan Sampel

3.1 Teknik Pengambilan Sampel

Ada dua teknik pengambilan sampel yaitu *non-probability sampling* dan *probability sampling*.

a. *Non-probability sampling*

Adalah suatu prosedur pengambilan sampel yang tidak memperhatikan kaidah-kaidah peluang (*probability*). Biasanya tergantung pada kebijakan dan pengalaman serta subyektifitas dari si peneliti. *Bias* dan *sampling error* pengambilan sampel ini tidak dapat ditentukan berdasarkan sampel yang terpilih, sehingga kurang dapat dipertanggungjawabkan untuk analisis secara statistik.

Jenis pengambilan sampel *non-probability sampling*:

Convenience Sampling. Pengambilan sampel dengan cara ini tidak mewakili secara normal dari target populasi karena unit sampel hanya dipilih berdasarkan *conveniently/readily available*. Contoh: Seorang peneliti ingin mengetahui kadar bakteri yang terdapat dalam 1 liter air di sungai Ciliwing. Maka yang diambil sebagai sampel hanya air yang ada di pinggir sungai saja.

Judgement/Purposive sampling. Pendekatan ini digunakan saat sampel yang diambil berdasarkan pada penilaian yang pasti (*expert judgement*) mengenai populasi secara keseluruhan (harus mempunyai pengetahuan yang cukup mengenai populasi). Contoh: Seorang peneliti ingin mengetahui kadar bakteri yang terdapat dalam 1 liter air di sungai Ciliwing. Maka yang diambil sebagai sampel tidak hanya air yang ada di pinggir sungai saja melainkan beberapa bagian sungai dengan pertimbangan keterwakilan..

Quota sampling yaitu pengambilan sampel dimana jumlah sampel telah ditentukan terlebih dahulu. Pengambil sampel tinggal memilih sampai jumlah tersebut dan biasanya tanpa kerangka sampel. Pengambilan sampel semacam ini sering digunakan dalam *public opinion survey*. Contoh: Populasi 55% pria dan 45% wanita. Dari sampel 100 orang berarti 55 pria dan 45 wanita. Pemilihan sampelnya sendiri tergantung penilaian peneliti.

Snowball sampling. Biasanya digunakan untuk *hidden population*. Responden diminta memberikan nama dan kontak dari anggota lain dari target populasi. Asumsinya bahwa sesama anggota saling mengenal Contohnya: GAY, pengguna narkoba dan *hackers*.

b. *Probability Sampling*

Adalah suatu prosedur pengambilan sampel yang memperhatikan kaidah-kaidah peluang (*probability*), sehingga *bias* dan *sampling error* pengambilan sampel ini dapat ditentukan berdasarkan sampel yang terpilih. Dalam bab 1 telah disebutkan bahwa hanya sebagian kecil dari unit di dalam populasi yang akan diteliti di dalam survei sampel. Oleh karena itu hasil *sampling* hanya bisa untuk menduga nilai populasinya (*parameter*).

Ada beberapa macam *probability sampling*, antara lain:

- Sampel Acak Sederhana (*Simple Random Sampling*), yaitu bila setiap unit dalam populasi diberi peluang sama untuk terpilih. Metode ini merupakan metode yang cukup mudah dan biasa digunakan pada populasi yang memuat karakteristik unit (unit) bersifat relatif homogen.

- Sistematik Sampling (*Systematic Sampling*), yaitu suatu metode pengambilan sampel secara random untuk unit sampel yang pertama dan unit-unit sampel selanjutnya dipilih secara sistematis.
- Sampel Acak Berlapis (*Stratified Random Sampling*), yaitu metode pemilihan sampel dimana berdasarkan suatu informasi (data) unit-unit di dalam populasi dibuat stratifikasi. Diusahakan nilai-nilai unit di dalam suatu kelompok cukup homogen, sedangkan antar lapisan heterogen. Kemudian dari setiap lapisan yang dibentuk, dipilih sejumlah sampel secara random.
- Sampel Acak Berkelompok (*Cluster Sampling*), yaitu prosedur sampling di mana unit terkecil dalam populasi merupakan kumpulan dari elemen-elemen. Di dalam *cluster* biasanya heterogen namun antar *cluster* homogen. Kemudian kita memilih sebuah sampel yang anggotanya adalah *cluster-cluster* sehingga bukan lagi sebuah sampel yang anggotanya adalah unit-unit analisis terkecil.

Bahasan selanjutnya pada modul ini hanya dibatasi pada pengambilan sampel berpeluang, yaitu: Sampel Acak Sederhana, Sistematik Sampling dan Sampel Acak Berlapis saja.

3.2 Kemungkinan Sampel

Kita mempunyai banyak pilihan kumpulan unit yang bisa diambil, karena hanya sebagian yang akan kita pilih dari unit yang ada dalam populasi. Tiap kumpulan unit yang mungkin akan terambil sebagai sampel yang menghasilkan nilai pendugaan yang berbeda. Oleh karena itu, apabila kita melakukan pengambilan sampel, harus dicari suatu cara untuk dapat mengukur tingkat kecermatan dari penduga yang dihasilkan. Apabila nilai penduga mempunyai kemungkinan cukup besar nilainya akan mendekati nilai populasi, maka tentunya hasil pengambilan sampel yang kita lakukan dapat dikatakan cukup baik, dan kurang baik apabila terjadi sebaliknya.

Permasalahannya adalah bagaimana kita dapat melakukan pengambilan sampel tersebut, sehingga kita bisa memperkirakan tingkat kecermatannya. Cara yang bisa digunakan adalah dengan menggunakan kaidah-kaidah peluang untuk penarikan unit ke dalam sampel sehingga aspek keacakan dapat terpenuhi.

Pada saat kita mempergunakan metode sampel berpeluang, maka setiap kita menarik unit sebagai anggota sampel, kita tidak mengetahui lebih dahulu unit mana yang akan terpilih. Sebagai gambaran seandainya kita mempunyai 4 unit di dalam populasi, misalnya A, B, C dan D, maka apabila kita gunakan peluang yang sama untuk menarik unit-unit tersebut, masing-masing akan mempunyai peluang yang sama untuk terpilih yaitu $\frac{1}{4}$.

Bila kita memilih 2 unit sebagai sampel, dan setiap unit dapat terpilih lebih dari sekali, kita dapatkan kemungkinan sampel sebagai berikut:

AA	BA	CA	DA
AB	BB	CB	DB
AC	BC	CC	DC
AD	BD	CD	DD

Jadi seandainya pada penarikan pertama (setelah diundi), kita dapatkan unit C, maka pada penarikan sampel kedua bisa kita dapatkan unit A atau B atau C atau D. Sehingga setiap unit yang terpilih akan mempunyai 4 pasangan yang mungkin. Pasangan-pasangan yang mungkin terpilih dinamakan kemungkinan sampel. Dalam gambaran di atas dinamakan pengambilan sampel dengan ulangan (*With Replacement*), karena setiap unit bisa terpilih lebih dari sekali. Sehingga banyaknya kemungkinan sampelnya (*all possible sample*) adalah sebesar N^n .

Seandainya cara penarikan unit tersebut kita ubah, yaitu unit yang sudah terpilih tidak boleh dipilih lagi pada pemilihan selanjutnya dan sampel AB dan BA kita anggap sama, maka kemungkinan sampelnya menjadi AB, AC, AD, BC, BD & CD. Jadi kita mempunyai 6 kemungkinan contoh. Cara penarikan semacam ini dinamakan penarikan sampel tanpa ulangan (*Without Replacement*). Sehingga banyaknya kemungkinan sampelnya (*all possible sample*) adalah sebesar ${}^N C_n$.

3.3 Penyimpangan Nilai Dugaan Dari Nilai Populasi

Karena unit yang diteliti hanya sebagian kecil dari populasi maka dengan sendirinya nilai penduga (*estimator*) tidak harus sama dengan nilai populasinya. Sebagai ilustrasi, seandainya nilai dari masing-masing unit adalah:

$$A = 4 \quad B = 1 \quad C = 2 \quad D = 3$$

dan kita mengambil 2 unit tanpa ulangan, jika AC terpilih sebagai sampel, maka:

- rata-rata sampel (penduga rata-rata populasi) adalah:

$$\bar{y} = \frac{4+2}{2} = 3,0$$

- sedangkan rata-rata populasinya adalah:

$$\bar{Y} = \frac{4+1+2+3}{4} = \frac{10}{4} = 2,5$$

Terlihat bahwa nilai dugaan di atas tidak sama dengan nilai populasinya. Nilai-nilai dugaan untuk masing-masing kemungkinan sampel adalah sebagai berikut:

Kemungkinan Sampel	Nilai \bar{y}	$\bar{y} - \bar{Y}$	$(\bar{y} - \bar{Y})^2$
AB	$\frac{4+1}{2} = 2,5$	0	0
AC	$\frac{4+2}{2} = 3,0$	0,5	0,25
AD	$\frac{4+3}{2} = 3,5$	1,0	1,0
BC	$\frac{1+2}{2} = 1,5$	-1,0	1,0
BD	$\frac{1+3}{2} = 2,0$	-0,5	0,25
CD	$\frac{2+3}{2} = 2,5$	0	0

Dari gambaran di atas terlihat bahwa hasil dugaan \bar{y} dari kemungkinan sampel AC, AD, BC atau AD nilainya berbeda dengan rata-rata populasi \bar{Y} . Bagaimana mengukur besar kecilnya kemungkinan penyimpangan nilai \bar{y} dari \bar{Y} ? Caranya adalah dengan menghitung variansnya. Dalam ilustrasi di atas variansnya adalah:

$$V(\bar{y}) = \frac{1}{K} \sum_{i=1}^K (\bar{y} - \bar{Y})^2$$

$$= 1/6 (0 + 0,25 + 1,0 + 1,0 + 0,25 + 0) = 1/6 (2,5) = 0,417$$

Nilai varians yaitu $V(\bar{y})$ dinyatakan sebagai rata-rata nilai $(\bar{y} - \bar{Y})^2$ untuk seluruh kemungkinan sampel. Selanjutnya perhatikan ilustrasi di atas bila pengambilan unit dengan ulangan.

Kemungkinan Sampel	Nilai \bar{y}	$\bar{y} - \bar{Y}$	$(\bar{y} - \bar{Y})^2$
AA	4,0	1,5	2,25
AB	2,5	0,0	0,00
AC	3,0	0,5	0,25
AD	3,5	1,0	1,00
BA	2,5	0,0	0,00
BB	1,0	-1,5	2,25
BC	1,5	-1,0	1,00
BD	2,0	-0,5	0,25
CA	3,0	0,5	0,25
CB	1,5	-1,0	1,00
CC	2,0	-0,5	0,25
CD	2,5	0,0	0,00
DA	3,5	1,0	1,00
DB	2,0	-0,5	0,25
DC	2,5	0,0	0,00
DD	3,0	0,5	0,25
Jumlah	40,0	0,0	10,00

Sehingga nilai varians, bila pengambilan unit dilakukan dengan ulangan adalah

$$V(\bar{y}) = \frac{1}{K} \sum_{i=1}^K (\bar{y} - \bar{Y})^2 = \frac{10}{16} = 0,625$$

Apabila kita perhatikan perbedaan nilai-nilai $\bar{y} - \bar{Y}$ pada penarikan sampel tanpa ulangan berkisar antara $-1,0$ s/d $+1,0$, sedangkan dalam penarikan sampel dengan ulangan berkisar antara $-1,5$ s/d $+1,5$, sehingga kita dapat menyatakan bahwa hasil dugaan menggunakan penarikan sampel tanpa ulangan mempunyai peluang yang lebih besar mendekati nilai populasinya dibandingkan dengan penarikan sampel dengan ulangan. Hal tersebut tergambar juga dengan besarnya nilai varians, dimana nilai varians untuk pengambilan sampel dengan ulangan lebih besar daripada nilai varians dalam pengambilan sampel tanpa ulangan. Akhirnya dapat ditarik kesimpulan bahwa makin kecil nilai varians, maka hasil dugaan dari sampel akan makin mendekati nilai populasinya.

Kemungkinan kesalahan nilai dugaan dari nilai populasinya dinamakan penyimpangan sampel atau *sampling error* (*SE*). Ukuran relatif besarnya kesalahan tersebut dinyatakan oleh :

$$SE(\bar{y}) = \sqrt{V(\bar{y})}$$

Dalam ilustrasi di atas nilai $SE(\bar{y})$ untuk penarikan sampel dengan ulangan adalah $\sqrt{0,625}$. Dan untuk penarikan sampel tanpa ulangan adalah $\sqrt{0,417}$

3.4 Kerangka Sampel

Keseluruhan unit dalam populasi akan membentuk kerangka sampel dan dari sinilah anggota sampel dipilih. Kerangka sampel bisa merupakan daftar dari orang, rumah tangga, perusahaan, catatan dalam sebuah file, kumpulan dokumen, atau berupa sebuah peta dimana telah tergambar unitnya secara jelas. Untuk bisa melakukan penarikan sampel secara acak, kita memerlukan kerangka sampel berupa daftar dari unit berikut keterangan tentang nama, alamat (identifikasi) dan keterangan-keterangan lain yang diperlukan. Persyaratan yang harus dipenuhi kerangka sampel adalah:

- Memiliki batas yang jelas, artinya setiap unit tidak saling tumpang tindih dengan unit lain.
- Lengkap dan *up to date*, artinya seluruh unit dalam populasi dalam keadaan terakhir harus didaftar.
- Dapat dikenali, artinya seluruh unit di dalam kerangka sampel dapat dikenal kembali melalui alamat atau petanya.

Jadi bila suatu penarikan sampel dilakukan dalam survei perbankan dengan responden adalah bank, maka kita harus mempunyai kerangka sampel berupa daftar seluruh bank yang ada serta keterangan yang diperlukan dalam wilayah penelitian menurut keadaan terakhir. Bank yang sudah tutup (dilikuidasi) harus dikeluarkan dari kerangka sampel, sedangkan bank yang baru harus dimasukkan ke dalam kerangka sampel lengkap dengan keterangan-keterangan yang diperlukan.

Apabila kerangka sampel belum tersedia dalam proses pemilihan unit sampel, maka sebagai kerangka sampel kita perlu mempersiapkan terlebih dahulu melalui data hasil pendaftaran secara lengkap (sensus) atau kalau data hasil sensus tidak tersedia bisa kita lakukan listing berupa pendaftaran secara lengkap terhadap unit-unit populasi yang akan dipilih sebagai sampel. Sebagai contoh pada suatu wilayah, sebuah survei akan dilakukan dengan responden rumah tangga dimana minimal salah satu anggota rumah tangganya menjadi nasabah sebuah bank atau sebut saja "*rumah tangga nasabah bank*". Seandainya belum tersedia daftar rumah tangga nasabah bank yang merupakan kerangka sampel, penyelenggara survei bisa melakukan pendaftaran (listing) terhadap seluruh rumah tangga di wilayah tersebut sehingga akan diperoleh daftar rumah tangga nasabah bank yang selanjutnya dapat digunakan sebagai dasar pengambilan sampel.

3.5 Pemilihan Sampel Secara Acak (Menggunakan TAR)

Untuk mempermudah penarikan sampel secara acak, bisa kita gunakan komputer, kalkulator atau tabel angka random (TAR). Penggunaan komputer untuk mendapatkan angka acak biasanya sudah tersedia paket programnya. Pada kalkulator yang lengkap, biasanya bisa digunakan untuk mendapatkan angka acak. Apabila tidak tersedia kedua-

duanya, maka cara mendapatkan angka acak adalah dengan menggunakan TAR. Contoh dari tabel angka random dapat dilihat pada lampiran.

Sebagai gambaran cara penggunaan TAR adalah sebagai berikut. Seandainya kita memilih sampel sebanyak $n = 10$ unit dari $N = 80$ unit dalam populasi, karena $N = 80$ unit terdiri dari 2 digit, maka yang kita lakukan adalah:

- kita pilih secara acak halaman TAR yang akan digunakan, misalnya halaman 1
- Pilihlah 2 kolom yang berdekatan secara random, misalnya kolom 3 dan 4
- Pilihlah baris sebagai titik mulai penarikan sampel secara random pula, misalnya baris ke-10.

Sehingga angka acak pertama yang berada di kolom 3-4, baris ke-10 adalah 60. Angka acak terpilih apabila angka acak tersebut lebih kecil dari N . Karena $60 < 80$, maka merupakan angka acak terpilih yang pertama. Angka acak terpilih berikutnya dilakukan dengan pembacaan angka acak dari atas ke bawah tetap pada kolom 3 dan 4. Sehingga apabila pemilihan unit tanpa ulangan angka acak terpilih selanjutnya adalah 18, 62, 42, 36, 29, 49, 08, 16 dan 34. Seandainya waktu penarikan angka acak tersebut sampai baris terakhir (baris 35) belum cukup memenuhi kebutuhan sampel, maka pindahlah ke kolom-kolom berikutnya dan mulailah dari baris pertama. Dalam contoh di atas bila pada kolom 3-4 dan baris 35 belum memenuhi 10 unit sampel, maka pindahlah ke kolom 5-6, baris pertama dan pilihlah angka acak seperti cara sebelumnya.

Setelah angka acak yang diperlukan sudah terpilih, maka unit-unit dalam populasi dengan nomor-nomor urut sesuai dengan angka acak yang terpilih akan dimasukkan sebagai anggota sampel.

Tabel 2. TABEL ANGKA RANDOM

Halaman 01

NO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	3	4	7	3	8	6	3	6	9	6	4	7	3	6	6	1	4	6	9	8	6	3	7	1
2	9	7	7	4	2	4	6	7	6	2	4	2	8	1	1	4	5	7	2	0	4	2	5	3	3
3	1	6	7	6	6	2	2	7	6	6	5	6	5	0	2	6	7	1	0	7	3	2	9	0	7
4	1	2	5	6	8	5	9	9	2	6	9	6	9	6	6	8	2	7	3	1	0	5	0	3	7
5	5	5	5	9	5	6	3	5	6	4	3	8	5	4	8	2	4	6	2	2	3	1	6	2	4
6	1	6	2	2	7	7	9	4	3	9	4	9	5	4	4	3	5	4	8	2	1	7	3	7	9
7	8	4	4	2	1	7	5	3	3	1	5	7	2	4	5	5	0	6	8	8	7	7	0	4	7
8	6	3	0	1	6	3	7	8	5	9	1	6	9	5	5	5	6	7	1	9	9	8	1	0	5
9	3	3	2	1	1	2	3	4	2	9	7	8	6	4	5	6	0	7	8	2	5	2	4	2	0
10	5	7	6	0	8	6	3	2	4	4	0	9	4	7	2	7	9	6	5	4	4	9	1	7	4
11	1	8	1	8	0	7	9	2	4	6	4	4	1	7	1	6	5	8	0	9	7	9	8	3	8
12	2	6	6	2	3	8	9	7	7	5	8	4	1	6	0	7	4	4	9	9	8	3	1	1	4
13	2	3	4	2	4	0	6	4	7	4	8	2	9	7	7	7	7	7	8	1	0	7	4	5	3
14	6	2	3	6	2	8	1	9	9	5	5	0	9	2	2	6	1	1	9	7	0	0	5	6	7
15	3	7	8	5	9	4	3	5	1	2	8	3	3	9	5	0	0	8	3	0	4	2	3	4	0
16	7	0	2	9	1	7	1	2	1	3	4	0	3	3	2	0	3	8	2	6	1	3	8	9	5
17	5	6	6	2	1	8	3	7	3	5	9	6	8	3	5	0	8	7	7	5	9	7	1	2	2
18	9	9	4	9	5	7	2	2	7	7	8	8	4	2	9	5	4	5	7	2	1	6	6	4	3
19	1	6	0	8	1	5	0	4	7	2	3	3	2	7	1	4	3	4	0	9	4	5	5	9	3
20	3	1	1	6	9	3	3	2	4	3	5	0	2	7	8	9	8	7	1	9	2	0	1	5	3
21	6	8	3	4	3	9	1	3	7	0	5	5	7	4	3	0	7	7	4	0	4	4	2	2	7
22	7	4	5	7	2	5	6	5	7	6	5	9	2	9	9	7	6	8	6	0	7	1	9	1	3
23	2	7	4	2	3	7	8	6	5	3	4	8	5	5	9	0	6	9	7	2	9	6	5	7	6
24	0	0	3	9	6	8	2	9	6	1	6	6	3	7	3	2	2	0	3	0	7	7	8	4	5
25	2	9	9	4	9	8	9	4	2	4	6	8	4	9	6	9	1	0	8	2	5	3	7	9	9
26	1	6	9	0	8	3	6	6	5	9	8	3	6	2	6	4	1	1	1	2	6	7	1	9	0
27	1	1	2	7	9	4	7	5	0	6	0	6	0	9	1	9	7	4	6	6	0	2	9	4	3
28	3	5	2	4	1	0	1	6	2	0	3	3	3	2	5	1	2	6	3	8	7	9	7	8	4
29	3	8	2	3	1	6	8	6	3	8	4	2	3	8	9	7	0	1	5	0	8	7	7	5	6
30	3	1	9	6	2	5	9	1	4	7	9	6	4	4	3	3	4	9	1	3	3	4	8	6	8
31	6	6	6	7	4	0	6	7	1	4	6	4	0	5	7	1	9	5	8	6	1	1	0	5	6
32	1	4	9	0	8	4	4	5	1	1	7	5	7	3	8	8	0	5	9	0	5	2	2	7	4
33	6	8	0	5	5	1	1	8	0	0	3	3	9	6	0	2	7	5	1	9	0	7	6	0	6
34	2	0	4	6	7	8	7	3	9	0	9	7	5	1	4	0	1	4	0	2	0	4	0	2	3
35	6	4	1	9	5	8	9	7	7	9	1	5	0	6	1	5	9	3	2	0	0	1	9	0	1

Contoh SOAL:

Seorang peneliti ingin mengetahui seberapa besar respon konsumen terhadap produk baru yang ditawarkan oleh perusahaan A. Berdasarkan data yang ada, diketahui terdapat 100 konsumen yang terdaftar dalam member card. Apabila peneliti itu ingin mengambil sampel sebanyak 10 konsumen, nomor urut berapa saja yang terpilih sebagai sampel, jika pengambilannya dimulai dari halaman 1 baris 5 kolom 5?

JAWAB:

$$N = 100$$

$$n = 10$$

Karena N terdiri dari 3 digit, maka angka random yang diambil adalah 3 kolom. Jatuhkan pensil secara acak ke Tabel Angka Random. Misalkan jatuh pada baris 5 kolom 5 maka angka random akan diambil dari kolom 5, 6 dan 7 pada baris ke 5. Kemudian angka random yang diambil adalah yang kurang dari atau sama dengan 100. Nomor urut sampel yang terpilih adalah:

$R_1 = 79$	$R_6 = 78$
$R_2 = 94$	$R_7 = 83$
$R_3 = 60$	$R_8 = 15$
$R_4 = 26$	$R_9 = 59$
$R_5 = 68$	$R_{10} = 42$

3.6 Latihan

1. Dalam banyak penelitian, survei lebih dipilih dibandingkan dengan sensus. Jelaskan mengapa demikian!
2. Setiap survei pasti akan menghasilkan kesalahan sampling dan kesalahan non sampling. Jelaskan hubungan kedua kesalahan tersebut dengan besaran sampel (n) yang digunakan
3. Jika diketahui suatu populasi berukuran $N=10$ yaitu A, B, C, D, E, F, G, H, I dan J dengan nilai variabel y berturut-turut adalah 3, 5, 1, 6, 3, 7, 9, 10, 4 dan 8. Sampel berukuran $n=4$ dipilih secara acak sederhana tanpa pengembalian (SRS-WOR) dari populasi tersebut.
 - a. Pilih keempat sampel tersebut dengan menggunakan Tabel Angka Random halaman 1 baris 3 kolom 3.
 - b. Hitung penduga rata-rata populasinya (\bar{y}) dan penduga standar errornya $SE(\bar{y})$.

3.7 Rangkuman

Teknik pengambilan sampel yang menggunakan kaidah-kaidah probabilitas disebut *probability sampling* dan sebaliknya disebut *non-probability sampling*. Teknik sampling yang termasuk dalam *probability sampling* adalah sampling acak sederhana, sistematis sampling, sampel acak berlapis dan sampel acak berkelompok.

Pengambilan sampel secara acak sederhana menggunakan tabel angka random baik dengan pengembalian (*with replacement*) maupun tanpa pengembalian (*without replacement*).

Sebelum melakukan penarikan sampel secara acak diperlukan sebuah kerangka sampel (*sampling frame*). Selain berisi unit populasi yang akan diteliti keadaan terakhir dengan batasan yang jelas, juga memuat keterangan-keterangan yang relevan dengan fokus penelitian sebagai dasar pengelompokan atau pembuatan strata.

Dengan menggunakan *probability sampling* maka dapat dihitung standar error dari penelitian yang dilakukan.

Bab IV. Sampel Acak Sederhana

4.1 Definisi Sampel Acak Sederhana

Suatu sampel dinamakan sampel acak sederhana (*simple random sampling*) bila setiap unit dalam populasi diberi peluang sama untuk terpilih. Metode ini merupakan metode yang cukup mudah dan biasa digunakan pada populasi yang memuat karakteristik unit (unit) bersifat relatif homogen. Bila kita mempunyai populasi dengan $N = 500$ unit, maka setiap unit diberikan peluang $1/500$ untuk dapat terpilih pertama. Ada dua metode penarikan sampel acak sederhana yaitu sampel acak sederhana dengan *ulangan* (*Simple Random Sampling With Replacement –SRSWR*) dan sampel acak sederhana tanpa *ulangan* (*Simple Random Sampling Without Replacement –SRSWOR*). Dalam SRSWR setiap unit dalam populasi dapat dipilih lebih dari sekali dalam sampel, sedangkan dalam SRSWOR hanya boleh terpilih sekali saja. Pada praktik di lapangan SRSWOR lebih sering digunakan daripada SRSWR.

Bila suatu survei akan dilakukan pada populasi yang terdiri dari N unit dan akan dipilih sampel sebanyak n unit secara SRS, maka prosedurnya adalah kita memilih angka random (AR) sejumlah n dengan syarat $AR \leq N$, maka unit-unit populasi yang terdapat di dalam kerangka sampel dengan nomor urut sesuai dengan AR terpilih merupakan sampel terpilih. Sebagai ilustrasi jika jumlah unit di dalam populasi yaitu $N=1000$ unit sedangkan jumlah sampel yang akan dipilih yaitu $n=100$, seandainya angka random terpilihnya adalah:

$AR_1 = 145$	$AR_2 = 056$	$AR_3 = 675$	$AR_4 = 324$
$AR_5 = 801$	$AR_6 = 287$	$AR_7 = 004$	$AR_8 = 098$
⋮	⋮	⋮	⋮
$AR_{97} = 989$	$AR_{98} = 451$	$AR_{99} = 777$	$AR_{100} = 610$

maka sampel yang terpilih adalah unit-unit yang terdapat di dalam kerangka sampel yang mempunyai nomor urut 145, 056, 675, 324, 801, 287, 004, 098, ..., 989, 451, 777 dan 610

4.2 Metode Penduga Rata-Rata dan Varians

Di dalam bab sebelumnya telah dijelaskan bahwa hasil suatu survei sampel digunakan untuk menduga nilai populasinya. Beberapa notasi yang perlu diperhatikan selanjutnya adalah:

N	= Jumlah unit dalam populasi
n	= banyaknya unit sampel yang ditarik/dipilih
Y	= Nilai total suatu karakteristik dalam populasi
\bar{Y}	= Rata-rata nilai suatu karakteristik dalam populasi
\hat{Y}	= Penduga nilai total populasi berdasarkan sampel
\bar{y}	= Penduga nilai rata-rata populasi (\bar{Y}) berdasarkan sampel
y_i	= Nilai suatu karakteristik pada unit ke- i ($i = 1, 2, 3, \dots, N$) pada populasi dan $i = 1, 2, 3, \dots, n$ pada sampel.

$V(\bar{y})$

\bar{y}

$v(\bar{y})$ = Variance nilai-nilai bila nilai populasi diketahui
 $v(\bar{y})$ = Penduga varians bila nilai populasi tidak diketahui

Penduga rata-rata nilai populasi dalam SRSWR maupun SRSWOR adalah:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Penduga rata-rata tersebut adalah suatu variabel yang nilainya tergantung pada kemungkinan sampel yang terpilih, maka tingkat pencaran nilai-nilai penduga rata-rata tersebut diukur dengan besarnya nilai varians [$V(\bar{y})$] atau standard error [$SE(\bar{y})$].

Karena dalam survei sampel nilai populasinya tidak diketahui, maka besarnya varians tersebut kita duga dengan penduganya yaitu $v(\bar{y})$ dan dihitung dari nilai-nilai unit sampelnya.

Di dalam SRSWR

$$V(\bar{y}) = \frac{\dagger^2}{n}; \rightarrow \dagger^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N}$$

$$v(\bar{y}) = \frac{s^2}{n}; \rightarrow s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

sedangkan dalam SRSWOR adalah:

$$V(\bar{y}) = \frac{N-n}{N} \frac{S^2}{n}; \rightarrow S^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}$$

$$v(\bar{y}) = \frac{N-n}{N} \frac{s^2}{n}$$

Untuk melakukan pendugaan nilai total \hat{Y} adalah $\hat{Y} = N\bar{y}$

Dengan penduga variansnya adalah $v(\hat{Y}) = N^2 v(\bar{y})$.

Untuk menduga $SE(\bar{y})$ adalah $se(\bar{y}) = \sqrt{v(\bar{y})}$

Dan untuk $se(\hat{Y}) = \sqrt{v(\hat{Y})} = N\sqrt{v(\bar{y})}$.

Di dalam sampel acak sederhana, n/N dinamakan fraksi sampel, yaitu berapa bagian unit sampel yang ditarik dari seluruh unit dalam populasi, biasanya kalau nilainya kurang dari 5%, maka fraksi sampel diabaikan. Sedangkan N/n dinamakan *inflation factor* atau faktor pengali dalam estimasi populasi.

CONTOH: Suatu penelitian dilakukan untuk mengetahui rata-rata jumlah anggota rumah tangga di suatu wilayah. Sebuah sampel yang terdiri dari 10 rumah tangga dipilih dari populasi yang terdiri dari 100 rumah tangga dengan SRSWOR. Dari sampel tersebut diketahui bahwa jumlah anggota rumah tangga yang dimiliki adalah sebagai berikut: 1, 3, 2, 5, 4, 6, 3, 4, 2, 5. Berdasarkan data tersebut perkirakan rata-rata jumlah anggota rumah tangganya beserta *standar error*nya! Dan Berapa total penduduk yang ada di wilayah tersebut?

JAWAB:

$$N=100$$

$$n=10$$

$y_1=1, y_2=3, y_3=2, y_4=5, y_5=4, y_6=6, y_7=3, y_8=4, y_9=2$ dan $y_{10}=5$

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{1+3+\dots+5}{10} = \frac{35}{10} = 3,5$$

$$s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n-1} = \frac{(1-3,5)^2 + \dots + (5-3,5)^2}{10-1} = 2,5$$

$$v(\bar{y}) = \frac{N-n}{N} \frac{s^2}{n} = \frac{100-10}{100} \frac{2,5}{10} = 0,225$$

$$se(\bar{y}) = \sqrt{v(\bar{y})} = \sqrt{0,225} = 0,4743$$

$$\hat{Y} = N\bar{y} = 100(3,5) = 350$$

- ❖ Rata-rata jumlah anggota rumah tangga yang ada di wilayah tersebut diperkirakan sebesar 3,5 dengan *standar error* sebesar 0,4743. Dengan total penduduk sebanyak 350 orang.

4.3 Penentuan Ukuran Sampel

Ada beberapa rumus yang dapat digunakan untuk menentukan jumlah sampel yaitu:

1. Menggunakan *Margin Of Error* (MOE)

MOE adalah batas kesalahan maksimum yang masih bisa ditoleransi. MOE biasanya dinotasikan dengan huruf **d**, dan diperoleh dari rumus:

$$d = Z_{r/2} se(\bar{y})$$

Dari bentuk hubungan tersebut, kemudian dapat diturunkan menjadi rumus ukuran sampel yang diperlukan untuk memperkirakan nilai parameter sesuai dengan tingkat ketelitian yang dikehendaki. Rumus yang diperoleh dari penurunan tersebut adalah sebagai berikut:

$$n = \frac{N Z_{r/2}^2 S^2}{N d^2 + Z_{r/2}^2 S^2} \implies \text{WOR}$$

$$n = \frac{Z_{r/2}^2 S^2}{d^2} \implies \text{WR}$$

CONTOH: Untuk mengetahui rata-rata jumlah bensin yang digunakan untuk perjalanan dari Citayam ke Jakarta, seorang peneliti berencana mengambil sampel dari pengguna motor yang bertempat tinggal di Citayam. Berdasarkan data dari survey

sebelumnya, diketahui ada sebanyak 120 pengguna motor dengan varians penggunaan bensin per harinya sebesar 0,06 liter. Apabila peneliti ingin menggunakan α sebesar 5 persen dan membatasi kesalahan sampling yang masih bisa ditoleransi sebesar 10 persen, berapa ukuran sampel yang harus diambil?

JAWAB:

$$n = \frac{120 (1,96)^2 0,06}{120 (0,1)^2 + (1,96)^2 0,06} = \frac{27,6595}{1,4305} = 19,3356 \approx 20$$

2. Rumus Bernoulli

Salah satu rumus yang paling banyak digunakan adalah rumus *Bernoulli*, yaitu :

$$N \geq \frac{(Z_{\alpha/2})^2 pq}{e^2}$$

dimana : N = jumlah sampel
 $Z_{\alpha/2}$ = distribusi normal (Z) yang diperoleh dari tabel
 p = besarnya proporsi yang diterima
 q = besarnya proporsi yang ditolak
 e = besarnya kesalahan yang diizinkan

3. Rumus Slovin

Penentuan jumlah sampel dengan menggunakan rumus *slovin*, yaitu:

$$n \geq \frac{N}{1 + Nr^2}$$

dimana : n = jumlah sampel yang akan diambil
 N = jumlah populasi pengamatan
 e = tingkat ketelitian yang digunakan

Contoh:

Berapa ukuran sampel minimum yang harus diambil dari populasi yang berukuran

—a. 1000 dengan taraf signifikansi = 0,05

b. 45.250 dengan taraf signifikansi = 0,01

Jawab:

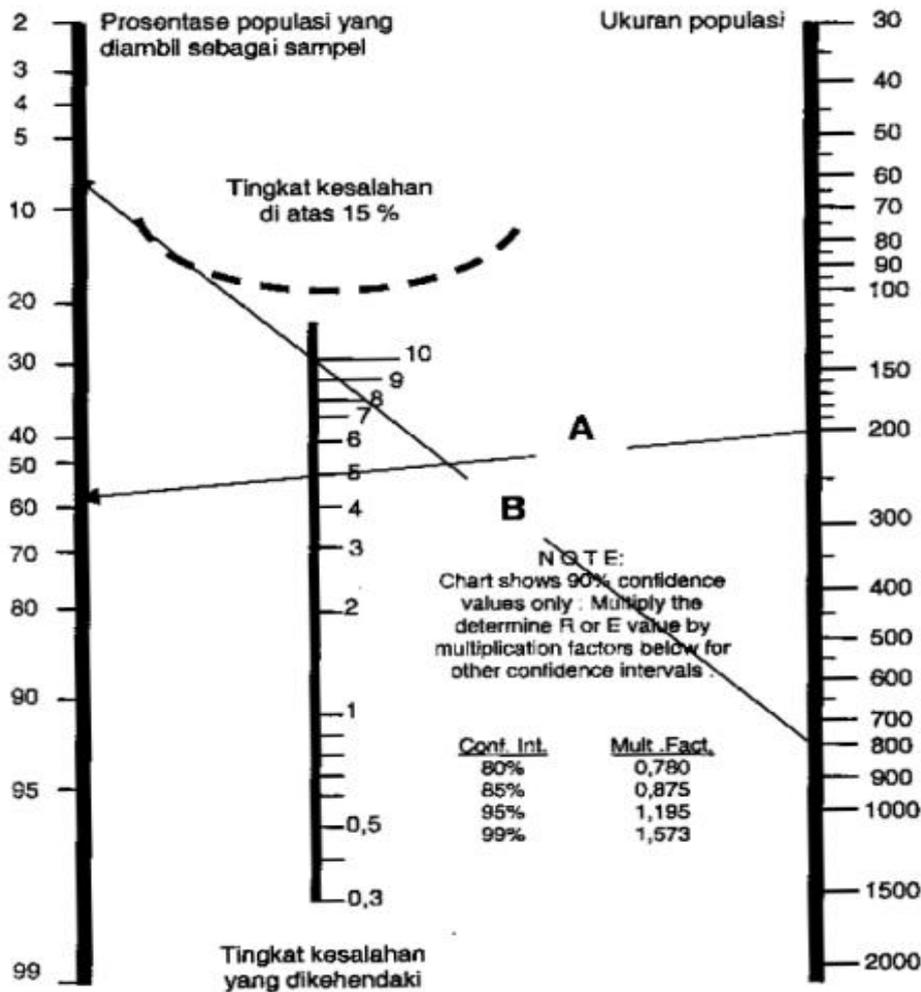
$$a. n = \frac{N}{1 + Nr^2} = \frac{1000}{1 + (1000) \times (0,05)^2} = 285,7143 \approx 286 \text{ (dibulatkan ke atas)}$$

$$b. n = \frac{N}{1 + Na^2} = \frac{45250}{1 + (45250) \times (0,01)^2} = 8.190,045 \approx 8.191 \text{ (dibulatkan ke atas)}$$

4. Nomogram Harry King

Nomogram Harry King dapat menentukan jumlah sampel n dari N populasi dengan tingkat kesalahan yang diinginkan. Adapun nomogram tersebut sebagaimana gambar dibawah ini.

Nomogram Harry King



Contoh:

Misal populasi berjumlah 200 dan tingkat kesalahan 5%, maka persentase populasi yang diambil sebagai sampel adalah 58% sehingga ukuran sampelnya adalah $200 \times 58\% \times 1,195 = 138,62$

4.4 Latihan

- Sebuah sample acak sederhana yang terdiri atas 30 rumah tangga dipilih dari sebuah kota yang berisi 960 rumah tangga. Jumlah anggota untuk setiap rumah tangga dalam sample adalah sebagai berikut:

5 6 3 3 2 3 3 3 4 4 3 2 7 4 3
5 4 4 3 3 4 3 3 1 2 4 3 4 2 4

Perkirakan rata-rata jumlah anggota rumah tangga beserta se-nya dan total penduduk di kota tersebut!

2. Dari 150 rumah tangga ingin diketahui rata-rata konsumsi perhari dari setiap rumah tangga. Menurut perkiraan rata-rata konsumsi perhari diharapkan menyimpang tidak lebih dari satu terhadap rata-rata populasinya. Nilai σ^2 tidak diketahui, namun berdasarkan survei pendahuluan diperoleh sekitar 9. Jika diinginkan tingkat kepercayaan 95%, Berapa ukuran sample yang harus diambil?

JAWAB:

1. $N = 960$
 $n = 30$

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{5 + 6 + \dots + 4}{30} = \frac{104}{30} = 3,4667$$

$$s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n-1} = \frac{(5 - 3,4667)^2 + \dots + (4 - 3,4667)^2}{30-1} = 1,4988$$

$$v(\bar{y}) = \frac{N-n}{N} \frac{s^2}{n} = \frac{960-30}{960} \frac{1,4988}{30} = 0,0484$$

$$se(\bar{y}) = \sqrt{v(\bar{y})} = \sqrt{0,0484} = 0,22$$

$$\hat{Y} = N\bar{y} = 960(3,4667) = 3328$$

- ❖ Rata-rata jumlah anggota rumah tangga yang ada di wilayah tersebut diperkirakan sebesar 3,4667 orang dengan *standar error* sebesar 0,22. Total jumlah penduduknya sebanyak 3.328 orang.

2. $N = 150$
 $S^2 = 9$
 $d = 1$
 $= 5 \%$
maka:

$$n = \frac{150 (1,96)^2 9}{120 (1)^2 + (1,96)^2 9} = \frac{5186,16}{154,5744} = 33,55 \approx 34$$

- ❖ Untuk mengetahui rata-rata konsumsi dari rumah tangga di wilayah tersebut dibutuhkan sebanyak 34 rumah tangga sampel.

4.5 Rangkuman

Suatu sampel dinamakan sampel acak sederhana (*simple random sampling*) bila setiap unit dalam populasi diberi peluang sama untuk terpilih.

Metode penarikan sampel acak sederhana dapat dilakukan dengan dua cara yaitu sampel acak sederhana dengan pengembalian (*Simple Random Sampling With Replacement – SRSWR*) dan sampel acak sederhana tanpa pengembalian (*Simple Random Sampling Without Replacement – SRSWOR*). Dalam SRSWR setiap unit dalam populasi memiliki peluang untuk dapat dipilih lebih dari sekali dalam sampel, sedangkan dalam SRSWOR hanya memiliki peluang terpilih sekali saja. Pada praktik di lapangan SRSWOR lebih sering digunakan daripada SRSWR.

Bab V. Sampel Sistematis (*Systematic Sampling*)

5.1 Definisi Sampel Sistematis

Metode pengambilan sampel sistematis adalah metode untuk mengambil sampel secara sistematis dengan interval (jarak) tertentu dari suatu kerangka sampel yang telah diurutkan. Misalkan sebuah populasi terdiri dari N unit diberi nomor 1 sampai N dalam beberapa susunan. Untuk memilih sebuah sampel berukuran n unit, kita ambil sebuah unit dari k unit yang pertama, selanjutnya mengambil setiap kelipatan k .

Keuntungan dari penggunaan sampel sistematis adalah lebih cepat dan mudah pelaksanaannya karena kita cukup mengambil satu angka random yang selanjutnya tinggal mengambil kelipatannya saja. Selain itu sampel sistematis tersebar lebih merata, sehingga kemungkinan besar menghasilkan sampel yang lebih representatif dan lebih efisien dari pada SRS.

Sebagai contoh mengambil sebuah sampel yang besarnya 5% dari suatu populasi yang terdiri dari 10.000 unit, maka sampel yang diambil adalah sebanyak 500 unit. Apabila kita menggunakan *simple random sampling* maka kita membutuhkan sebanyak 500 buah angka random dan memilih unit dalam populasi yang bernomor sesuai dengan angka random yang terpilih. Hal ini tentu sangat tidak praktis dalam penerapannya.

Beda halnya ketika kita menggunakan sampel sistematis, dengan sampel yang besarnya seperti di atas, kita cukup memilih sebuah angka random antara 1 dan 20, kemudian mengambil setiap angka yang selisihnya dengan angka sebelumnya adalah 20. Jadi, jika angka random yang terpilih adalah angka 5, maka unit-unit yang terpilih adalah unit-unit dengan nomor 5, 25, 45, 65 dan seterusnya sampai 1995. Dalam hal ini $N/n = 10.000/500 = 20$ disebut *sampling interval*. Pemilihan sampel semacam ini disebut *systematic sampling*. Walaupun tidak sama dengan *simple random sampling*, tetapi merupakan Metode Penarikan Sampel yang bisa diterima karena peluang bagi setiap unit diketahui dengan pasti dan kita bisa menghitung *sampling error*nya.

5.2 Metode Penduga Rata-Rata dan Varians

Untuk memperoleh penduga rata-rata dan varians dari sampel sistematis dapat didekati dengan menggunakan rumus yang ada pada *simple random sampling*. Penduga rata-rata nilai populasi dalam sampel sistematis adalah:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Dengan varians:

$$V(\bar{y}) \approx \frac{N-n}{N} \frac{S^2}{n}; \rightarrow S^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}$$

$$v(\bar{y}) \approx \frac{N-n}{N} \frac{s^2}{n}$$

$$\hat{Y} \qquad \hat{Y} = N\bar{y}$$

Untuk melakukan pendugaan nilai total adalah

Dengan penduga variansnya adalah $v(\hat{Y}) = N^2 v(\bar{y})$

Untuk menduga $SE(\bar{y})$ adalah $se(\bar{y}) = \sqrt{v(\bar{y})}$

Dan untuk $se(\hat{Y}) = \sqrt{v(\hat{Y})} = N \sqrt{v(\bar{y})}$.

CONTOH: Suatu penelitian dilakukan untuk mengetahui rata-rata mie instan yang dikonsumsi oleh mahasiswa kelas 2A dalam seminggu. Bila diketahui di dalam kelas tersebut terdapat 36 orang mahasiswa, kemudian diambil sebanyak 6 orang mahasiswa dengan menggunakan sampel sistematis. Tentukan mahasiswa nomor urut berapa saja yang terpilih sebagai sampel bila $AR_1=2$! Kemudian bila diketahui dari 6 orang mahasiswa tersebut masing-masing mengkonsumsi 3, 2, 5, 4, 6, 4 bungkus mie instan, perkirakan rata-rata mie instan yang dikonsumsi oleh mahasiswa kelas 2A dalam seminggu beserta *standar error*-nya!

JAWAB:

$$I = N/n = 36/6 = 6$$

Karena $AR_1 = 2$, maka:

$$R_1 = 2 \qquad R_3 = 2 + 12 = 14 \qquad R_5 = 2 + 24 = 26$$

$$R_2 = 2 + 6 = 8 \qquad R_4 = 2 + 18 = 20 \qquad R_6 = 2 + 30 = 32$$

Kemudian

$$N=36$$

$$n=6$$

$$y_1=3, y_2=2, y_3=5, y_4=4, y_5=6, y_6=4$$

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{3+2+\dots\dots+4}{6} = \frac{24}{6} = 4$$

$$s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n-1} = \frac{(3-4)^2 + \dots\dots + (4-4)^2}{6-1} = 2$$

$$v(\bar{y}) \approx \frac{N-n}{N} \frac{s^2}{n} \approx \frac{36-6}{36} \frac{2}{6} \approx 0.278$$

$$se(\bar{y}) = \sqrt{v(\bar{y})} = \sqrt{0,278} = 0,527$$

$$\hat{Y} = N\bar{y} = 36(4) = 144$$

- ❖ Rata-rata mie instan yang dikonsumsi oleh mahasiswa kelas 2A diperkirakan sebanyak 4 bungkus per minggu dengan *standar error* sebesar 0,527. Atau banyaknya mie instan yang dikonsumsi adalah 144 bungkus per minggu.

5.3 Latihan

1. Suatu penelitian dilakukan untuk mengetahui total hasil panen padi di suatu wilayah. Dari total 290 petak sawah, 29 petak sawah dipilih dengan *systematic random sampling* dan datanya adalah sebagai berikut:

Petak Sawah	Hasil panen padi (ton)								
1	0.0	7	0.5	13	2.3	19	6.0	25	8.2
2	0.9	8	3.1	14	3.5	20	5.4	26	5.4
3	0.0	9	2.8	15	2.4	21	2.3	27	6.5
4	0.0	10	2.7	16	3.8	22	2.9	28	6.6
5	0.3	11	2.8	17	4.1	23	2.1	29	4.1
6	0.1	12	2.6	18	4.9	24	6.3		

Perkirakan total hasil panen di wilayah tersebut beserta se-nya!

JAWAB:

$$N = 290$$

$$n = 29$$

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{0,0 + 0,9 + \dots + 4,1}{29} = \frac{92,6}{29} = 3,1931$$

$$s^2 = \frac{\sum_1^n (y_i - \bar{y})^2}{n - 1} = \frac{(0,0 - 3,1931)^2 + \dots + (4,1 - 3,1931)^2}{29 - 1} = 5,1914$$

$$v(\bar{y}) \approx \frac{N - n}{N} \frac{s^2}{n} \approx \frac{290 - 29}{290} \frac{5,1914}{29} \approx 0,1611$$

$$se(\bar{y}) = \sqrt{v(\bar{y})} = \sqrt{0,1611} = 0,4014$$

$$\hat{Y} = N\bar{y} = 290(3,1931) = 926$$

- ❖ Rata-rata hasil panen di wilayah tersebut diperkirakan sebesar 3,1914 ton per hektar dengan *standar error* sebesar 0,1611 per hektar. Atau total hasil panen nya adalah sebesar 926 ton.
2. Menurut saudara, apakah *Systematic sampling* selalu lebih efisien dibanding *Simple Random Sampling*? Jelaskan!

5.4 Rangkuman

Metode pengambilan sampel sistematis adalah metode untuk mengambil sampel secara sistematis dengan interval (jarak) tertentu dari suatu kerangka sampel yang telah diurutkan.

Keuntungan dari penggunaan sampel sistematis adalah lebih cepat dan mudah pelaksanaannya karena kita cukup mengambil satu angka random yang selanjutnya tinggal mengambil kelipatannya saja. Selain itu sampel sistematis tersebar lebih merata, sehingga kemungkinan besar menghasilkan sampel yang lebih representatif dan lebih efisien dari pada SRS.

Bab VI. Penutup

6.1 Simpulan

Data dan informasi statistik yang didapatkan dari hasil survei berguna untuk melihat perubahan-perubahan, menganalisa dan akhirnya sebagai dasar menentukan kebijakan-kebijakan. Penelitian dengan menggunakan metode survei yang menghasilkan informasi dari data sampel memberikan beberapa keuntungan antara lain yaitu efisiensi waktu, biaya dan tenaga kerja dibandingkan dengan pengumpulan data dengan metode sensus.

Jenis pengambilan sampel ada yang sifatnya probabilistik dan non probabilistik tergantung dari karakteristik unit populasi yang akan diteliti. Terdapat 4 (empat) jenis pengambilan sampel yang sifatnya probabilistik yaitu SRS, *Systematic Sampling*, *Stratified Random Sampling* dan *Cluster Sampling*.

Sebelum kita melakukan pengambilan sampel yang sifatnya probabilistik, kita perlu mengetahui seluruh kemungkinan sampel kemudian barulah melakukan penarikan sampel dan setelah melakukan pendataan dengan sampel-sampel tersebut kita perlu menghitung penyimpangan nilai dugaan dari nilai populasinya.

Kerangka sampel adalah keseluruhan unit dalam populasi. akan membentuk kerangka sampel dan dari sinilah anggota sampel dipilih. Untuk bisa melakukan penarikan sampel secara acak, kita memerlukan kerangka sampel berupa daftar dari keseluruhan unit populasi sebagai dasar penarikan sampel. Kerangka sampel bisa dibuat berdasarkan pengalaman hasil sensus atau kita melakukan listing terlebih dahulu.

Sampel acak sederhana (*simple random sampling*) dan sistematis sampling digunakan jika setiap unit dalam populasi diberi peluang sama untuk terpilih. Jika unit populasinya tidak terlalu banyak dan dapat diurutkan, biasanya digunakan sistematis sampling.

6.2 Tindak Lanjut

Berbekal hasil belajar mata diklat Metode Penarikan Sampel dengan mempergunakan modul ini, diharapkan peserta dapat menerapkan metode penarikan sampel yang tepat jika di unit kerjanya atau instansinya melakukan kegiatan pengumpulan data melalui tehnik survei.

Daftar Pustaka

Cohran, W.G. 1963. *Sampling Techniques*, Second Edition.
New York : John Willey & Sons Inc.

Dajan, Anto. 1986. *Pengantar Metode Statistik Jilid II*, Cetakan ke 11.
Jakarta : LP3ES.

Murthy, M. N., *Sampling Teory Methods*, Statistical Publishing Society 1977

Singh, Daroga. 1986. *Theory and Analysis of Sampel Survei Designs*.
New Delhi : Wiley Eastern Limited.

William W. Hines and Douglas C.Montgomery. 1990. *Probabilita dan Statistik dalam ilmu Rekayasa dan Manajemen, Edisi ke dua*.
Depok : Penerbit Universitas Indonesia, UI-Press.

•
•
•



Modul 03

Pengolahan Data

.....



DAFTAR ISI

DAFTAR ISI	i
DAFTAR GAMBAR	ii
Bab I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Deskripsi Singkat	1
1.3 Hasil Belajar	2
1.4 Indikator Hasil Belajar	2
1.5 Materi Pokok	2
1.6 Manfaat	2
Bab II Pengenal Data Statistik dan Paket Program Kerja SPSS	3
2.1 Jenis Data Statistik	3
2.2 Cara Kerja SPSS	5
Bab III Persiapan Data	7
3.1 Pemasukan Data	7
3.2 Pendefinisian Variabel	8
3.3 Pengelolaan Data Sebelum Dilakukan Analisis	11
3.4 Rangkuman	25
3.5 Latihan	26
Bab IV Statistik Deskriptif	29
4.1 Menganalisis Statistik Deskriptif	29
4.2 Menganalisis Statistik Frekuensi	31
4.3 Menganalisis Statistik Eksplorasi Data	34
4.4 Rangkuman	38
4.5 Latihan	39
Bab V Uji Validitas dan Reliabilitas	41
5.1. Pengujian Validitas	41
5.2. Pengujian Reliabilitas	42
Bab VI Penutup	47
6.1 Simpulan	47
6.2 Tindak Lanjut	47
LAMPIRAN	49
DAFTAR PUSTAKA	61

DAFTAR GAMBAR

Gambar 1. Alur Pengolahan Data SPSS	5
Gambar 2. Tampilan Data Editor	7
Gambar 3. Tampilan Data Setelah Diinput di SPSS	8
Gambar 4. Tampilan Variabel View.....	8
Gambar 5. Tampilan Variabel View dengan Nama dan Label yang Telah Terisi	9
Gambar 6. Tampilan Kotak Dialog Value Labels	9
Gambar 7. Kotak Dialog Value Labels yang Telah Diisi.....	10
Gambar 8. Pilihan Jenis Skala Pengukuran Data	10
Gambar 9. Tampilan <i>Role Variable</i> pada SPSS 18	11
Gambar 10. Kotak Dialog Split File.....	12
Gambar 11. Tampilan Submenu Merge File	13
Gambar 12. Kotak Dialog Pemilihan Data Set untuk Add Cases	13
Gambar 13. Kotak Dialog Add Cases	14
Gambar 14. Kotak Dialog Pemilihan Data Set untuk Add Variables	15
Gambar 15. Kotak Dialog Add Variables	15
Gambar 16. Kotak Dialog Select Cases	17
Gambar 17. Kotak dialog Compute Variable	18
Gambar 18. Kotak dialog Type and Label	18
Gambar 19. Kotak dialog Recode into Same Variables	20
Gambar 20. Kotak dialog Old and New Values	20
Gambar 21. Recode into Same Variables Setelah Pengisian	21
Gambar 22. Kotak dialog Recode into Different Variables	22
Gambar 23. Recode into Different Variables: Old and New Values.....	22
Gambar 24. Recode into Different Variables Setelah Pengisian.....	23
Gambar 25. Tampilan Setelah Menggunakan Recode into Different Variables	23
Gambar 26. Kotak dialog Automatic Recode.....	24
Gambar 27. Tampilan Data Setelah Menggunakan Automatic Recode	25
Gambar 29. Kotak Dialog Descriptives.....	30
Gambar 30. Kotak Dialog Option pada Descriptives	30
Gambar 31. Output Descriptives Statistics.....	30

Gambar 32. Kotak Dialog Frequences	31
Gambar 33. Kotak Dialog Frequencies Statistics	32
Gambar 34. Kotak Dialog Frequencies Charts	32
Gambar 35. Kotak Dialog Frequencies: Format	33
Gambar 36. Output Statistik Frekuensi	33
Gambar 37. Kotak Dialog Explore	34
Gambar 39. Tampilan Kotak Dialog Plot Eksplorasi	36
Gambar 40. Tampilan Kotak Dialog Option Eksplorasi	38
Gambar 41. Tampilan Menu Compare Means	Error! Bookmark
Gambar 42. Kotak Dialog One-Sample T Test	Error! Bookmark
Gambar 43. Kotak Dialog One-Sample T Test: Option	Error! Bookmark
Gambar 44. Output Analisis Uji-t Satu Sampel	Error! Bookmark
Gambar 45. Kotak Dialog Independent-Samples T Test	Error! Bookmark
Gambar 46. Tampilan Menu Compare Means	Error! Bookmark
Gambar 47. Output Analisis Uji-t Dua Sampel Bebas	Error! Bookmark
Gambar 48. Kotak Dialog Uji-t Sampel Berpasangan	Error! Bookmark
Gambar 49. Output Analisis Uji-t Dua Sampel Bebas	Error! Bookmark
Gambar 50. Tampilan Menu Compare Means	Error! Bookmark
Gambar 39. Kotak Dialog ANOVA	Error! Bookmark
Gambar 51. Tampilan Option ANOVA	Error! Bookmark
Gambar 52. Kotak Dialog <i>Post-Hoc</i> pada ANOVA	Error! Bookmark
Gambar 53. Output ANOVA	Error! Bookmark
Gambar 57. Tampilan Kotak Dialog Korelasi Dua Variabel	Error! Bookmark
Gambar 58. Tampilan Kotak Dialog Korelasi Parsial	Error! Bookmark
Gambar 59. Tampilan Submenu Regression	Error! Bookmark
Gambar 60. Kotak Dialog Regresi Linier	Error! Bookmark
Gambar 61. Kotak Dialog Linear Regression: Statistics	Error! Bookmark

Bab I Pendahuluan

1.1 Latar Belakang

Di dalam perhitungan statistik, tidak semuanya mudah diselesaikan dengan perhitungan manual. Berbagai jenis dan banyaknya data membuat semakin pentingnya program komputer untuk penyelesaiannya. SPSS (*Statistical Package for the Social Sciences*) adalah sebuah program pada komputer yang digunakan untuk membuat analisis statistika. SPSS pertama kali dirilis pada tahun 1968 dan diciptakan oleh Norman Nie. Pada akhir tahun 2009, perusahaan pembuat SPSS diakuisisi oleh IBM.

Sekarang ini berkembang SPSS berbasis Windows sehingga di kenal dengan SPSS for windows. Pertama kali muncul versi windows adalah SPSS for windows versi 6.00, hingga kini SPSS selalu memperbaharui versinya. Pada dasarnya pengoperasian SPSS memiliki kesamaan dalam berbagai versi, perbedaannya pada fasilitas tambahan. Selain itu, SPSS merupakan software statistik yang paling populer, fasilitasnya sangat lengkap dibandingkan dengan software lainnya. Oleh karena itu, diharapkan dengan penggunaan SPSS dapat memberikan kemudahan dan ketepatan dalam mengolah data.

SPSS memiliki banyak kegunaan bagi pengguna seperti peneliti ekonomi, peneliti kesehatan, perusahaan survei, pemerintah, peneliti pendidikan, organisasi pemasaran, dan sebagainya. Ada beberapa teknik statistik yang dapat digunakan untuk menganalisis data. Tujuan dari analisis data adalah untuk mendapatkan informasi yang relevan yang terdapat dalam data tersebut dan menggunakan hasilnya untuk memecahkan suatu masalah dan membantu dalam menetapkan kebijakan dalam perencanaan pembangunan.

1.2 Deskripsi Singkat

Materi Paket Program Statistik membahas tentang pengolahan data statistik dengan menggunakan software pengolah SPSS yang bertujuan untuk mempermudah interpretasi suatu informasi yang dijelaskan dengan perhitungan statistik.

1.3 Hasil Belajar

Setelah membaca modul Paket Program Statistik ini peserta pelatihan dapat mengerti dan mampu melakukan proses pengolahan, penyusunan laporan dan analisis statistik secara deskriptif dan inferensia dengan menggunakan paket program SPSS.

1.4 Indikator Hasil Belajar

Setelah mempelajari modul ini secara tuntas, peserta dapat:

1. memahami tentang pengolahan, penyusunan laporan dan analisis tabulasi data berupa frekuensi, kuartil dan lainnya.
2. menerapkan penghitungan, pengolahan, penyusunan laporan dan analisis Statistik Deskriptif seperti ukuran rata-rata, median, modus dan lainnya.
3. mengolah, melakukan penguraian, penyusunan dan analisis eksplorasi data seperti diagram akar batang dan daun (steam-leaf-plot) dan box-plot dan lainnya.
4. menerapkan penghitungan, pengolahan, penyusunan laporan dan analisis statistik inferensia uji perbedaan dengan SPSS.
5. menerapkan penghitungan, pengolahan, penyusunan laporan dan analisis statistik inferensia uji asosiasi dengan SPSS

1.5 Materi Pokok

Materi pokok yang dibahas dalam modul Paket Program Statistik ini adalah:

- a. Pengenalan data statistik dan paket program kerja SPSS
- b. Persiapan data
- c. Statistik deskriptif
- d. Statistik inferensia

1.6 Manfaat

Berbekal hasil belajar pada modul Paket Program Statistik, peserta diharapkan mampu meningkatkan kemampuan, kemauan, dan tanggungjawab sebagai fungsional statistisi dalam melaksanakan tugas-tugas yang berkaitan dengan jabatannya guna peningkatan kerja instansinya.

Bab II Pengenalan Data Statistik dan Paket Program Kerja SPSS

2.1 Jenis Data Statistik

Pengertian data dalam statistik agak berbeda dengan pengertian data pada umumnya. Menurut sifatnya, terdapat dua jenis data yaitu data numerik dan non numerik. Jenis data sangat menentukan jenis analisis statistik yang akan dilakukan.

1. Data Kualitatif

Secara sederhana, data kualitatif dapat disebut sebagai data yang bukan berupa angka. Ciri utama data kualitatif adalah tidak dapat dilakukan operasi matematika terhadapnya, seperti penjumlahan, pengurangan dan sebagainya.

Data kualitatif dibagi menjadi dua :

- Data Nominal

Data tipe nominal adalah data yang paling 'rendah' dalam level pengukuran data. Hal ini dikarenakan ukurannya yang paling sederhana, dimana angka yang diberikan kepada objek mempunyai arti sebagai label saja, dan tidak menunjukkan tingkatan dan peringkat apapun. Misal dalam pengisian data jenis kelamin, laki-laki dikategorikan sebagai '1' dan perempuan sebagai '2'. Namun, ini bukan berarti angka 2 akan lebih besar nilainya dibandingkan dengan 1. Contoh lain misalnya tentang jenis olah raga yakni tenis, basket dan renang. Kemudian masing-masing anggota set di atas diberikan angka, misalnya tenis (1), basket (2) dan renang (3). Jelas kelihatan bahwa angka yang diberikan tidak menunjukkan bahwa tingkat olah raga basket lebih tinggi dari tenis ataupun tingkat renang lebih tinggi dari tenis. Angka yang diberikan hanya berfungsi sebagai label saja.

- Data Ordinal

Data ordinal adalah data kualitatif dengan level yang lebih 'tinggi' dari data nominal. Jika pada data nominal, semua data kategori dianggap setara, maka pada data ordinal terdapat tingkatan (preferensi) data. Misal pada data jenis kelamin diatas, laki-laki dianggap setara dengan perempuan. Pada data ordinal terdapat data dengan urutan lebih tinggi dan lebih rendah. Contohnya adalah data tentang sikap seseorang terhadap produk tertentu, ada yang 'tidak suka', 'suka' dan 'sangat suka'. Angka yang diberikan untuk data

ordinal menunjukkan nilai ranking dari objek saja, bukan nilai absolutnya. Misal 1=tidak suka, 2=suka dan 3=sangat suka. Contoh lain adalah ukuran bentuk seperti 1=kecil, 1=sedang dan 3=besar.

2. Data Kuantitatif

Data kuantitatif dapat disebut sebagai data berupa angka dalam arti sebenarnya. Jadi berbagai operasi matematika dapat dilakukan pada data kuantitatif. Data kuantitatif juga dapat dibagi menjadi dua bagian, yaitu :

- Data Interval

Data Interval menempati level pengukuran data yang lebih “tinggi” dari data ordinal karena selain bisa bertingkat urutannya, juga urutan tersebut bisa dikuantitatifkan. Seperti pengukuran temperatur sebuah ruangan pembakaran roti dari PT PASTI JOSS. Interval Temperatur ruang tersebut:

- Cukup Panas jika temperatur antara 50°C – 80°C
- Panas jika temperatur antara 80°C – 110°C
- Sangat Panas jika temperatur antara 110°C – 140°

Dalam kasus di atas, data temperatur bisa dikatakan data interval karena data mempunyai interval (jarak) tertentu, yaitu 30°C. Namun, di sini data interval tidak mempunyai titik nol yang absolut. Misal pada pengukuran temperatur, seperti pernyataan bahwa ‘air membeku pada 0°C’. Pernyataan di atas bersifat relatif, karena 0°C hanya sebagai tanda saja. Dalam pengukuran °F, air membeku bukan pada 0°F, namun pada 32°F. Dengan demikian, juga tidak bisa dikatakan bahwa suhu 100°F adalah dua kali lebih panas dari suhu 50°F

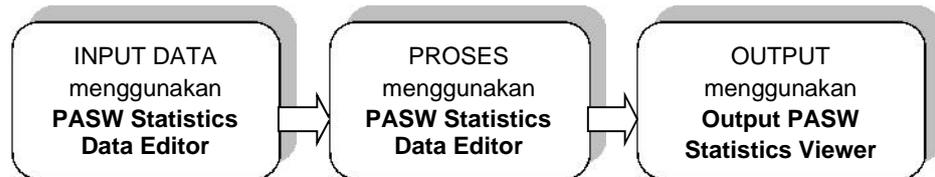
- Data Rasio

Data Rasio adalah data dengan tingkat pengukuran paling “tinggi” di antara jenis data lainnya. Data Rasio adalah data bersifat angka dalam arti sesungguhnya (bukan kategori seperti pada data nominal dan ordinal) dan bisa dioperasikan secara matematika. Perbedaan dengan data interval adalah bahwa data rasio mempunyai titik nol dalam arti sesungguhnya. Oleh karena ada titik nol, maka ukuran rasio dapat dibuat perkalian ataupun pembagian. Angka pada skala rasio dapat menunjukkan nilai sebenarnya dari objek yang diukur. Misal ada 4 orang pengemudi, A, B, C dan D yang mempunyai pendapatan masing-masing perhari Rp 10.000, Rp 30.000, Rp 40.000 dan Rp 50.000. Dengan ukuran rasio, dapat dikatakan bahwa pendapatan pengemudi C adalah 4 kali pendapatan pengemudi A. Pendapatan D adalah 5 kali pendapatan A. Pendapatan C adalah $\frac{4}{3}$ kali pendapatan B. Dengan kata lain, rasio antara C dan A adalah 4 : 1, rasio antara D dan A adalah 5 : 1, sedangkan rasio antara C dan B adalah 4 :3.

2.2 Cara Kerja SPSS

SPSS adalah singkatan dari *Statistical Package for the Social Sciences*. SPSS merupakan aplikasi untuk menganalisis data statistik, terutama statistik untuk ilmu-ilmu sosial, ekonomi, dan lainnya. Adapun Versi SPSS yang digunakan dalam modul ini adalah SPSS versi 18, atau PASW 18 SPSS.

Proses pengolahan data pada SPSS dapat digambarkan sebagai berikut:



Gambar 1. Alur Pengolahan Data SPSS

Penjelasan :

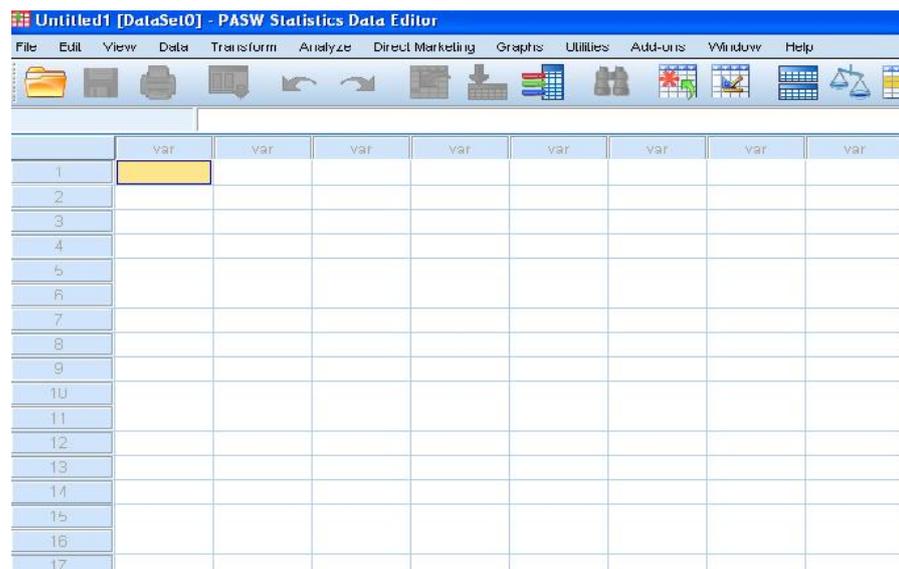
1. Data yang akan diproses dimasukkan lewat *Dataset PASW Statistics Data Editor* yang otomatis muncul di layar saat SPSS dijalankan.
2. Data yang telah diinput kemudian diproses, juga lewat **PASW Statistics Data Editor**.
3. Hasil pengolahan data muncul di layar (*window*) yang lain yaitu **Output PASW Statistics Viewer**.

Bab III Persiapan Data

Sebelum melakukan analisis terhadap data menggunakan aplikasi SPSS, hal pertama yang harus kita lakukan adalah memasukkan data ke dalam SPSS dan mendefinisikan variabel-variabel yang akan dipakai untuk proses analisis selanjutnya. Namun terkadang, seorang peneliti tidak harus melewati proses pemasukan data dan pendefinisian variabel dikarenakan analisis yang akan dilakukan tidak membutuhkan data baru. Artinya, analisis akan dilakukan pada data yang telah tersedia. Yang perlu menjadi perhatian, pada data yang tersedia tersebut, tidak seluruh isi di dalamnya akan dianalisis. Sehingga, diperlukan pengelolaan data sebelum dilakukan analisis.

3.1 Pemasukan Data

Di bawah ini adalah tampilan ketika pertama kali masuk ke dalam software SPSS, yakni DATA EDITOR.

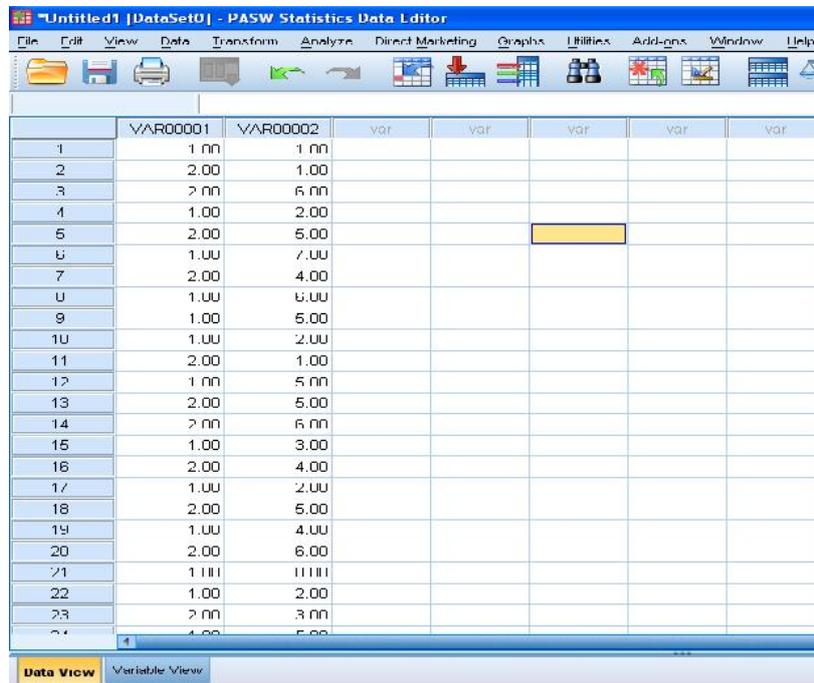


Gambar 2. Tampilan Data Editor

Pada dasarnya Input data/*Data Entry* bisa dilakukan dengan dua cara:

1. Input langsung
2. Memindahkan data dari Ms.Excel ke SPSS dengan cara *copy* dan *paste* atau dengan memilih menu **File-Open-Data**, pada pilihan **Files of Type**, pilih Excel.

Berikut adalah tampilan setelah data diinput di SPSS:



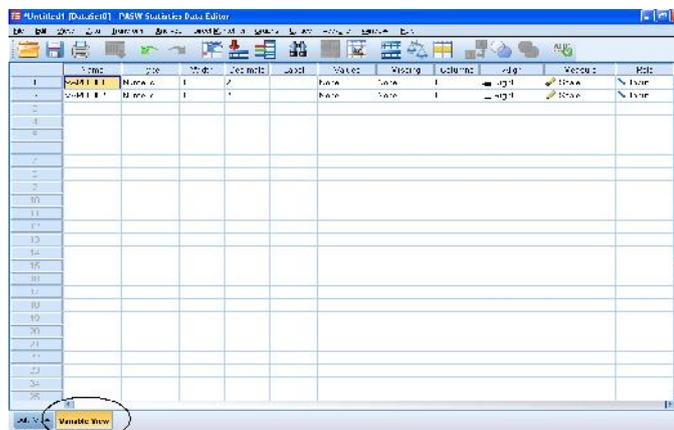
Gambar 3. Tampilan Data Setelah Diinput di SPSS

3.2 Pendefinisian Variabel

3.2.1 Memberi Nama Variabel

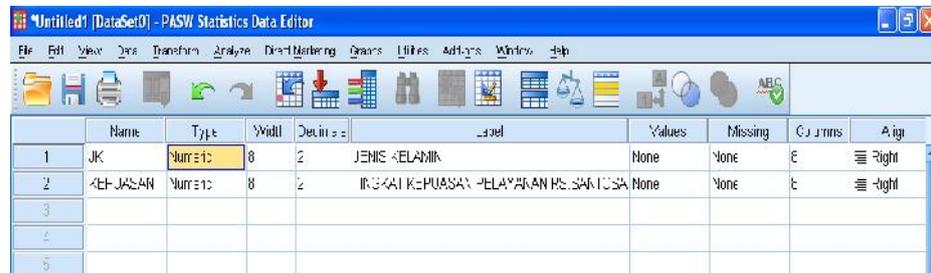
Setelah diinput di SPSS, langkah selanjutnya adalah memberi nama untuk setiap variabel. Langkah-langkahnya adalah sebagai berikut:

1. Klik *sheet* **Variabel View**, selanjutnya format SPSS akan berbentuk sebagai berikut.



Gambar 4. Tampilan Variabel View

Untuk contoh di atas, beri nama JK pada var00001, dan KEPUASAN pada var00002. Kemudian pada kolom label, beri keterangan tentang variabel. Ketik JENIS KELAMIN untuk label variabel pertama, dan TINGKAT KEPUASAN PELAYANAN RS.SANTOSA untuk variabel kedua. Selanjutnya, pada *sheet* **Variable View** akan tampil keterangan tentang variabel sebagai berikut:



Gambar 5. Tampilan Variable View dengan Nama dan Label yang Telah Terisi

3.2.2 Memberi Keterangan Variabel Kategorik

Apabila suatu variabel berskala nominal maupun ordinal, maka harus diberikan keterangan tambahan. Langkah-langkahnya adalah sebagai berikut:

1. Klik kolom **Values** pada sheet **Variable View**, maka akan muncul kotak dialog sebagai berikut.

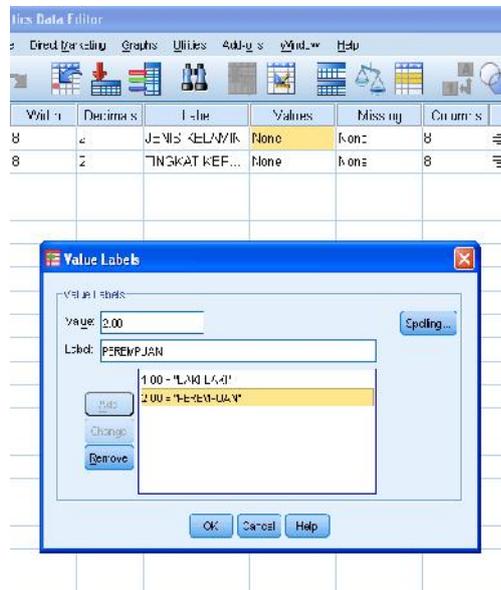


Gambar 6. Tampilan Kotak Dialog Value Labels

Kolom **values** berguna untuk memberikan keterangan apabila data yang diinput berskala nominal atau ordinal.

2. Oleh karena variabel jenis kelamin dikelompokkan sebagai data nominal yang terdiri atas laki-laki dan perempuan, maka pada sheet **Data View**, bilangan 1 merupakan label untuk jenis kelamin laki-laki dan label bilangan 2 adalah label untuk jenis kelamin

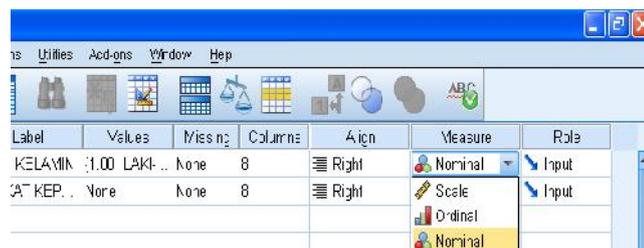
perempuan. Pada kotak dialog **Value Labels**, beri nilai 1 pada kolom **Value** dan laki-laki pada kolom **Label** lalu klik **Add**. Begitu pula selanjutnya untuk jenis kelamin perempuan. Hasilnya adalah sebagai berikut:



Gambar 7. Kotak Dialog Value Labels yang Telah Diisi

3.2.3 Menentukan Jenis Variabel

Jenis variabel akan menentukan jenis analisis yang akan dipakai. Oleh karena itu, kolom **Measure** sangatlah penting untuk diperhatikan. Pada kolom **Measure**, definisikan jenis data yang akan diinput. Misal, untuk variabel jenis kelamin, skala datanya dirubah dari Scale menjadi Nominal dengan mengklik kolom **Measure** dan memilih datanya menjadi berskala nominal. Berikut adalah tampilannya.



Gambar 8. Pilihan Jenis Skala Pengukuran Data

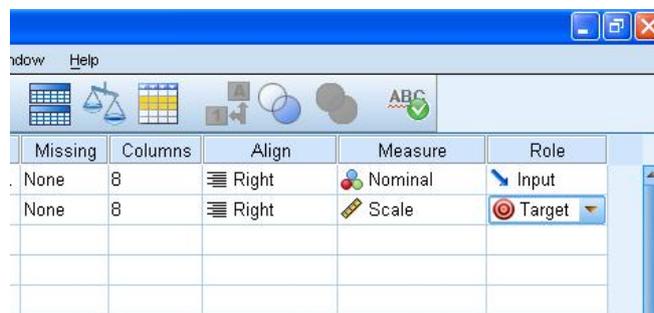
3.2.4 Menentukan Peran Variabel

Pada SPSS 18, disediakan fasilitas untuk mendefinisikan peran variabel. Adapun peran yang tersedia adalah:

1. Input. Variabel akan digunakan sebagai masukan (misalnya, prediktor, variabel bebas).

2. Target. Variabel akan digunakan sebagai output atau target (misalnya, variabel dependen).
3. Both. Variabel akan digunakan sebagai input maupun output.
4. None. Variabel tidak memiliki tugas peranan.
5. Partisi. Variabel akan digunakan untuk partisi data ke sampel terpisah untuk pelatihan, pengujian, dan validasi.
6. Split.

Dalam contoh ini, Variabel JK adalah variabel independen dan KEPUASAN adalah variabel dependen. Sehingga pada kolom Role, variabel JK didefinisikan sebagai Input dan variabel KEPUASAN didefinisikan sebagai Target. Tampilannya akan menjadi seperti di bawah ini.



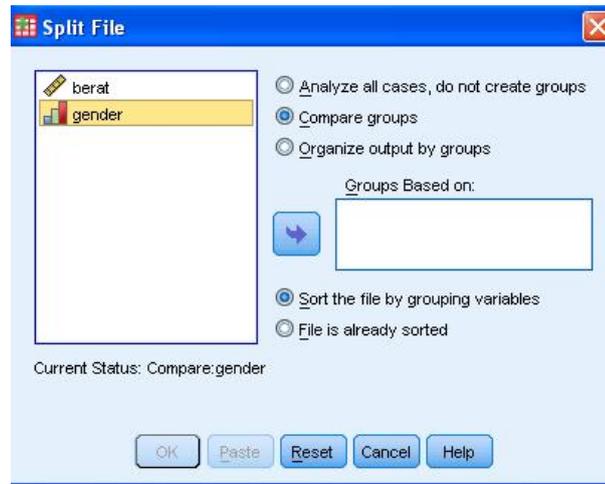
Gambar 9. Tampilan Role Variable pada SPSS 18

3.3 Pengelolaan Data Sebelum Dilakukan Analisis

3.3.1 Split File

Split file adalah fasilitas untuk memisah isi file dengan kriteria tertentu. Misalnya ada sebuah file yang berisikan data berat badan, dimana data berat badan wanita dan pria berselang-seling. Untuk kasus yang sangat banyak dan diinginkan variabel dengan label tertentu dipisah, dalam hal ini kasus berat dengan gender pria dan wanita ingin dipisah, bisa digunakan perintah SPLIT FILE dengan langkah-langkah:

1. Klik menu DATA, kemudian pilih **Split File**. Tampak pada layar:



Gambar 10. Kotak Dialog Split File

Pengisian:

- Karena akan memisahkan file dalam group, maka klik pilihan **Organize Output by Groups**
- Karena pembagian berdasarkan gender, pindahkan variabel gender ke dalam kotak **Group Based On** dengan cara mengklik .
- Karena pada mula-mula file masih acak antara gender pria dan wanita, maka klik mouse pada pilihan **Sort file by grouping variables**.

Jika pengisian kotak dialog di atas telah selesai, klik OK. Pada DATA EDITOR, akan tampak data berat badan wanita dan pria tidak lagi selang-seling.

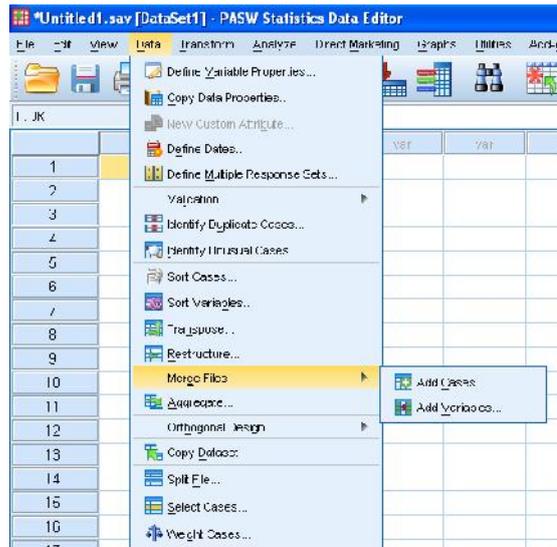
3.3.2 Merge Files

Peneliti dapat menggabungkan data dari dua file dalam dua cara yang berbeda, yakni:

- Menggabungkan dataset aktif dengan dataset lain yang berisi variabel yang sama tapi kasus yang berbeda.
- Menggabungkan dataset aktif dengan dataset lain yang memiliki kasus yang sama tetapi variabel yang berbeda.

Untuk melakukan penggabungan data, langkah-langkahnya:

- Klik menu **Data**, dan pilih **Merge File..**

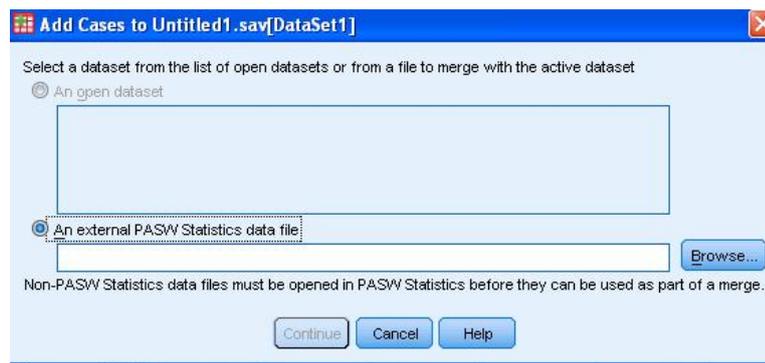


Gambar 11. Tampilan Submenu Merge File

- Pilih **Add Cases** apabila yang ingin ditambah adalah jumlah kasusnya namun jumlah variabel tetap (yang bertambah adalah jumlah baris). Sebaliknya, pilih **Add Variable** apabila yang ingin ditambah adalah jumlah variabelnya namun jumlah kasus tetap (yang bertambah jumlah kolomnya).

1. Menambah Kasus

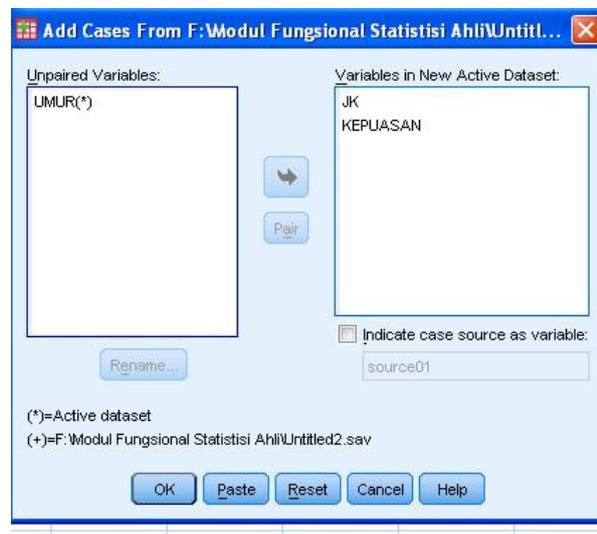
Apabila ingin menambah kasus (baris) dengan mempertahankan jumlah variabel (kolom), maka fasilitas **Merge Files** yang dipilih adalah **Add Cases**. Maka di layar akan muncul:



Gambar 12. Kotak Dialog Pemilihan Data Set untuk Add Cases

Kolom pertama, **An Open Dataset**, diisi bila kasus yang ingin ditambahkan ada pada dataset lain yang aktif di window. Sedangkan apabila kasus yang ingin ditambahkan masih tersimpan di file yang sedang tidak aktif/dibuka, maka isilah kolom **An external PASW Statistics data file** dengan

mengklik tombol **Browse**. Apabila file yang diinginkan telah terpilih klik **Continue**. Sehingga di layar akan muncul:



Gambar 13. Kotak Dialog Add Cases

Unpaired Variables, yaitu kumpulan variabel yang tidak berpasangan antara dataset aktif dan dataset lain yang ingin digunakan untuk penggabungan. Variabel dari dataset aktif diidentifikasi dengan tanda bintang (*), sedangkan variabel dari dataset lain diidentifikasi dengan tanda tambah (+). Misalkan dalam kolom Unpaired Variables ini ada variabel UMUR(*). Artinya, antara dua dataset tersebut terdapat variabel yang hanya ada di satu dataset saja, yaitu variabel UMUR, dimana variabel UMUR berasal dari data set yang aktif (dilihat dari tanda (*)).

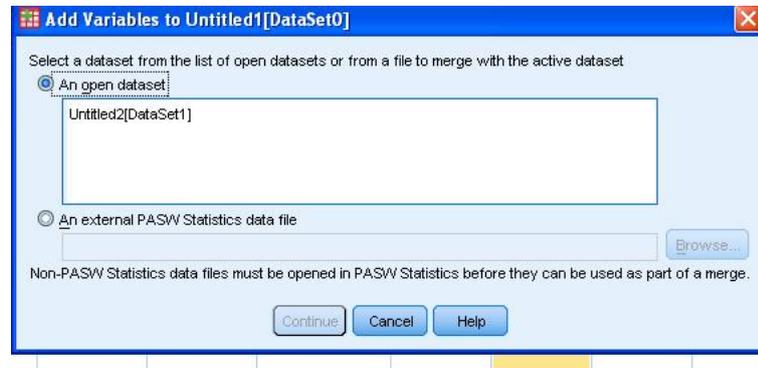
Variables in New Active Dataset, yaitu kumpulan variabel yang sama-sama ada di kedua data set yang akan digabungkan dan akan muncul di dataset yang telah digabungkan nanti.

Indicate case Source as variables, ditandai apabila peneliti ingin menunjukkan sumber dari setiap kasus. Akan muncul variabel baru yang memiliki nilai 0 untuk kasus yang bersumber dari dataset aktif dan memiliki nilai 1 untuk kasus dari data set lain / file data eksternal. Apabila semua telah terisi, klik **OK**.

2. Menambah Variabel

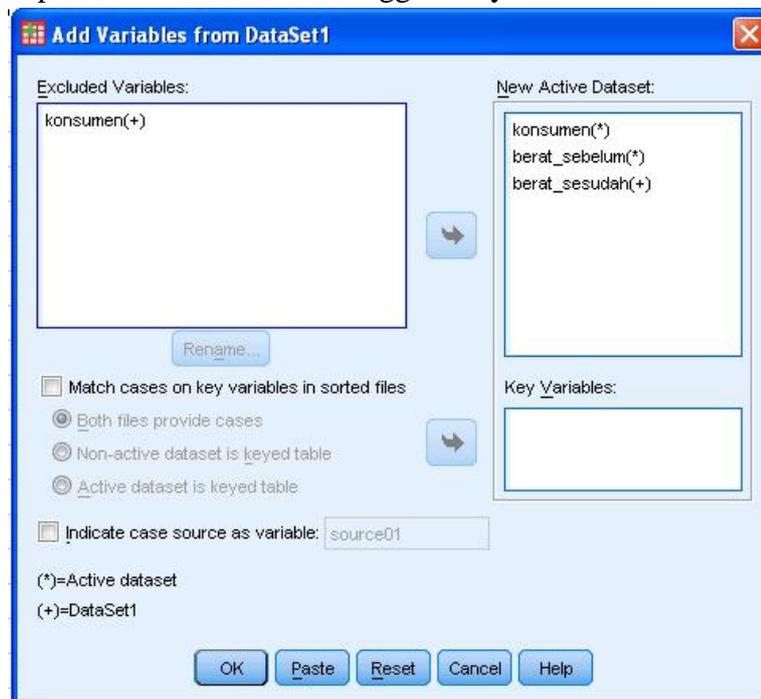
Apabila ingin menambah variabel (kolom) dengan mempertahankan jumlah kasus (baris), maka fasilitas **Merge Files** yang dipilih adalah **Add Variabel**. Contohnya apabila

Anda ingin membandingkan data pengukuran berat badan sebelum dan sesudah mengonsumsi obat pelangsing, dimana jumlah kasusnya adalah sama, namun anda ingin menggabungkan file berat sebelum dan berat sesudah. Klik **Data - Merge Files - Add Variabel** Maka di layar akan muncul:



Gambar 14. Kotak Dialog Pemilihan Data Set untuk Add Variables

Kolom pertama, **An Open Dataset**, diisi bila variabel yang ingin ditambahkan ada pada dataset lain yang aktif di window. Sedangkan apabila variabel yang ingin ditambahkan masih tersimpan di file yang sedang tidak aktif/dibuka, maka isilah kolom **An external PASW Statistics** data file dengan mengklik tombol **Browse**. Apabila file yang diinginkan telah terpilih klik **Continue**. Sehingga di layar akan muncul:



Gambar 15. Kotak Dialog Add Variables

Indicate cases source as variable. Pilihan ini berguna untuk membuat variabel baru yang menunjukkan sumber data untuk setiap kasus. Variabel ini memiliki nilai 0 untuk kasus-kasus dari dataset aktif dan nilai 1 untuk kasus dari file data eksternal.

Excluded Variables. Secara *default*, daftar ini berisi nama-nama variabel yang terdapat di kedua dataset. Variabel dari dataset aktif diidentifikasi dengan tanda bintang (*). Variabel dari dataset lain diidentifikasi dengan tanda tambah (+). Jika Anda ingin menyertakan sebuah **excluded variables** dengan nama ke dalam file gabungan nanti, Anda dapat mengubah nama variabel tersebut dengan mengklik **RENAME** dan menambahkannya ke daftar variabel untuk dimasukkan.

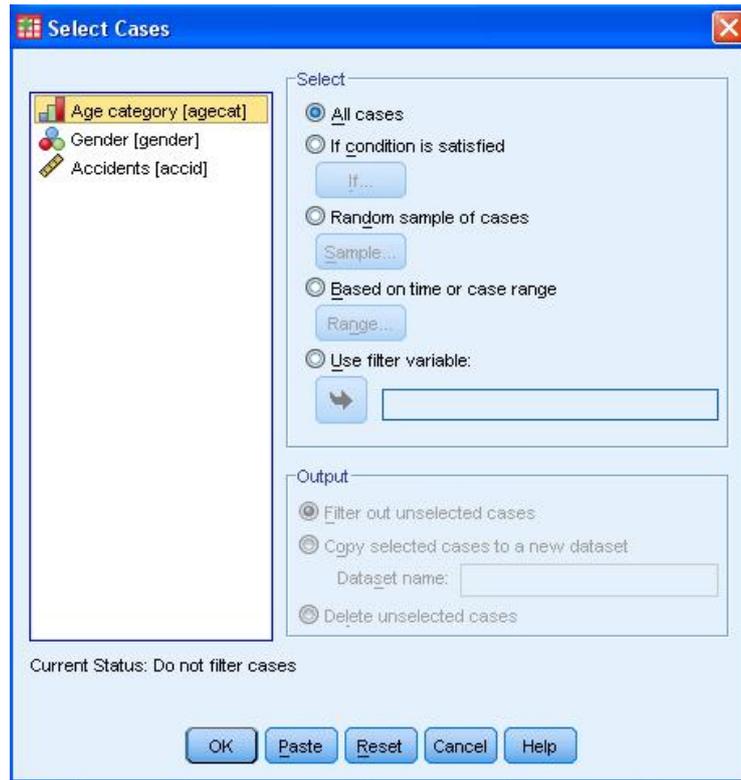
New Active Dataset. Secara *default*, semua nama variabel yang unik di kedua dataset termasuk dalam daftar.

Key Variables. Jika beberapa kasus di salah satu dataset tidak memiliki kasus yang cocok dalam dataset lainnya (yaitu, beberapa kasus yang hilang dalam satu dataset), gunakan variabel kunci untuk mengidentifikasi kasus dari dua dataset. Variabel kunci harus memiliki nama yang sama di kedua dataset dan telah diurutkan secara *Ascending* (gunakan fasilitas *sortcase*).

3.3.3 Select Case

Fasilitas ini dapat digunakan untuk menyeleksi isi file (kasus) berdasarkan kriteria tertentu. Misalnya terdapat suatu file data yang berisi data kecelakaan lalu lintas di suatu kota. Namun yang akan dianalisis saat ini hanyalah kecelakaan lalu lintas orang yang umurnya di atas 17 tahun. Sehingga tidak semua data dalam file tersebut akan diikuti dalam analisis selanjutnya. Untuk itu, fasilitas **SELECT CASE** bisa digunakan dengan langkah-langkah sebagai berikut:

1. Klik menu **Data**, kemudian pilih **Select Cases**, maka di layar akan muncul kotak dialog seperti di bawah ini:



Gambar 16. Kotak Dialog Select Cases

pada kolom **Select**, terdapat beberapa pilihan:

- All cases**, jika seleksi untuk seluruh kasus.
- If**, jika seleksi berdasar kriteria tertentu.
- Random**, jika seleksi berdasar bilangan acak (random).
- Time or case range**, jika seleksi berdasarkan range tertentu.
- Filter variable**, jika seleksi berdasarkan variable tertentu.

Untuk contoh kasus di atas, pada kolom **Select**, yang dipilih adalah **If**, karena yang akan dianalisis adalah data kecelakaan orang-orang yang umurnya lebih dari 17 tahun.

Pada kolom **Output**, juga terdapat beberapa pilihan antara lain:

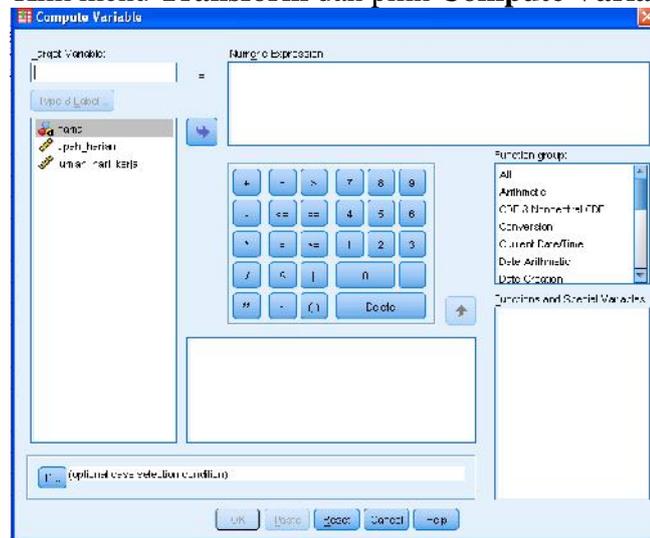
- Filter out unselected cases**, kasus yang tidak terpilih akan tetap muncul di dalam file yang sama, namun tidak akan diikuti pada analisis selanjutnya.
- Copy selected cases to new dataset**, kasus-kasus yang terpilih akan di-copy ke file yang baru, sehingga file yang lama/asli tidak akan mengalami perubahan.
- Delete unselected cases**, kasus yang tidak terpilih akan dihapus dari file lama/asli.

3.3.4 Compute

Fasilitas ini berfungsi menambah variabel baru yang berisi hasil perhitungan (*compute*) berdasarkan data dari variabel lama. Misalkan

awalnya dalam file data telah tersedia variabel “upah harian” dan “jumlah hari kerja”. Sedangkan yang dibutuhkan dalam analisis adalah variabel “total upah” yang bisa didapatkan dari perkalian antara “upah harian” dan “jumlah hari kerja”. Langkah-langkahnya adalah sebagai berikut:

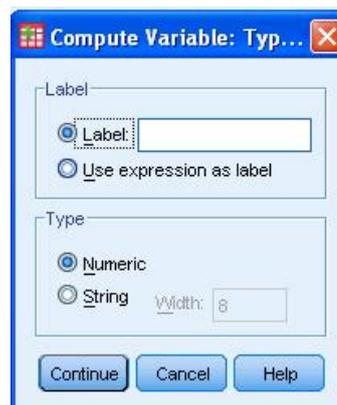
- Klik menu **Transform** dan pilih **Compute Variable**.



Gambar 17. Kotak dialog Compute Variable

Pengisian:

- **Target Variable**, isilah nama variabel baru yang akan dihitung, dalam contoh ini: **total_upah**. Kemudian tekan tombol **Type & Label** hingga tampak di layar:



Gambar 18. Kotak dialog Type and Label

Pada kolom **Label**, ketik keterangan untuk variabel baru. Misalkan ketik **Total upah pekerja**. Pada kolom **Type**, ketik tipe

data dari variabel baru. Karena pada contoh di atas variabel baru bersifat numerik, maka pada kolom Type, tandai **Numeric**.

- Pada **Numeric Expression**, isilah operasi matematika yang diinginkan. Masukkan variabel yang digunakan untuk membentuk variabel baru dengan mengklik . Untuk contoh di atas, pada kotak **Numeric Expression** akan terisi **upah_harian*jumlah_hari_kerja**.
- Klik OK, maka variabel baru dengan nama **Total_upah** akan muncul di layar **Dataset**.
- Apabila dalam pembentukan variabel baru diperlukan persyaratan, maka klik tombol **IF** di pojok kiri bawah kotak dialog **Compute Variabel** tersebut. Misal, apabila variabel **Total_Upah** hanya akan dihitung untuk pekerja yang hari kerjanya >25 hari. Sehingga untuk pekerja yang bekerja kurang atau sama dengan 25 hari, total upahnya tidak akan dihitung.

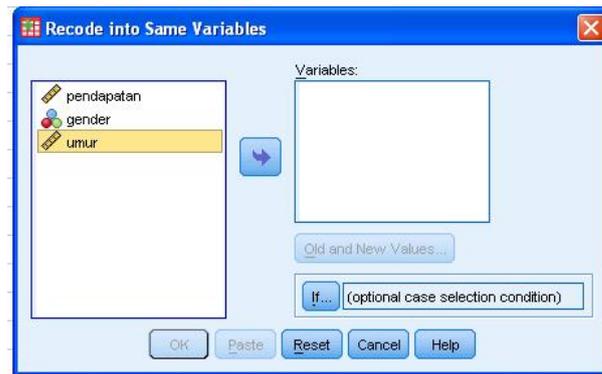
3.3.5 Recode Variable

Recode Variable adalah fasilitas untuk membuat kode ulang atas data yang telah tersedia. Misal, untuk kasus umur. Pada data yang tersedia, variabel umur dinyatakan dengan umur sebenarnya. Sementara, untuk analisis yang akan dilakukan, umur hanya dinyatakan dalam 3 kategori, yaitu <15 tahun, 15-50 tahun dan >50 tahun. Kondisi ini menuntut kita untuk mengodekan ulang variabel umur. Pada kondisi inilah, transformasi variabel sangat diperlukan. Pada menu **Transform** di SPSS 18, terdapat tiga pilihan pengodean ulang, yakni **Recode into Same Variables**, **Recode into Different Variables**, dan **Automatic Recode**.

Recode into Same Variables

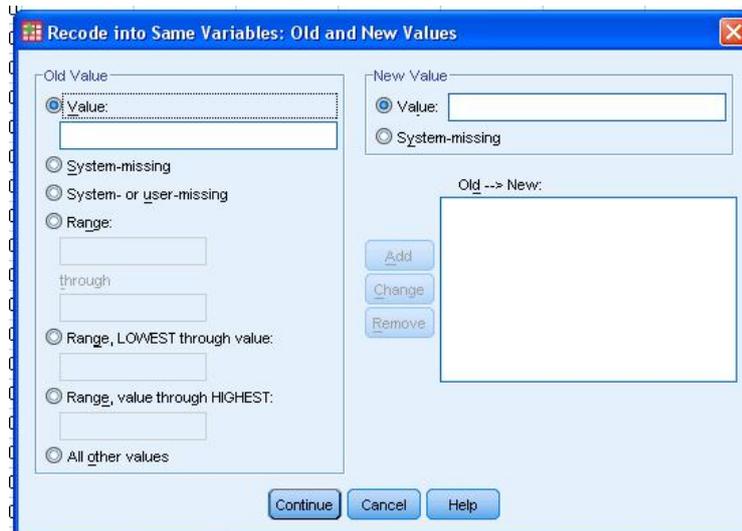
Fasilitas ini membantu peneliti untuk melakukan pengodean ulang variabel, dimana jenis variabel yang dikodekan ulang bisa saja berubah (numerik tetap numerik, string tetap string, numerik menjadi string begitu pula sebaliknya), namun perubahan tersebut hanya akan terjadi pada variabel yang sama. Artinya, tidak ada penambahan variabel baru karena yang berubah adalah nilai-nilai dalam variabel yang bersangkutan. Misal, untuk kasus umur di atas. Langkah-langkahnya adalah sebagai berikut:

1. Pilih menu **Transform**, kemudian pilih **Recode into Same Variables**, maka akan muncul kotak dialog sebagai berikut:



Gambar 19. Kotak dialog Recode into Same Variables

Pada kolom **Variables**, isikan variabel yang ingin dikode ulang. Dalam kasus ini, variabel umur. Kemudian Klik tombol **Old and New Values**, sehingga muncul kotak dialog:



Gambar 20. Kotak dialog Old and New Values

Karena yang ingin dirubah adalah variabel umur, dimana variabel ini ingin dijadikan ke dalam 3 kategori umur (dengan skala ordinal), maka

- pada kolom **Old Value**: Tandai pilihan **Range, LOWEST through value** dan ketik 15.
- Pada kolom **New Value**: Tandai pilihan **Value** dan ketik angka 1.
- Klik tombol **Add**. Artinya, semua umur 14 tahun ke bawah akan berubah menjadi kode 1 (Umur 15 tahun tidak ikut berubah menjadi kode 1)

- Selanjutnya, ulangi prosedur di atas untuk mengodekan ulang umur 15-50 tahun dengan menandai pilihan **Range**, dan mengetik 15 pada kolom pertama dan 50 pada kolom kedua. Pada kolom **New Value**: Tandai pilihan **Value** dan ketik angka 2, kemudian klik tombol **Add**.
- Untuk umur >50 tahun, pada **Old Value**: Tandai pilihan **Range**, **value HIGHEST** dan ketik 50. Pada **New Value**: Tandai pilihan **Value** dan ketik angka 3. Kemudian klik tombol **Add**. (Umur 50 tahun masuk ke kode 2, bukan kode 3.) Sehingga di layar akan tampak:



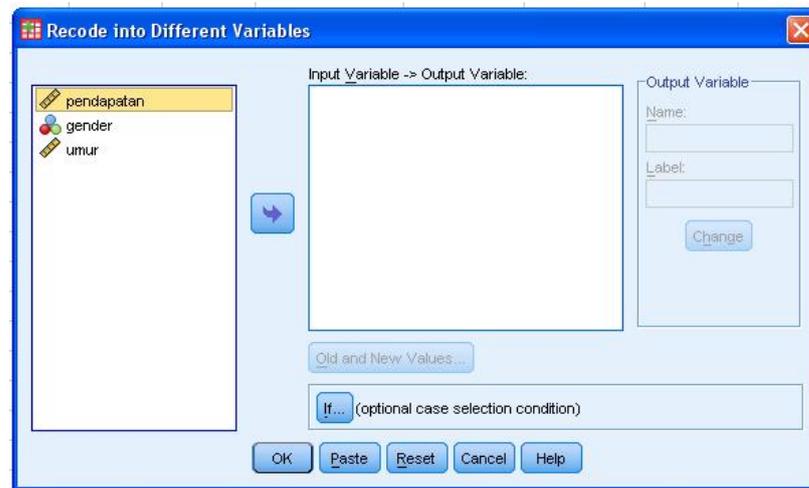
Gambar 21. Recode into Same Variables Setelah Pengisian

- Klik **Continue**
2. Klik **OK**. Maka pada **DATA EDITOR**, data dalam variabel umur telah dinyatakan dalam 3 kode, yaitu 1 untuk umur <15 tahun, 2 untuk umur 15-50 tahun, dan 3 untuk umur >50 tahun. Perlu diingat bahwa perubahan kode terjadi di dalam variabel umur sendiri, sehingga tidak ada variabel baru yang terbentuk.

Recode into Different Variables

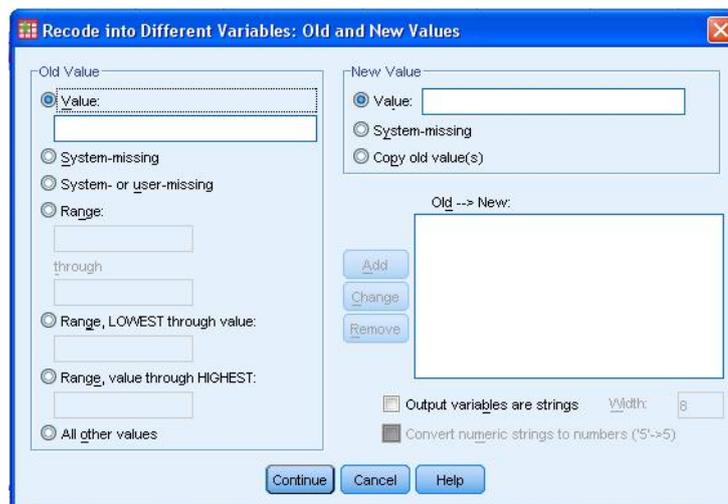
Fasilitas ini pada dasarnya sama dengan fasilitas **Recode into Different Variables** di atas, namun pengodean ulang dilakukan dengan menghasilkan variabel baru. Sehingga, setelah pengkodean ulang dengan fasilitas ini dilakukan, variabel lama masih bertahan dan muncul variabel baru yang merupakan pengodean ulang dari variabel lama tersebut. Misal, untuk kasus umur di atas. Langkah-langkahnya adalah sebagai berikut:

1. Pilih menu **Transform**, kemudian pilih **Recode into Different Variables**, maka akan muncul kotak dialog sebagai berikut



Gambar 22. Kotak dialog Recode into Different Variables

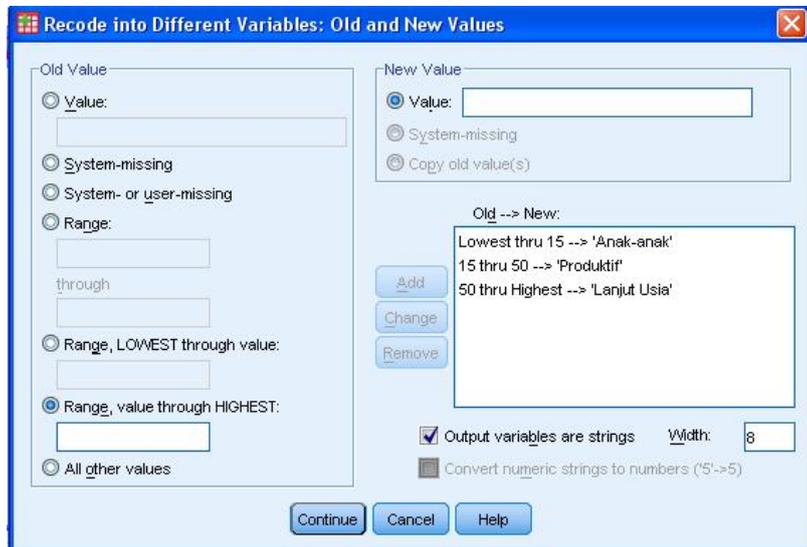
2. Pada kolom **Input Variable** → **Output Variable**, isikan variabel yang ingin dikode ulang. Dalam kasus ini, variabel umur.
3. Pada Kolom **Output Variable**, ketiklah nama variabel baru yang merupakan pengodean ulang dari variabel sebelumnya. Misal untuk kasus umur di atas, pada kolom **Name** ketiklah **Rentang Umur**, pada kolom **Label**, ketiklah **Rentang Umur Pekerja**. Klik **Change**.
4. Klik tombol **Old and New Values**, sehingga muncul kotak dialog:



Gambar 23. Recode into Different Variables: Old and New Values

Prosedur pengisian kolom **Old Value** kali ini akan sama seperti **Recode into Same Variables** (lihat bahasan sebelumnya).

Namun yang perlu diperhatikan pada kolom **New Value** adalah apabila variabel numeric akan diubah ke dalam bentuk string, maka pilihan **Output variables are strings** harus ditandai, barulah kolom **New Value** bisa diisi dengan karakter huruf. Sehingga hasilnya akan tampak pada layar:



Gambar 24. Recode into Different Variables Setelah Pengisian

Klik **Continue** untuk kembali ke kotak dialog utama.

1. Klik **OK**. Sehingga pada layar akan tampak sebuah variabel baru bernama **Rentang** seperti di bawah ini:

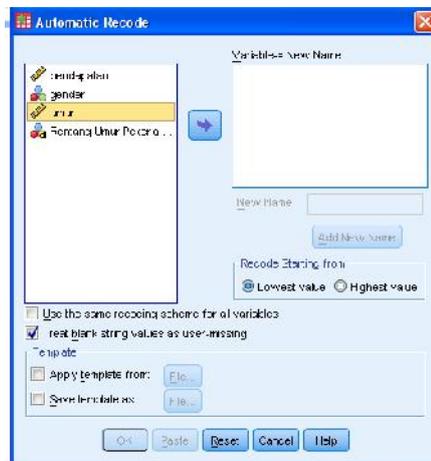
	pendapatan	gender	umur	Rentang	var
1	72.00	1.00	25.00	Produktif	
2	68.00	2.00	34.00	Produktif	
3	78.65	1.00	36.00	Produktif	
4	79.25	2.00	39.00	Produktif	
5	60.00	1.00	57.00	Lanjut Usia	
6	88.15	2.00	26.00	Produktif	
7	60.00	1.00	29.00	Produktif	
8	55.00	2.00	25.00	Produktif	
9	84.56	1.00	34.00	Produktif	
10	88.25	1.00	36.00	Produktif	
11	76.22	1.00	39.00	Produktif	
12	77.89	1.00	57.00	Lanjut Usia	
13	79.02	2.00	26.00	Produktif	
14	80.21	2.00	29.00	Produktif	
15	82.65	2.00	57.00	Lanjut Usia	
16	82.53	2.00	26.00	Produktif	
17	92.56	2.00	29.00	Produktif	

Gambar 25. Tampilan Setelah Menggunakan Recode into Different Variables

Automatic Recode

Pada SPSS 18, terdapat fasilitas Automatic Recode. Pada dasarnya, fasilitas ini memungkinkan Anda untuk mengodekan ulang variabel berdasarkan urutannya. Apabila variabel yang akan dikodekan itu adalah umur (numerik), maka **Automatic Recode** akan menghasilkan variabel baru yang datanya berupa ranking dari umur yang terekam di variabel lama. Untuk bisa menggunakan fasilitas ini, langkah-langkahnya:

1. Pilih menu **Transform**, kemudian pilih **Automatic Recode**, maka akan muncul kotak dialog sebagai berikut



Gambar 26. Kotak dialog Automatic Recode

2. Pada kolom **Variable**→**New Name**, masukkan variabel yang akan dibuat ulang kodenya, dalam hal ini variabel “umur”.
3. Ketik nama variabel baru yang ingin dimunculkan, untuk contoh di atas ketiklah **Urutan** dan klik pilihan **Add New Name**.
4. Pilih alternatif pengodean ulangannya pada kolom **Recode Starting from**. Untuk contoh di atas pilihlah **Lowest Value**. Ini berarti, umur-umur yang terekam dalam variabel umur akan diurutkan dari nilai terkecil hingga terbesar. Urutan itulah yang akan diisikan pada variabel baru, yaitu variabel “Urutan”.
5. Karena tidak ada *missing value*, klik **OK**. Sehingga pada layar akan tampak sebuah variabel baru bernama **Urutan** seperti di bawah ini:

	pendapatan	gencor	umur	Rentang	Urutan	var	var
1	72.00	1.00	25.00	Produktif	4		
2	63.00	2.00	34.00	Produktif	7		
3	73.00	1.00	26.00	Produktif	0		
4	73.25	2.00	39.00	Produktif	9		
5	63.00	1.00	57.00	Larjut Usia	1		
6	83.15	2.00	26.00	Produktif	5		
7	63.00	1.00	29.00	Produktif	6		
8	55.00	2.00	25.00	Produktif	4		
9	84.66	1.00	34.00	Produktif	7		
10	83.25	1.00	36.00	Produktif	8		
11	73.22	1.00	29.00	Produktif	9		
12	77.89	1.00	57.00	Larjut Usia	1		
13	73.02	2.00	26.00	Produktif	5		
14	83.21	2.00	29.00	Produktif	6		
15	82.85	2.00	57.00	Larjut Usia	1		
16	82.83	2.00	26.00	Produktif	5		
17	92.66	2.00	29.00	Produktif	6		
18	63.00	1.00	12.00	Anak-anak	1		
19	05.12	1.00	13.00	Anak-anak	2		
20	84.66	1.00	26.00	Produktif	5		
21	73.22	1.00	54.00	Larjut Usia	0		
22	77.89	1.00	80.00	Larjut Usia	2		
23	63.00	2.00	21.00	Produktif	3		
24	83.21	2.00	21.00	Produktif	3		

Gambar 27. Tampilan Data Setelah Menggunakan Automatic Recode

3.4 Rangkuman

Hal pertama yang dilakukan dalam penggunaan SPSS adalah memasukkan data. Pemasukan data dapat dilakukan dengan dua cara: 1) Input langsung; dan 2) Memindahkan data dari Ms.Excel ke SPSS.

Kemudian setelah data di input, tahap selanjutnya adalah mendefinisikan variabel. Dalam mendefinisikan variabel perlu dilihat beberapa tahap, diantaranya: memberi nama variabel, memberi keterangan untuk variabel kategorik, menentukan jenis variabel, dan menentukan peran variabel.

Setelah variabel selesai didefinisikan, maka data-data yang sudah dimasukkan dapat dilakukan pengelolaan data, diantaranya: 1) *Split File* – data dipisah berdasarkan kategori tertentu; 2) *Merge File* – data digabung dengan data dari file yang berbeda; 3) *Select Case* - menyeleksi data berdasarkan kriteria tertentu; 4) *Compute* – Melakukan perhitungan antar variabel sehingga diperoleh variabel baru dalam data; dan 5) *Recode* - membuat kode ulang atas data yang telah tersedia.

3.5 Latihan

Sebagai latihan, cobalah untuk memasukkan data-data pada tabel di bawah ini ke dalam editor SPSS. Tentukan tipe data dan variabel sesuai dengan karakteristik data yang ada di dalamnya.

Data Hasil Sensus Pertanian 2013 Kabupaten Puncak Jaya, Papua

Nama Kepala RT	Umur	Jenis Kelamin	Jumlah ART	
			Laki-Laki	Perempuan
Tainggen Kogoya	37	Laki-Laki	2	3
Manur Murib	38	Laki-Laki	3	2
Marlin Enumbi	40	Perempuan	3	3
Putradani Magi	41	Laki-Laki	4	3
Yon Labane	46	Laki-Laki	4	3
Isak Enumbi	45	Laki-Laki	4	3
Arlinus Enumbi	38	Laki-Laki	3	2
Erven Enumbi	35	Laki-Laki	5	3
Mitenus Murib	37	Laki-Laki	3	3
Belmin Murib	40	Laki-Laki	2	3
Pinranus Tabuni	44	Laki-Laki	3	3
Amin Tabuni	42	Laki-Laki	7	3
Ayuwarak Tabuni	39	Laki-Laki	4	3
Hati Kogoya	40	Laki-Laki	3	3
Yeli Murib	39	Laki-Laki	4	5

Setelah data dimasukkan ke dalam editor SPSS, lakukan proses-proses di bawah ini secara berurutan.

1. Pisahkan data berdasarkan jenis kelamin
2. Hitung jumlah total anggota rumah tangga dengan menggunakan menu *compute*. Beri nama variabel baru: Total_ART

3. Lakukan pengkodean ulang ke dalam variabel yang sama untuk:
 - a. Kategori Jenis Kelamin → kode 1 untuk jenis kelamin Laki-laki; dan kode 2 untuk jenis kelamin perempuan.
 - b. Kategori Umur → kode 1 untuk usia <40; kode 2 untuk usia 40-45; dan kode 3 untuk usia >45
4. Tambahkan variabel berikut pada file yang sudah dibuat diatas dengan menu *Add Variabel!*

Nama Kepala RT	Komoditi Palawija Utama	Jumlah Produksi (kg)
Tainggeng Kogoya	Ubi Jalar	900
Manur Murib	Ubi Kayu	345
Marlin Enumbi	Jagung	115
Putradani Magi	Jagung	195
Yon Labane	Kedelai	135
Isak Enumbi	Ubi Kayu	353
Arlinus Enumbi	Jagung	165
Erven Enumbi	Ubi Kayu	518
Mitenus Murib	Ubi Jalar	1045
Belmin Murib	Ubi Jalar	850
Pinranus Tabuni	Kedelai	133
Amin Tabuni	Ubi Kayu	462
Ayuwarak Tabuni	Jagung	151
Hati Kogoya	Kedelai	302
Yeli Murib	Ubi Jalar	1250

Bab IV Statistik Deskriptif

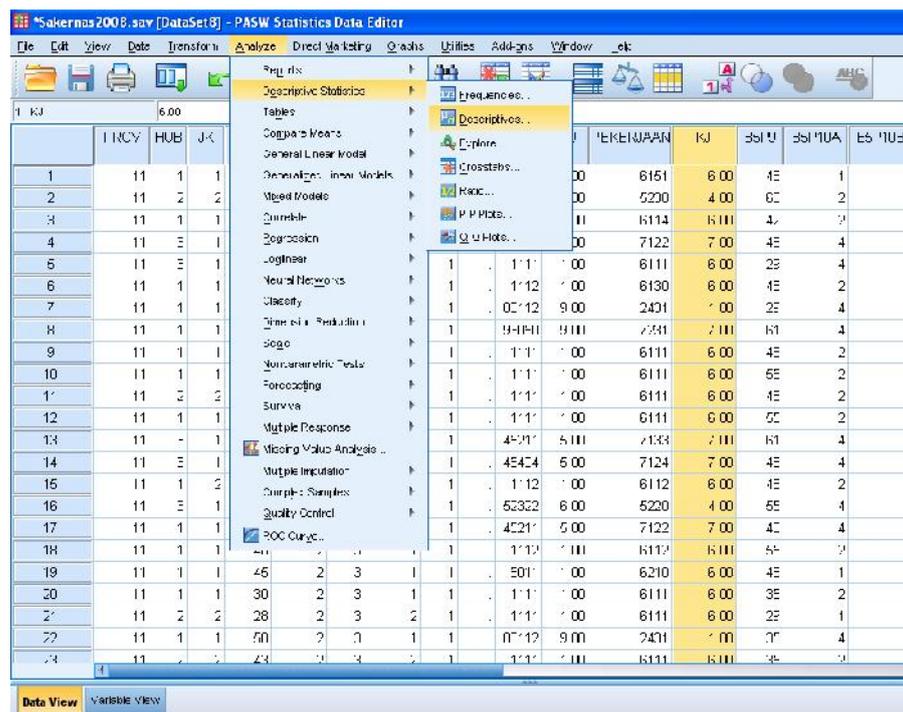
Deskripsi atau penggambaran sekumpulan data secara visual dapat dilakukan dalam dua bagian

1. Deskripsi dalam bentuk tulisan/teks, yang terdiri atas bagian-bagian yang penting yang menggambarkan isi data secara keseluruhan
2. Deskripsi dalam bentuk gambar/grafik, kegunaanya adalah untuk melengkapi deskripsi berupa teks.

4.1 Menganalisis Statistik Deskriptif

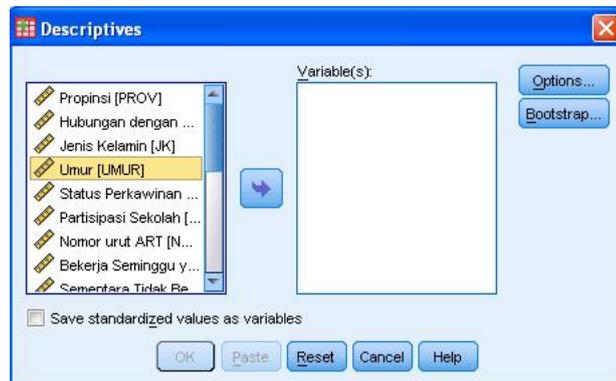
Langkah-langkah melakukan analisis deskriptif adalah sebagai berikut:

1. Klik **Analyze** pada menu utama, maka akan tampil sub menu.
2. Klik **Descriptive Statistics** dan akan muncul tampilan sebagai berikut.



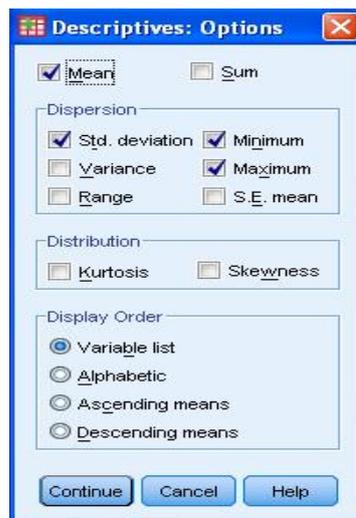
Gambar 28. Tampilan Menu Descriptive Statistic

Klik **Descriptives** maka akan tampil kotak dialog:



Gambar 29. Kotak Dialog Descriptives

Pilih Variabel yang akan dianalisis, kemudian klik OK (dalam contoh ini adalah UMUR). SPSS akan memproses data anda sesuai dengan program yang ada, jika ingin lebih dikembangkan maka klik tombol **Options**, sehingga akan muncul kotak dialog berikut:



Gambar 30. Kotak Dialog Option pada Descriptives

Pilih ukuran statistik, dispersi atau distribusi yang diinginkan, dengan mengklik pada **Check Box**. Kemudian Klik continue dan kembali pada kotak dialog sebelumnya (descriptives)

3. Klik OK, tunggu sebentar hingga proses berjalan, hasilnya akan ditampilkan pada Output window seperti berikut ini.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Umur	102552750	15	98	38.61	13.683
Valid N (listwise)	102552750				

Gambar 31. Output Descriptives Statistics

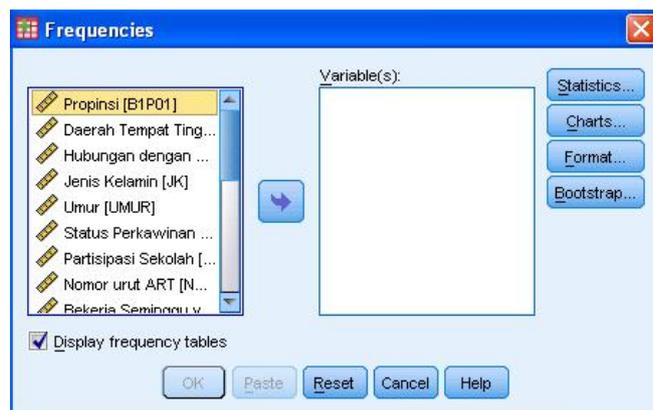
Analisis Output:

Dari output di atas, dapat dilihat bahwa jumlah sampel yang diteliti adalah 102.552.750 sampel. Rata-rata umur sampel adalah 38,61 tahun yang artinya rata-rata umur dari sampel berkisar antara 38 sampai 39 tahun dengan standar deviasi sebesar 13.683, dimana umur terendah dari sampel adalah 15 tahun dan umur tertingginya adalah 98 tahun.

4.2 Menganalisis Statistik Frekuensi

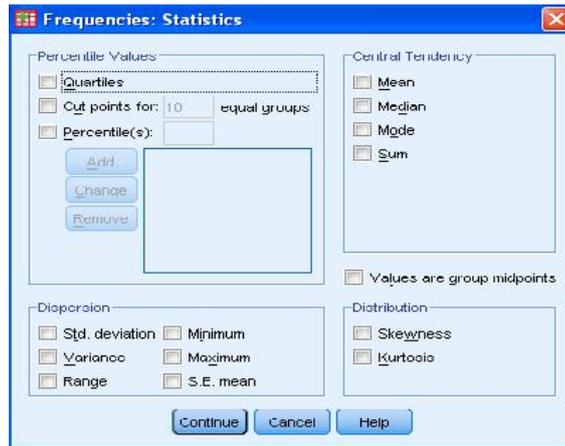
Langkah-langkah untuk melakukan analisis statistik frekuensi adalah sebagai berikut:

1. Masukkan data yang akan di analisis atau import data pada file lain
2. Klik **Analyze** pada menu Utama maka akan tampil sub menu.
3. Klik **Descriptive Statistics** maka akan muncul sub menu berikutnya Klik **Frequencies**, maka akan tampil kotak dialog



Gambar 32. Kotak Dialog Frequencies

Pilih Variabel yang akan kita analisis (misalkan Partisipasi Sekolah). Kemudian klik OK, maka SPSS akan memproses data anda sesuai dengan program yang ada. Jika ingin lebih dikembangkan, maka klik tombol **Options Statistic**, akan muncul kotak dialog berikutnya (**Frequencies Statistics**).

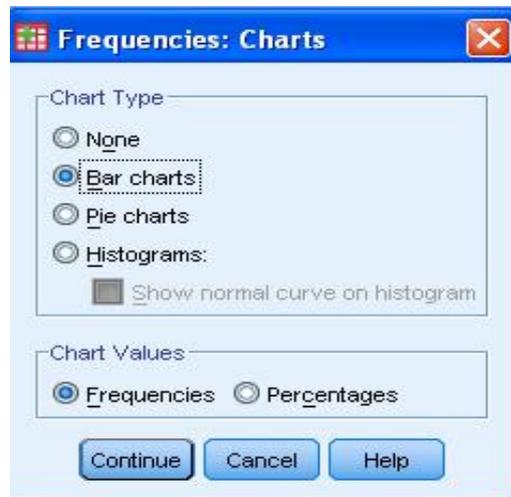


Gambar 33. Kotak Dialog Frequencies Statistics

Pilih ukuran statistik yang diinginkan dengan mengklik button **Check Box** yang disediakan kotak seperti **Percentile Values**, **Central Tendency**, **Dispersion** dan **Distribution**.

Kemudian Klik **continue** dan kembali pada kotak dialog sebelumnya (**frequencies statistics**).

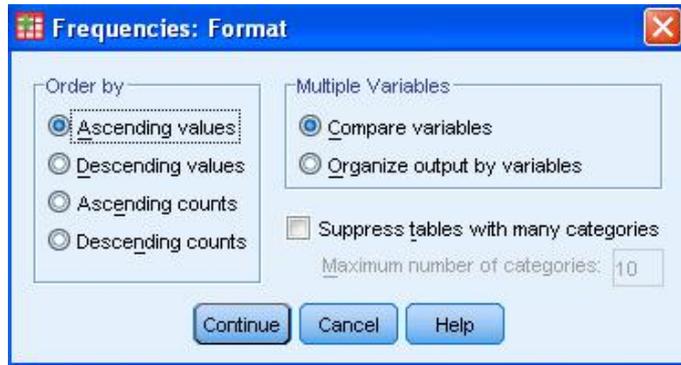
4. Klik **Options Chart** maka akan muncul:



Gambar 34. Kotak Dialog Frequencies Charts

Pilih salah satu bentuk type chart (Bar, Pie, atau Histogram), Klik **continue** dan kembali pada kotak dialog sebelumnya (**frequencies statistics**).

5. Klik **Options Format** maka akan muncul:



Gambar 35. Kotak Dialog Frequencies: Format

Pilih salah satu **Order by** (**Ascending values**, **Descending values**, **Ascending counts**, **Descending counts**) serta **Multiple Variables** (**Compare variables**, **Organize output by variables**), Klik **continue** dan kembali pada kotak dialog sebelumnya (**frequencies statistics**)

6. Klik **OK**, tunggu sebentar hingga proses selesai, hasilnya akan ditampilkan pada Output window sebagai berikut.

Statistics		
Partisipasi Sekolah		
N	Valid	102552750
	Missing	0
Percentiles	25	3.00
	50	3.00
	75	3.00

Partisipasi Sekolah					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Tidak/Belum pernah Sekolah	5468130	5.3	5.3	5.3
	Masih Bersekolah	1651484	1.6	1.6	6.9
	Tidak Bersekolah Lagi	95433136	93.1	93.1	100.0
	Total	102552750	100.0	100.0	

Gambar 36. Output Statistik Frekuensi

Analisis Output:

Tabel (atas) digunakan untuk memvalidasi apakah terdapat data yang tidak diisi oleh responden (*missing value*) atau tidak. Berdasarkan tabel di atas, diketahui bahwa tidak terdapat *missing value*. Hal ini bisa dilihat dari baris **Missing** pada kolom **Statistics** di atas yang menunjukkan nilai 0 untuk variabel Partisipasi Sekolah.

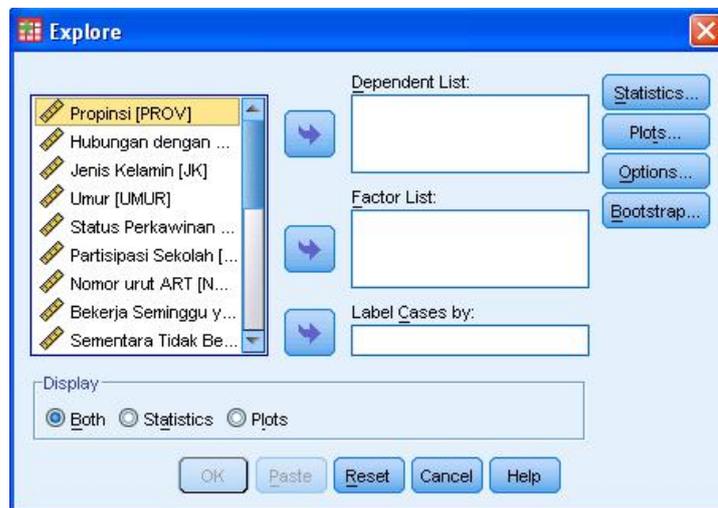
Tabel Partisipasi Sekolah menunjukkan bahwa dari 102.552.750 responden yang diwawancarai, terdapat 5.468.130 (5,3%) orang responden yang tidak/belum pernah sekolah, 1.651.484 (1,6%) orang responden yang masih sekolah dan 95.433.136 (93,1%) orang responden yang sudah tidak bersekolah lagi.

4.3 Menganalisis Statistik Eksplorasi Data

Eksplorasi data pada dasarnya berguna untuk menganalisis data lebih dalam dibandingkan tabel deskriptif maupun frekuensi yang telah dibahas sebelumnya. Fasilitas eksplorasi data digunakan untuk menguji ada tidaknya data ekstrim (*outlier*) dengan menggunakan **Boxplots** dan diagram **Steam and Leaf**. Selain itu, fasilitas ini juga bisa digunakan untuk menguji normalitas data, homogenitas varians, deskripsi statistik dan plot lainnya.

Langkah-langkah untuk melakukan analisis statistik eksplorasi adalah sebagai berikut.

1. Masukkan data yang akan di analisis atau import data pada file lain
2. Klik **Analyze** pada menu Utama maka akan tampil sub menu.
3. Klik **Descriptive Statistics** maka akan muncul sub menu berikutnya
4. Klik **Explore** maka akan tampil kotak dialog seperti di bawah ini:



Gambar 37. Kotak Dialog Explore

Semua variabel pada file data akan ditampilkan pada daftar variable. Pindahkan satu variabel numerik atau lebih ke kotak **Dependent List**. Klik tombol **Ok** untuk mendapatkan analisis default, yaitu menampilkan **Boxplots**, **Steam-leaf Plots** dan deskripsi statistik dasar untuk masing-masing variabel.

Pada fasilitas ini, bisa dilakukan pemisahan analisis untuk grup-grup *case* berdasarkan value-value dari satu atau beberapa variabel numerik atau string yang dijadikan variabel faktor. Untuk membuat variabel faktor, pindahkan satu atau beberapa variabel numerik (khususnya variabel kategori) ke kotak **Factor List**. Misalnya “gajiawal” dan “gajiskrg” sebagai variabel dependen sedangkan “jk” sebagai variabel faktor. Maka analisis akan dilakukan terhadap masing-masing grup secara terpisah, sebanyak kategori dari variabel “jk”. Jika lebih dari satu variabel yang menjadi variabel faktor, maka analisis terpisah untuk masing-masing dependen akan dihasilkan sesuai dengan banyaknya variabel faktor.

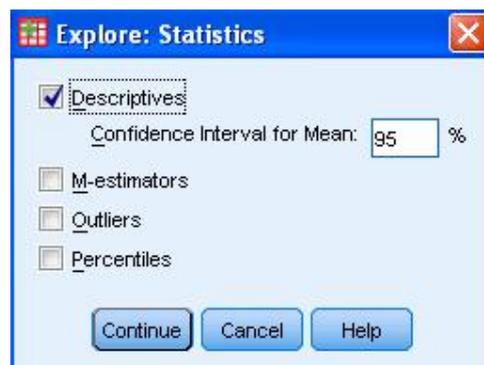
Secara opsional, dapat juga memberikan label dengan value-value variabel yang dipindahkan ke kotak **Label Cases By**. Bila variabel ini merupakan string panjang, maka hanya 15 karakter pertama saja dari valuenya yang digunakan sebagai label.

Pada kotak **Display** pilihlah salah satu alternatif berikut:

- **Both**, Menampilkan plot dan statistik. Pilihan ini merupakan *default*
- **Statistics**, Menampilkan statistik saja
- **Plots**, Menampilkan plot saja

4.3.1 Statistik Eksplorasi

Untuk mendapatkan **robust estimator** atau menampilkan *outliers*, persentil atau tabel frekuensi, pada kotak **Display** dikotak dialog utama tentukan pilihan pada **Both** atau **Statistics**. Kemudian klik tombol **Statistics**, maka akan ditampilkan kotak dialog:



Gambar 38. Kotak Dialog Statistik Eksplorasi

Minimal sebuah statistik harus dipilih dari beberapa pilihan:

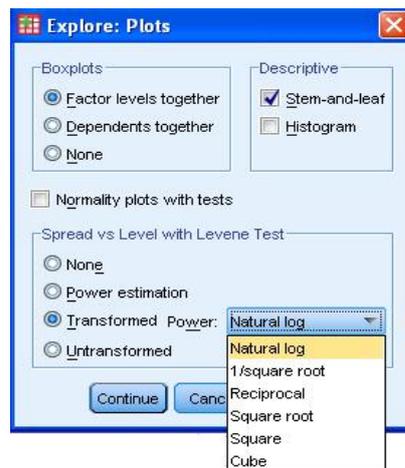
- **Descriptives**, menampilkan mean, interval konfidensi dari mean, median, 5% trimmed mean, error standard, variansi standard deviasi minimum, maksimum, range (jangkauan) dan range Interkuartil, skewness (kemiringan) dengan error standarnya, dan

kurtosis (kelancipan) dengan standardnya, Dalam keadaan default, check box ini dipilih

- **M-estimators**, menampilkan estimator lokasi robust maksimum - likelihood, yakni Huber's M-estimator ($c = 1,339$), Andrew's wave estimator ($c = 1,34 \text{ phi}$), Hampel's redescending M-estimator ($a = 1,7$, $b = 3,4$ dan $c = 8,5$) dan Tukey's biweight estimator ($c = 4,685$)
- **Outliers**, menampilkan 5 nilai data terkecil dan 5 nilai data terbesar yang dalam output dilabelkan dengan Extreme Values.
- **Percentiles**, menampilkan persentil ke 5, 10, 25, 75, 90, dan 95. Juga menampilkan Tukey's hinges, yaitu persentil ke 25 dan 75.

4.3.2 Plot Eksplorasi

Untuk membuat histogram, plot dan uji normalitas, atau plot *Spread vs Level* beserta statistik Levene's, maka pada kotak **Display** di kotak dialog utama tentukan pilihan pada **Both** atau **plots**, Kemudian klik tombol **Plots**. maka akan ditampilkan kotak dialog:



Gambar 39. Tampilan Kotak Dialog Plot Eksplorasi

Pada kotak **Boxplots**, pilihlah salah satu alternatif:

- **Factor Level Together**, untuk setiap variabel dependen yang diberikan, pilihan ini akan menampilkan boxplot untuk masing-masing grup. Apabila tidak ada variabel faktor yang diseleksi maka akan ditampilkan boxplot untuk keseluruhan data. Pilihan ini merupakan default.
- **Dependents Together**, untuk setiap grup yang diberikan, pilihan ini akan menampilkan Boxplot untuk masing-masing variabel dependen. Pilih alternatif ini jika hendak membandingkan variabel-variabel untuk sebuah grup tersendiri.
- **None**, Tidak menampilkan boxplot

Pada kotak **Descriptive**, pilih salah satu alternatif atau bisa keduanya.:

- **Steam-and-leaf**, menampilkan Steam - Leaf plot pada output file. Dalam keadaan default, check box ini dipilih.
- **Histogram**, *range* dari nilai-nilai data akan dibagi ke dalam interval yang sama, dan banyaknya *case* dari masing-masing interval akan ditampilkan.

Untuk plot probabilitas normal dan probabilitas *detrended*, terdapat pilihan **Normality Plots With Tests**, pilihan ini akan menghasilkan plot probabilitas normal dan probabilitas *detrended*. Selain itu, akan menghitung statistik Shapiro-Wilks dan statistik kolmogorov-Smirnov dengan taraf signifikansi liliefors untuk uji normalitas. Statistik Shapiro-Wilks tidak akan dihasilkan bila ukuran sampel melebihi 50. Pada test ini, H_0 adalah data berdistribusi normal. Artinya bila $p\text{-value} < \alpha$, maka tolak H_0 yang artinya data tidak berdistribusi normal.

Lewat kotak Spread Vs Level With Levene Test, bisa ditampilkan uji Levene untuk homogenitas varians dan slope garis regresi untuk seluruh plot Spread vs Level. Jika tidak ada variabel faktor yang diseleksi, maka tidak akan ada plot Spread-vs-Level yang dihasilkan. Pilih salah satu alternatif ini:

- **None**, tidak menampilkan uji Levene dan plot Spread-vs-Level, Pilihan ini merupakan default.
- **Power estimation**, Untuk masing-masing grup, akan dihasilkan plot logaritma natural (ln) dari median terhadap logaritma natural range interkuartil. Gunakan metode ini untuk menentukan transformasi yang paling cocok.
- **Transformed**, pada pilihan ini, data akan ditransformasikan menurut power yang ditentukan oleh user (sebaiknya berdasarkan harga power yang dihasilkan lewat pilihan Power estimation). Pilihan ini juga akan menampilkan plot range interkuartil dan median dari data hasil transformasi.
- **Untransformed**, Tidak ada transformasi data (Transformasi dengan power-1)

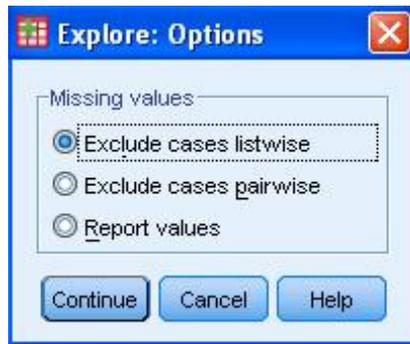
Untuk memilih power yang akan digunakan dalam transformasi data, pilih salah satu alternatif berikut dari daftar drop-down:

- Natural Log (Ln), data ditransformasi dengan Ln. Sehingga $x^* = \ln(x)$
- 1/squareroot, data ditransformasi dg kebalikan akar-kuadratnya $x^* = \frac{1}{\sqrt{x}}$
- Reciprocal, data ditransformasi dengan kebalikannya $x^* = 1/x$
- Square root, data ditransformasi dengan akar-kuadratnya $x^* = \sqrt{x}$

- Square, data ditransformasi dengan kuadratnya ($x^*=x^2$)
- Cube, data ditransformasi dengan dipangkatkan 3 ($x^*=x^3$)

4.3.3 Option-Option Prosedur Eksplorasi

Untuk menangani keberadaan missing value, klik tombol **Option** pada kotak dialog utama, maka akan ditampilkan kotak dialog :



Gambar 40. Tampilan Kotak Dialog Option Eksplorasi

Pada kotak Missing Value dapat memilih salah satu dari alternatif berikut:

- **Exclude Cases Listwise**, kasus-kasus yang dinyatakan *missing* untuk sembarang variabel dependen maupun variabel faktor akan dikeluarkan dalam seluruh analisis. Pilihan ini merupakan *default*.
- **Exclude Cases Pairwise**, kasus-kasus yang tidak mempunyai missing value untuk variabel dalam suatu sel tertentu akan diikutsertakan dalam analisis sel tersebut. Akan tetapi kasus-kasus tersebut boleh mempunyai missing value untuk variabel-variabel di luar sel yang sedang dianalisis.
- **Report Value**, *missing value* untuk variabel faktor akan diperlakukan sebagai kategori tersendiri. semua output akan dihasilkan untuk kategori tambahan ini. Tabel frekuensi akan memasukkan kategori untuk missing value.

4.4 Rangkuman

Deskripsi atau penggambaran sekumpulan data secara visual dapat dilakukan dalam dua bagian, yaitu: 1) Deskripsi dalam bentuk tulisan/teks; dan 2) Deskripsi dalam bentuk gambar/grafik.

Dalam analisisnya, dibagi umum menjadi analisis statistik deskriptif, analisis statistik frekuensi, dan analisis statistik eksplorasi data. Masing-masing analisis memiliki tujuan dan keunggulan tersendiri. Analisis deskriptif sangat membantu dalam meringkas perbandingan beberapa variabel data skala dalam satu tabel dan dapat digunakan untuk melakukan pengamatan outlier/penyimpangan data. Analisis frekuensi sangat berguna untuk memperoleh ringkasan suatu variabel

individual. Sementara analisis eksplorasi data melengkapinya dengan berbagai tampilan visual dan ringkasan numerik data. Pada analisis ini variabel dependen harus data skala dan variabel grouping berupa data nominal atau ordinal (dalam kategori).

4.5 Latihan

Buka kembali file yang telah dibuat pada latihan sebelumnya seperti berikut ini

Nama Kepala RT	Umur	JK	Jumlah ART		Total ART	Komoditi Palawija	Produksi (kg)
			L	P			
Tainggen Kogoya	1	1	2	3	5	Ubi Jalar	900
Manur Murib	1	1	3	2	5	Ubi Kayu	345
Marlin Enumbi	2	2	3	3	6	Jagung	115
Putradani Magi	2	1	4	3	7	Jagung	195
Yon Labane	3	1	4	3	7	Kedelai	135
Isak Enumbi	2	1	4	3	7	Ubi Kayu	353
Arlinus Enumbi	1	1	3	2	5	Jagung	165
Erven Enumbi	1	1	5	3	8	Ubi Kayu	518
Mitenus Murib	1	1	3	3	6	Ubi Jalar	1045
Belmin Murib	2	1	2	3	5	Ubi Jalar	850
Pinranus Tabuni	2	1	3	3	6	Kedelai	133
Amin Tabuni	2	1	7	3	10	Ubi Kayu	462
Ayuwarak Tabuni	1	1	4	3	7	Jagung	151
Hati Kogoya	2	1	3	3	6	Kedelai	302
Yeli Murib	1	1	4	5	9	Ubi Jalar	1250

1. Lakukan analisis deskriptif terhadap variabel umur !
2. Kemudian lakukan analisis frekuensi terhadap variabel Total ART!

Bab V Uji Validitas dan Reliabilitas

5.1. Pengujian Validitas

Validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan atau kesahihan sesuatu instrumen (Arikunto, 2002: 144). Sebuah instrumen dikatakan valid apabila mampu mengukur apa yang diinginkan dan dapat mengungkap data dari variabel yang diteliti secara tepat. Tinggi rendahnya validitas instrumen menunjukkan sejauh mana data yang terkumpul tidak menyimpang dari gambaran tentang validitas yang dimaksud.

Cara yang dipakai dalam menguji tingkat validitas adalah dengan variabel internal, yaitu menguji apakah terdapat kesesuaian antara bagian instrumen secara keseluruhan. Untuk mengukurnya menggunakan analisis butir. Pengukuran pada analisis butir yaitu dengan cara skor-skor yang ada kemudian dikorelasikan dengan menggunakan Rumus korelasi *product moment* yang dikemukakan oleh Pearson dalam Arikunto, (2002: 146) sebagai berikut:

$$r_{xy} = \frac{\sum xy - \left\{ \sum x \right\} \left\{ \sum y \right\}}{N} \sqrt{\left[\frac{\sum x^2 - \left(\sum x \right)^2}{N} \right] \left[\frac{\sum y^2 - \left(\sum y \right)^2}{N} \right]}$$

dengan pengertian

- r_{xy} : koefisien korelasi antara x dan y r_{xy}
- N : Jumlah Subyek
- X : Skor item
- Y : Skor total
- X : Jumlah skor items
- Y : Jumlah skor total
- X^2 : Jumlah kuadrat skor item
- Y^2 : Jumlah kuadrat skor total

(Suharsimi Arikunto, 2002 : 146)

Kesesuaian harga r_{xy} diperoleh dari perhitungan dengan menggunakan rumus diatas dikonsultasikan dengan tabel harga regresi moment dengan korelasi harga r_{xy} lebih besar atau sama dengan regresi tabel, maka butir instrumen tersebut valid dan jika r_{xy} lebih kecil dari regresi tabel maka butir instrumen tersebut tidak valid.

5.2. Pengujian Reliabilitas

Reliabilitas adalah sesuatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data karena instrumen tersebut sudah baik (Arikunto, 2002: 154). Pada penelitian ini untuk mencari reliabilitas instrumen menggunakan rumus alpha , karena instrumen dalam penelitian ini berbentuk angket atau daftar pertanyaan yang skornya merupakan rentangan antara 1-5 dan uji validitas menggunakan item total, dimana untuk mencari reliabilitas instrumen yang skornya bukan 1 dan 0, misalnya angket atau soal bentuk uraian maka menggunakan rumus alpha :

Reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya, maksudnya apabila dalam beberapa pelaksanaan pengukuran terhadap kelompok yang sama diperoleh hasil yang relatif sama (Syarifuddin Azwar, 2000 : 3). Dalam penelitian ini, uji reliabilitas dilakukan dengan menggunakan teknik Formula Alpha Cronbach dan dengan menggunakan program SPSS 15.0 for windows.

Rumus :

$$= \frac{k}{k-1} \left(1 - \frac{\sum S^2 j}{S^2 x} \right)$$

Keterangan :

- = koefisien reliabilitas alpha
- k = jumlah item
- Sj = varians responden untuk item I
- Sx = jumlah varians skor total

Indikator pengukuran reliabilitas menurut Sekaran (2000: 312) yang membagi tingkatan reliabilitas dengan kriteria sebagai berikut :Jika alpha:

1. 0,8-1,0 = Reliabilitas baik
2. 0,6-0,799 = Reliabilitas diterima
3. kurang dari 0,6 = Reliabilitas kurang baik/ditolak

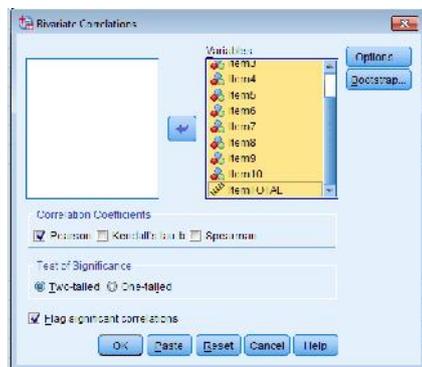
Contoh: Mengukur Validitas Kuesioner

Misalkan terdapat 10 item pertanyaan dengan skala data (jawaban) 1 s.d 4 sebagai berikut:

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	ItemTOTAL
1	3.00	4.00	3.00	4.00	4.00	3.00	3.00	3.00	3.00	3.00	33.00
2	4.00	3.00	3.00	4.00	3.00	3.00	3.00	3.00	3.00	3.00	32.00
3	2.00	2.00	1.00	3.00	2.00	2.00	3.00	1.00	2.00	3.00	21.00
4	3.00	4.00	4.00	3.00	3.00	3.00	4.00	3.00	3.00	4.00	34.00
5	3.00	4.00	3.00	3.00	3.00	4.00	3.00	4.00	4.00	3.00	34.00
6	3.00	2.00	4.00	4.00	3.00	4.00	4.00	3.00	4.00	4.00	35.00
7	2.00	3.00	3.00	4.00	4.00	4.00	3.00	4.00	2.00	3.00	32.00
8	1.00	2.00	2.00	1.00	2.00	2.00	1.00	3.00	4.00	3.00	21.00
9	4.00	2.00	3.00	3.00	4.00	2.00	1.00	1.00	4.00	4.00	28.00
10	3.00	3.00	3.00	4.00	4.00	4.00	4.00	4.00	3.00	3.00	36.00
11	4.00	4.00	3.00	4.00	4.00	3.00	4.00	4.00	4.00	3.00	36.00
12	3.00	2.00	1.00	2.00	3.00	1.00	1.00	2.00	3.00	3.00	21.00

Maka langkah-langkah untuk menguji validitas pertanyaan tersebut adalah:

1. Jumlahkan semua item kedalam sebuah variabel ItemTOTAL
2. Klik Menu Analyze – Correlate – Bivariate sehingga tampil



3. Pindahkan semua variabel ke kotak Variables, kemudian centang Pearson dan pilih Two Tiled dan terakhir klik OK

4. Output yang dihasilkan adalah sbb:

		Correlations										
		Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item TOTAL
Item1	Pearson Correlation	1	.327	.392	.549	.569	.093	.240	-.008	.301	.029	.534
	Sig. (2-tailed)		.299	.207	.064	.053	.774	.453	.981	.341	.928	.074
Item2	Pearson Correlation	.327	1	.497	.445	.435	.498	.578*	.644*	.033	-.321	.706*
	Sig. (2-tailed)	.299		.100	.148	.157	.099	.049	.024	.918	.310	.010
Item3	Pearson Correlation	.392	.497	1	.561	.469	.733**	.592*	.500	.344	.408	.871**
	Sig. (2-tailed)	.207	.100		.058	.124	.007	.043	.098	.274	.188	.000
Item4	Pearson Correlation	.549	.445	.561	1	.656*	.685*	.750**	.369	-.219	-.082	.791**
	Sig. (2-tailed)	.064	.148	.058		.021	.014	.005	.237	.495	.801	.002
Item5	Pearson Correlation	.569	.435	.469	.656*	1	.393	.253	.362	.040	-.104	.645*
	Sig. (2-tailed)	.053	.157	.124	.021		.206	.428	.248	.902	.747	.024
Item6	Pearson Correlation	.093	.498	.733**	.685*	.393	1	.752**	.751**	.030	.026	.839**
	Sig. (2-tailed)	.774	.099	.007	.014	.206		.005	.005	.926	.935	.001
Item7	Pearson Correlation	.240	.578*	.592*	.750**	.253	.752**	1	.551	-.152	-.088	.780**
	Sig. (2-tailed)	.453	.049	.043	.005	.428	.005		.064	.638	.786	.003
Item8	Pearson Correlation	-.008	.644*	.500	.369	.362	.751**	.551	1	.139	-.412	.686*
	Sig. (2-tailed)	.981	.024	.098	.237	.248	.005	.064		.666	.184	.014
Item9	Pearson Correlation	.301	.033	.344	-.219	.040	.030	-.152	.139	1	.104	.215
	Sig. (2-tailed)	.341	.918	.274	.495	.902	.926	.638	.666		.747	.502
Item10	Pearson Correlation	.029	-.321	.408	-.082	-.104	.026	-.088	-.412	.104	1	.018
	Sig. (2-tailed)	.928	.310	.188	.801	.747	.935	.786	.184	.747		.956
ItemTO TAL	Pearson Correlation	.534	.706*	.871**	.791**	.645*	.839**	.780**	.686*	.215	.018	1
	Sig. (2-tailed)	.074	.010	.000	.002	.024	.001	.003	.014	.502	.956	

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Analisis butir pertanyaan (setiap nilai r-hitung)

Perhatikan masing-masing nilai korelasi item pertanyaan dengan item total. Oleh karena kita menggunakan sampel 12 responden maka derajat kebebasan (*degrees of freedom*) adalah $n-2=10$. Sekarang cari nilai tabel r dengan derajat bebas (df) = 10 pada Alpha 0,05 diperoleh 0,5760. Inilah batasan (*cut off*) nilai yang kita pakai. Nilai korelasi harus lebih besar dari 0,5760.

pada item 1, (r -hitung = 0,534) < (r -tabel = 0,576), maka tidak valid
 pada item 2, (r -hitung = 0,706) > (r -tabel = 0,576), maka valid
 pada item 3, (r -hitung = 0,871) > (r -tabel = 0,576), maka valid
 pada item 4, (r -hitung = 0,791) > (r -tabel = 0,576), maka valid
 pada item 5, (r -hitung = 0,645) > (r -tabel = 0,576), maka valid

pada item 6, (r-hitung = 0, 839) > (r-tabel = 0,576), maka valid
 pada item 7, (r-hitung = 0, 780) > (r-tabel = 0,576), maka valid
 pada item 8, (r-hitung = 0, 686) > (r-tabel = 0,576), maka valid
 pada item 9, (r-hitung = 0, 215) < (r-tabel = 0,576), maka tidak valid
 pada item 10, (r-hitung = 0,018) < (r-tabel = 0,576), maka tidak valid

b. Dari hasil analisis di atas, didapat bahwa :

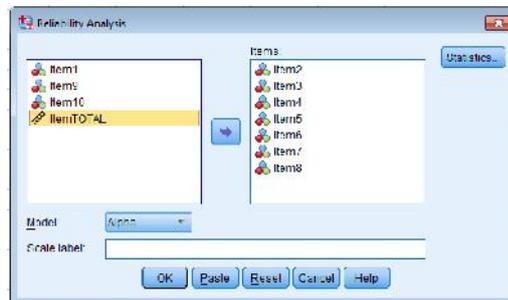
1. Item 2, item 3, item 4, item 5, item 6, item 7 dan item 8, valid, karena nilai r hitung > r tabel atau *Significant (two tailed)* < 0,05
2. Sedangkan skor 1, skor 9 dan skor 10 tidak valid, sehingga diperlukan perbaikan pada item-item skor tersebut. Berarti kita hanya menggunakan tujuh pertanyaan dalam penelitian (yang valid).

Contoh: Mengukur Reliabilitas Kuesioner

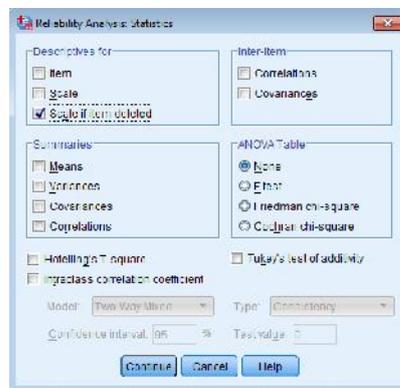
Dengan menggunakan contoh data di atas, maka berdasarkan hasil uji validitas, maka uji reliabilitas hanya dilakukan pada 7 (tujuh) item pertanyaan saja karena yang lainnya tidak valid.

Langkah-langkah untuk menguji reliabilitas adalah:

1. Klik Menu **Analyze – Scale – Reliability Analysis** sehingga tampil



2. Klik tab **Statistics..** sehingga tampil



3. Beri tanda centang pada **Scale if item deleted** lalu klik **Continue** kemudian **Ok**.
4. Output yang dihasilkan adalah sbb:

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item2	17.9167	22.811	.654	.883
Item3	18.0833	21.902	.709	.877
Item4	17.5833	21.720	.733	.874
Item5	17.5833	24.811	.515	.897
Item6	17.9167	20.629	.842	.860
Item7	18.0000	19.818	.753	.872
Item8	17.9167	21.356	.670	.882

Berdasarkan tabel output diatas maka dapat disimpulkan bahwa ke-tujuh item pertanyaan mempunyai reliabilitas baik karena semuanya mempunyai nilai alph diatas 0,8.

Bab VI Penutup

6.1 Simpulan

1. Indikator hasil belajar dari modul ini adalah : setelah mengikuti diklat ini peserta memahami:
 - a. cara pengolahan, penyusunan laporan dan analisis tabulasi data berupa deskriptif, frekuensi, dan eksplorasi data
 - b. cara perhitungan, pengolahan, penyusunan laporan dan analisis uji perbedaan
 - c. cara perhitungan, pengolahan, penyusunan laporan dan analisis uji asosiasi
2. Pengolahan data dengan SPSS, dimulai dengan mempersiapkan data sesuai dengan jenis datanya.
3. Setelah data siap, data dapat dikelola dengan beberapa tahap seperti mendefinisikan variabel, memodifikasi data, dan mentransformasi data.
4. Data yang telah siap, selanjutnya dapat dianalisis deskriptif diantaranya analisis deskripsi, analisis frekuensi, dan eksplorasi data. Analisis deskriptif bertujuan melihat gambaran data secara umum.
5. Dalam analisis inferensia, didapatkan gambaran yang lebih mendalam tentang data yang diolah. Analisisnya dapat dengan Uji t Untuk Satu Sampel (One-Sample T Test), Uji t Untuk Dua Sampel Bebas (Independent-Sample T Test), Uji t Untuk Dua Sampel Berpasangan (Paired-Samples T Test), dan Uji Perbedaan lebih dari dua Sampel (One-Way ANOVA).

6.2 Tindak Lanjut

Setelah selesai mengikuti Pendidikan dan Pelatihan, diharapkan alumni diklat lebih mampu dan berkesempatan meningkatkan peran sertanya dalam penyelesaian laporan pekerjaannya. Untuk memenuhi harapan tersebut, perlu dilakukan beberapa tindak lanjut sebagai berikut:

1. Meneliti, mempelajari, dan lebih memperhatikan keterkaitan antara data-data yang dimiliki dan jenis pengolahan SPSS yang dibutuhkan untuk mendapatkan analisis data yang relevan dan bermanfaat
2. Apabila peserta sering menggunakan *software* SPSS, maka perlu mempelajari keahlian dan keterampilan penggunaan fungsi-fungsi statistik lain yang terdapat di dalam SPSS serta memperbaharui versi dari SPSS.
3. Meningkatkan dan membantu rekan-rekan dalam menyebarkan ilmu dan keahlian di dalam penggunaan *software* SPSS, sehingga dapat membantu pengolahan dalam penyusunan laporan atau penelitian.

4. Melakukan tindak lanjut 1, 2, dan 3 secara teratur, baik secara periodik atau terus menerus, sesuai dengan tugas pokok dan fungsinya dengan memperhatikan kebijakan di lingkungan kerja dan instansinya.

LAMPIRAN

Latihan. 3.5

1. Pisahkan data berdasarkan jenis kelamin

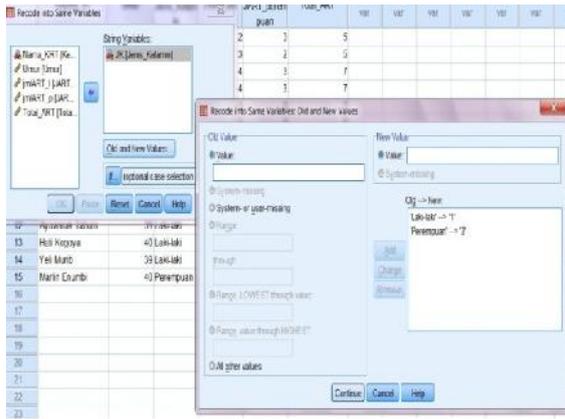
	Kepala_RT	Umur	Jenis_Kelamin	JART_laki	JART_perempuan
1	Tainggen Kogoya	37	Laki-laki	2	3
2	Manur Murib	38	Laki-laki	3	2
3	Putradani Mugi	41	Laki-laki	4	3
4	Yon Labane	46	Laki-laki	4	3
5	Isak Enumbi	45	Laki-laki	4	3
6	Arlinus Enumbi	38	Laki-laki	3	2
7	Erven Enumbi	35	Laki-laki	5	3
8	Mitenus Murib	37	Laki-laki	3	3
9	Belmin Murib	40	Laki-laki	2	3
10	Pimianus Tabuni	44	Laki-laki	3	3
11	Amin Tabuni	42	Laki-laki	7	3
12	Ayuwarak Tabuni	39	Laki-laki	4	3
13	Hati Kogoya	40	Laki-laki	3	3
14	Yeli Murib	39	Laki-laki	4	5
15	Marlin Enumbi	40	Perempuan	3	3

2. Hitung jumlah total anggota rumah tangga dengan menggunakan menu *compute*. Beri nama variabel baru: Total_ART

The screenshot shows the 'Compute Variable' dialog box with 'Total_ART' as the target variable and the expression 'JART_laki + JART_perempuan'. A 'Compute Variable Type...' dialog is also open, showing 'Label' as 'Total ART' and 'Type' as 'Numeric'. The background data table now includes the 'Total_ART' column.

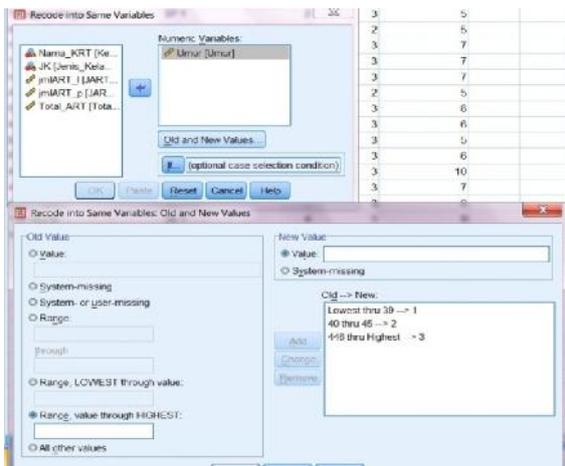
	Kepala_RT	Umur	Jenis_Kelamin	JART_laki	JART_perempuan	Total_ART
1	Tainggen Kogoya	37	Laki-laki	2	3	5.00
2	Manur Murib	38	Laki-laki	3	2	5.00
3	Putradani Mugi	41	Laki-laki	4	3	7.00
4	Yon Labane	46	Laki-laki	4	3	7.00
5	Isak Enumbi	45	Laki-laki	4	3	7.00
6	Arlinus Enumbi	38	Laki-laki	3	2	5.00
7	Erven Enumbi	35	Laki-laki	5	3	8.00
8	Mitenus Murib	37	Laki-laki	3	3	6.00
9	Belmin Murib	40	Laki-laki	2	3	5.00
10	Pimianus Tabuni	44	Laki-laki	3	3	6.00
11	Amin Tabuni	42	Laki-laki	7	3	10.00
12	Ayuwarak Tabuni	39	Laki-laki	4	3	7.00
13	Hati Kogoya	40	Laki-laki	3	3	6.00
14	Yeli Murib	39	Laki-laki	4	5	9.00
15	Marlin Enumbi	40	Perempuan	3	3	6.00

3. Lakukan pengkodean ulang ke dalam variabel yang sama untuk:
 - a. Kategori Jenis Kelamin → kode 1 untuk jenis kelamin Laki-laki; dan kode 2 untuk jenis kelamin perempuan
 - b. Kategori Umur → kode 1 untuk usia <40; kode 2 untuk usia 40-45; dan kode 3 untuk usia >45



persiapan data 1.sav [DataSet1] - PASW Statistics Data Editor

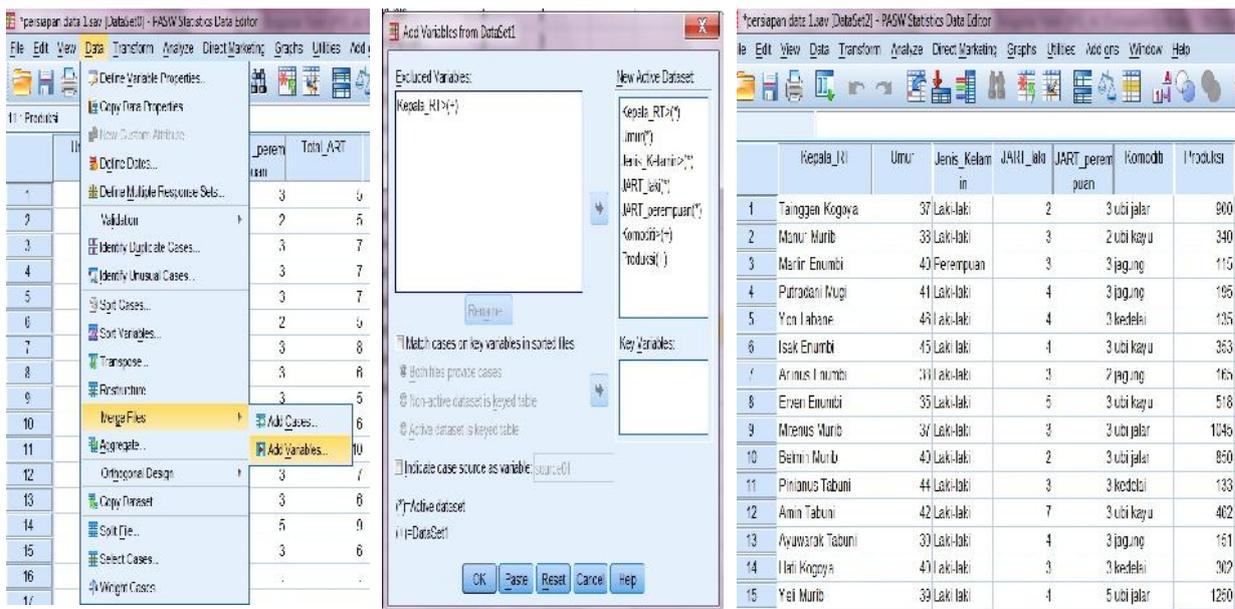
	Kepala_RT	Umur	Jenis_Kelamin	JART_laki	JART_perempuan	Total_ART
1	Tainggeng Kogoya	37	1	2	3	5
2	Manur Murib	36	1	3	2	5
3	Putradani Mugi	41	1	4	3	7
4	Yon Labane	46	1	4	3	7
5	Isak Enumbi	45	1	4	3	7
6	Arlinus Enumbi	36	1	3	2	5
7	Erven Enumbi	35	1	5	3	8
8	Mbenus Murib	37	1	3	3	6
9	Belmin Murib	40	1	2	3	5
10	Pinianus Tabuni	44	1	3	3	6
11	Amin Tabuni	42	1	7	3	10
12	Ayuwarak Tabuni	39	1	4	3	7
13	Hati Kogoya	40	1	3	3	6
14	Yeli Murib	39	1	4	5	9
15	Marlin Enumbi	40	2	3	3	6



persiapan data 1.sav [DataSet1] - PASW Statistics Data Editor

	Kepala_RT	Umur	Jenis_Kelamin	JART_laki	JART_perempuan	Total_ART
1	Tainggeng Kogoya	1	1	2	3	5
2	Manur Murib	1	1	3	2	5
3	Putradani Mugi	2	1	4	3	7
4	Yon Labane	2	1	4	3	7
5	Isak Enumbi	3	1	4	3	7
6	Arlinus Enumbi	2	1	3	2	5
7	Erven Enumbi	1	1	5	3	8
8	Mbenus Murib	1	1	3	3	6
9	Belmin Murib	1	1	2	3	5
10	Pinianus Tabuni	2	1	3	3	6
11	Amin Tabuni	2	1	7	3	10
12	Ayuwarak Tabuni	2	1	4	3	7
13	Hati Kogoya	1	1	3	3	6
14	Yeli Murib	2	1	4	5	9
15	Marlin Enumbi	1	2	3	3	6

4. Tambahkan variabel berikut pada file yang sudah dibuat diatas dengan menu *Add Variabel*



Latihan 4.5

1. Lakukan analisis deskriptif terhadap variabel umur !

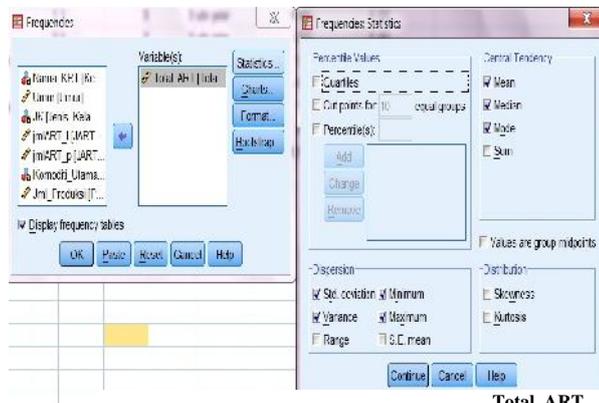


Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Umur	15	1	3	1.60	.632
Valid N (listwise)	15				

Dari output di atas, dapat dilihat bahwa jumlah sampel yang diteliti adalah 15 sampel. Rata-rata umur (yang sudah dikategori) sampel adalah 1,6. Artinya kategori umur penduduk cenderung mendekati kategori 2 (penduduk berusia 40-45 tahun).

2. Kemudian lakukan analisis frekuensi terhadap variabel Total ART!



Statistics

Total_ART		
N	Valid	15
	Missing	0
Mean		6.6000
Median		6.0000
Mode		5.00 ^a
Std. Deviation		1.50238
Variance		2.257
Minimum		5.00
Maximum		10.00

a. Multiple modes exist. The smallest value is shown

Total_ART

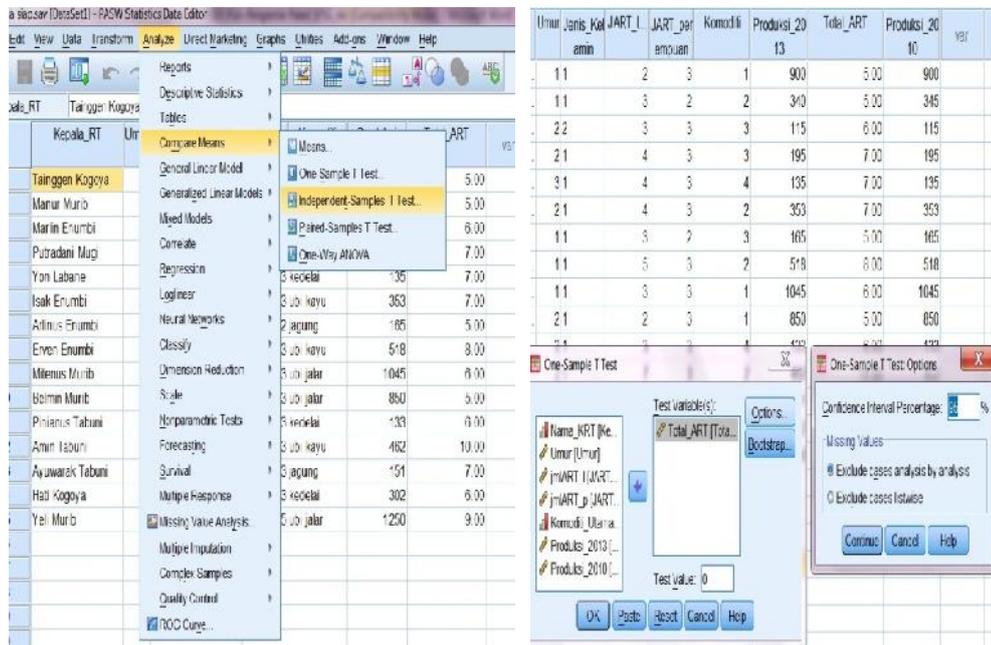
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5.00	4	26.7	26.7	26.7
	6.00	4	26.7	26.7	53.3
	7.00	4	26.7	26.7	80.0
	8.00	1	6.7	6.7	86.7
	9.00	1	6.7	6.7	93.3
	10.00	1	6.7	6.7	100.0
	Total	15	100.0	100.0	

Pada output statistik terlihat tidak terdapat data *missing*. Total ART minimum adalah 5, sementara total ART maksimum adalah 10. Tabel Total_ART menunjukkan bahwa dari 15 sampel, masing-masing terdapat 6,7% untuk Total ART sebanyak 8, 9, dan 10 orang. Dan masing-masing terdapat 26,7% untuk Total ART sebanyak 5, 6, dan 7 orang.

Latihan 5.5

1. Dari data sampel yang telah ada sebelumnya. Ingin diketahui kebenaran dugaan bahwa rata-rata jumlah anggota rumah tangga (Total_ART) lebih

dari 5 orang. Dengan tingkat kepercayaan 95%, buktikanlah dugaan tersebut!



H_0 = Rata-rata jumlah anggota rumah tangga kurang dari 5 orang

H_1 = Rata-rata jumlah anggota rumah tangga lebih dari 5 orang

One-Sample Statistics

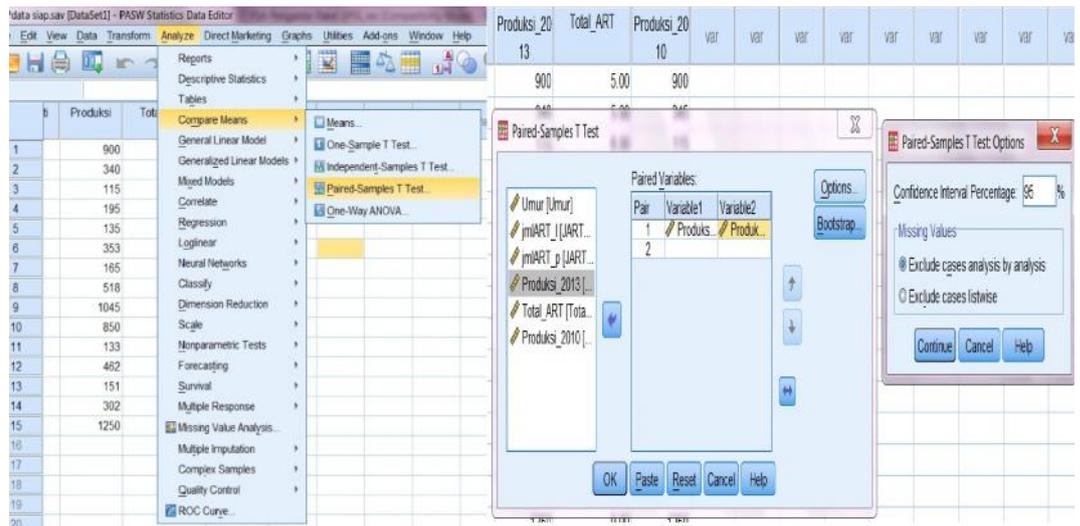
	N	Mean	Std. Deviation	Std. Error Mean
Total_ART	15	6.6000	1.50238	.38791

One-Sample Test

	Test Value = 5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Total_ART	4.125	14	.001	1.60000	.7680	2.4320

Dari output diketahui nilai statistik uji $t = 4,125$, derajat bebas $df=14$, nilai signifikansi sig. (2-tailed) = 0.001. Tingkat signifikansi yang digunakan pada pengujian ini $\alpha=5\%$ (0,05), maka untuk uji dua sisi menjadi 2,5%. Hipotesis nol di atas ditolak karena $\text{sig.}(2 \text{ tailed})/2 < 0.025$. Oleh karena itu, dengan tingkat kepercayaan 95% dapat disimpulkan bahwa rata-rata Total ART tersebut lebih dari 5 orang

- Apabila diketahui data hasil produksi komoditi tahun 2010 (seperti soal) Hitunglah dugaan bahwa terdapat kenaikan jumlah produksi komoditi tanaman palawija di Kabupaten Puncak Jaya, dengan tingkat kesalahan 5%!



H_0 = Kenaikan jumlah produksi komoditi tanaman palawija tidak signifikan

H_1 = Kenaikan jumlah produksi komoditi tanaman palawija signifikan

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean	
Pair 1					
	Produksi_2010	461.27	15	373.145	96.346
	Produksi_2013	460.93	15	373.258	96.375

Paired Samples Correlations

	N	Correlation	Sig.	
Pair 1				
	Produksi_2010 & Produksi_2013	15	1.000	.000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)	
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
Pair 1	Produksi_2010 - Produksi_2013	.333	1.291	.333	-.382	1.048	1.000	14	.334

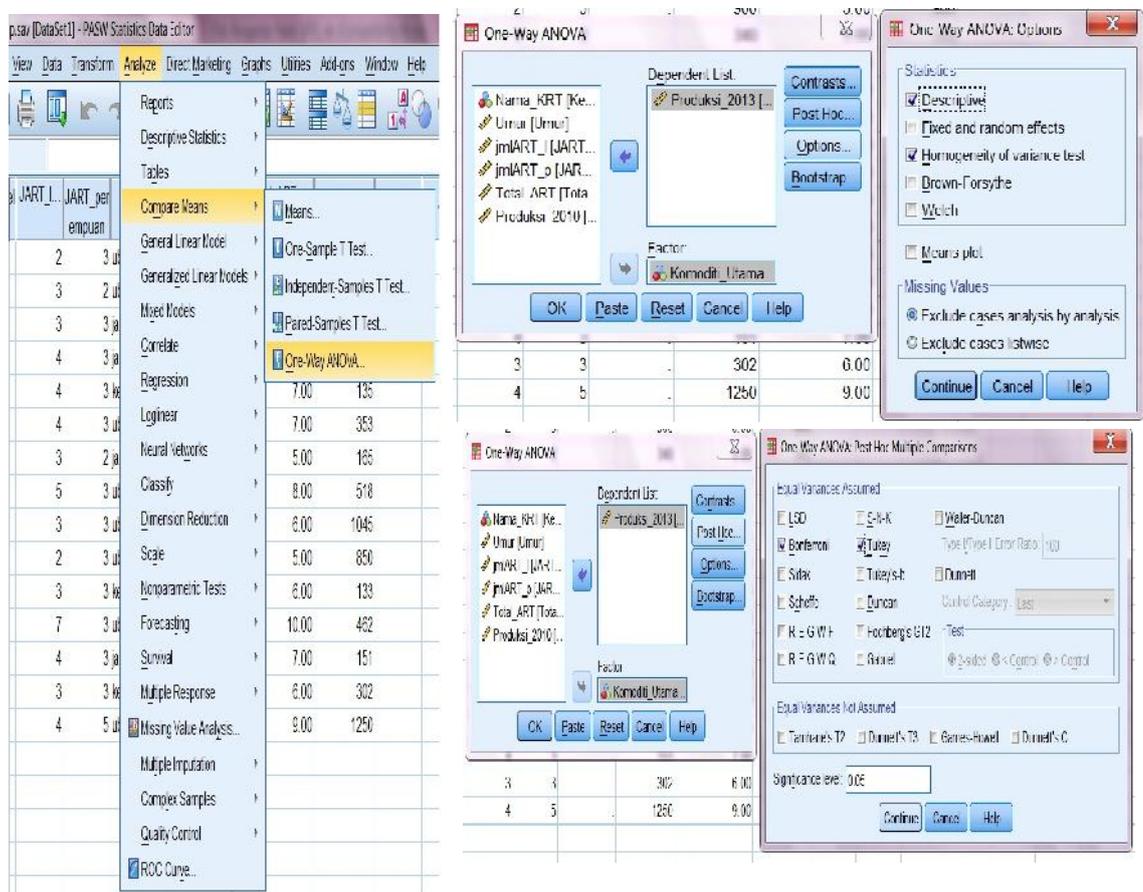
Tabel Paired Samples Statistics menunjukkan rata-rata dan standar deviasi jumlah produksi palawija tahun 2010 dan produksi tahun 2013. Pada tahun 2010, rata-rata produksi tanaman palawija adalah 461.27 kg dengan standar deviasi 373.145. Pada tahun 2013, rata-rata produksi tanaman palawija 460.93 kg dan standar deviasi 373.258.

Paired Samples Correlations menunjukkan korelasi antara dua variabel yang menunjukkan nilai 1.000 dengan nilai probabilitas di bawah 0.05 (nilai sidnifikansi output adalah 0.000). hal ini menyatakan bahwa

korelasi antara produksi palawija tahun 2010 dan tahun 2013 adalah sangat erat dan benar-benar berhubungan secara nyata.

Paired Samples Test menunjukkan p-value dan statistik uji t. Terlihat bahwa t hitung adalah 1,000 dengan probabilitas 0,334. Taraf nyata yang digunakan adalah 5%, sehingga untuk uji duasisi menjadi 2,5%. Dikarenakan $0,334/2 > 0.025$ maka tidak tolak H_0 . Sehingga dapat disimpulkan bahwa produksi palawija tahun 2010 dan tahun 2013 relatif sama.

3. Jika sampel dikelompokkan berdasarkan jenis komoditinya. Hitunglah dengan tingkat keyakinan 95%, dugaan bahwa keempat kelompok memiliki rata-rata jumlah produksi tanaman palawija yang berbeda



H_0 = Keempat kelompok memiliki rata-rata jumlah produksi yang sama

H_1 = Keempat kelompok memiliki rata-rata jumlah produksi yang berbeda

Descriptives

Produksi_2013

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		

ubi jalar	4	1011.25	179.368	89.684	725.84	1296.66	850	1250
ubi kayu	4	418.25	86.110	43.055	281.23	555.27	340	518
jagung	4	156.50	33.201	16.601	103.67	209.33	115	195
kedelai	3	190.00	97.000	56.003	-50.96	430.96	133	302
Total	15	460.93	373.258	96.375	254.23	667.64	115	1250

Descriptives. Pada tabel pertama ini, terdapat ringkasan statistik dari keempat sampel. Contoh, untuk Tanaman Ubi Jalar:

- o Rata-rata jumlah produksinya pada tahun 2013 adalah 1011,25 kg.
- o Produksi minimum adalah 850 kg dan maksimum 1250 kg.

Dengan tingkat kepercayaan 95% rata-rata produksi ada pada rentang 725,84 kg sampai 1296,66 kg.

Test of Homogeneity of Variances

Produksi_2013

Levene Statistic	df1	df2	Sig.
3.465	3	11	.055

Test of Homogeneity of Variances. Hasil output, *Levene Test* hitung adalah 3,465 dengan nilai probabilitas 0,055. Oleh karena probabilitas > 0,05, maka H_0 diterima, atau keempat varians adalah sama. Dengan demikian, asumsi kesamaan varians untuk uji ANOVA sudah terpenuhi

ANOVA

Produksi_2013

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1809614.433	3	603204.811	47.096	.000
Within Groups	140888.500	11	12808.045		
Total	1950502.933	14			

ANOVA. Terlihat bahwa F hitung adalah 47.096 dengan probabilitas 0,000. Oleh karena probabilitas (p-value) < 0.05, maka H_0 ditolak, atau rata-rata produksi keempat kelompok tanaman palawija tersebut memang berbeda

Multiple Comparisons

Dependent Variable:Produksi_2013

(I) Komoditi_Utama	(J) Komoditi_Utama	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound

Tukey HSD	ubi jalar	ubi kayu	593.000*	80.025	.000	352.16	833.84
		— jagung	854.750*	80.025	.000	613.91	1095.59
		kedelai	821.250*	86.437	.000	561.11	1081.39
	ubi kayu	ubi jalar	-593.000*	80.025	.000	-833.84	-352.16
		— jagung	261.750*	80.025	.032	20.91	502.59
		kedelai	228.250	86.437	.092	-31.89	488.39
	— jagung	ubi jalar	-854.750*	80.025	.000	-1095.59	-613.91
		— ubi kayu	-261.750*	80.025	.032	-502.59	-20.91
		kedelai	-33.500	86.437	.979	-293.64	226.64
	kedelai	ubi jalar	-821.250*	86.437	.000	-1081.39	-561.11
		— ubi kayu	-228.250	86.437	.092	-488.39	31.89
		jagung	33.500	86.437	.979	-226.64	293.64
Bonferroni	ubi jalar	ubi kayu	593.000*	80.025	.000	336.27	849.73
		— jagung	854.750*	80.025	.000	598.02	1111.48
		kedelai	821.250*	86.437	.000	543.95	1098.55
	ubi kayu	ubi jalar	-593.000*	80.025	.000	-849.73	-336.27
		— jagung	261.750*	80.025	.045	5.02	518.48
		kedelai	228.250	86.437	.138	-49.05	505.55
	— jagung	ubi jalar	-854.750*	80.025	.000	-1111.48	-598.02
		— ubi kayu	-261.750*	80.025	.045	-518.48	-5.02
		kedelai	-33.500	86.437	1.000	-310.80	243.80
	kedelai	ubi jalar	-821.250*	86.437	.000	-1098.55	-543.95
		— ubi kayu	-228.250	86.437	.138	-505.55	49.05
		jagung	33.500	86.437	1.000	-243.80	310.80

*. The mean difference is significant at the 0.05 level.

Multiple Comparisons.

Apabila nilai p -value (Sig.) > 0.05, maka H_0 diterima, artinya tidak ada perbedaan rata-rata yang nyata antar kelompok tanaman. Cara lain adalah memperhatikan ada tidaknya tanda “*” pada kolom *Mean Difference*. Apabila ada, artinya kelompok tanaman palawija tersebut berbeda signifikan. Dalam kasus ini terlihat rata-rata/Mean dari produksi ubi jalar berbeda signifikan/nyata dengan ubi kayu, jagung, dan kedelai. Rata-rata produksi ubi kayu berbeda nyata dari ubi jalar dan jagung, namun tidak berbeda nyata dengan kedelai. Untuk produksi jagung, rata-rata produksinya berbeda nyata dengan ubi jalar dan ubi kayu, tetapi tidak berbeda nyata dengan kedelai. Dan produksi kedelai terlihat berbeda secara nyata dengan ubi jalar, namun tidak berbeda nyata dengan ubi kayu dan jagung.

Produksi_2013

Komoditi_Utama	N	Subset for alpha = 0.05		
		1	2	3
Tukey HSD ^{a,b}				
jagung	4	156.50		
kedelai	3	190.00	190.00	
ubi kayu	4		418.25	
ubi jalar	4			1011.25
Sig.		.977	.078	1.000

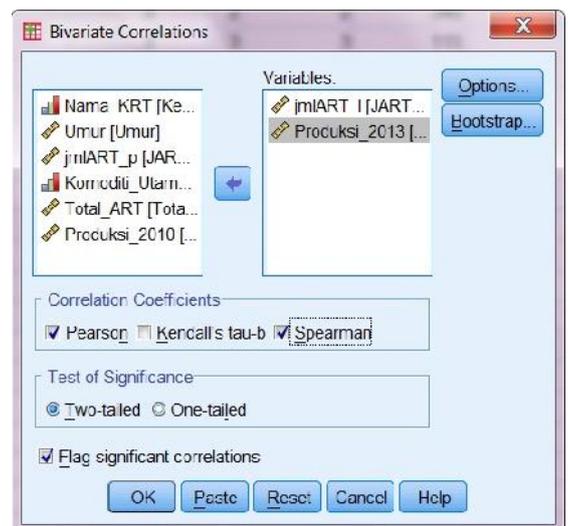
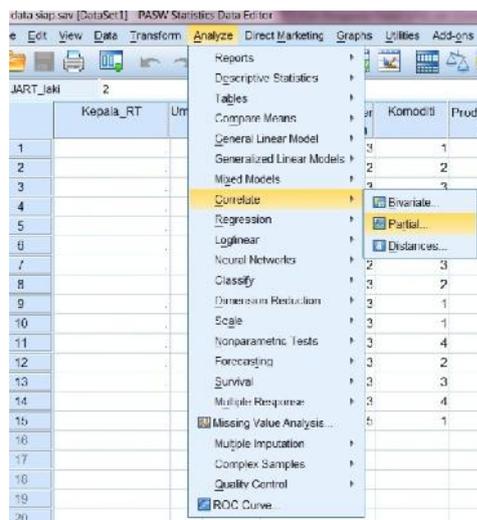
Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 3,692.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Homogeneous Subsets. Jika test Tukey dan Bonferroni untuk menguji kelompok mana saja yang memiliki perbedaan nyata, maka dalam Homogeneous Subset justru akan dicari grup/subset mana saja yang mempunyai perbedaan rata-rata yang tidak berbeda secara signifikan. Output di atas menunjukkan bahwa rata-rata jumlah produksi jagung dan kedelai tidak memiliki perbedaan yang nyata, begitu pula dengan produksi kedelai dan ubi kayu. Rata-rata jumlah produksi dari ubi jalar yang dari semuanya terlihat berbeda signifikan. (hasil dari **Tukey** dan **Bonferroni** dengan **Homogeneous Subset** selalu saling melengkapi).

4. Buktikan dugaan bahwa terdapat hubungan nyata antara jumlah anggota rumah tangga laki-laki yang dimiliki dengan jumlah produksi tanaman palawija



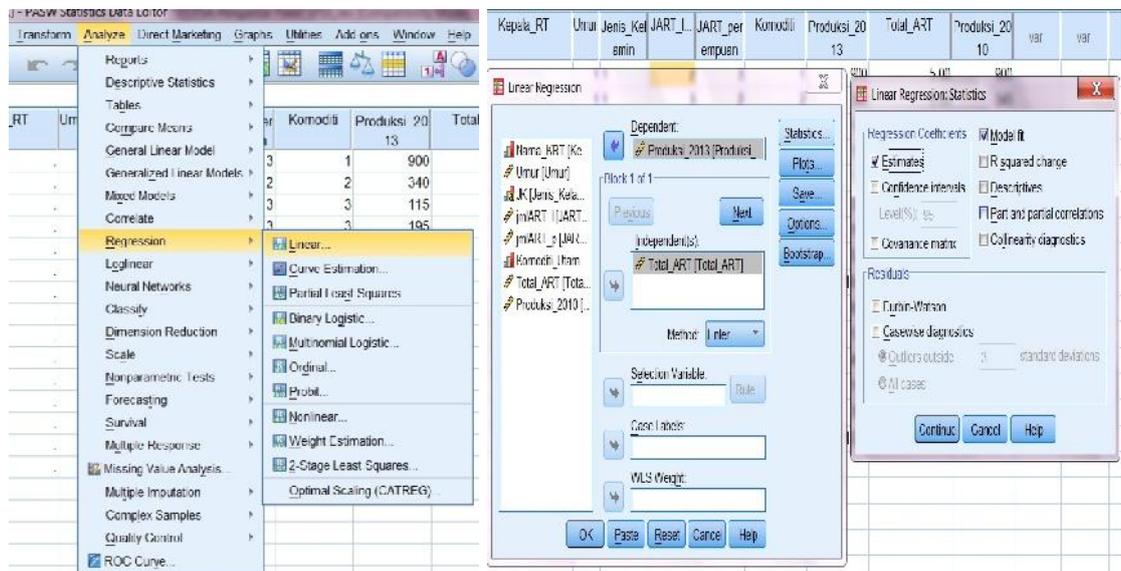
H_0 = Tidak ada hubungan antara jumlah ART laki-laki dengan jumlah produksi tanaman palawija rumah tangga pertanian

H_1 = Ada hubungan antara jumlah ART laki-laki dengan jumlah produksi tanaman palawija rumah tangga pertanian

		jmlART_1	Produksi_2013
jmlART_1	Pearson Correlation	1	-.143
	Sig. (2-tailed)		.610
	N	15	15
Produksi_2013	Pearson Correlation	-.143	1
	Sig. (2-tailed)	.610	
	N	15	15

Hasil output SPSS menunjukkan bahwa nilai korelasi sebesar -0,143 dengan nilai sig 0,610 . nilai sig > 0,05 maka terima H₀. Yang berarti tidak ada hubungan signifikan antara jumlah anak laki-laki suatu rumah tangga pertanian dengan jumlah produksi tanaman palawija yang dihasilkan.

- Hitunglah seberapa besar pengaruh total ART yang dimiliki rumah tangga pertanian terhadap jumlah produksi tanaman palawija, dengan tingkat keyakinan 95 % !



Model	Variables Entered	Variables Removed	Method
1	Total_ART ^b		Enter

- All requested variables entered.
- Dependent Variable: Produksi_2013

Tabel Variables Entered Removed menunjukkan metode regresi linier yang dipilih, yaitu Enter. Pemilihan metode memungkinkan Anda menentukan bagaimana variabel independen (Total_ART) dimasukkan untuk analisis. Metode Enter memasukkan semua variabel independen sekaligus untuk dianalisis. Pada analisis regresi linier satu variabel independen, perbedaan antar metode tidak tampak.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.136 ^a	.018	-.057	383.770

Tabel Model Summary menunjukkan nilai koefisien korelasi (R) yang menunjukkan tingkat hubungan antarvariabel (0,136). R Square atau koefisien determinasi memiliki arti 0,018 jumlah produksi tanaman palawija dapat dijelaskan dari variabel Total_ART. Adjusted R Square sama dengan R Square dengan menyesuaikan numerator maupun denominator dengan derajat kebebasan masing-masing. Std Error of the Estimated mengukur dispersi titik-titik pasangan X dan Y dari garis duga regresi.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35866.239	1	35866.239	.244	.630 ^a
	Residual	1914636.694	13	147279.746		
	Total	1950502.933	14			

a. Predictors: (Constant), Total_ART

b. Dependent Variable: Produksi_2013

Tabel Anova memaparkan uji kelinieran.

Hipotesis:

H₀ = Model linier antara variabel Total_ART dengan variabel Produksi_2013 tidak signifikan

H₁ = Model linier antara variabel Total_ART dengan variabel Produksi_2013 signifikan

Nilai Sig (0,630) > 0,05, maka H₀ diterima. Artinya model linier tersebut tidak signifikan. Apabila model sudah tidak signifikan maka tidak perlu dilanjutkan ke garis regresi. Namun untuk membuktikan, silahkan dilihat output dibawah.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
		1	(Constant)	238.580		
	Total_ART	33.690	68.270	.136	.493	.630

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	238.580	461.347		.517	.614
	Total_ART	33.690	68.270	.136	.493	.630

a. Dependent Variable: Produksi_2013

Pada tabel Coefficient, dipaparkan nilai konstanta a dan b dari persamaan linier: $Y = 238,580 + 33,690 x$

Hipotesis uji koefisien a

H_0 = koefisien a tidak signifikan

H_1 = koefisien a signifikan

Nilai sig(0,614) < 0,05, maka H_0 ditolak. Menunjukkan bahwa koefisien a tidak signifikan

Hipotesis uji koefisien b

H_0 = koefisien b tidak signifikan

H_1 = koefisien b signifikan

Nilai sig(0,630) < 0,05, maka H_0 ditolak. Menunjukkan bahwa koefisien b tidak signifikan

Hasil diatas membuktikan lebih jelas bahwa tidak ada pengaruh antara Total ART yang dimiliki suatu rumah tangga pertanian dengan jumlah produksi tanaman palawija yang dihasilkan.

DAFTAR PUSTAKA

- Santoso, Singgih. 2008. *SPSS 16*. Jakarta: PT Elex Media Komputindo.
- Sofyan Yamin dan Heri Kurniawan. 2009. *SPSS COMPLETE Teknik Analisis Statistik Terlengkap dengan Software SPSS*. Jakarta: Salemba Infotek.
- Trihendradi, T. 2010. *Step by Step SPSS 18 Analisis Data Statistik*. Yogyakarta: ANDI.
- Uyanto, Stanislaus S. 2009. *SPSS Pedoman Analisis Data dengan SPSS*. Yogyakarta: Graha Ilmu.
- Wahyono, Teguh. 2006. *Analisis Data Statistik dengan SPSS 14*. Jakarta: PT Elex Media Komputindo.



Penyajian Data



DAFTAR ISI

DAFTAR ISI	i
DAFTAR TABEL.....	ii
DAFTAR GAMBAR	iii
Bab I Pendahuluan	1
1.1 Latar Belakang.....	1
1.2 Deskripsi Singkat.....	1
1.3 Hasil Belajar (Tujuan Pembelajaran Umum)	1
1.4 Indikator Hasil Belajar (Tujuan Pembelajaran Khusus).....	1
1.5 Materi Pokok	2
1.6 Manfaat.....	2
Bab II Ukuran Statistik.....	3
2.1 Ukuran Pemusatan.....	3
2.2 Ukuran Penyebaran.....	11
Bab III Pengertian Penyajian Data.....	17
3.1 Pengertian Penyajian Data.....	17
3.2 Maksud Dan Tujuan Penyajian Data	17
3.3 Bentuk-Bentuk Penyajian Data	18
3.4 Rangkuman.....	18
Bab IV Penyajian Data Dalam Bentuk Tabel	19
4.1 Tujuan Penyajian Tabel.....	19
4.2 Jenis-Jenis Tabel.....	19
4.2.1 Tabel Satu Arah (<i>One Way Table</i>)	21
4.2.2 Tabel Dua arah (<i>Two Way Table</i>).....	22
4.2.3 Tabel Tiga Arah (<i>Three Way Table</i>)	23
4.3 Rangkuman.....	24
4.4 Soal-Soal.....	
Bab V Penyajian Data Dalam Bentuk Grafik	25
5.1 Tujuan Penyajian Grafik.....	25
5.2 Jenis-Jenis Grafik	25
5.2.1. Grafik dari Tabel Distribusi Frekuensi	25
a. Histogram.....	26
b. Poligon Frekuensi	26
c. Ogive.....	26
5.2.2. Grafik dari Bentuk Tabel.....	31
a. Grafik Garis (<i>Line Chart</i>)	32
b. Grafik Batang (<i>Bar Chart</i>).....	32
c. Grafik Lingkaran <i>Pie Chart</i>).....	40
d. Diagram Gambar (<i>Pictogram</i>).....	45
e. Diagram Peta (<i>Cartogram</i>)	46
5.3 Rangkuman.....	47
5.4 Soal-Soal.....	50
Bab VI Penutup.....	51
6.1 Rangkuman.....	51
6.2 Implikasi	51
6.3 Tindak Lanjut	52
Daftar Pustaka	53

DAFTAR TABEL

	Hal.
Tabel 1 Menghitung Arithmetic Mean Upah per Minggu dari 260 Buruh suatu Pabrik	4
Tabel 2 Menghitung Rata-rata Hitung Tertimbang Tingkatan Nilai Seorang Mahasiswa dalam suatu Semester	6
Tabel 3 Indeks Harga 8 Kelompok Komoditi Utama.....	7
Tabel 4 Menghitung Median Upah per Minggu dari 260 Buruh suatu Pabrik.....	9
Tabel 5 Jumlah Penduduk DKI Jakarta Menurut Kab/Kota, 2012	21
Tabel 6 Jumlah Penduduk DKI Jakarta Menurut Kab/Kota dan Jenis Kelamin Tahun 2012	22
Tabel 7 Jumlah Usia 10 Tahun ke Atas di DKI Jakarta menurut Kab/Kota, Jenis Kelamin, dan Kemampuan Membaca dan Menulis	23
Tabel 8 Perkembangan Jumlah Sepeda Motor di DKI Jakarta, 2008 – 2012	33
Tabel 9 Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008-2012	34
Tabel 10 .Perkembangan Persentase Kendaraan Bermotor di DKI Jakarta, Tahun 2008-2012 (Juta)	37
Tabel 11 Perkembangan Nilai Ekspor dan Impor Melalui DKI Jakarta, 2009 – 2012 (000 US\$)	39
Tabel 12 Harga Beras Grosir di Pasar Induk Cipinang DKI Jakarta, Juli – Desember 2012 (Rp/Kg)	40
Tabel 13 Nilai Ekspor Produk DKI Jakarta menurut Negara Tujuan, 2012 (Juta US \$).....	41

DAFTAR GAMBAR

	Hal.
Grafik 1	Histogram Distribusi Frekuensi Berat Badan..... 30
Grafik 2	Poligon Distribusi Frekuensi Berat Badan..... 30
Grafik 3	Kurva Ogive..... 31
Grafik 4	Perkembangan Jumlah Sepeda Motor di DKI Jakarta, 2008 – 2012 (Juta) 34
Grafik 5	Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta) 35
Grafik 6	Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta) 36
Grafik 7	Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta) 38
Grafik 8	Perkembangan Nilai Ekspor dan Impor Melalui DKI Jakarta, 2009 – 2012 (000 US \$)..... 39
Grafik 9	Harga Beras Grosir di Pasar Induk Cipinang DKI Jakarta, Juli – Desember 2012 (Rp/kg) 40
Grafik 10	Nilai Ekspor Produk DKI Jakarta menurut Negara Tujuan, 2012 (Juta US \$) 41
Grafik 11	Jumlah Penduduk DKI Jakarta menurut Kab/Kota dan Jenis Kelamin, 2012 41
Grafik 12	Jumlah Penduduk DKI Jakarta menurut Jenis Kelamin dan Kab/Kota, 2012. 42
Grafik 13	Jumlah Penduduk DKI Jakarta menurut Kab/Kota dan Jenis Kelamin, 2012 42
Grafik 14	Persentase Penduduk DKI Jakarta menurut Jenis Kelamin dan Kab/Kota, 2012 43
Grafik 15	Jumlah Penduduk DKI Jakarta menurut Kab/Kota, 2012 43
Grafik 16	Persentase Nilai Ekspor Produk DKI Jakarta menurut Negara Tujuan, 2012 44

Grafik 17	Kepadatan Penduduk Per Km ² menurut Kabupaten/Kota di DKI Jakarta, 2012	46
Grafik 18	Peta Potensi Ikan Perairan Indonesia, 2011.....	47

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis statistika terdiri dari rangkaian kegiatan dimulai dari pengumpulan data, pengolahan data, penyajian data, analisis data, dan pengambilan kesimpulan. Tahap penyusunan dan penyajian data sangat penting untuk memudahkan analisis dan pembacaan data hasil pengolahan.

Penyajian Data merupakan upaya untuk menyusun dan mengatur data dalam bentuk yang sederhana dan menarik perhatian pembaca. Selain itu juga sangat diperlukan dalam menyimpulkan dan memberikan makna terhadap hasil pengolahan data, baik interpretasinya maupun mencari implikasinya yang relevan.

Meski tampaknya mudah dan sederhana namun dalam prakteknya masih banyak orang yang menemui kesulitan bagaimana menyajikan data secara tepat sehingga mudah dicerna dan dipahami oleh konsumen. Data harus disajikan ke dalam suatu bentuk yang dapat memberikan arti dan tidak membebani pembaca atau konsumen berfikir dalam melihat data yang disajikan. Suatu hal yang harus diperhatikan dalam penyajian data adalah siapa yang memerlukannya dan untuk apa data tersebut disajikan.

1.2 Deskripsi Singkat

Penyajian Data merupakan salah satu kegiatan dalam pembuatan laporan hasil penelitian yang telah dilakukan sesuai dengan tujuan yang diinginkan.

1.3 Hasil Belajar (Tujuan Pembelajaran Umum)

Setelah mengikuti pelatihan peserta diharapkan mampu memahami kegunaan tabel dan grafik serta dapat menyajikan data dalam bentuk tabel, ukuran statistik dan grafik.

1.4 Indikator Hasil Belajar (Tujuan Pembelajaran Khusus)

Setelah mempelajari modul ini peserta diharapkan dapat :

1. Membuat tabel dan grafik
 2. Membaca tabel dan grafik dengan benar
-

3. Menjelaskan manfaat penyajian tabel dan grafik
4. Membedakan kegunaan masing-masing jenis tabel dan grafik

1.5 Materi Pokok

1. Ukuran Statistik
2. Pengertian Penyajian Data
3. Penyajian Data Dalam Bentuk Tabel
4. Penyajian Data Dalam Bentuk Grafik

1.6 Manfaat

Sesuai dengan sifatnya, bahan ajar ini diharapkan dapat memberikan kemudahan dan membantu peserta dalam memperoleh pemahaman mengenai bagaimana menyajikan data dari hasil proses pelaksanaan pengolahan data di lapangan yang diharapkan nantinya dapat menjadi acuan dalam mengembangkan dan mengaplikasikannya.

Bab II Ukuran Statistik

Setelah mempelajari materi ukuran pemusatan, peserta diharapkan dapat menjelaskan pengertian ukuran pemusatan, menghitung ukuran-ukuran pemusatan, serta memahami dan mampu melakukan perhitungan ukuran statistik dan membuat analisis sederhana.

2.1 Ukuran Pemusatan

2.1.1 Rata-rata Hitung (*Arithmetic Mean*)

Arithmetic Mean (rata-rata hitung) dikenal dalam kehidupan sehari-hari dengan kata yang singkat yakni "rata-rata". Rata-rata (*average*) oleh "Crom" dan "Smith" dalam bukunya yang berjudul "Business Statistics" disebutkan sebagai ukuran kecenderungan pusat. karena nilai-nilai individu disekitarnya cenderung untuk mengelompok dekat dengan nilai rata-ratanya.

Rata-rata hitung dari sekumpulan nilai pengamatan/observasi adalah sama dengan jumlah seluruh nilai pengamatan tersebut dibagi dengan jumlah pengamatannya (observasinya).

Dalam kehidupan sehari-hari disadari atau tidak sering didengar perkataan rata-rata, misalnya: rata-rata pendapatan perkapita; rata-rata gaji perbulan pegawai negeri; rata-rata tingkat kecelakaan perbulan di DKI Jakarta; dan lain-lain.

Kita membedakan tiga macam rata-rata hitung dimana cara perhitungannya yang berbeda yakni:

1. Rata-rata hitung untuk data tidak berkelompok;
2. Rata-rata hitung untuk data berkelompok;
3. Rata-rata hitung tertimbang (*Weighted Arithmetic Mean*).

a. Rata-rata Hitung Data Tidak Berkelompok

Rata-rata hitung adalah ukuran lokasi yang diperoleh dengan cara menjumlahkan semua nilai, kemudian membaginya dengan banyaknya nilai yang bersangkutan.

Rata-rata hitung adalah ukuran lokasi yang dapat menunjukkan nilai rata-rata dari setiap individu.

Dirumuskan sebagai:

$$(I) \quad x = \frac{1}{n} \sum_{i=1}^n x_i \quad \dots\dots\dots$$

di mana :

\bar{x} dibaca eks bar atau nilai rata-rata.

$$X_i = X_1; X_2; \dots; X_n$$

$\sum_{i=1}^n x_i$ dibaca sigma eks-i, i bergerak mulai dari 1 sampai n.

(n = banyaknya individu).

X_i = nilai karakteristik individu ke i, dan $i = 1, 2, 3, \dots, n$

Nilai karakteristik ini bisa tinggi badan, berat badan, umur peserta Kursus Statistik, gaji karyawan perbulan, banyaknya karyawan pada perusahaan, dan lain-lain.

Catatan :

1. Rumus (I) diatas disebut Rumus Pokok Rata-rata hitung atau Rumus Dasar. Semua rumus rata-rata hitung lahir dan beranjak dari Rumus (I)
2. Nilai rata-rata hitung merupakan nilai rata-rata setiap individu.
3. Nilai rata-rata hitung dapat "mewakili" dengan baik dan memuaskan, jika datanya relatif homogen (merata).
4. Rata-rata hitung kurang baik dipakai, jika terdapat data yang ekstrim (terlalu kecil atau terlalu besar). Dalam praktek agar tidak mengganggu data yang ekstrim biasanya di buang tidak ikut diolah).
5. Ada cara lain, bagaimana agar data yang ekstrim tidak mengganggu, tanpa membuang data yang ekstrim tersebut.

Contoh:

Nilai ujian matematika dari sepuluh murid suatu Sekolah Dasar adalah:
90, 65, 55, 40, 70, 45, 80, 30, 75, 50.

Jawab:

Rata-rata hitung

$$\bar{X} = \frac{\sum_{I=1}^{10} X_I}{10} = \frac{90 + 65 + \dots + 75 + 50}{10} = 60$$

b. Rata-rata Hitung Data Berkelompok

Arti data berkelompok adalah bahwa nilai-nilainya tidak lagi merupakan nilai-nilai individu seperti dalam contoh tersebut dimuka (nilai ujian matematika), melainkan sudah dikelompokkan dalam klas-klas tertentu dalam suatu distribusi frekuensi.

Contoh:

Tabel-1. Menghitung Arithmetic Mean Upah per Minggu dari 260 Buruh suatu Pabrik.

Kelas interval (ribuan Rp)	Jumlah buruh (f)	Nilai tengah (x)	f . x
2,0 - 3,9	12	2,95	35,40
4,0 - 5,9	19	4,95	94,05
6,0 - 7,9	39	6,95	271,05
8,0 - 9,9	70	8,95	626,50

10,0 - 11,9	52	10,95	569,40
12,0 - 13,9	24	12,95	310,80
14,0 - 15,9	21	14,95	313,95
16,0 - 17,9	15	16,95	254,25
18,0 - 19,9	8	18,95	151,60
Jumlah	260	-	2.627,00

Dari perhitungan arithmetic mean di atas, mempunyai anggapan bahwa nilai tengah dari setiap klas adalah nilai rata-rata dari semua individu yang termasuk dalam klas tersebut.

Sebagai contoh, dalam tabel di atas sejumlah 12 orang buruh pada klas pertama upah per minggunya berkisar 2,0 - 3,9 (ribu rupiah), tetapi untuk tiap-tiap buruh kita tidak tahu besar upah perminggunya, oleh karena itu dianggap bahwa 12 buruh yang mempunyai upah per minggu 2,0 - 3,9 mempunyai upah rata-rata 2,95 (ribu rupiah). Perkalian 12 dengan 2,95 (=35,40) adalah merupakan taksiran jumlah upah per minggu dari 12 buruh tersebut, sedangkan jumlah keseluruhan dari kolom $f \times m$ (2.627,00) adalah taksiran jumlah keseluruhan upah per minggu dari 260 buruh di pabrik tersebut.

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{2.627,00}{260} = 10,10$$

Jadi rata-rata upah per minggunya adalah 10,10 (ribu rupiah).

c. Rata-rata Hitung Tertimbang

Perhitungan rata-rata hitung tertimbang dilakukan dengan mengkalikan tiap-tiap data yang akan dicari rata-ratanya dengan timbangan yang ditentukan. Jumlah hasil perkalian tersebut kemudian dibagi dengan jumlah timbangannya.

Dinyatakan dalam rumus adalah sebagai berikut:

ditentukan. $\bar{X} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$ di mana: W adalah timbangan yang

Contoh:

Tabel-2. Menghitung Rata-rata Hitung Tertimbang Tingkatan Nilai Seorang Mahasiswa dalam suatu Semester

Ujian	Tingkatan Nilai	Timbangan	X.W
Tugas:			
No. 1	81	2	162
No. 2	75	2	150
No. 3	68	2	136
No. 4	90	2	180
No. 5	40	2	80
No. 6	71	2	142
No. 7	60	2	120
No. 8	58	2	116
No. 9	90	2	180
No.10	73	2	146
Kuis:			
No. 1	75	10	750
No. 2	84	10	840
No. 3	79	10	790
	87	10	870
Ujian akhir:			
No. 1	91	40	3.640
	-	100	8.302

$$\bar{X} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i} = \frac{8.302}{100} = 83,02$$

Dalam penentuan nilai rata-rata seorang murid sebaiknya menggunakan rata-rata tertimbang dengan cara menggabungkan tingkatan nilai yang tidak sama pentingnya (bobotnya).

Nilai rata-rata tidak dihitung dengan menjumlahkan tingkatan nilai 10 menit test, 60 menit ujian dan ujian akhir kemudian dibagi 15, melainkan sebaiknya tiap kelompok tingkatan diberikan timbangan yang berbeda untuk menentukan nilai rata-rata akhir.

Misalnya test 10 menit diberi timbangan 2, ujian 60 menit diberi timbangan 10 dan ujian akhir diberi timbangan 40.

Dalam menentukan rata-rata tertimbang, sebaiknya diberikan timbangan yang benar-benar layak.

2.1.2 Rata-rata Ukur (Geometric Mean)

Perhitungan pada rata-rata ukur (geometric mean) didasarkan pada seluruh nilai pengamatan, ini berarti bahwa rata-rata ukur dipengaruhi oleh semua nilai variabel. Dalam hal ini apabila terdapat nilai-nilai ekstrim, maka pengaruhnya dapat diperkecil, bila dibandingkan dengan rata-rata hitung.

Jadi yang dimaksud dengan rata-rata ukur dari sekumpulan nilai pengamatan, adalah sama dengan hasil perkalian nilai-nilai tersebut pangkat satu dibagi jumlah pengamatannya.

Rata-rata ukur dari sekumpulan nilai pengamatan, dengan data tidak berkelompok mempunyai rumus sebagai berikut:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n}$$

Metode sederhana untuk menghitung rata-rata ukur adalah dengan menggunakan logaritma. jika kita menggunakan logaritma diperoleh:

$$\log G = \frac{\sum_{i=1}^n \log x_i}{n}$$

di mana: G = rata-rata ukur
 xi = nilai pengamatan
 n = jumlah pengamatan

Contoh:

Hitunglah rata-rata ukur dari indeks harga grosir dari 8 kelompok komoditi utama yaitu 107, 132, 120, 116, 130, 126, 116, dan 122.

Tabel-3. Indeks Harga 8 Kelompok Komoditi Utama

Indeks harga (xi)	Log xi
132	2,1206
120	2,0792
116	2,0645
130	2,1139
126	2,2004
116	2,0645
122	2,0864
Jumlah	16,6568

$$\log G = \frac{\sum_{i=1}^n \log x_i}{n} = \frac{16,6568}{8} = 2,0824$$

Maka :

$$G = \text{antilog } 2,0824 \\ = 120,9$$

2.1.3 Median

Median ini disebut juga rata-rata letak (positional measure) karena perhitungan median didasarkan pada letak dari nilainya. Median didapatkan dengan menyusun nilai-nilai variabel dalam bentuk array (urutan) dan kemudian mendapatkan nilai tengahnya.

Median adalah ukuran rata-rata juga atau ukuran tendensi sentral. Banyak ahli statistik yang menamakan median ini sebagai ukuran dari item tengah (size of the middle item), sesudah item-itemnya disusun menurut perurutan besar jumlahnya. Ini berarti bahwa di atas item tengah terdapat jumlah item "yang di atas atau yang lebih besar" yang sama besarnya dengan item-item "yang dibawah atau yang lebih kecil" daripada item tengahnya.

Median yang akan diuraikan ini harus diartikan sebagai: "nilai dari item tengah atau kasus tengah".

a. Median Data Tidak Berkelompok

Setelah semua nilai kita susun dari yang kecil ke yang besar, maka posisi median dapat kita tentukan sebagai berikut:

1. Untuk jumlah pengamatan (n) ganjil:

Kalau n ganjil, maka median sama dengan nilai X yang ke $\frac{x_{n+1}}{2}$ setelah semua nilai disusun dalam bentuk array.

Contoh:

$$x_1, x_2, x_3, \dots, x_{\frac{n-1}{2}}, x_{\frac{n+1}{2}}, x_{\frac{n+3}{2}}, \dots, x_n$$

Median

Persentase murid-murid yang lulus dari lima buah Sekolah Lanjutan Tingkat Atas adalah: 50, 74, 76, 89, dan 93. Hitunglah mediannya.

$$\begin{array}{cccccc} 50\% & 74\% & 76\% & 89\% & 93\% \\ X_1 & X_2 & X_3 & X_4 & X_5 \end{array}$$

Disini n = 5 (ganjil), Jadi Md sama dengan X ke $-(5+1)/2 = X_3$. Dari daftar X_3 adalah 76%. Jadi median adalah 76%.

2. Untuk jumlah pengamatan (n) genap:

Jika diurutkan dalam notasi, dapat dinyatakan sebagai berikut:

$$x_1, x_2, x_3, \dots, x_{\frac{n}{2}}, x_{\frac{n+2}{2}}, \dots, x_n$$

Median

Maka median sama dengan nilai X yang ke $n/2$ atau nilai X yang ke $(n+2)/2$.

Contoh:

Jumlah keluarga dari 8 rumah tangga adalah: 7, 2, 4, 5, 4, 8, 6, 6.

Tentukan nilai mediannya ?

Jawab:

Semua nilai di atas setelah diurutkan:

$$\begin{array}{cccccccc} 2 & 4 & 4 & 5 & 6 & 6 & 8 & 8 \\ X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 \end{array}$$

Karena $n = 8$ (genap),

Maka Median sama dengan: $\frac{X_n + X_{(n+2)/2}}{2}$

(nilai rata-rata dari dua nilai yang terletak di tengah-tengah).

Berarti nilai median adalah = $Md = (5 + 6)/2 = 5,5$.

b. Median Data Berkelompok

Apabila tiap-tiap nilai individunya tidak diberikan, tetapi hanya diberikan kelompok-kelompok nilai dalam klas-klas seperti dalam distribusi frekuensi, maka cara mencari nilai median tidak bergitu sederhana lagi seperti mencari nilai median untuk data tidak berkelompok.

Perhitungan median untuk data berkelompok, berbentuk distribusi frekuensi dimana diperoleh dengan jalan interpolasi (penyisipan) di dalam salah satu klas-klas dari distribusi itu.

Contoh :

Tabel-4. Menghitung Median Upah per Minggu dari 260 Buruh suatu Pabrik.

Interval klas (ribuan Rp)	Jumlah buruh (f)	Frek. Kumulatif
2,0 - 3,9	12	12
4,0 - 5,9	19	31
6,0 - 7,9	39	70
8,0 - 9,9	70	140
10,0 - 11,9	52	192
12,0 - 13,9	24	216
14,0 - 15,9	21	237
16,0 - 17,9	15	252
18,0 - 19,9	8	260
T o t a l	n=260	260

Adapun formula/rumus median untuk data berkelompok adalah:

$$Md = L_{Md} + \left(\frac{\frac{n}{2} - F_{LMd}}{f_{Md}} \right) i$$

keterangan:

Md = nilai median yang hendak kita hitung.

L_{md} = batas bawah nyata klas yang mengandung median

n = jumlah frekuensi dalam distribusi

F_{LMd} = frek. kumulatif sebelum batas bawah klas yang mengandung median.

f_{md} = frekuensi dari klas yang mengandung median

i = luas klas (interval class)

Berdasarkan data di atas, maka nilai median adalah:

$$Md = 7,95 + \left(\frac{\frac{260}{2} - 70}{70} \right) 2$$

$$Md = 7,95 + 1,71 = 9,66$$

Jadi separoh dari jumlah buruh yang bekerja pada pabrik tersebut mendapat upah kurang atau sama dengan 9,66 ribu rupiah atau Rp 9.660,- per minggunya.

2.1.4 Modus (Mode)

Perkataan modus berasal dari kata "Mode". Dalam pengertian sehari-hari diartikan sebagai nilai variabel yang tercatat berjumlah paling banyak. Kata ini kiranya sudah tidak asing lagi pada masa sekarang ini. Dalam dunia kewanitaan misalnya kita mengenal mode pakaian, mode potongan rambut, artinya adalah pakaian atau potongan rambut yang sedang banyak penggemarnya, sesuatu yang sedang disukai (in fashion).

Juga dalam ilmu statistik kata "modus atau mode" menunjukkan pengertian yang kurang lebih sama dengan pengertian dalam kehidupan sehari-hari, yakni menunjukkan sesuatu karakteristik yang banyak frekuensinya dan karena karakteristik itu dinyatakan dalam bilangan (nilai), maka kira-kira modus/mode dapat diberi batasan sebagai berikut:

"Mode dari suatu kumpulan nilai adalah suatu nilai dari kumpulan nilai tersebut yang paling sering terjadi".

Untuk data diskrit, modus/mode adalah nilai variabel yang paling sering terjadi.

Contoh:

Carilah modus/mode dari besarnya keluarga dari sepuluh rumahtangga sebagai berikut: 7, 4, 4, 6, 5, 6, 6, 9, 5 dan 6.

Kita nyatakan dalam bentuk yang lebih sederhana:

Besar keluarga	Frekuensi
4	2
5	2
6	4
7	1
9	1
T o t a l	10

Dari tabel tersebut diketahui bahwa yang paling banyak frekuensinya adalah nilai 6. Jadi nilai modus = 6.

2.2 Ukuran Penyebaran

Konsep lain yang sangat penting dalam statistik adalah ukuran penyebaran/dispersi/variasi. Mean, median dan modus hanya menunjukkan ciri-ciri karakteristik dari distribusi frekuensi, yaitu tendensi sentral. Ukuran penyebaran (dispersi) adalah suatu ukuran yang menyatakan atau memberikan gambaran tentang penyebaran sekelompok data.

Contoh

Tiga kelompok data terdiri dari:

- a. 50, 50, 50, 50, 50 (homogen)
rata-rata hitung = 50
 - b. 50, 40, 30, 60, 70 (heterogen)
rata-rata hitung = 50
 - c. 100, 40, 80, 20, 10 (heterogen)
rata-rata hitung = 50
- Kelompok c lebih heterogen dibandingkan kelompok b

2.2.1 Manfaat dan Kegunaan dari Ukuran Penyebaran

Ukuran penyebaran dapat membantu kita dalam mempelajari sifat yang penting dari suatu distribusi, yaitu sampai seberapa jauh data tersebut bervariasi satu dengan yang lain dan dari rata-ratanya. Untuk memperoleh gambaran tentang keadaan suatu kelompok data, maka tidak cukup hanya dengan mempelajari/mengetahui ukuran lokasinya saja, tetapi diperlukan juga ukuran dispersi.

Dalam kehidupan sehari-hari sering mendengar data statistik, misal rata-rata upah bulanan karyawan suatu perusahaan makanan adalah Rp 2.000.000,-. Setiap mendengar kata rata-rata, maka akan membayangkan sekelompok nilai "di sekitar" rata-rata tersebut. Ada yang sama dengan rata-rata, ada yang lebih kecil, dan ada yang lebih besar dari rata-rata tersebut. Dengan kata lain, ada penyebaran atau dispersi dari nilai-nilai tersebut, baik terhadap nilai lainnya maupun terhadap rata-ratanya (terhadap rata-rata hitung, median, atau modus). Contoh di atas belum memberikan gambaran yang jelas dari upah para pegawai tersebut. Beberapa orang mungkin hanya menerima upah Rp 1.500.000,- sedangkan beberapa orang lainnya

akan menerima Rp 3.000.000,- atau lebih. Pada beberapa distribusi, kasus-kasusnya mungkin menggerombol disekitar nilai rata-rata, sedangkan pada distribusi lainnya kasus-kasusnya menyebar secara luas/lebar. Oleh karena itu, perlulah ditentukan penyebaran dari nilai-nilai individual yang menyebar di atas dan di bawah tendensi sentralnya.

2.2.2 Jenis-jenis Ukuran Penyebaran

Ada beberapa macam ukuran dispersi, misalnya nilai jarak (*range*) dan simpangan baku (*standard deviation*). Diantara ukuran dispersi tersebut simpangan baku yang sering dipergunakan, khususnya untuk keperluan analisis data.

Nilai rata-rata seperti mean atau median hanya menitikberatkan pada pusat data, tapi tidak memberikan informasi tentang sebaran nilai pada data tersebut. Kita tidak akan menyeberangi sebuah sungai begitu saja jika kita tahu kedalaman rata-ratanya 2 meter. Selanjutnya, dispersi dapat membandingkan sebaran data dari dua informasi distribusi nilai. Misalnya, untuk membandingkan tingkat produktivitas dari dua perusahaan. Meskipun kita mengetahui bahwa produksi rata-rata dari dua perusahaan sepatu adalah 100 pasang sehari, namun kita tidak dapat langsung mengatakan bahwa tingkat produksi kedua perusahaan identik. Kita perlu melihat bagaimana sebaran nilai (jumlah produksi harian) dari kedua perusahaan tersebut. Mungkin perusahaan pertama cenderung lebih homogen, artinya jumlah produksi harian tidak jauh dari kisaran rata-rata. Mungkin pula perusahaan kedua cenderung memiliki tingkat distribusi produksi yang lebih heterogen, yang berarti bahwa jumlah produksi harian sangat beragam dan menyebar jauh dari sekitar rata-rata.

2.2.3 Range (Nilai Jarak)

Di antara ukuran variasi yang paling sederhana dan paling mudah dihitung adalah nilai jarak (*range*). Apabila suatu kelompok nilai (data) sudah disusun menurut urutan yang terkecil (X_1) sampai dengan yang terbesar (X_n), maka untuk menghitung nilai jarak dipergunakan rumus berikut:

$$NJ = X_n - X_1$$

dimana:

NJ = nilai jarak (range)

X_n = Nilai maksimum

X_1 = Nilai minimum

Contoh:

Carilah range dari data berikut:

50 40 30 60 70

Urutkan data dari terkecil sampai terbesar menjadi:

30 40 50 60 70

diperoleh: $X_1 = 30$ dan $X_5 = 70$

sehingga,

$$NJ = X_5 - X_1$$

$$NJ = 70 - 30$$

$$= 40$$

Contoh:

Misal range nilai materi statistika peserta kursus kelas A adalah $90-83 = 7$, dan range nilai pelajaran statistika peserta kursus kelas B adalah $95-78 = 17$.

Kecilnya range dari kelas A menunjukkan bahwa nilai materi statistika yang didapat oleh kelas A lebih merata (homogen) dari pada kelas B.

2.2.4 Simpangan Baku (Standar Deviasi)

Simpangan baku merupakan salah satu ukuran dispersi yang paling banyak dipergunakan, sebab mempunyai sifat-sifat matematis (*mathematical property*) yang sangat penting dan berguna sekali untuk pembahasan teori dan analisis.

Simpangan baku merupakan salah satu ukuran dispersi yang diperoleh dari akar kuadrat positif varians. Varians adalah rata-rata hitung dari kuadrat simpangan setiap pengamatan terhadap rata-rata hitungnya.

Jika mempunyai suatu populasi dengan jumlah elemen sebanyak N dan sampel sebanyak n elemen, selanjutnya nilai suatu karakteristik tertentu dikumpulkan (umur, hasil penjualan, harga barang, produksi barang, nilai ujian), maka akan memperoleh sekumpulan nilai observasi sebagai berikut:

Populasi : $X_1, X_2, \dots, X_i, \dots, X_N$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = \text{rata-rata sebenarnya dari X (rata-rata populasi)}$$

Sampel : $X_1, X_2, \dots, X_i, \dots, X_n$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \text{rata-rata perkiraan (estimasi) dari } X \text{ (rata-rata sampel)}$$

\bar{X} adalah perkiraan dari μ

Simpangan baku (varians) terbagi ke dalam varians populasi dan varians sampel. Simbol dari varians populasi adalah \dagger^2 yang merupakan varians sebenarnya dari X . Rumusnya adalah:

$$\dagger^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \sim)^2$$

dimana $(X_i - \sim)$ adalah simpangan (deviasi) dari observasi terhadap rata-rata sebenarnya. Sedangkan varians sampel (S^2) dirumuskan sebagai berikut:

$$S^2 = \frac{1}{n \pm 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

atau,

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Dimana $(X_i - \bar{X})$ adalah simpangan (deviasi) dari observasi terhadap rata-rata sampel. Rumus dan simbol dari simpangan baku populasi adalah:

$$\dagger = \sqrt{\frac{\sum_{i=1}^N (X_i - \sim)^2}{N}}$$

Atau,

$$\dagger = \sqrt{\frac{1}{N} \left[\sum_{i=1}^N X_i^2 - \frac{\left(\sum_{i=1}^N X_i \right)^2}{N} \right]}$$

Dimana σ merupakan simpangan baku dari X .

Pada prakteknya, pengumpulan data yang hanya didasarkan atas sampel tidak menghasilkan varians atau simpangan baku yang sebenarnya, tetapi hanya suatu perkiraan saja dengan rumus sebagai berikut:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

Atau,

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Catatan: S = simpangan baku perkiraan (S perkiraan dari σ)/ simpangan baku sampel

Contoh

Hitung simpangan baku upah perminggu dari 11 buruh suatu perusahaan (ribuan rupiah)

Buruh	Upah (X)	$(X_i - \bar{X})^2$	Buruh	Upah (X)	$(X_i - \bar{X})^2$
A	11,10	10,9561	G	14,00	1,1681
B	11,50	8,4681	H	16,00	2,5281
C	12,00	5,8081	I	17,25	8,0685
D	12,20	4,8841	J	18,50	16,7281
E	12,45	3,8416	K	20,00	31,2481
F	13,50	0,8281			
			Jumlah	158,50	93,5241

Simpangan data dari tabel di atas sebagai berikut:

$$\begin{aligned} S &= \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \\ &= \sqrt{\frac{93,5241}{11}} \\ &= 3,06 \end{aligned}$$

BAB III Pengertian Penyajian Data

Data statistik yang telah dikumpulkan dan diolah, selanjutnya disajikan dalam bentuk yang mudah dimengerti, sehingga dapat dipergunakan sebagai dasar analisis atau interpretasi data statistik.

Bentuk penyajian data tersebut tergantung dari tujuan pengumpulan data dan kepada siapa data tersebut akan disajikan, sehingga perlu dipilih bentuk penyajian yang sesuai dengan kebutuhan dari pengguna data.

Bentuk penyajian data yang ditujukan bagi masyarakat umum, tentunya berbeda dengan bentuk penyajian data yang ditujukan bagi para peneliti atau ilmuwan. Penyajian data yang ditujukan bagi masyarakat umum harus sederhana serta lebih mudah dipahami dan dimengerti.

3.1 Pengertian Penyajian Data

Penyajian data merupakan salah satu kegiatan statistik yang dilakukan setelah tahap pengumpulan dan pengolahan data. Penyajian data dibuat untuk memberikan deskripsi mengenai data yang telah dikumpulkan dan memudahkan untuk pengambilan keputusan dalam pembuatan laporan hasil penelitian yang telah dilakukan sehingga sesuai dengan tujuan yang diinginkan.

3.2 Maksud dan Tujuan Penyajian Data

Maksud dan tujuan penyajian data adalah untuk :

1. Membandingkan beberapa variabel, beberapa kategori dalam variabel atau satu variabel pada waktu dan tempat yang berbeda.
2. Meramalkan perubahan yang terjadi dengan berjalannya waktu (*time series*).
3. Mengetahui adanya hubungan dua variabel atau lebih.
4. Memberikan informasi yang akurat hasil sebuah penelitian.

3.3 Bentuk-Bentuk Penyajian Data

Menurut bentuknya penyajian data dapat dibagi menjadi 4 bagian yaitu :

- a. Penyajian data dalam bentuk **tabel**.
- b. Penyajian data dalam bentuk **grafik**.
- c. Penyajian data dalam bentuk **ukuran-ukuran statistik**.
- d. Penyajian data dalam bentuk **uraian tertulis**.

Penyajian data dalam bentuk uraian tertulis biasanya digunakan pada artikel-artikel di majalah atau surat kabar.

3.4 Rangkuman

Penyajian data merupakan salah satu kegiatan statistik yang dilakukan setelah tahap pengumpulan dan pengolahan data.

Bentuk penyajian data yang ditujukan bagi masyarakat umum, tentunya berbeda dengan bentuk penyajian data yang ditujukan bagi para peneliti atau ilmuwan. Penyajian data yang ditujukan bagi masyarakat umum harus sederhana serta lebih mudah dipahami dan dimengerti.

BAB IV Penyajian Data Dalam Bentuk Tabel

Sebelum kita membahas lebih banyak mengenai bentuk-bentuk penyajian tabel dan cara pembuatannya, ada baiknya kita mengetahui dahulu tujuan serta manfaat penyajian data dalam bentuk tabel. Diharapkan dengan memahami tujuan dan manfaatnya, diharapkan akan mempermudah dalam membuat tabel dan membedakan jenis-jenis tabel.

4.1 Tujuan Penyajian Tabel

Tabel merupakan penyajian data dalam bentuk angka-angka yang disusun secara sistematis menurut baris dan kolom.

Tujuan penyajian data dalam bentuk tabel adalah :

1. Memudahkan dalam analisis data
2. Informasi yang disajikan lebih lengkap
3. Mudah dimengerti oleh pengguna data
4. Bentuk yang paling umum dan efektif dalam menyajikan informasi statistik

4.2 Jenis-Jenis Tabel

Jenis-jenis penyajian tabel adalah sebagai berikut:

1. Tabel satu arah (*One way table*)
2. Tabel dua arah (*Two way table*)
3. Tabel tiga arah (*Three way table*)

Hal-hal Yang Perlu Diperhatikan Dalam Penyusunan Tabel :

- a. Judul tabel, sebaiknya menjelaskan seluruh karakteristik yang ada pada tabel, yaitu mengindikasikan isi tabel tentang apa, bagaimana diklasifikasikan dan dimana (tempat) serta kapan (waktu) data yang disajikan tersebut dikumpulkan. Judul tabel usahakan singkat mungkin tapi jelas.
-

- b. Judul Kolom, menentukan susunan rincian dalam kolom pertama (stub). Jika rincian dalam kolom pertama (stub) terlalu panjang dapat dibuat dua baris tetapi masih terpisah dengan rincian yang lain.
- c. Nomor Kolom, masing-masing kolom dibuat nomor untuk mempermudah mengidentifikasi setiap rincian kolom.
- d. Isi Tabel, memeriksa nilai-nilai dalam setiap sel. Bila diperlukan nilai jumlah, maka perlu dikoreksi jumlahnya (terhadap baris maupun kolom). Menyediakan sub jumlah bila diperlukan, dan tiap sel dalam tabel harus ada isian
- e. Sumber Data, diperlukan bila sebagian atau seluruh data dalam tabel diambil dari publikasi lain (ini merupakan etika).
- f. Catatan kaki, kadang-kadang diperlukan untuk menjelaskan suatu angka atau suatu baris/kolom tertentu, atau untuk mengingatkan pembaca akan keterbatasan suatu angka, misalnya angka sementara atau angka tidak lengkap (data tahunan, tapi hanya 11 bulan).

Contoh :

Tabel ... Jumlah Penduduk DKI Jakarta Menurut Kab/Kota Tahun 2012	
Kab/Kota	Jumlah Penduduk *)
(1)	(2)
Judul Kolom	Isi Tabel

Sumber : Jakarta Dalam Angka Tahun 2013

*) Angka Sementara

← Judul Tabel

← Judul Kolom

← Nomor Kolom

Isi Tabel

← Sumber Data

↳ Catatan Kaki

Beberapa jenis penyajian tabel dapat diuraikan berikut ini:

4.2.1 Tabel Satu Arah (*One Way Table*)

Tabel satu arah adalah tabel yang dirinci menurut 1 (satu) karakteristik.

Isian setiap sel hanya menjelaskan kategori tertentu dari satu karakteristik.

Misalnya :

- Data penduduk menurut kabupaten/kota
- Data pegawai menurut kelompok umur

Contoh :

Tabel 5. Jumlah Penduduk DKI Jakarta Menurut Kab/Kota, 2012

Kab/Kota	Jumlah Penduduk
(1)	(2)
Kep. Seribu	23.129
Jakarta Selatan	2.052.645
Jakarta Timur	2.764.864
Jakarta Pusat	1.061.339
Jakarta Barat	2.214.401
Jakarta Utara	1.645.029
DKI Jakarta	9.761.407

Sumber : Jakarta Dalam Angka Tahun 2013

Isi sel bernilai 1.061.339 (kolom 2) merupakan jumlah penduduk di “Jakarta Pusat” sebanyak 1.061.339 jiwa. Dalam hal ini isian setiap sel hanya menjelaskan satu karakteristik saja yaitu kabupaten/kota.

4.2.2 Tabel Dua Arah (*Two Way Table*)

Tabel dua arah adalah tabel yang dirinci menurut 2 (dua) karakteristik.

Isian setiap sel menjelaskan rincian dari 2 (dua) karakteristik.

Misalnya :

- Data penduduk menurut kabupaten/kota dan jenis kelamin
- Data pegawai menurut kelompok umur dan pendidikan yang ditamatkan

Contoh :

**Tabel 6. Jumlah Penduduk DKI Jakarta
Menurut Kab/Kota dan Jenis Kelamin, Tahun 2012**

Kab/Kota	Laki-laki	Perempuan	Total
(1)	(2)	(3)	(4)
Kep Seribu	11.854	11.275	23.129
Jakarta Selatan	1.054.501	998.144	2.052.645
Jakarta Timur	1.423.261	1.341.603	2.764.864
Jakarta Pusat	542.784	518.555	1.061.339
Jakarta Barat	1.144.264	1.070.137	2.214.401
Jakarta Utara	849.725	795.304	1.645.029
DKI Jakarta	5.026.389	4.735.018	9.761.407

Sumber : Jakarta Dalam Angka Tahun 2013

Isi sel bernilai 795.304 berarti banyaknya penduduk di “Jakarta Utara” dan berjenis kelamin “Perempuan” adalah 795.304 jiwa. Dalam hal ini isian setiap sel menjelaskan dua karakteristik yaitu wilayah dan jenis kelamin.

4.2.3 Tabel Tiga Arah (*Three Way Table*)

Tabel tiga arah adalah tabel yang dirinci menurut 3 (tiga) karakteristik. Isian setiap sel menjelaskan rincian dari 3 (tiga) karakteristik.

Misalnya:

- Data penduduk menurut kabupaten/kota, jenis kelamin, dan kelompok umur
- Data pegawai menurut kelompok umur, pendidikan yang ditamatkan, dan jenis kelamin

Contoh :

Tabel 7. Jumlah Penduduk Usia 10 Tahun ke Atas di DKI Jakarta menurut Kab./Kota, Jenis Kelamin, dan Kemampuan Membaca dan Menulis, 2012

Kab/Kota	Dapat Membaca/Menulis			Tidak Dapat Membaca/Menulis		
	Laki-laki	Perempuan	Jumlah	Laki-laki	Perempuan	Jumlah
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Kep Seribu	8.771	8.230	17.001	124	331	455
Jakarta Selatan	891.810	870.269	1.762.079	2.750	9.015	11.765
Jakarta Timur	1.156.187	1.117.173	2.273.360	3.450	6.450	9.900
Jakarta Pusat	381.476	377.653	759.129	425	3.404	3.829
Jakarta Barat	1.004.350	950.147	1.954.497	2.859	20.483	23.342
Jakarta Utara	700.588	692.260	1.392.848	4.387	16.948	21.335
DKI Jakarta	4.143.182	4.015.732	8.158.914	13.995	56.631	70.626

Sumber: Jakarta Dalam Angka Tahun 2013

Isi sel bernilai 377.653 berarti jumlah penduduk "perempuan" usia 10 tahun ke atas di "Jakarta Pusat" yang "dapat membaca dan menulis" adalah sebanyak 377.653 orang. Dalam hal ini isian setiap sel menjelaskan tiga karakteristik yaitu kabupaten/kota, jenis kelamin, dan kemampuan membaca dan menulis.

4.3 Rangkuman

Tabel merupakan penyajian data dalam bentuk angka-angka yang disusun secara sistematis menurut baris dan kolom. Penyajian data dalam bentuk

tabel adalah bentuk yang paling umum dan efektif dalam menyajikan informasi statistik.

Jenis-jenis penyajian tabel adalah sebagai berikut:

1. Tabel Satu Arah (*One Way Table*), tabel yang dirinci menurut 1 (satu) karakteristik
2. Tabel Dua Arah (*Two Way Table*), tabel yang dirinci menurut 2 (dua) karakteristik
3. Tabel Tiga Arah (*Three Way Table*), tabel yang dirinci menurut 3 (tiga) karakteristik

BAB V Penyajian Data Dalam Bentuk Grafik

Grafik merupakan bentuk penyajian data secara fisual. Penyajian data dalam bentuk fisual pada umumnya lebih mudah dibaca dibandingkan dengan melihat angka-angka dalam tabel.

Dasar pembuatan grafik itu sendiri adalah data-data yang telah disajikan ke dalam bentuk tabel. Dalam hal ini tabel yang akan dibuat grafiknya sebaiknya dibatasi pada tabel dua arah agar grafiknya menarik dan mudah dibaca. Jika tabel yang akan dibuat grafiknya merupakan tabel lebih dari dua arah, maka sebaiknya tabel tersebut harus dipecah sehingga menjadi beberapa tabel dua arah.

5.1 Tujuan Penyajian Grafik

Secara umum penyajian data dalam bentuk grafik bertujuan untuk mempertegas dan memperjelas penyajian data.

Secara khusus, tujuan penyajian data dalam bentuk grafik adalah:

- Dapat menunjukkan suatu perkembangan data dari waktu ke waktu atau perbandingan antara dua hal atau lebih.
- Umumnya lebih mudah membaca secara fisual daripada membaca angka-angka dalam tabel.
- Tampilan lebih menarik dibandingkan tabel.
- Mempercepat pengertian/pemahaman sehingga dapat ditarik kesimpulan.
- Mengurangi kejenuhan melihat angka-angka.
- Menunjukkan gambaran secara umum/menyeluruh dengan mudah.

5.2 Jenis-Jenis Grafik

Grafik dibedakan menjadi 2 (dua) macam, yaitu :

5.2.1 Grafik Tabel Distribusi Frekuensi

Grafik yang dibuat berdasarkan tabel distribusi frekuensi adalah:

a. Histogram

Histogram merupakan grafik dari distribusi frekuensi suatu variabel. Tampilan histogram berupa petak-petak empat persegi panjang. Sebagai sumbu horizontal (absis, sumbu x) boleh memakai tepi-tepi kelas (*class boundaries*), batas-batas kelas (*class limits*) atau nilai-nilai variabel yang diobservasi, sedang sumbu vertikal (ordinat, sumbu y) menunjukkan frekuensi. Untuk distribusi bergolong/kelompok yang menjadi absis adalah nilai tengah dari masing-masing kelas.

b. Poligon Frekuensi

Poligon frekuensi merupakan grafik dari distribusi frekuensi bergolong suatu variabel. Tampilan poligon berupa garis-garis patah yang diperoleh dengan cara menghubungkan puncak dari masing-masing nilai tengah kelas. Jadi absisnya adalah nilai tengah dari masing-masing kelas.

c. Kurva Ogive

Ogive merupakan grafik dari distribusi frekuensi kumulatif suatu variabel. Dalam suatu ogive, yang digunakan sebagai absis adalah batas kelas (*class boundaries*), sedangkan sebagai sumbu vertikal adalah frekuensi kumulatif. Untuk suatu tabel distribusi frekuensi, dapat dibuat *ogive* frekuensi kumulatif “kurang dari” (positif) dan frekuensi kumulatif “lebih dari” (negatif).

Proses pembuatan grafik yang dibuat berdasarkan tabel distribusi frekuensi adalah sebagai berikut:

Penyusunan Data

Untuk mempermudah pekerjaan khususnya dalam melakukan penataan data adalah dengan cara mengurutkannya/*Sorting (Ascending/Descending)*.

Disamping itu dalam membaca tabel statistik orang tidak selalu memerlukan gambaran yang terlalu panjang dan terinci, tetapi cukup dapat terlihat kecenderungan dari data (nilai ekstrim) secara umum. Untuk itu seringkali disusun dalam kelompok – kelompok dalam tabel yang disebut Tabel Distribusi Frekuensi.

Tujuan pengelompokkan data ke dalam Distribusi Frekuensi adalah guna memperoleh gambaran yang sederhana, jelas dan sistematis mengenai peristiwa yang dinyatakan dalam angka-angka.

Membuat tabel distribusi frekuensi berarti mendistribusikan data kedalam beberapa kelas atau kategori, kemudian menentukan banyaknya individu yang termasuk kelas tertentu yang disebut frekuensi kelas (Class Frequency).

Ada beberapa ketentuan dan istilah sehubungan dengan tabel Distribusi Frekuensi :

a. Batas Kelas (*Class Limits*)

Dua buah nilai yang membatasi suatu kelas. Nilai yang dibawah disebut “batas bawah” kelas (*LCL / Lower Class Limits*), sedangkan yang diatas disebut “batas atas” kelas (*UCL / Upper Class Limits*).

b. Nilai Tengah (*Class Mark, Mid Point*)

Nilai tengah dari kelas, yaitu : $(\text{Batas bawah} + \text{Batas Atas}) / 2$

c. Selang Kelas (*Class Interval*)

Selang antara batas atas dengan batas sebelumnya dari kelas yang berurutan, atau selisih antara batas bawah sebelumnya dari kelas yang berurutan. Dalam tabel distribusi frekuensi besarnya selang kelas pada umumnya sama.

d. Rentang (*Range / R*)

Selisih antara nilai yang terbesar dan yang terkecil dari data (Max-Min).

e. Banyaknya Kelas = Rentang : Selang Kelas - $K = 1 + 3,322 \log n$

Dalam tabel dist. freq. Banyaknya kelas sebaiknya antara 8 s.d 20 kelas.

f. Titik Awal (*Starting Point*)

Batas bawah dari kelas pertama untuk starting point biasanya dipilih angka yang “ baik “ (kelipatan dari 5 atau berakhir dengan 0).

g. Gap

Selisih antara batas suatu kelas dengan batas bawah dari kelas berikutnya. Bila data mentah terdiri dari angka-angka maka gap =1, bila data mentah terdiri dari angka-angka satu desimal maka gap adalah 0,1. Bila data mentah terdiri dari angka-angka dua desimal maka gap adalah 0,01.

Tahapan yang dilakukan dalam membuat tabel distribusi frekuensi:

1. Tentukan besarnya rentang. (Nilai Maximum – Nilai Minimum)
2. Tentukan banyaknya kelas : $k = 1 + 3,322 \log n$
 k = banyaknya kelas, n = jumlah data
3. Tentukan besarnya selang kelas (*class interval*).
 $I = \text{Jarak} / (1 + 3,3 \log n)$
4. Tentukan besarnya interval kelas /selang kelas/lebar kelas (LK)/panjang kelas dengan cara membagi range dan Jumlah Kelas / JK
 $LK = R / JK$
 $LK = LCL_{i+1} - LCL_i$
 $LK = UCL_{i+1} - UCL_i$
5. Penentuan $LCL_i = X_{\min}$ (tidak harus)
6. Tabel Distribusi
7. Grafik
 - Histogram
 - Poligon frekuensi
 - Ogive

Bentuk Tabel Distribusi Frekuensi

Kelas	Selang Kelas	Batas Kelas	Mid Point	Frekuensi	
				Absolut	Relatif
	$LCL - UCL$	$LCB - UCB$	$\frac{LCL+UCL}{2}$		

				n	%
--	--	--	--	---	---

Contoh Soal :

Berikut ini adalah 30 data berat badan anggota suatu gymnasium dalam kg, buatlah tabel distribusi frekuensi dari data di bawah

30	60	30	56	60	54
42	35	42	40	47	58
70	80	48	68	39	40
50	75	68	38	67	45
65	66	50	54	46	55

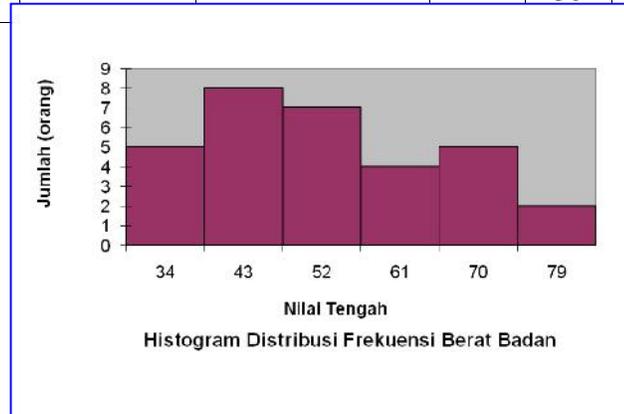
Langkah-langkah

1. Urutkan
2. $R = X_{\max} - X_{\min}$
 $R = 80 - 30 = 50$
3. $JK = 1 + 3,3 \log 30 = 5,8 = 6$
4. $LK = R/JK$
 $= 50/6 = 8,3 = 9$

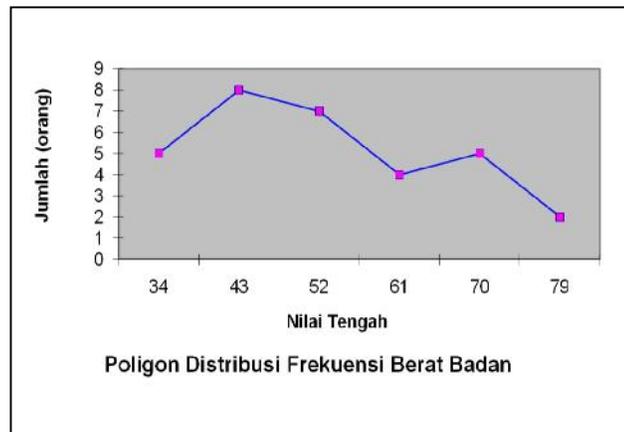
Tabel Distribusi Frekuensi_nya :

Kelas	Selang Kelas	Batas Kelas	Mi	Frek	
				Abs	Rel (%)
1	30 – 38	29.5 - 38.5	34	4	13.33
2	39 – 47	38.5 - 47.5	43	8	26.67
3	48 – 56	47.5 - 56.5	52	7	23.33
4	57 – 65	56.5 - 65.5	61	4	13.33
5	66 – 74	65.5 - 74.5	70	5	16.67

6	75 - 83	74.5 - 83.5	79	2	6.67
				30	100



Grafik 1. Histogram Distribusi Frekuensi Berat Badan



Grafik 2. Poligon Distribusi Frekuensi Berat Badan

Tabel Kurang Dari

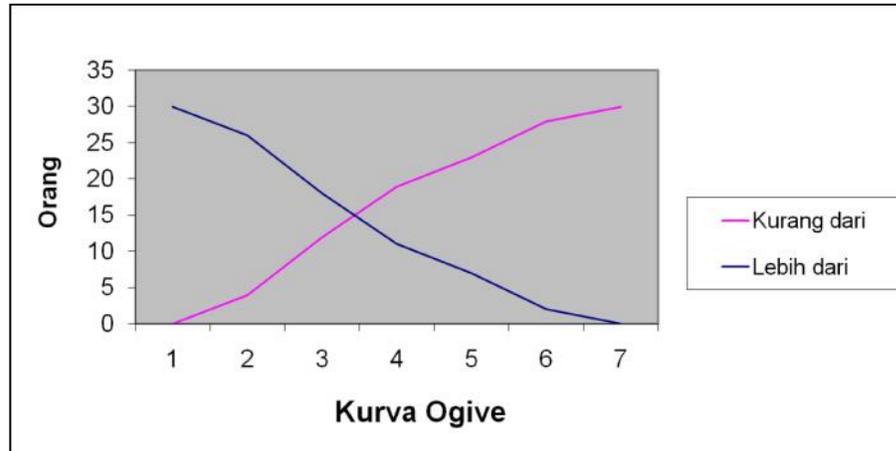
Kurang dari	Frekuensi	
	Abs	Rel (%)
29.5	0	0.0
38.5	4	13.3
47.5	12	40.0
56.5	19	63.3

Tabel Lebih Dari

Lebih dari	Frekuensi	
	Abs	Rel (%)
29.5	30	100.0
38.5	26	86.7
47.5	18	60.0
56.5	11	36.7

65.5	23	76.7
74.5	28	93.3
83.5	30	100.0

65.5	7	23.3
74.5	2	6.7
83.5	0	0.0



Grafik 3. Kurva Ogive

5.2.2 Grafik dari Bentuk Tabel

Jenis grafik inilah yang akan dibahas pada materi penyajian data.

Secara garis besar jenis penyajian grafik/diagram dibedakan atas :

- Grafik Garis (*Line chart*)
- Grafik Batang (*Bar chart*)
- Grafik Lingkaran (*Pie chart*)
- Diagram Gambar (*Pictogram*)
- Diagram Peta (*Cartogram*)

Agar tujuan membuat grafik dapat mencapai sasaran maka dalam memilih grafik yang sesuai perlu diperhatikan hal-hal sebagai berikut :

- Dari sudut penganalisaan/penarikan kesimpulannya.

Apakah yang ingin dilihat, "perkembangan data" atau "perbandingan data".

Bila ingin menganalisa perkembangan data dari beberapa periode tertentu, maka lebih tepat data digambarkan dalam bentuk grafik garis (*line chart*). Bila ingin menganalisa perbandingan data dari beberapa kejadian atau keadaan pada saat tertentu, maka lebih tepat data digambarkan dalam bentuk grafik batang/balok (*bar chart*), grafik lingkaran (*pie chart*), atau diagram gambar (*pictogram*).

- b. Dari bentuk/macam data yang disajikan dalam tabel yang ingin dibuat grafiknya.

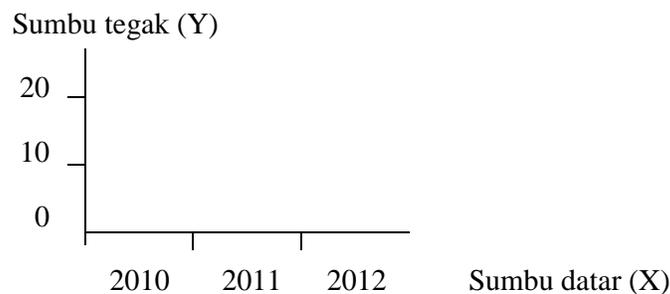
Ada data menurut deret waktu, data dalam dua periode atau lebih yang tidak menurut deret waktu, dan sebagainya.

- c. Harus disesuaikan dengan tingkat pengetahuan konsumen data.

Agar grafik yang kita buat dapat dimengerti oleh konsumen data, maka grafik harus dibuat sedemikian rupa (misalnya gambar-gambar yang menarik yang dapat membantu membaca angka-angka statistik yang disajikan bagi masyarakat awam) sehingga penyajian data tersebut akan mencapai sasaran.

Ketentuan dalam pembuatan grafik:

1. Perbandingan sumbu tegak (y) dan sumbu datar (x) tidak terlalu jauh, kecuali grafik lingkaran tidak menggunakan sumbu.
2. Ukuran grafik tidak terlalu besar, tinggi, dan pendek
3. Kelengkapan grafik terdiri dari:
 - Nomor grafik dicantumkan sebelum judul.
 - Judul grafik menyatakan jenis, tempat, dan periode data.
 - Sumber data
 - Catatan kaki/keterangan jika diperlukan.



Beberapa jenis penyajian grafik dapat diuraikan berikut ini:

a. Grafik Garis (*Line Chart*)

Grafik garis adalah grafik yang menggambarkan perkembangan data. Pola atau kecenderungan data dapat dengan mudah kita ketahui dari grafik garis, yaitu dengan melihat arah garis yang menghubungkan titik-titik pada grafik tersebut.

Grafik ini dapat dibagi atas :

- 1) Grafik garis tunggal (*single line chart*)
- 2) Grafik garis berganda (*multiple line chart*)
- 3) Grafik garis komponen berganda (*multiple component line chart*)
- 4) Grafik garis komponen berganda dalam persentase (*percentage multiple component line chart*)
- 5) Grafik garis saling menimbang (*balance line chart*)
- 6) Grafik garis interval (*interval line chart*)

1) Grafik Garis Tunggal (*Single Line Chart*)

Grafik garis tunggal biasanya digunakan untuk memberikan gambaran perkembangan satu jenis data. Misalnya ingin melihat perkembangan data jumlah sepeda motor dari tahun 2008 sampai dengan tahun 2012. Grafik garis tunggal dibuat berdasarkan tabel satu arah seperti tampak pada tabel berikut ini:

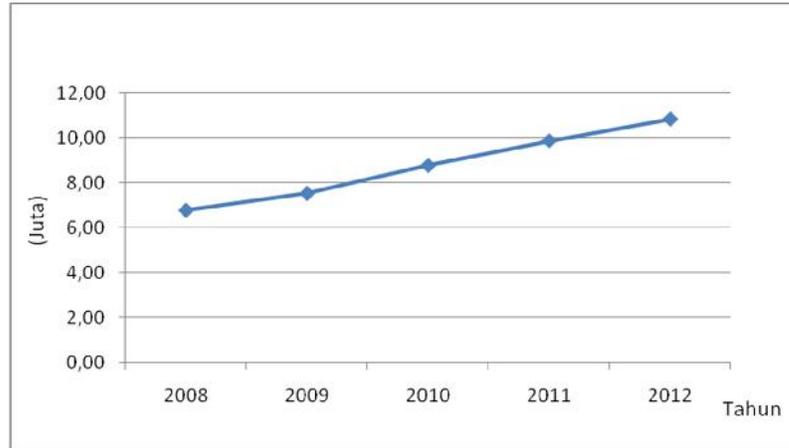
Tabel 8. Perkembangan Jumlah Sepeda Motor di DKI Jakarta, 2008 – 2012

Tahun	Sepeda Motor
(1)	(2)
2008	6.765.723
2009	7.518.098
2010	8.764.130
2011	9.861.973

2012	10.825.973
------	------------

Sumber: Jakarta Dalam Angka Tahun 2013

Berdasarkan Tabel 4 di atas dapat dibuat grafik garis sebagai berikut:



Grafik 4. Perkembangan Jumlah Sepeda Motor di DKI Jakarta, 2008 – 2012 (Juta)

2) Grafik Garis Berganda (*Multiple Line Chart*)

Grafik garis ini memberikan gambaran perkembangan dari beberapa data dalam satu tabel, misalnya ingin melihat perkembangan jumlah kendaraan bermotor (mobil penumpang, bis, dan truk) sejak tahun 2008 hingga 2012 seperti pada tabel dua arah berikut :

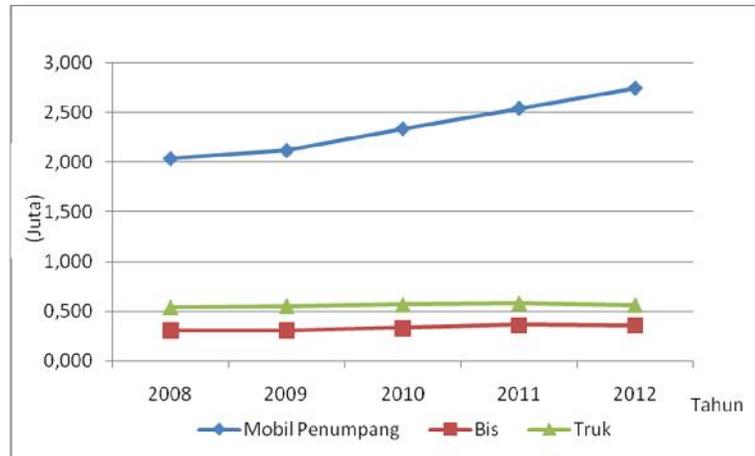
Tabel 9. Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008 – 2012

Tahun	Mobil Penumpang	Bis	Truk	Jumlah
(1)	(2)	(3)	(4)	(5)
2008	2.034.943	308.528	538.731	2.882.202
2009	2.116.282	309.385	550.924	2.976.591
2010	2.334.883	332.779	565.727	3.233.389

2011	2.541.351	363.710	581.290	3.486.351
2012	2.742.414	358.895	561.918	3.663.227

Sumber: Jakarta Dalam Angka Tahun 2013

Berdasarkan Tabel 5 di atas dapat dibuat grafik garis sebagai berikut:



Gambar 5. Grafik Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta)

3) Grafik Garis Komponen Berganda (*Multiple Component Line Chart*)

Grafik garis ini bertujuan untuk melihat perkembangan dari beberapa jenis data, sekaligus untuk melihat perkembangannya secara kumulatif. Sebelum kita membuat grafiknya perlu dibuat terlebih dahulu tabel yang merupakan jumlah kumulatif pada tiap-tiap tahun secara berurutan menurut masing-masing jenis datanya. Misalnya untuk data pada Tabel 5, dilakukan perhitungan sebagai berikut:

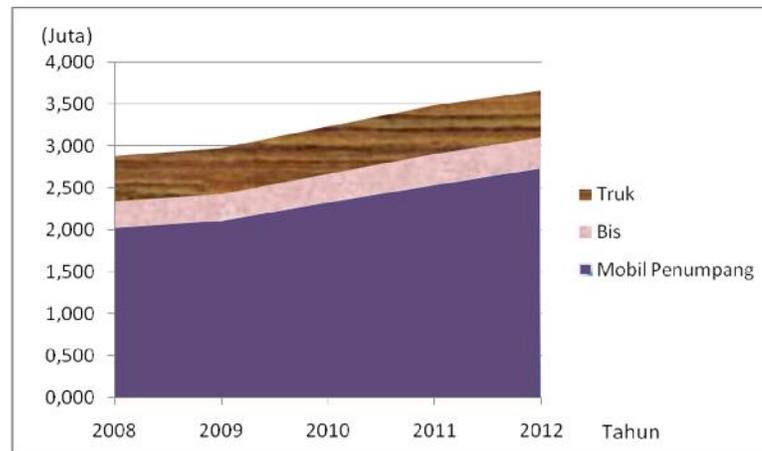
Diolah dari Tabel 5.

Tahun	Mobil Penumpang	Bis	Truk
(1)	(2)	(3)	(4)
2008	2,035	0,309	0,539
2009	2,116	0,309	0,551
		2,426	2,977

2010		0,333	0,566
	2,335	2,668	3,233
2011		0,364	0,581
	2,541	2,905	3,486
2012		0,359	0,562
	2,742	3,101	3,663

Catatan :

Cara penghitungan adalah sebagai berikut, misalnya untuk tahun 2008, angka **2,343** diperoleh dengan menjumlahkan angka 2,035 dengan 0,309, dan angka **2,882** diperoleh dengan menjumlahkan angka **2,343** dengan 0,539 dan seterusnya.



Grafik 6. Perkembangan Jumlah Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta)

4) Grafik Garis Komponen Berganda Dalam Persentase (Percentage Multiple Component Line Chart)

Cara membuat grafik ini pada prinsipnya sama dengan pembuatan grafik garis komponen berganda, hanya nilai yang diplotkan dalam bentuk persentase. Data pada Tabel 5 diubah dalam bentuk persentase (%), yaitu dengan menghitung nilai setiap sel masing-masing jenis kendaraan bermotor terhadap jumlah pada periode yang bersangkutan, sehingga data akan menjadi sebagai berikut :

Tabel 10. Perkembangan Persentase Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta)

Tahun	Mobil Penumpang	Bis	Truk	Jumlah
(1)	(2)	(3)	(4)	(6)
2008	70,60	10,70	18,69	100,00
2009	71,10	10,39	18,51	100,00
2010	72,21	10,29	17,50	100,00
2011	72,89	10,43	16,67	100,00
2012	74,86	9,80	15,34	100,00

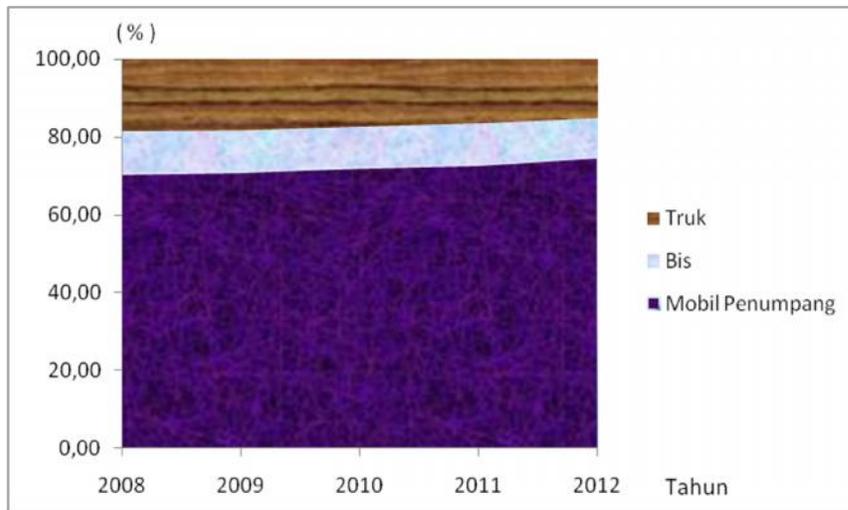
Sumber : Diolah dari Tabel 5.

Setelah itu perlu disusun tabel kumulatif berdasarkan Tabel 6 di atas.

Tahun	Mobil Penumpang	Bis	Truk
(1)	(2)	(3)	(4)
2008		10,70	18,69
	70,60	81,30	99,99
2009		10,39	18,51
	71,10	81,49	100,00
2010		10,29	17,50
	72,21	82,50	100,00
2011		10,43	16,67
	72,89	83,32	99,99
2012		9,80	15,34
	74,86	84,66	100,00

Diolah dari Tabel 6

Cara meletakkan (memplotkan) titik pada grafik, dilakukan seperti pada penyusunan Grafik 3 yaitu dengan menghubungkan nilai-nilai yang telah dikumulatifkan.



Grafik 7. Perkembangan Persentase Kendaraan Bermotor di DKI Jakarta, 2008 – 2012 (Juta)

5) Grafik Garis Saling Menimbang (*Balance Line Chart*)

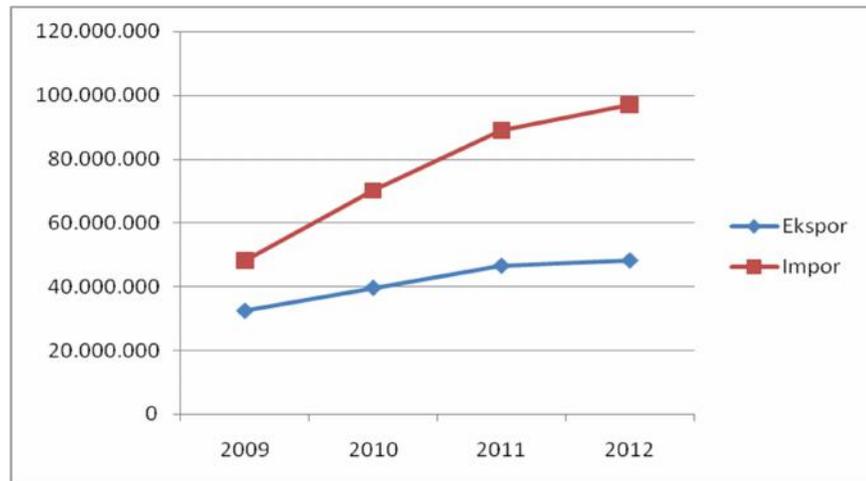
Kegunaan grafik ini adalah untuk melihat perkembangan dari dua kelompok data yang saling berlawanan, sehingga perkembangan selisih antara dua kelompok data tersebut dapat diketahui. Misalnya data tentang pendapatan dan pengeluaran, ekspor dan impor, dan sebagainya.

Contoh :

Tabel 11. Perkembangan Nilai Ekspor dan Impor Melalui DKI Jakarta, 2009 – 2012 (000 US \$)

Tahun	Ekspor	Impor
(1)	(2)	(3)
2009	32.536.510	48.099.308
2010	39.648.257	70.069.085
2011	46.476.171	88.874.102
2012	48.134.849	96.926.336

Sumber : Jakarta Dalam Angka Tahun 2013



Grafik 8. Perkembangan Nilai Ekspor dan Impor Melalui DKI Jakarta, 2009 – 2012 (000 US \$)

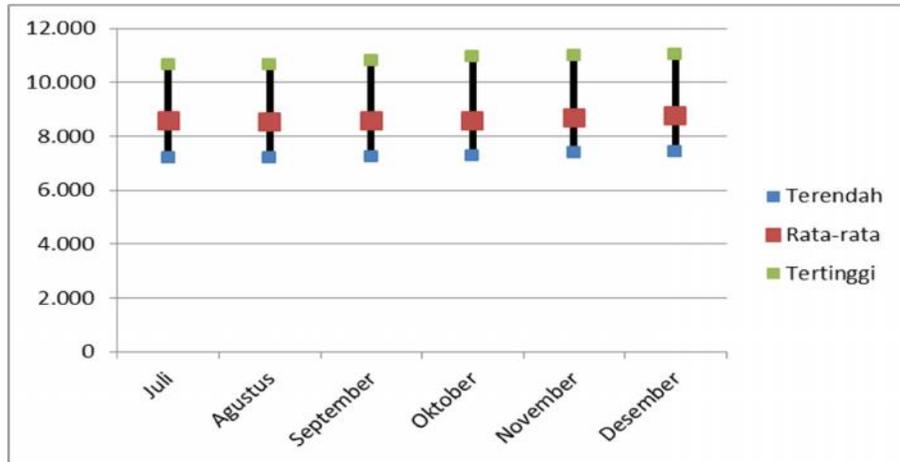
6) Grafik Garis Interval (*Interval Line Chart*)

Grafik ini untuk menggambarkan perkembangan data yang mempunyai dua nilai. Misalkan mengenai data harga beras tertinggi dan terendah suatu daerah, keuntungan terbesar dan terkecil dari suatu penjualan dan sebagainya.

Tabel 11. Harga Beras Grosir di Pasar Induk Cipinang DKI Jakarta, Juli – Desember 2012 (Rp/kg)

Bulan	Terendah	Rata-rata	Tertinggi
(1)	(2)	(3)	(4)
Juli	7.192	8.572	10.674
Agustus	7.202	8.543	10.648
September	7.240	8.560	10.820
Oktober	7.277	8.585	10.977
November	7.393	8.670	11.000
Desember	7.435	8.768	11.019

Sumber : Jakarta Dalam Angka Tahun 2013



Grafik 9. Harga Beras Grosir di Pasar Induk Cipinang DKI Jakarta, Juli – Desember 2012 (Rp/kg)

b. Grafik Batang (*Bar Chart*)

Kegunaan grafik batang adalah untuk menggambarkan perbandingan data antara data periode yang satu dengan periode lainnya, atau antara data yang satu dengan data lainnya pada periode yang berbeda.

Grafik batang mempunyai berbagai bentuk, yaitu:

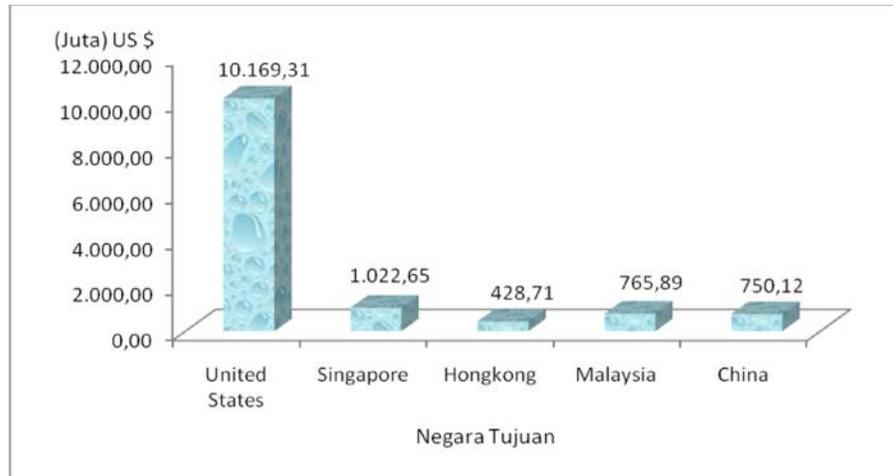
1) Grafik Batang Tunggal (*Single Bar Chart*)

Grafik batang ini untuk menggambarkan perbandingan beberapa data pada periode yang sama (berasal dari tabel satu arah).

Tabel 13. Nilai Ekspor Produk DKI Jakarta menurut Negara Tujuan, 2012 (Juta US \$)

Negara Tujuan	Nilai Ekspor
(1)	(2)
United States	10.169,3
Singapore	1.022,6
Hongkong	437,9
Malaysia	2.612,8
China	301,0

Sumber: Jakarta Dalam Angka Tahun 2013



Grafik 10. Nilai Ekspor Produk DKI Jakarta menurut Negara Tujuan, 2012 (Juta US \$)

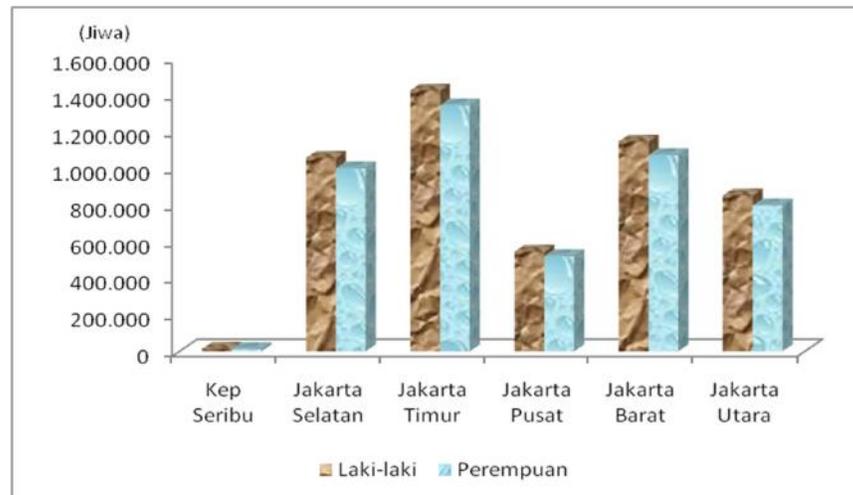
2) Grafik Batang Berganda (Multiple Bar Chart)

Grafik ini menggambarkan perbandingan beberapa data yang dirinci menurut beberapa kategori. Cara pembuatannya adalah sebagai berikut, misalnya akan digambarkan data Tabel 2, maka penggambaran batang untuk setiap kategori harus berimpit untuk setiap kategori, dan dipisahkan (diberi jarak) dengan kategori lainnya.

Contoh: Diolah dari Tabel 2



Grafik 11. Jumlah Penduduk DKI Jakarta menurut Kab/Kota dan Jenis Kelamin, 2012



Grafik 12. Jumlah Penduduk DKI Jakarta menurut Jenis Kelamin dan Kab/Kota, 2012

3) Grafik Batang Komponen Berganda (*Multiple Component Bar Chart*)

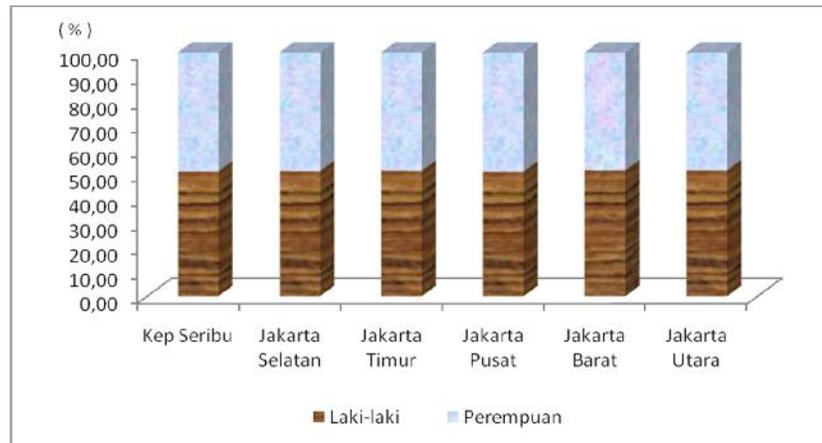
Grafik ini menggambarkan perbandingan data yang dirinci menurut beberapa kategori sekaligus dapat menggambarkan perbandingan jumlah datanya. Cara pembuatannya dengan membuat nilai kumulatif pada masing-masing periode. Sebagai contoh, berdasarkan Tabel 2 dapat dibuat nilai kumulatifnya sebagai berikut :



Gambar 13. Grafik Jumlah Penduduk DKI Jakarta menurut Kab/Kota dan Jenis Kelamin, 2012

4) Grafik Batang Komponen Berganda Dalam Persentase (*Percentage Multiple Component Bar Chart*)

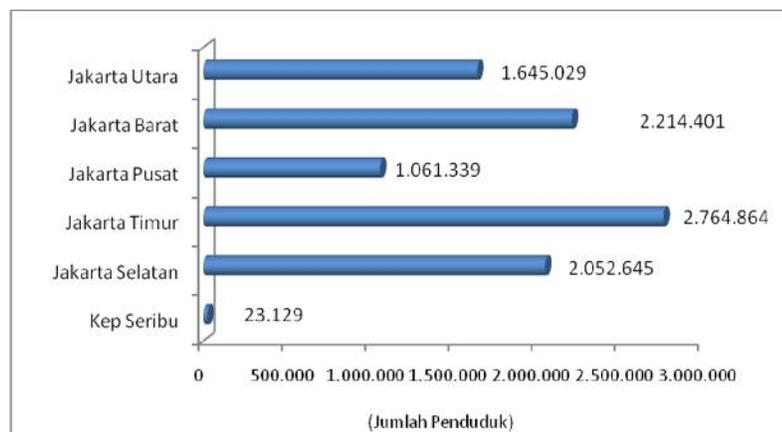
Grafik ini menggambarkan perbandingan data dalam persentase yang telah dirinci menurut beberapa kategori. Cara penggambarannya sama seperti grafik batang komponen setelah nilai-nilai datanya diubah dalam persentase. Sebagai contoh data dari Tabel 2, setelah diubah dalam persentase dan dibuat kumulatif untuk setiap jenis kelamin, maka grafiknya adalah sebagai berikut :



Grafik 14. Persentase Penduduk DKI Jakarta menurut Jenis Kelamin dan Kab/Kota, 2012

5) Grafik Batang Mendatar (*Horizontal Bar Chart*)

Grafik ini biasanya untuk menggambarkan perbandingan antara kategori yang satu dengan kategori lainnya pada suatu periode tertentu (bukan kategori dalam deret waktu). Agar lebih mudah membandingkan data-data tersebut, maka nilai data disusun dari yang terbesar ke nilai yang terkecil atau sebaliknya. Sebagai contoh digunakan data dari Tabel 1.



Grafik 15. Jumlah Penduduk DKI Jakarta menurut Kab/Kota, 2012

c. Grafik Lingkaran (*Pie Chart*)

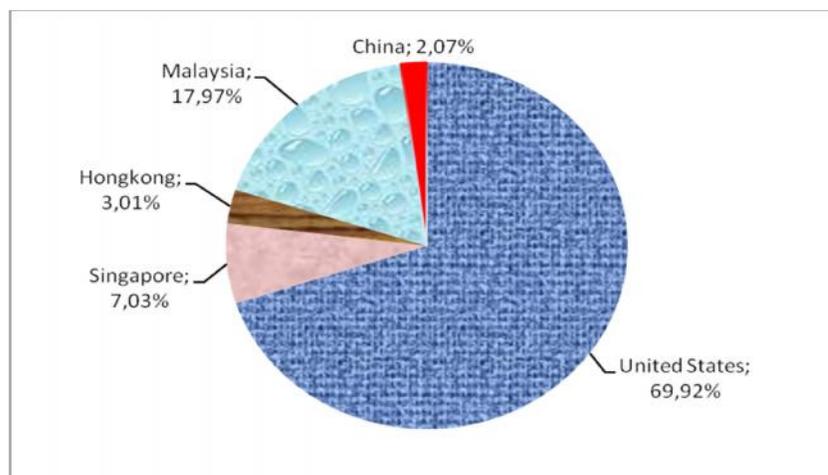
Kegunaan grafik lingkaran pada dasarnya sama dengan grafik batang tunggal, yaitu untuk memberi gambaran mengenai perbandingan beberapa data. Perbedaannya adalah pada grafik lingkaran perbandingan tersebut dilihat dari nilai persentasenya, sedangkan grafik batang tunggal adalah nilai mutlaknya. Diagram lingkaran dibuat berdasarkan tabel persentase satu arah, langkah-langkahnya adalah sebagai berikut :

- ⊙ Gambarkan sebuah lingkaran lalu dibagi-bagi dalam beberapa sektor.
- ⊙ Tiap sektor merupakan kategori data yang terlebih dahulu diubah ke dalam "derajat"

Contoh:

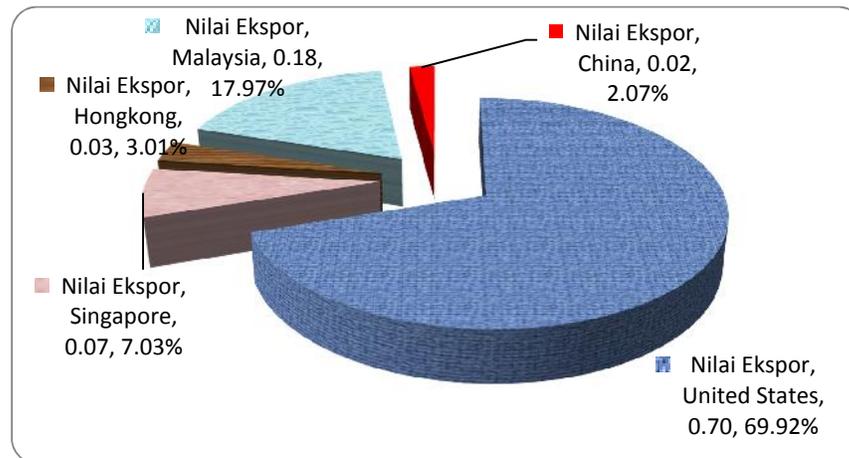
Dari Tabel 9 dapat dibuat persentase dan derajat sektornya sebagai berikut.

Negara Tujuan	Nilai Ekspor	(%)	Derajat
(1)	(2)	(3)	(4)
United States	10.169,30	69,92	251,72
Singapore	1.022,60	7,03	25,31
Hongkong	437,90	3,01	10,84
Malaysia	2.612,80	17,97	64,68
China	301	2,07	7,45
Jumlah	14.543,60	100,00	360,00



Grafik 16. Persentase Nilai Ekspor Produk DKI Jakarta menurut Negara Tujuan, 2012

Atau dengan bentuk grafik lingkaran lainnya seperti:



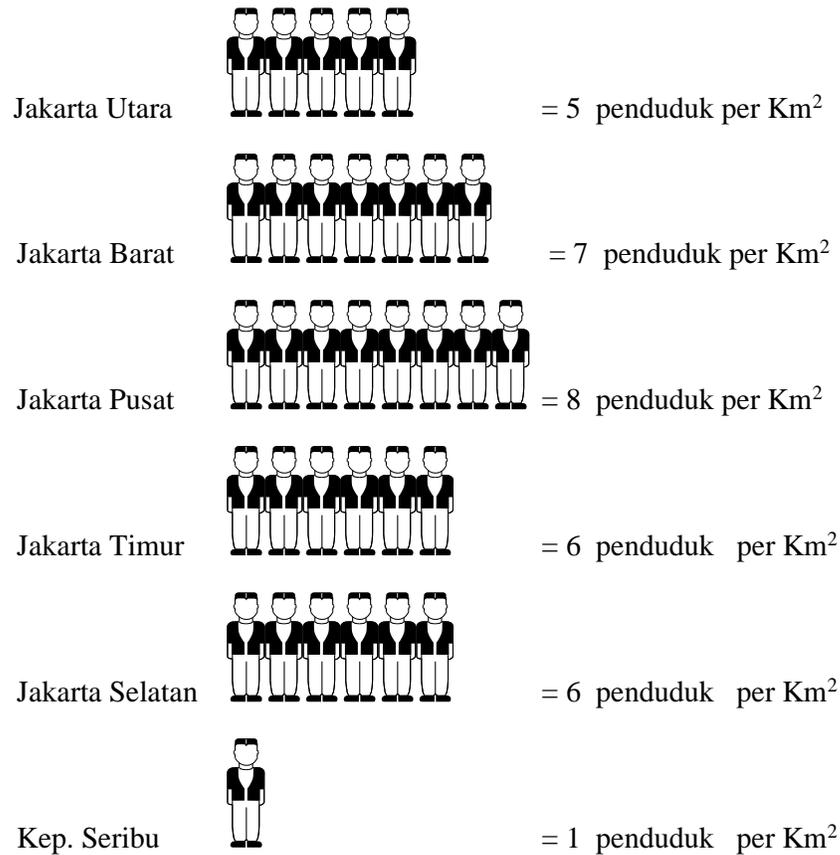
d. Diagram Gambar (*Pictogram*)

Diagram gambar sering dipakai untuk mendapatkan gambaran secara kasar dari suatu persoalan dan sebagai penyajian visual yang baik bagi orang awam, lebih-lebih jika simbol yang digunakan cukup menarik. Sebagaimana tersebut dalam kutipan berikut :

"Penyajian grafik yang paling menarik ialah penyajian secara piktografis, yang dapat disajikan secara tepat serta mengena. Bentuk diagram sedemikian itu sebetulnya tidak memberikan perbandingan yang memuaskan pada pembacanya, tetapi penyajian sedemikian itu sangat menarik perhatian. Efek lukisannya sangat mengesankan" (Anto Dajan, 1986, hal. 71).

Dalam grafik gambar, setiap satuan jumlah tertentu dibuat sebuah simbol sesuai dengan macam datanya. Misalnya untuk data mengenai jiwa, penduduk dan pegawai dibuat gambar orang, satu gambar untuk tiap 5000 jiwa. Sedangkan untuk data bangunan, gedung sekolah dan bangunan lain dibuat gambar gedung, misalnya satu gedung menyatakan 1000 buah, dan masih banyak contoh lain lagi. Kesulitan yang dihadapi ialah ketika menggambarkan bagian simbol untuk satuan yang tidak penuh.

Contoh:



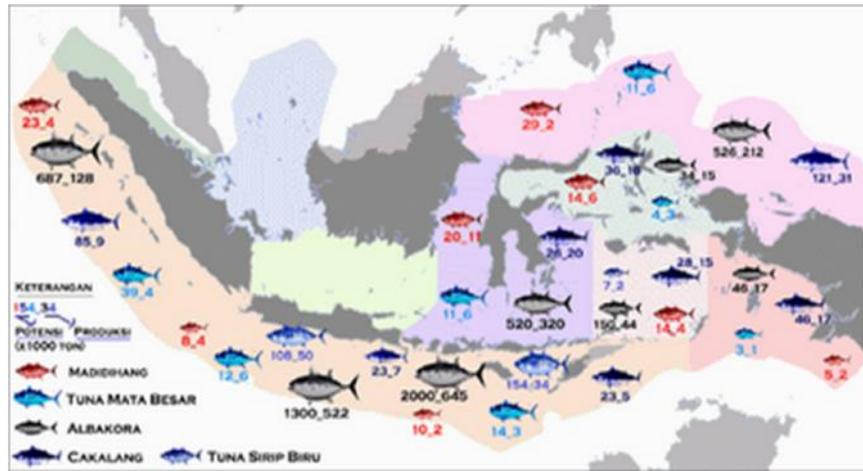
Sumber : Proyeksi Penduduk Hasil Sensus Penduduk 2010, DKI Jakarta

Gambar 17. Grafik Kepadatan Penduduk Per Km² Menurut Kabupaten/Kota di DKI Jakarta, 2012

e. Diagram Peta (*Cartogram*)

Diagram peta (*cartogram*) adalah diagram yang memberikan keterangan tempat dari jenis data yang ada pada sebuah peta. Misalnya potensi ikan di perairan Indonesia, maka data tersebut ditulis dalam peta yang dibuat, dan ditempatkan sesuai dengan daerah/lokasi tersebut. Dalam pembuatannya, digunakan peta geografis atau peta wilayah di mana terdapat data yang akan disajikan. Diagram peta (*cartogram*) dapat juga digambarkan dengan menambahkan grafik lainnya ke dalam peta, misalnya grafik gambar, batang dan lingkaran.

Contoh :



Grafik 18. Peta Potensi Ikan Perairan Indonesia, Tahun 2011

4.3 Rangkuman

Grafik merupakan bentuk penyajian data secara fisual. Penyajian data dengan grafik dianggap lebih komunikatif karena dalam waktu singkat dapat diketahui karakteristik dari data yang disajikan.

Kelebihan penyajian data dalam bentuk grafik diantaranya adalah :

1. Lebih mudah diingat.
2. Tampilan penyajian lebih menarik.
3. Informasi visual dan dapat diperbandingkan.
4. Menyajikan perubahan hubungan variabel.

Agar tujuan membuat grafik dapat mencapai sasaran, maka perlu dipilih grafik yang sesuai dengan analisisnya, yaitu:

- Apabila ingin menganalisa perkembangan data dari beberapa periode tertentu, maka lebih tepat data digambarkan dalam bentuk grafik garis (*line chart*).
- Apabila ingin menganalisa perbandingan data dari beberapa kejadian atau keadaan pada saat tertentu, maka lebih tepat data digambarkan dalam bentuk grafik batang/balok (*bar chart*), grafik lingkaran (*pie chart*), atau diagram gambar (*pictogram*).

Grafik dibedakan menjadi 2 (dua) macam, yaitu :

1. Grafik Tabel Distribusi Frekuensi

- a. Histogram
- b. Poligon Frekuensi
- c. Kurva Ogive

2. Grafik dari Bentuk Tabel

a. Grafik Garis (*Line chart*)

1. Grafik garis tunggal (*Single line chart*)

Grafik ini memberikan gambaran perkembangan satu jenis data.

2. Grafik garis berganda (*Multiple line chart*)

Grafik ini untuk memberikan gambaran perkembangan dari beberapa jenis data.

3. Grafik garis komponen berganda (*Multiple component line chart*)

Grafik ini untuk melihat perkembangan dari beberapa jenis data, sekaligus melihat perkembangannya secara kumulatif.

4. Grafik garis komponen berganda dalam persentase (*Percentage multiple component line chart*)

Prinsipnya sama dengan pembuatan grafik garis komponen berganda, hanya nilai yang diplotkan dalam bentuk persentase.

5. Grafik garis saling menimbang (*Balance line chart*)

Grafik ini untuk melihat perkembangan dari dua kelompok data yang saling berlawanan, sehingga perkembangan selisih antara dua kelompok data tersebut dapat diketahui.

6. Grafik garis interval (*Interval line chart*)

Grafik ini untuk menggambarkan perkembangan data yang mempunyai dua nilai.

b. Grafik Batang (*Bar chart*)

1. Grafik batang tunggal (*Single bar chart*)

Grafik ini untuk menggambarkan perbandingan beberapa data pada periode yang sama.

2. Grafik batang berganda (*Multiple bar chart*)

Grafik ini menggambarkan perbandingan beberapa data yang dirinci menurut beberapa kategori.

3. Grafik batang komponen berganda (*Percentage multiple component bar chart*)

Grafik ini menggambarkan perbandingan data yang dirinci menurut beberapa kategori sekaligus dapat menggambarkan perbandingan jumlah datanya.

4. Grafik batang komponen berganda dalam persentase (*Percentage multiple component bar chart*)

Grafik ini menggambarkan perbandingan data dalam persentase yang telah dirinci menurut beberapa kategori.

5. Grafik batang saling menimbang (*Balance bar chart*)

Grafik ini biasanya untuk menggambarkan perbandingan antara kategori yang satu dengan kategori lainnya pada suatu periode tertentu.

6. Grafik batang mendatar (*Horizontal bar chart*)

Grafik ini biasanya untuk menggambarkan perbandingan antara kategori yang satu dengan kategori lainnya pada suatu periode tertentu (bukan kategori dalam deret waktu).

c. Grafik Lingkaran (*Pie chart*)

Kegunaan grafik lingkaran pada dasarnya sama dengan grafik batang tunggal, yaitu untuk memberi gambaran mengenai perbandingan beberapa data.

d. Diagram Gambar (*Pictogram*)

Biasanya digunakan untuk mendapatkan gambaran secara kasar dari suatu persoalan dan sebagai penyajian visual yang baik bagi orang awam, lebih-lebih jika simbol yang digunakan cukup menarik.

e. Diagram Peta (*Cartogram*)

Diagram peta adalah diagram yang memberikan keterangan tempat dari jenis data yang ada pada sebuah peta.

4.4 Soal-Soal

1. Mengapa penyajian data dalam bentuk grafik lebih menarik dibandingkan dalam bentuk tabel?
2. Sebutkan tujuan penyajian data dalam bentuk grafik.
3. Agar tujuan membuat grafik dapat mencapai sasaran, jelaskan hal-hal yang perlu diperhatikan dari sudut penganalisaan/penarikan kesimpulan.
4. Mengacu pada soal No. 3 pada Bab III buatlah grafik masing-masing untuk keberangkatan dalam negeri dan keberangkatan luar negeri

BAB VI Penutup

6.1 Rangkuman

Dalam kegiatan statistik, data yang telah dikumpulkan dan diolah harus disajikan. Penyajian data dibuat untuk memberikan deskripsi mengenai data yang telah dikumpulkan sehingga memudahkan dalam analisa dan pengambilan keputusan.

Tabel merupakan bentuk yang paling umum dan efektif dalam menyajikan informasi statistik. Selain informasi yang disajikan lebih lengkap, penyajian data dalam bentuk tabel diharapkan dapat memudahkan analisis dan dimengerti oleh pengguna data. Penyajian data melalui tabel dapat dibuat dalam bentuk tabel satu arah, dua arah, atau tiga arah sesuai dengan karakteristik data yang tersedia.

Untuk mengurangi kejenuhan melihat angka-angka, grafik dapat dijadikan sebagai alternatif dalam menyajikan hasil olahan data. Grafik ditampilkan dalam bentuk fisualisasi gambar sehingga pada umumnya lebih mudah dibaca dibandingkan dengan melihat angka-angka dalam tabel. Grafik juga dapat menunjukkan suatu perkembangan data dari waktu ke waktu atau perbandingan beberapa jenis data sehingga mempercepat pengertian/pemahaman dan dapat ditarik kesimpulan.

Untuk dapat menganalisa perkembangan data dari waktu ke waktu dapat digunakan grafik garis. Sedangkan untuk menganalisa perbandingan data dari beberapa kejadian atau keadaan pada saat tertentu, maka lebih tepat digambarkan dalam bentuk grafik batang, grafik lingkaran, atau diagram gambar.

Jika ingin mengetahui frekuensi sebaran data dari masing-masing kelas interval dapat digunakan tabel distribusi frekuensi. Tabel distribusi frekuensi memuat susunan data dalam suatu tabel yang telah diklasifikasikan menurut kelas-kelas atau kategori tertentu agar dapat diperoleh gambaran yang sederhana, jelas dan sistematis mengenai peristiwa yang dinyatakan dalam angka-angka. Dari tabel distribusi frekuensi dapat dibuat jenis grafik histogram, poligon frekuensi, dan ogive.

6.2 Implikasi

Dampak yang diharapkan bagi peserta setelah menerima mata diklat ini adalah dapat memahami tentang bagaimana proses pembuatan tabel dan grafik dalam hal penyajian data. Dan nantinya bahan ajar ini dapat dijadikan sebagai acuan dalam membuat publikasi statistik setelah pelaksanaan kegiatan dilapangan setelah proses pengumpulan dan pengolahan data.

6.3 Tindak Lanjut

Untuk lebih menguasai apa yang telah diketahui terkait dengan disiplin ilmu statistik khususnya mata diklat penyajian data sebaiknya harus sering diulang baik dalam pemahaman maupun aplikasinya termasuk dalam hal praktek pembuatan tabel dan grafik dalam publikasi.

Daftar Pustaka

- Anto Dayan (1990), Pengantar Metode Statistik Jilid II, Jakarta, LP3S.
- J. Supranto, MA (2000), Statistik Teori & Aplikasi Jilid I, Erlangga, Jakarta.
- J. Supranto, MA (1992), Statistik Teori & Aplikasi Jilid II, Erlangga, Jakarta.
- Murray R. Spiegel, Larry J. Stephens, Statistik, Edisi Ketiga, Schaum's OutLines
- RK. Sembiring (1995), Analisis Regresi, Penerbit ITB, Bandung.
- Ronald E. Walpole (1992), Pengantar Statistika, Gramedia, Jakarta.
- Suharyadi, Purwanto SK (2004), Statistika untuk Ekonomi dan Keuangan Modern Jilid 2, Penerbit Salemba 4, Jakarta.
-



Modul 05



Teknik Analisis Data



DAFTAR ISI

DAFTAR ISI	i
DAFTAR GAMBAR	iii
Bab I Pendahuluan	7
1.1 Latar Belakang	7
1.2 Deskripsi Singkat	7
1.3 Hasil Belajar (Tujuan Pembelajaran Umum).....	7
1.4 Indikator Hasil Belajar (Tujuan Pembelajaran Khusus)	7
1.5 Materi Pokok.....	8
1.6 Manfaat	8
Bab II Analisis Korelasi	9
2.1 Koefisien Korelasi	9
2.2 Koefisien Korelasi Parsial.....	9
Bab III Persamaan Regresi Linier Berganda	13
3.1 Persamaan Regresi Linier Sederhana.....	13
3.2 Persamaan Regresi Linier Berganda	13
3.3 Perhitungan Koefisien Regresi.....	16
3.4 Rangkuman	23
3.5 Soal-Soal	23
Bab IV Korelasi Dalam Regresi Linier Berganda	27
4.1 Koefisien Determinasi Berganda	27
4.2 Koefisien Non Determinasi.....	31
4.3 Koefisien Korelasi Linier Berganda	31
4.5 Rangkuman	32
4.6 Soal-Soal	33
Bab V. Pengujian Model Regresi Berganda	39
5.1 Uji Signifikansi secara Parsial	39
5.2 Uji Signifikansi secara Simultan.....	41
5.3 Uji Koefisien Korelasi Regresi Linier Berganda	46
5.4 Rangkuman	50
5.5 Soal-Soal	53
Bab VI. Asumsi Dan Pelanggaran Asumsi Regresi Berganda	55
6.1 Multikolinieritas.....	56
6.2 Heteroskedastisitas.....	59
6.3 Koefisien <i>rank</i> korelasi dari <i>Spearman</i> :	60
6.4 Autokorelasi	62
6.5 Uji <i>Durbin Watson</i>	62
6.6 Rangkuman	64
6.7 Soal-Soal	64
Bab VII. Cara Memilih Independen Variabel	67

7.1 <i>Backward Elimination</i>	68
7.2 <i>Forward Inclusion</i>	69
7.3 <i>Stepwise</i>	71
7.4 Rangkuman.....	71
8.1 Kesimpulan.....	73
8.2 Implikasi	73
8.3 Tindak Lanjut	73
Aplikasi Komputer	74
Latihan.....	76
Jawaban	80
Daftar Pustaka	89
Lampiran.....	90

DAFTAR GAMBAR

Judul Gambar	Halaman
Gambar 1. Kotak Dialog Ms Excel	17
Gambar 2. Kotak Dialog Tools	17
Gambar 3. Kotak Dialog Regression	17
Gambar 4. Kotak Dialog Summary <i>Output</i>	18
Gambar 5. Kotak Dialog Perhitungan Ms Excel	69
Gambar 6. Kotak Dialog SPSS. 17	82
Gambar 7. Kotak Dialog Analyze	83
Gambar 8. Kotak Dialog Logistic Regressions	83
Gambar 9. Kotak Dialog Categorical	84
Gambar 10. Kotak Dialog Logistic Regression Options	84
Gambar 11. Kotak Dialog Method	85
Gambar 12. Kotak Dialog Excel	90

Bab I Pendahuluan

1.1 Latar Belakang

Estimasi nilai suatu variabel tertentu, dapat dilakukan dengan analisis regresi. Jika variabel pertama dipengaruhi oleh variabel yang lainnya, maka estimasi nilai variabel pertama didapatkan dari persamaan atau fungsi dari variabel yang lainnya. Estimasi digunakan untuk menentukan target dalam pengambilan keputusan suatu kebijakan. Estimasi yang salah dapat mengakibatkan kesalahan dalam pengambilan keputusan suatu kebijakan. Dengan demikian pemahaman tentang korelasi dan regresi merupakan suatu yang penting.

Modul korelasi dan regresi membahas tentang cara menentukan estimasi dengan benar. Modul ini berguna bagi peserta khususnya maupun yang telah menduduki fungsional statistisi dan yang berkaitan dengan pengambilan keputusan dalam suatu kebijakan.

Pada dasarnya korelasi dan regresi digunakan dalam berbagai aspek kehidupan manusia, misalnya dalam bidang ekonomi, sosial, kesehatan dan lainnya. Regresi biasanya digunakan untuk mengestimasi nilai pada masa yang akan datang.

1.2 Deskripsi Singkat

Mata diklat ini membahas regresi dan korelasi.

1.3 Hasil Belajar (Tujuan Pembelajaran Umum)

Setelah mengikuti pembelajaran ini, Peserta mampu menerapkan korelasi dan regresi secara benar dalam kasus-kasus riil.

1.4 Indikator Hasil Belajar (Tujuan Pembelajaran Khusus)

Setelah mempelajari materi ini secara khusus, peserta dapat:

1. menghitung regresi linier berganda;
2. menghitung korelasi;
3. melakukan pengujian terhadap model regresi linier berganda;
4. mengatasi permasalahan akibat pelanggaran asumsi dalam regresi linier berganda;

5. melakukan pengujian terhadap koefisien korelasi.

1.5 Materi Pokok

Materi pokok yang dibahas dalam modul ini adalah:

1. Analisis Korelasi
2. Persamaan Regresi Linier Berganda
3. Korelasi dalam Regresi Linier Berganda
4. Kesalahan Baku/*Standar Error* Regresi Linier Berganda
5. Interval/Selang Kepercayaan
6. Pengujian Model Regresi Berganda
7. Asumsi dan Pelanggaran Asumsi Regresi Linier Berganda
8. Cara Memilih Independen Variabel

1.6 Manfaat

Manfaat Modul Korelasi dan Regresi Linier Berganda ini adalah:

1. Sebagai bahan acuan bagi peserta dalam proses belajar dan pembelajaran regresi linier berganda dan korelasi.
2. Sebagai bahan peserta diklat untuk memahami penggunaan analisis regresi linier berganda dalam kasus-kasus riil.
3. Sebagai bahan peserta diklat untuk memahami jenis-jenis regresi.
4. Sebagai bahan peserta diklat untuk menghitung regresi linier berganda.
5. Sebagai bahan peserta diklat untuk memahami jenis-jenis korelasi.
6. Sebagai bahan peserta diklat untuk menghitung korelasi parsial dalam regresi linier berganda.
7. Sebagai bahan peserta diklat untuk memahami asumsi dan pelanggaran asumsi pada regresi linier berganda.

Bab II Analisis Korelasi

Di dunia ini, kita tidak dapat hidup sendiri tetapi memerlukan hubungan dengan orang lain (dengan tetangga, rekan kantor, kawan sekolah, pegawai bank, petugas pajak, dan lain sebagainya). Hubungan ini pada umumnya dilakukan dengan maksud tertentu (*J. Supranto, 2008*).

Untuk melihat/mengukur seberapa kuatnya hubungan antara kejadian yang satu dengan kejadian yang lain atau perubahan nilai dua variabel digunakan analisis korelasi. Misalnya, hubungan perubahan nilai pada harga (X) dengan perubahan nilai pada hasil penjualan (Y).

2.1 Koefisien Korelasi

Hubungan dua variabel ada yang positif dan negatif. Hubungan X dan Y dikatakan positif apabila kenaikan (penurunan) X pada umumnya diikuti oleh kenaikan (penurunan) Y. Sebaliknya dikatakan negatif bila kenaikan (penurunan) X pada umumnya diikuti oleh penurunan (kenaikan) Y. Kuat atau tidaknya hubungan antara X dan Y apabila dapat dinyatakan dengan fungsi linier (paling tidak mendekati), diukur dengan suatu nilai yang disebut koefisien korelasi (r). Nilai koefisien korelasi ini paling sedikit -1 dan paling besar 1 ($-1 \leq r \leq 1$) (*J. Supranto, 2008*).

Cara menghitung r adalah sebagai berikut:

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}}$$

2.2 Koefisien Korelasi Parsial

Korelasi yang kuat antara Y dengan salah satu variabel, katakanlah X_2 , mungkin saja semata-mata disebabkan oleh kenyataan bahwa Y dan X_2 berhubungan dengan variabel lain X_1 . Korelasi yang sebenarnya antara Y dengan X_2 , hanya dapat diamati bila pengaruh X_1 dibuat tetap/konstan. Hal ini disebut dengan koefisien korelasi parsial yang mengukur korelasi Y dengan X_2 sementara X_1 dibuat tetap yang dilambangkan oleh $r_{YX_2.X_1}$.

Koefisien korelasi parsial Y dan X_2 , jika X_1 konstan dapat dihitung dengan rumus :

$$r_{YX_2.X_1} = \frac{r_{YX_2} - r_{YX_1} \cdot r_{X_1X_2}}{\sqrt{(1 - r_{YX_1}^2)(1 - r_{X_1X_2}^2)}}$$

Koefisien korelasi parsial Y dan X_1 , jika X_2 konstan, dihitung dengan rumus :

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2} \cdot r_{X_1X_2}}{\sqrt{(1 - r_{YX_2}^2)(1 - r_{X_1X_2}^2)}}$$

Sedangkan koefisien korelasi parsial X_1 dan X_2 , jika Y konstan dihitung dengan rumus :

$$r_{X_1X_2.Y} = \frac{r_{X_1X_2} - r_{YX_1} \cdot r_{YX_2}}{\sqrt{(1 - r_{YX_1}^2)(1 - r_{YX_2}^2)}}$$

keterangan:

r_{YX_1} = koefisien korelasi linier sederhana 2 variabel yaitu Y dan X_1

r_{YX_2} = koefisien korelasi linier sederhana 2 variabel yaitu Y dan X_2

$r_{X_1X_2}$ = koefisien korelasi linier sederhana 2 variabel yaitu X_1 dan X_2

Catatan :

Cara menghitung koefisien korelasi linier sederhana antara 2 variabel dapat dilihat pada materi regresi linier sederhana

Misalnya Y = hasil penjualan; X_1 = biaya iklan/advertensi; X_2 = pendapatan masyarakat, maka :

$r_{YX_1.X_2}$ = Koefisien korelasi parsial antara Y dengan X_1 jika X_2 konstan. Jadi dapat dikatakan pengaruh pendapatan terhadap hasil penjualan tidak diperhitungkan.

$r_{YX_2.X_1}$ = Koefisien korelasi parsial antara Y dan X_2 jika X_1 konstan. Jadi pengaruh iklan terhadap hasil penjualan tidak diperhitungkan.

$r_{X_1X_2.Y}$ = Koefisien korelasi parsial antara X_1 dan X_2 jika Y konstan. Jadi dapat dikatakan pengaruh hasil penjualan terhadap pendapatan dan biaya iklan tidak diperhitungkan.

Contoh 1:

Teori permintaan menyatakan bahwa permintaan suatu produk akan ditentukan oleh harga barang itu sendiri dan pendapatan seseorang. Hukum permintaan juga menyatakan bahwa apabila harga barang meningkat, maka permintaan menurun, sehingga hubungan antara permintaan dan harga barang negatif. Hubungan antara pendapatan dengan permintaan bisa negatif dan positif. Terhadap barang normal terdapat hubungan yang positif, artinya apabila pendapatan meningkat, permintaan terhadap barang normal juga meningkat, sebaliknya untuk barang inferior, terdapat hubungan yang negatif yaitu apabila pendapatan meningkat, permintaan terhadap barang inferior justru menurun. Berdasarkan teori tersebut, Suryani (2003) melakukan penelitian di salah satu supermarket untuk mengetahui hubungan dan pengaruh variabel harga dalam ribuan/liter (X_1) dan pendapatan dalam jutaan/bulan (X_2) terhadap permintaan minyak goreng dalam liter (Y). Berikut adalah hasil penelitiannya.

No	(Y)	(X_1)	(X_2)
1	3	8	10
2	4	7	10
3	5	7	8
4	6	7	5
5	6	6	4
6	7	6	3
7	8	6	2
8	9	6	2
9	10	5	1
10	10	5	1
Jumlah	68	63	46

$$r_{YX_1} = -0,931 \quad r_{YX_2} = -0,951 \quad r_{X_1X_2} = 0,901$$

maka :

$$r_{YX_2 \cdot X_1} = \frac{-0,951 - (-0,931 \cdot 0,901)}{\sqrt{(1 - (-0,931)^2)(1 - (0,901)^2)}} = \frac{-0,111}{0,158} = -0,71$$

$$r_{YX_1 \cdot X_2} = \frac{-0,931 - (-0,951 \cdot 0,901)}{\sqrt{(1 - (-0,951)^2)(1 - (0,901)^2)}} = \frac{-0,074}{0,134} = -0,55$$

$$r_{X_1X_2.Y} = \frac{0,901 - (-0,931 \cdot -0,951)}{\sqrt{(1 - (-0,931)^2)(1 - (-0,951)^2)}} = \frac{0,016}{0,113} = 0,143$$

Nilai $r_{YX_1.X_2}$ dan $r_{YX_2.X_1}$ mendekati 1 berarti menunjukkan hubungan antara Y dengan X_1 erat begitu pula hubungan Y dengan X_2 erat. Sedangkan hubungan X_1 dan X_2 cukup lemah. Apabila hubungan antara variabel-variabel yang bebas adalah kuat maka akan terjadi *multikolinieritas* yang akan menyebabkan koefisien determinasinya lemah (akan dijelaskan pada sub bab pelanggaran asumsi).

Bab III Persamaan Regresi Linier Berganda

Sebelum kita membahas lebih banyak mengenai Regresi Linier Berganda, ada baiknya kita mengingat kembali mengenai Regresi Linier Sederhana. Diharapkan dengan memahami Regresi linier Sederhana dengan baik maka akan mempermudah dalam memahami Regresi Linier Berganda, karena pemahaman Regresi Linier Sederhana adalah prasyarat dalam memahami Regresi Linier Berganda.

3.1 Persamaan Regresi Linier Sederhana

Regresi Linier Sederhana merupakan suatu alat ukur yang juga dapat digunakan untuk mengukur ada atau tidaknya korelasi antara dua variabel. Jika kita memiliki dua buah variabel atau lebih maka sudah selayaknya apabila kita ingin mempelajari bagaimana variabel-variabel itu berhubungan atau dapat diramalkan. Analisis regresi mempelajari hubungan yang dinyatakan dalam persamaan matematika yang menyatakan hubungan fungsional antara variabel-variabel.

Analisis regresi lebih akurat dalam melakukan analisis korelasi, karena pada analisis itu kesulitan dalam menunjukkan slop (tingkat perubahan suatu variabel terhadap variabel lainnya dapat ditentukan). Dengan demikian maka melalui analisis regresi, peramalan nilai variabel terikat pada nilai variabel bebas lebih akurat pula.

Persamaan regresi linier dari Y terhadap X dirumuskan sebagai berikut:

$$Y = a + b X$$

keterangan:

Y = variabel terikat

X = variabel bebas

a = intersep

b = koefisien regresi/slop

3.2 Persamaan Regresi Linier Berganda

Pada kehidupan sehari-hari banyak kasus yang memerlukan pengetahuan tentang hubungan tersebut, namun terkadang tidak hanya terbatas pada dua variabel saja. Sebagai contoh pada kasus

konsumsi, Teori Keynes (1883-1946) menyatakan bahwa yang memengaruhi konsumsi seseorang adalah pendapatannya. Contoh lain adalah keputusan investasi dari investor asing dipengaruhi tidak hanya tingkat suku bunga, tetapi indeks harga saham, tingkat inflasi, politik dan lain-lain.

Pada contoh diatas ternyata investasi tidak hanya dipengaruhi oleh satu variabel saja, tetapi oleh banyak variabel lainnya. Penting bagi kita untuk mengetahui hubungan antara suatu variabel dengan variabel lainnya, bagaimana pengaruhnya dan seberapa besar pengaruh setiap variabel terhadap variabel lain. Pada modul ini akan dibahas mengenai hubungan antara tiga (dua variabel bebas) atau lebih variabel yang dikenal dengan analisis regresi berganda, regresi digunakan untuk menduga hubungan statistika (hubungan yang mengandung error/kesalahan). Regresi tidak digunakan untuk menduga hubungan matematika (hubungan yang tidak mengandung error/kesalahan).

Contoh hubungan matematika :

1. $q = 1-p$
2. $w = \frac{(x - \bar{x})}{s}$
3. $x_1 = x_2 + x_3$, dimana $x_1 =$ jumlah pendapatan
 $x_2 =$ pendapatan dari upah
 $x_3 =$ pendapatan dari bukan upah

Regresi Linier Berganda ini sering digunakan untuk menganalisis hal-hal yang lebih kompleks. Model regresi linear berganda yang dalam hal ini hanya dua variabel bebas adalah $Y = b_0 + b_1X_1 + b_2X_2$.

Keterangan:

b_0 = Nilai y , jika $X_1 = X_2 = 0$

b_1 = Besarnya perubahan (kenaikan/penurunan) Y dalam satuan, Jika X_1 mengalami perubahan (kenaikan/penurunan) satu satuan, sedangkan X_2 konstan.

b_2 = Besarnya perubahan (kenaikan/penurunan) Y dalam satuan, Jika X_2 mengalami perubahan (kenaikan/penurunan) satu satuan, sedangkan X_1 konstan.

Contoh 2 :

Y = Pengeluaran untuk pembelian barang-barang tahan lama per minggu (ratusan rupiah) pada PT. Maju Mundur pada tahun 2010.

X_1 = Pendapatan per minggu (ribuan rupiah).

X_2 = Jumlah anggota rumahtangga (orang).

Persamaan regresinya : $Y = 3,92 + 2,50X_1 - 0,48X_2$

Maka, arti dari nilai-nilai model ini adalah :

$b_0 = 3,92$ \Rightarrow Besarnya pengeluaran untuk pembelian barang-barang tahan lama jika tidak ada pendapatan dan anggota rumahtangga.

$b_1 = 2,50$ \Rightarrow Besarnya kenaikan pengeluaran untuk pembelian barang-barang tahan lama jika pendapatan per minggu naik satu satuan, sedangkan anggota rumahtangganya konstan.

$b_2 = -0,48$ \Rightarrow Besarnya penurunan pengeluaran untuk pembelian barang-barang tahan lama jika anggota rumah tangga naik satu satuan, sedangkan pendapatan per minggunya konstan.

Sedangkan bentuk persamaan regresi dengan 3 (tiga) variabel independen adalah :

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

Sehingga bentuk umum persamaan regresi untuk k variabel independen dapat dirumuskan sebagai berikut :

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

Misalkan dari data hasil penelitian untuk k variabel independen dapat dibuat dalam tabel sebagai berikut :

X_1	X_2	X_3	...	X_k	Y
X_{11}	X_{21}	X_{31}	...	X_{k1}	Y_1
X_{12}	X_{22}	X_{32}	...	X_{k2}	Y_2
X_{13}	X_{23}	X_{33}	...	X_{k3}	Y_3

⋮	⋮	⋮	⋮	⋮	⋮
X_{1n}	X_{2n}	X_{3n}	...	X_{kn}	Y_n

Untuk persamaan regresi dengan dua variabel independen nilai koefisien regresi yaitu b_1 dan b_2 serta nilai intersep b_0 dapat dihitung dengan mempergunakan alat hitung sederhana atau manual. Namun demikian untuk yang lebih dari dua variabel independen maka untuk memudahkan mencari koefisien regresinya dapat dibantu dengan program komputer baik menggunakan Spreadsheet 123, QualtroPro atau Excel.

3.3 Perhitungan Koefisien Regresi

Untuk memperoleh nilai koefisien regresi b_0 , b_1 dan b_2 dari persamaan $Y = b_0 + b_1X_1 + b_2X_2$ dapat digunakan metode *ordinary least square* (OLS) yang telah dikemukakan pada regresi sederhana.

Apabila dinyatakan dalam bentuk persamaan matriks, akan diperoleh rumus sebagai berikut :

$$\underline{Y} = \underline{X} \cdot \underline{b}$$

dimana :

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \underline{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}; \underline{X} = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \vdots & \vdots & \vdots \\ n & X_{1n} & X_{2n} \end{bmatrix}$$

Jadi,

$$X^T \underline{X} \underline{b} = X^T \underline{Y}$$

$$\begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} \cdot X_{2i} \\ \sum X_{2i} & \sum X_{1i} \cdot X_{2i} & \sum X_{2i}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{1i} \cdot Y_i \\ \sum X_{2i} \cdot Y_i \end{bmatrix}$$

Di dapat persamaan :

$$(a) \sum Y_i = nb_0 + b_1 \sum X_{1i} + b_2 \sum X_{2i}$$

$$(b) \sum X_{1i} \cdot Y_i = b_0 \sum X_{1i} + b_1 \sum X_{1i}^2 + b_2 \sum X_{1i} \cdot X_{2i}$$

$$(c) \sum X_{2i} \cdot Y_i = b_0 \sum X_{2i} + b_1 \sum X_{1i} \cdot X_{2i} + b_2 \sum X_{2i}^2$$

Contoh 3 :

Berdasarkan data contoh 1, cobalah hitung koefisien regresinya!

Jawaban :

Untuk mendapatkan koefisien regresi, sesuai dengan persamaan (a), (b) dan (c) perlu dihitung lebih dahulu nilai-nilai sebagai berikut :

$\sum Y \cdot X_1$	$\sum Y \cdot X_2$	$\sum X_1^2$	$\sum X_2^2$	$\sum X_1 \cdot X_2$
24	30	64	100	80
28	40	49	100	70
35	40	49	64	56
42	30	49	25	35
36	24	36	16	24
42	21	36	9	18
48	16	36	4	12
54	18	36	4	12
50	10	25	1	5
50	10	25	1	5
409	239	405	324	317

Dengan memasukan nilai-nilai hasil perhitungan di atas dengan persamaan (a), (b) dan (c) diperoleh persamaan sebagai berikut :

$$68 = 10b_0 + 63b_1 + 46b_2 \quad \dots\dots(1)$$

$$409 = 63b_0 + 405b_1 + 317b_2 \quad \dots\dots(2)$$

$$239 = 46b_0 + 317b_1 + 324b_2 \quad \dots\dots(3)$$

Untuk mendapatkan nilai koefisien regresi b_0 , b_1 dan b_2 dapat dilakukan dengan substitusi antar persamaan.

Persamaan (1) dan (2), persamaan (1) dapat dikalikan dengan $-6,3$ supaya penjumlahan persamaan (1) dan (2) menghasilkan nilai b_1 nol. Persamaan tersebut menjadi :

$$\begin{array}{rcll}
 -428,4 & = & -63b_0 & -396,9b_1 & -289,8b_2 & \dots\dots(1) \times (-6,3) \\
 409 & = & 63b_0 & +405b_1 & +317b_2 & \dots\dots(2) \\
 \hline
 -19,4 & = & 0 & +81b_1 & +27,2b_2 & \dots\dots(4)
 \end{array}$$

Persamaan (1) dan (3), dengan mengalikan persamaan (1) dengan $-4,6$ dan hasilnya :

$$\begin{array}{rcll}
 -312,8 & = & -46b_0 & -289,8b_1 & -211,6b_2 & \dots\dots(1) \times (-4,6) \\
 239 & = & 46b_0 & +317b_1 & +324b_2 & \dots\dots(3) \\
 \hline
 -73,8 & = & 0 & +27,2b_1 & +112,4b_2 & \dots\dots(5)
 \end{array}$$

Untuk mendapatkan nilai b_2 dapat substitusi persamaan (4) dan (5), dengan mengalikan persamaan (4) dengan $-3,36$ dan hasilnya sebagai berikut :

$$\begin{array}{rcll}
 65,15 & = & 0 & -27,2b_1 & -91,34b_2 & \dots\dots(4) \times (-3,36) \\
 -73,8 & = & 0 & -27,2b_1 & +112,4b_2 & \dots\dots(5) \\
 \hline
 -8,65 & = & 0 & 0 & +21,06b_2 & \dots\dots(6)
 \end{array}$$

Dari persamaan (6), maka nilai b_2 adalah $-8,65/21,06 = -0,41$. Setelah menemukan nilai b_2 , maka b_1 dapat dicari dengan mempergunakan persamaan (4) atau (5).

$$\begin{aligned}
 -19,4 &= 8,1b_1 + 27,2b_2 & \Rightarrow & -19,4 = 8,1b_1 + 27,2(-0,41) \\
 -19,4 &= 8,1b_1 - 11,18 \\
 8,1b_1 &= -19,4 + 11,18 \\
 8,1b_1 &= -8,22 \\
 b_1 &= \frac{-8,22}{8,1} = -1,015
 \end{aligned}$$

Nilai b_1 dan b_2 sudah ditemukan, maka nilai b_0 dapat dicari dengan memasukan nilai koefisien regresi b_1 dan b_2 ke dalam persamaan (1) atau (2) atau (3), hasilnya sebagai berikut :

$$\begin{aligned}
 68 &= 10b_0 + 63b_1 + 46b_2 & \Rightarrow & 68 = 10b_0 + 63(-1,015) + 46(-0,41) \\
 68 &= 10b_0 - 63,96 - 18,90 \\
 10b_0 &= 63 + 92,86 \\
 b_0 &= \frac{150,86}{10} = 15,086
 \end{aligned}$$

Maka didapat persamaan regresinya :

$$y = 15,086 - 1,015X_1 - 0,41X_2$$

Dari persamaan tersebut, kita dapat melihat hubungan dan besar pengaruh variabel independen yaitu harga (X_1) dan pendapatan (X_2) terhadap permintaan minyak goreng (Y).

Variabel harga mempunyai koefisien regresi sebesar $-1,015$, jadi apabila harga minyak goreng naik Rp. 1000,- maka permintaan minyak goreng setiap keluarga akan turun 1,015 (satu koma nol satu lima) liter per bulan dengan syarat kondisi lain tidak berubah.

Variabel pendapatan mempunyai nilai koefisien regresi $-0,41$ ini menunjukkan bahwa apabila pendapatan naik 1 (satu) juta, maka permintaan minyak goreng turun 0,41 liter per bulan, berarti minyak goreng adalah produk yang bersifat inferior yaitu produk yang permintaannya menurun ketika pendapatan masyarakat meningkat.

Melakukan perhitungan seperti pada persamaan (1) sampai (6) untuk mendapatkan koefisien regresi memang relatif membosankan, disamping diperlukan ketelitian yang baik. Oleh sebab itu, dikembangkan beberapa cara yang lebih mudah yaitu dengan cara sebagai berikut :

$$\begin{aligned} A &= n \sum X_1 Y - \sum X_1 \sum Y \\ B &= n \sum (X_2)^2 - (\sum X_2)^2 \\ C &= n \sum X_1 X_2 - \sum X_1 \sum X_2 \\ D &= n \sum X_2 Y - \sum X_2 \sum Y \\ E &= n \sum (X_1)^2 - (\sum X_1)^2 \\ F &= EB - C^2 \end{aligned}$$

Dari persamaan tersebut diperoleh nilai koefisien regresi untuk b_0 , b_1 dan b_2 dengan cara sebagai berikut :

$$\begin{aligned} b_1 &= \frac{AB - CD}{F} \\ b_2 &= \frac{DE - AC}{F} \\ b_0 &= \frac{\sum Y - b_1 \sum X_1 - b_2 \sum X_2}{n} \end{aligned}$$

Dari contoh 2. kita dapat menghitung kembali koefisien regresi dengan cara :

$$\begin{aligned}
A &= n \sum X_1 Y - \sum X_1 \sum Y = 10(409) - 63(68) = -194 \\
B &= n \sum (X_2)^2 - (\sum X_2)^2 = 10(324) - (46)^2 = 1124 \\
C &= n \sum X_1 X_2 - \sum X_1 \sum X_2 = 10(317) - 63(46) = 272 \\
D &= n \sum X_2 Y - \sum X_2 \sum Y = 10(239) - 46(68) = -738 \\
E &= n \sum (X_1)^2 - (\sum X_1)^2 = 10(405) - (63)^2 = 81 \\
F &= EB - C^2 = 81(1124) - (272)^2 = 17060
\end{aligned}$$

$$\begin{aligned}
b_1 &= \frac{AB - CD}{F} = \frac{(-194 \times 1124) - (272 \times -38)}{17060} = -1,015 \\
b_2 &= \frac{DE - AC}{F} = \frac{(-738 \times 81) - (-194 \times 272)}{17060} = -0,41 \\
b_0 &= \frac{\sum Y - b_1 \sum X_1 - b_2 \sum X_2}{n} = \frac{68 - (-1,015 \times 63) - (-0,41 \times 46)}{10} = 15,086
\end{aligned}$$

Metode di atas relatif lebih mudah dibandingkan dengan lainnya, tetapi hasilnya akan sama. Saat ini koefisien regresi dapat dihitung dengan mudah dengan bantuan program komputer (menggunakan *spreadsheet* 123, QualtroPro atau Excel).

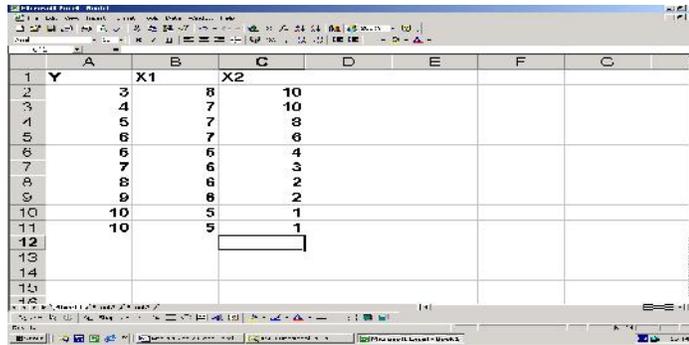
Contoh 4 :

Mencari koefisien regresi berganda dengan menggunakan Ms Excel.

Teknologi komputer akan mempermudah dalam mencari nilai koefisien regresi berganda. Berikut langkah-langkah untuk menghitung koefisien regresi dengan menggunakan Ms Excel.

Setelah komputer dihidupkan, klik *start*, klik *program*, dan pilih Ms Excel.

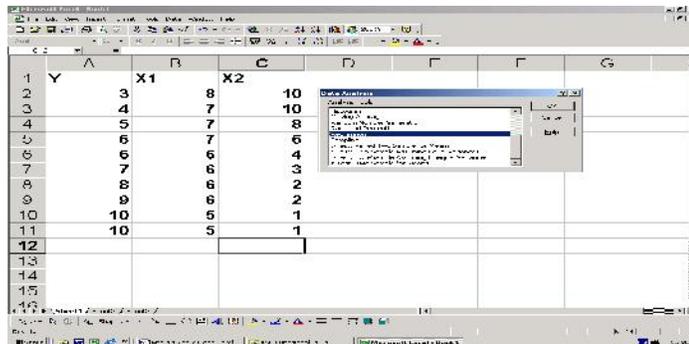
Setelah Ms Excel terbuka, klik pada *File*, dan pilih *New*. Maka akan mendapatkan *sheet* kosong, masukan data Y ke kolom A, data X_1 ke kolom B dan data X_2 ke kolom C, seperti pada gambar 1.



	A	B	C	D	E	F	G
1	Y	X1	X2				
2		3	8	10			
3		4	7	10			
4		5	7	8			
5		6	7	6			
6		6	6	4			
7		7	6	3			
8		8	6	2			
9		9	6	2			
10		10	5	1			
11		10	5	1			
12							
13							
14							
15							

Gambar 1. Kotak Dialog Ms Excel

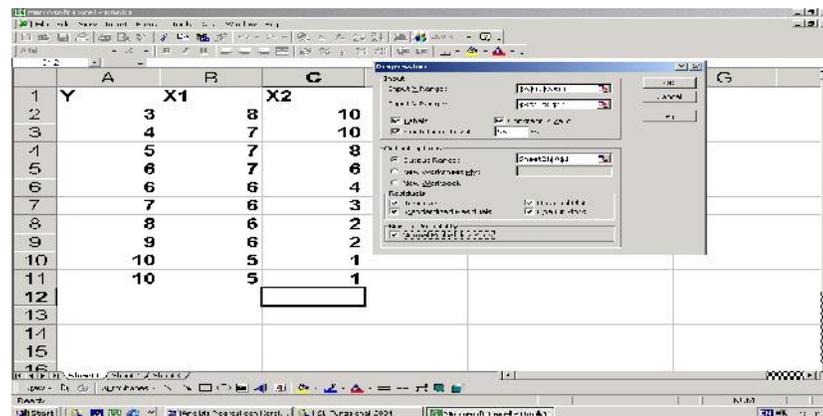
Setelah selesai meng-*entry* data, klik *Tools* dan pilih *Data Analysis*, sehingga keluar kotak dialog seperti pada gambar 2.



	A	B	C	D	E	F	G
1	Y	X1	X2				
2		3	8	10			
3		4	7	10			
4		5	7	8			
5		6	7	6			
6		6	6	4			
7		7	6	3			
8		8	6	2			
9		9	6	2			
10		10	5	1			
11		10	5	1			
12							
13							
14							
15							

Gambar 2. Kotak Dialog Tools

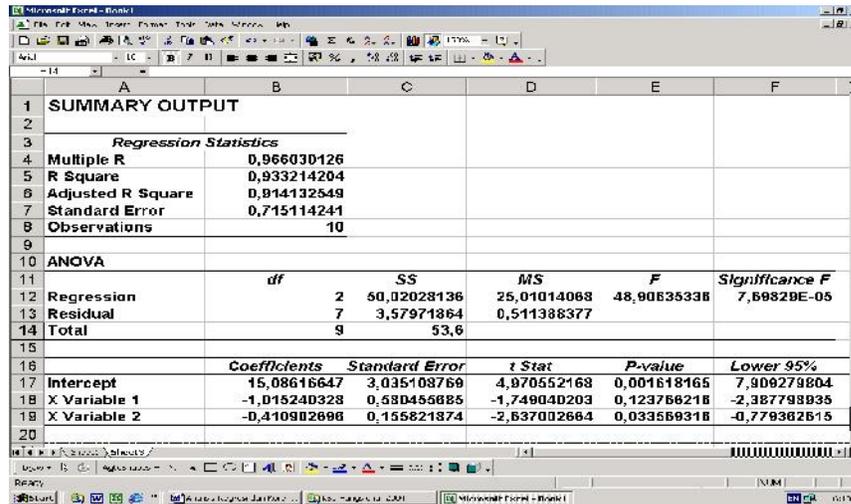
Kemudian pilih *Regression*, muncul kotak dialog seperti pada gambar 3. Isikan *range* data ke dalam kolom *input* yang tersedia. Pada kotak input terdapat kata “*Input Y Range*”, masukan range data Y. Begitu juga data X_1 dan X_2 . Setelah selesai tekan OK.



	A	B	C	D	E	F	G
1	Y	X1	X2				
2		3	8	10			
3		4	7	10			
4		5	7	8			
5		6	7	6			
6		6	6	4			
7		7	6	3			
8		8	6	2			
9		9	6	2			
10		10	5	1			
11		10	5	1			
12							
13							
14							
15							

Gambar 3. Kotak Dialog Regression

Setelah di klik OK, maka hasil regresi akan keluar seperti gambar 4 di bawah ini. Perhatikan pada kolom yang tertulis Coefficients, di sana terlihat intercept untuk b_0 yaitu 15,086166. X variable 1 untuk b_1 , yaitu -1,0152403 dan X untuk variable 2 untuk b_2 yaitu -0,4109027. Dengan mudah kita mendapatkan persamaan regresinya yaitu $y = 15,086 - 1,015X_1 - 0,41X_2$.



SUMMARY OUTPUT					
Regression Statistics					
Multiple R					
R Square					
Adjusted R Square					
Standard Error					
Observations					
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	50,02028136	25,01014068	48,90635338	7,59829E-05
Residual	7	3,57971864	0,511388377		
Total	9	53,6			
Coefficients					
		Standard Error	t Stat	P-value	Lower 95%
Intercept	15,08616647	3,035103769	4,970552168	0,001618165	7,909279804
X Variable 1	-1,015240328	0,580456885	-1,749040203	0,123768218	-2,387798935
X Variable 2	-0,410902698	0,155821874	-2,637002664	0,033569318	-0,779362615

Gambar 4. Kotak Dialog Summary Output

3.4 Rangkuman

Rangkuman dari bab ini adalah:

1. Analisis regresi mempelajari hubungan yang dinyatakan dalam persamaan matematika yang menyatakan hubungan fungsional antara variabel-variabel.
2. Hubungan antara tiga variabel (dua variabel bebas) atau lebih yang dikenal dengan analisis regresi linier berganda, juga digunakan untuk menduga hubungan statistika (hubungan yang mengandung error/kesalahan).
3. Bentuk umum persamaan regresi linier berganda untuk k variabel independen dapat dirumuskan
$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$
4. Memproleh nilai koefisien regresi b_0 , b_1 dan b_2 dari persamaan $Y = b_0 + b_1X_1 + b_2X_2$ digunakan metode *ordinary least square* (OLS) dinyatakan dalam bentuk persamaan matriks, diperoleh rumus $X^T \underline{X} \underline{b} = \underline{X}^T \underline{Y}$

3.5 Soal-Soal

1. Tabel di bawah menunjukkan jumlah hasil panen jagung dalam kg per are (Y) dan jumlah pupuk yang digunakan dalam kilogram per are (X). Data merupakan hasil dari tahun 1991-2000.

Tahun	n	Y_i	X_i
1991	1	44	5
1992	2	49	10
1993	3	59	12
1994	4	60	13
1995	5	67	18
1996	6	70	24
1997	7	80	29
1998	8	85	32
1999	9	87	37
2000	10	89	40

- a. Carilah persamaan regresinya!

- b. Dari persamaan regresi yang didapatkan, berapakah jumlah hasil panen ketika menggunakan pupuk 10 kg/are ?

2. Tabel berikut merupakan perluasan dari tabel sebelumnya dengan menambah variabel jumlah insektisida yang digunakan (X_2) gram/are.

Tahun	n	Y	X_1	X_2
1991	1	44	5	3
1992	2	49	10	4
1993	3	59	12	4
1994	4	60	13	5
1995	5	67	18	10
1996	6	70	24	11
1997	7	80	29	15
1998	8	85	32	16
1999	9	87	37	18
2000	10	89	40	24

- a. Carilah persamaan regresi bergandanya!
 b. Dari persamaan regresi yang didapatkan, berapakah jumlah hasil panen ketika menggunakan pupuk 10 kg/are dan insektisida 5 gram/are?
3. Pihak administrasi kampus STIS ingin mengetahui apakah Indeks Prestasi mahasiswa (Y) dapat diprediksi melalui nilai tes masuk saat penerimaan mahasiswa tersebut (X). Untuk itu, digunakan data 20 mahasiswa yang dipilih secara acak. Datanya sebagai berikut.

i	1	2	3	4	5	6	7	8	9	10
X_i	5,5	4,8	4,7	3,9	4,5	6,2	6,0	5,2	4,7	4,3
Y_i	3,1	2,3	3,0	1,9	2,5	3,7	3,4	2,6	2,8	1,6
i	11	12	13	14	15	16	17	18	19	20
X_i	4,9	5,4	5,0	6,3	4,6	4,3	5,0	5,9	4,1	4,7
Y_i	2,0	2,9	2,3	3,2	1,8	1,4	2,0	3,8	2,2	1,5

Berikut ringkasan perhitungan dari data tersebut : $\sum X_i = 100,0$;
 $\sum Y_i = 50,0$;
 $\sum X_i^2 = 509,12$; $\sum Y_i^2 = 134,84$; $\sum X_i Y_i = 257,66$.

- Carilah nilai perkiraan koefisien regresinya dan tentukan persamaan regresinya.
 - Berapakah nilai perkiraan IP untuk mahasiswa dengan nilai tes masuk $X = 5.0$?
 - Berapakah perkiraan besar perubahan nilai IP jika nilai tes masuk meningkat satu poin?
4. Data berikut diambil dari 20 kiriman yang datang ke gudang berupa drum yang berisi bahan kimia. Data mencakup banyaknya drum dalam kiriman (X_1), total berat kiriman (X_2 , dalam ratusan kg), dan lamanya waktu (dalam menit) yang dibutuhkan untuk memindahkan kiriman ke gudang (Y).

I	1	2	3	4	5	6	7	8	9	10
X_{i1}	7	18	5	14	11	5	23	9	16	5
X_{i2}	5,11	16,72	3,20	7,03	10,98	4,04	22,07	7,03	10,62	4,76
Y_i	58	152	41	93	101	38	203	78	117	44
I	11	12	13	14	15	16	17	18	19	20
X_{i1}	17	12	6	12	8	15	17	21	6	11
X_{i2}	11,02	9,51	3,79	6,45	4,60	13,86	13,03	15,21	3,64	9,57
Y_i	121	112	50	82	48	127	140	155	39	90

Ringkasan perhitungan dari data tersebut : $\sum x_1 x_2 = 502,40$;
 $\sum x_1 y = 4766,90$;
 $\sum x_2 y = 4436,24$; $\sum x_1^2 = 587,80$; $\sum x_2^2 = 500,66$

- Carilah nilai koefisien-koefisien regresinya dan tentukan persamaan regresinya.
 - Tentukan nilai perkiraan lamanya waktu memindahkan kiriman yang terdiri dari 12 drum dengan berat total 110 kg.
5. Studio “Dwayne” mengoperasikan studio foto khusus potret anak-anak di 21 kota di Australia. Perusahaan akan mengadakan perluasan ke kota-kota yang lain dan ingin menyelidiki apakah

penjualan (Y) dalam satu komunitas dapat diprediksi dari jumlah penduduk berusia 16 tahun (X1) dan pendapatan perkapita di masyarakat (X2), Tentukan persamaan regresinya?

Diperoleh Data :

case	X ₁	X ₂	Y
1	68,5	16,7	174,4
2	45,2	16,8	164,4
3	92,3	18,2	244,2
4	47,8	16,3	154,6
5	46,9	17,3	181,6
6	66,1	18,2	207,5
7	49,5	15,9	152,8
8	52	17,2	163,2
9	48,9	16,6	145,4
10	38,4	16	137,2
11	87,9	18,3	241,9
12	72,8	17,1	191,1
13	88,4	17,4	232
14	42,9	15,8	145,3
15	52,5	17,8	161,1
16	85,7	18,4	209,7
17	41,3	16,5	146,4
18	51,7	16,3	144
19	89,6	18,1	232,6
20	82,7	19,1	224,1
21	52,3	16	166,5

Bab IV Korelasi Dalam Regresi Linier Berganda

Pada materi regresi linier sederhana telah dibahas tentang korelasi antara dua variabel. Apabila kita mempunyai lebih dari 2 variabel yaitu variabel Y, X_1, X_2, \dots, X_k , maka jika ingin mengetahui kuatnya hubungan atau pengaruh antara variabel Y dengan beberapa variabel X lainnya (misalnya antara Y dengan X_1, X_2, \dots, X_k), maka kita harus menggunakan suatu koefisien korelasi.

4.1 Koefisien Determinasi Berganda

Koefisien Determinasi (Koefisien Penentuan) merupakan ukuran untuk mengetahui kesesuaian atau ketepatan hubungan antara variabel dependen atau tidak bebas (Y) dengan variabel independen atau bebas X (X_1, X_2, \dots, X_k) dalam suatu persamaan regresi. Dengan kata lain Koefisien Determinasi menunjukkan kemampuan variabel bebas X menerangkan atau menjelaskan variabel Y . Semakin besar nilai koefisien determinasi maka semakin baik kemampuan variabel X menerangkan atau menjelaskan variabel Y .

Koefisien Determinasi (R^2) ditentukan dengan menghitung proporsi varian yang diterangkan persamaan regresi (SSR) = $\sum (\hat{Y} - \bar{Y})^2$ terhadap Varian Total (SST) = $\sum (Y - \bar{Y})^2$ atau $R^2 = \frac{SSR}{SST}$

Cara lain menghitung Rumus Koefisien Determinasi (R^2) adalah :

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$$

Jika diketahui \hat{Y}_i adalah nilai taksiran Y untuk pengamatan ke i ,
 $\hat{Y}_i = f(X_i)$

$$\text{dan } \bar{Y} = \frac{1}{n} \sum_{i=1}^k Y_i \quad n = \text{jumlah sampel/data}$$

atau Rumus Koefisien Determinasi R^2 untuk 2 variabel bebas adalah :

$$R_{y.12}^2 = \frac{b_1 \sum x_{1i} y_i + b_2 \sum x_{2i} y_i}{\sum y_i^2}$$

$$\begin{aligned} \text{diketahui: } \quad \sum x_{1i}y_i &= \sum X_1Y - \frac{\sum X_1 \sum Y}{n} = \sum X_1Y - n\bar{X}_1\bar{Y} \\ \sum x_{2i}y_i &= \sum X_2Y - \frac{\sum X_2 \sum Y}{n} = \sum X_2Y - n\bar{X}_2\bar{Y} \\ \sum y_i^2 &= \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum Y_i^2 - n\bar{Y}^2 \end{aligned}$$

Nilai R^2 berkisar antara 0 sampai 1 dan digunakan untuk menyatakan berapa % variabilitas X_1, X_2, \dots, X_k secara serentak berpengaruh pada variabilitas Y .

Jika $R^2 = 1$ berarti variabel bebas X_1, X_2 , maupun X_k mampu menerangkan variabel tidak bebas Y sebesar 100%.

Jika $R^2 = 0$ berarti tidak ada variabel bebas X_1, X_2 , maupun X_k yang dapat menerangkan variabel tidak bebas Y .

$R^2 < 0,5$ maka variabel bebas X_1, X_2 , maupun X_k relatif kurang baik dalam menerangkan variabel tidak bebas Y . Dalam ilmu sosial seperti antropologi hampir tidak pernah menunjukkan R^2 lebih dari 0,5 tetapi variabel tetap dipakai.

$R^2 > 0,5$ maka variabel bebas X_1, X_2 , maupun X_k dapat menerangkan variabel Y dengan baik/kuat.

R^2 makin mendekati 1, makin baik X_1, X_2, \dots, X_k menerangkan Y atau makin "cocok" regresi X terhadap Y

Contoh 5 :

Hitunglah Koefisien Determinasi antara permintaan minyak goreng dengan harga minyak goreng dan pendapatan pada contoh 2, yang menghasilkan persamaan regresi $Y = 15,086 - 1,015X_1 - 0,41X_2$!

Jawab :

Dari soal diketahui bahwa :

$$\begin{aligned} \sum X_1Y &= 409 & \sum X_2Y &= 239 & \sum Y^2 &= 516 & n &= 10 \\ \sum Y &= 68 & \sum X_2 &= 46 & \sum X_1 &= 63 & b_1 &= -1,015 \\ & & & & & & b_2 &= -0,41 \end{aligned}$$

$$R^2 = \frac{b_1 \sum x_{1i}y_i + b_2 \sum x_{2i}y_i}{\sum y_i^2}$$

$$\sum x_{1i}y_i = \sum X_1Y - \frac{\sum X_1 \sum Y}{n} = 409 - \frac{63.68}{10} = -19,4$$

$$\sum x_{2i}y_i = \sum X_2Y - \frac{\sum X_2 \sum Y}{n} = 239 - \frac{46.68}{10} = -73,8$$

$$\sum y_i^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 516 - \frac{(68)^2}{10} = 53,6$$

$$R^2 = \frac{-1,015 \cdot -19,4 + -0,41 \cdot -73,8}{53,6} = 0,932$$

Ini berarti variabel bebas X_1 dan X_2 mampu menerangkan bahwa harga minyak goreng dan pendapatan dapat menjelaskan dengan baik terhadap permintaan minyak goreng sebesar 93,2%.

Contoh 6 :

Diketahui peubah nilai ekonomi makro (Y) dipengaruhi oleh jumlah jam belajar per minggu (X_1) dan nilai pengantar ekonomi (X_2) dengan data sebagai berikut :

Mahasiswa	Y	X_1	X_2
1	40	1	30
2	44	1	35
3	49	2	42
4	53	2	47
5	60	3	50
6	65	3	62
7	69	4	64
8	78	5	71
9	85	6	79
10	92	7	85

Berdasarkan data di atas tentukan hubungan matematis antara nilai ekonomi makro dengan jumlah jam belajar per minggu dan nilai pengantar ekonomi.

Jawaban :

Dari data di atas diketahui bahwa Y merupakan fungsi linier dari X_1 dan X_2 , $Y=f(X_1, X_2)$ sehingga persamaan regresi yang didapat akan seperti ini :

$$Y = b_0 + b_1X_1 + b_2X_2$$

Mahasiswa	Y	X ₁	X ₂	X ₁ .X ₁	X ₂ .X ₂	X ₁ .X ₂	X ₁ .Y	X ₂ .Y
1	40	1	30	1	900	30	40	1200
2	44	1	35	1	1225	35	44	1540
3	49	2	42	4	1764	84	98	2058
4	53	2	47	4	2209	94	106	2491
5	60	3	50	9	2500	150	180	3000
6	65	3	62	9	3844	186	195	4030
7	69	4	64	16	4096	256	276	4416
8	78	5	71	25	5041	355	390	5538
9	85	6	79	36	6241	474	510	6715
10	92	7	85	49	7225	595	644	7820
Jumlah ()	635	34	565	154	35045	2259	2483	38808

Persamaan normalnya ialah sebagai berikut :

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix}$$

$$\begin{bmatrix} 10 & 34 & 565 \\ 34 & 154 & 2259 \\ 565 & 2259 & 35045 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 635 \\ 2483 \\ 38808 \end{bmatrix}$$

Dengan metode *Cramer* didapatkan $b_0 = 20.638$; $b_1=3.742$; $b_2=0.533$ sehingga persamaan regresinya menjadi :

$$Y = 20.638 + 3.742 X_1 + 0.533 X_2$$

4.2 Koefisien Non Determinasi

Koefisien Non Determinasi adalah faktor –faktor selain variabel X yang tidak diukur atau diteliti dalam hal ini disebut dengan residu.

Rumus Koefisien Non Determinasi = $(1-R^2) \times 100\%$.

Contoh 7:

Koefisien Non Determinasi pada contoh 2 di atas sebesar $(1-0,932) \cdot 100\% = 6,8\%$. Hal ini menunjukkan bahwa permintaan minyak goreng tidak hanya dipengaruhi oleh kedua variabel saja tetapi juga oleh unsur lain misalnya selera, promosi dan lain-lain. Peran dari semua variabel yang tidak dimasukkan ini hanya sebanyak 6,8%.

4.3 Koefisien Korelasi Linier Berganda

Digunakan untuk mengetahui hubungan linear antara variabel tak bebas Y dengan variabel bebas X_1, X_2, \dots, X_k yang diukur dengan **koefisien korelasi linier berganda** (r) yang merupakan akar positif dari koefisien determinasi bergandanya.

$$r = \sqrt{R^2}$$

Nilai Korelasi $-1 \leq r \leq 1$

Apabila $r \approx \pm 1$ Korelasinya kuat (hubungan linear sempurna)

$r \approx 0$ Korelasinya lemah

(hubungan linear kurang sempurna atau tidak ada)

Contoh 8:

Berdasarkan contoh 2 maka nilai koefisien korelasi linier berganda adalah : $r = \sqrt{0,932} = 0,965$.

Karena koefisien korelasi liniernya sebanyak 0,965 mendekati 1 berarti hubungan linier antara permintaan minyak goreng dengan harga dan pendapatan sangat kuat.

4.5 Rangkuman

1. Koefisien Determinasi Berganda

Koefisien Determinasi (R^2) ditentukan dengan menghitung proporsi

varian yang diterangkan persamaan regresi (SSR) = $\sum (\hat{Y} - \bar{Y})^2$

terhadap Varians Total (SST) = $\sum (Y - \bar{Y})^2$ atau $R^2 = \frac{SSR}{SST}$

Cara lain menghitung Rumus Koefisien Determinasi (R^2) adalah :

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$$

Jika diketahui \hat{Y}_i adalah nilai taksiran Y untuk pengamatan ke i,

$$\hat{Y}_i = f(X_i)$$

dan $\bar{Y} = \frac{1}{n} \sum_{i=1}^k Y_i$ $n =$ jumlah sampel/data

atau Rumus Koefisien Determinasi R^2 untuk 2 variabel bebas adalah :

$$R_{y.12}^2 = \frac{b_1 \sum x_{1i} y_i + b_2 \sum x_{2i} y_i}{\sum y_i^2}$$

diketahui: $\sum x_{1i} y_i = \sum X_1 Y - \frac{\sum X_1 \sum Y}{n} = \sum X_1 Y - n \bar{X}_1 \bar{Y}$

$$\sum x_{2i} y_i = \sum X_2 Y - \frac{\sum X_2 \sum Y}{n} = \sum X_2 Y - n \bar{X}_2 \bar{Y}$$

$$\sum y_i^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum Y_i^2 - n \bar{Y}^2$$

2. Koefisien Non Determinasi

Rumus Koefisien Non Determinasi = $(1 - R^2) \times 100\%$.

3. Koefisien Korelasi Linier Berganda

$$r = \sqrt{R^2}$$

Nilai Korelasi $-1 \leq r \leq 1$

Apabila $r \approx \pm 1$ Korelasinya kuat (hubungan linear sempurna)

$r \approx 0$ Korelasinya lemah

(hubungan linear kurang sempurna atau tidak ada)

4. Koefisien Korelasi Parsial

Rumus koefisien korelasi parsial Y dan X_2 , jika X_1 konstan :

$$r_{YX_2 \cdot X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{(1 - r_{YX_1}^2)(1 - r_{X_1X_2}^2)}}$$

Rumus koefisien korelasi parsial Y dan X_1 , jika X_2 konstan :

$$r_{YX1.X2} = \frac{r_{YX1} - r_{YX2}r_{X1X2}}{\sqrt{(1 - r_{YX2}^2)(1 - r_{X1X2}^2)}}$$

Rumus koefisien korelasi parsial X_1 dan X_2 , jika Y konstan :

$$r_{X1X2.Y} = \frac{r_{X1X2} - r_{YX1}r_{YX2}}{\sqrt{(1 - r_{YX1}^2)(1 - r_{YX2}^2)}}$$

keterangan:

r_{YX1} = koefisien korelasi linier sederhana 2 variabel yaitu Y dan X_1

r_{YX2} = koefisien korelasi linier sederhana 2 variabel yaitu Y dan X_2

r_{X1X2} = koefisien korelasi linier sederhana 2 variabel yaitu X_1 dan X_2

4.6 Soal-Soal

1. Lihat kembali tabel pada *soal nomor 2 di Bab I*.
 - a. Dari data tersebut, hitunglah nilai koefisien determinasinya!
 - b. Apa yang bisa ditunjukkan dari koefisien determinasi tersebut ?
 - c. Seberapa besar pengaruh variabel-variabel lain di luar model terhadap hasil panen jagung yang tidak disertakan dalam model ?
 - d. Hitung pula besarnya koefisien korelasi liniernya dan apa artinya.

2. Lihat kembali *soal nomor 4 di Bab I*.
Diketahui nilai $SSE = 536,47$; dan $SST = 41.032,95$.
 - a. Seberapa besarkah keragaman waktu memindahkan kiriman ke gudang dapat dijelaskan oleh banyaknya drum dalam kiriman dan berat total kiriman tersebut ?
 - b. Bagaimanakah hubungan linier antara waktu yang dibutuhkan untuk memindahkan kiriman ke gudang dengan banyaknya drum dalam kiriman dan berat total kiriman terhadap ?

3. Data berikut merupakan *output* dari perhitungan regresi dari Excel. Data yang digunakan berupa uji coba untuk mengetahui hubungan kesukaan terhadap suatu produk kue (Y) dengan kelembutan isi (X_1) dan tingkat kemanisan (X_2).

Summary Output

<i>Regression Statistics</i>	
Multiple R	0.97573509
R Square	0.95205897
Adjusted R Square	0.94468343
Standard Error	2.69329652
Observations	16

Anova

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	1872.7	936.35	129.083	2.65826E-09
Residual	13	94.3	7.25		
Total	15	1967.0			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-Value</i>
Intercept	37.65	2.996	12.566	1.1996E-08
X ₁	4.425	0.301	14.695	1.7780E-09
X ₂	4.375	0.673	6.498	2.0110E-05

- a. Dari *output* di atas, seperti apa persamaan regresi yang terbentuk ?
 - b. Jika tingkat kelembutan produk kue diberi nilai 4 dan kemanisannya 6, berikanlah nilai perkiraan kesukaan pelanggan terhadap kue tersebut !
 - c. Seberapa besarkah kesukaan pelanggan terhadap kue dapat dijelaskan oleh lembut dan manisnya kue tersebut ?
 - d. Bagaimana hubungan linier antara kesukaan pelanggan dengan lembut dan manisnya kue ?
4. *Output* berikut merupakan hasil olahan dari program SPSS. Data digunakan untuk mengetahui hubungan tingkat kepuasan pasien (Y) dengan umur pasien (X₁), keparahan penyakit (X₂, angka indeks) dan kecemasan pasien (X₃, angka indeks) di suatu rumah

sakit. Pihak administrasi memilih secara acak 23 pasien dan berikut hasilnya.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.820 ^a	.673	.621	10.28945

a. Predictors: (Constant), X3, X1, X2

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4133.633	3	1377.878	13.014	.000 ^a
	Residual	2011.584	19	105.873		
	Total	6145.217	22			

a. Predictors: (Constant), X3, X1, X2

b. Dependent Variable: Y

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	162.876	25.776		6.319	.000
	X1	-1.210	.301	-.613	-4.015	.001
	X2	-.666	.821	-.177	-.811	.427
	X3	-8.613	12.241	-.157	-.704	.490

a. Dependent Variable: Y

- a. Dari *output* di atas, seperti apa persamaan regresi yang terbentuk ?
 - b. Seberapa besarkah tingkat kepuasan pasien dapat ditunjukkan oleh umur pasien, keparahan penyakit, dan kecemasan pasien ? Berikan komentar !
5. Seorang peneliti ingin mengetahui apakah ada hubungan antara pengeluaran, pendapatan, dan banyaknya anggota keluarga. Untuk keperluan tersebut diambil sampel sebanyak 7 rumah tangga. Data sebagai berikut :

Tabel Data pengeluaran, pendapatan, dan banyaknya anggota rumah tangga

VARIABEL	RUMAH TANGGA						
	I	II	III	IV	V	VI	VII
Pengeluaran (Y)	3	5	6	7	4	6	9
Pendapatan (X ₁)	5	8	9	10	7	7	11
Jumlah Anggota Keluarga (X ₂)	4	3	2	3	2	4	5

Sumber : Data Fiktif

Pertanyaan :

- a. Carilah nilai koefisien korelasinya dan koefisien determinasinya !
 - b. Jelaskan makna hubungannya !
6. Ada salah satu areal pembibitan di PT Tunas Agro telah diketahui bahwa pertumbuhan bibit kelapa sawit di MN (main nursery) kurang maksimal. Untuk mengetahui kurang maksimalnya pertumbuhan bibit kelapa sawit tersebut, maka perusahaan melakukan riset untuk mengetahui kuat lemahnya pengaruh hubungan antara dosis pemupukan (X₁) dengan curah hujan (X₂) terhadap pertumbuhan bibit kelapa sawit (Y).

Pada riset yang dilakukan kali ini akan menitikberatkan pengaruh curah hujan terhadap pertumbuhan bibit kelapa sawit dengan dosis pemupukan yang dikendalikan (konstan). Untuk itu perusahaan mengambil 12 sampel bibit kelapa sawit pada beberapa lokasi pembibitan di Main Nursery secara acak dan diperoleh data sebagai berikut:

Tabel Data intensitas curah hujan, pertumbuhan bibit kelapa sawit, dan dosis pemupukan

X ₁	X ₂	Y
5	26	1.2
5	97	1.24
5	47	1.3
10	88	1.33
10	97	1.42
10	75	1.5
20	88	1.57
20	75	1.61
20	20	1.74
40	88	1.81
40	53	1.89
40	75	1.96
dX₁=225	dX₂=829	dY=18.57

Sumber : Data Fiktif

Keterangan :

X_1 = Dosis Pupuk (gr)

X_2 = Curah Hujan (mm)

Y = Pertumbuhan (m)

Pertanyaan :

Berapakah koefisien korelasi parsialnya? Interpretasikan!

Bab V. Pengujian Model Regresi Berganda

Pada bab ini akan dibahas tentang pengujian terhadap model baik secara parsial maupun simultan (bersama-sama). Pengujian dugaan sementara (hipotesis) untuk melihat apakah variabel bebas mampu secara simultan atau parsial menjelaskan keragaman variabel tidak bebas Y. Di bawah ini akan dijelaskan pengujian terhadap model secara simultan dan parsial.

5.1 Uji Signifikansi secara Parsial

Pengujian ini dimaksudkan untuk menguji apakah suatu variabel bebas berpengaruh atau tidak secara nyata (signifikan) terhadap variabel tidak bebas. Untuk mengetahui apakah suatu variabel secara parsial berpengaruh nyata atau tidak digunakan statistik Uji t atau *t – Student*.

Dari model berikut akan dilakukan pengujian terhadap koefisien regresi :

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

Langkah-langkah pengujian:

1. Menentukan dugaan sementara (Hipotesis).

$$H_0 : b_k = 0$$

$$H_1 : b_k \neq 0$$

Suatu variabel bebas berpengaruh tidak nyata apabila nilai koefisiennya sama dengan nol, atau sebaliknya.

2. Menentukan daerah kritis (wilayah ditolaknya H_0).

Daerah kritis ditentukan oleh nilai tabel t dengan derajat bebas yaitu $(n-k)$, dan taraf nyata α , dimana n = jumlah sampel dan k = banyaknya variabel bebas dan tidak bebas. Sedangkan taraf nyata α yang biasa dipakai 1 %, 5 % atau lainnya. Dengan melakukan Uji 2 Arah, maka nilai t -hitung dapat dicari dengan menggunakan taraf nyata $\alpha / 2$.

3. Menentukan Nilai t – hitung.

$$\text{Nilai } t \text{ hitung} = \frac{b_k - b_k}{Sb_k}$$

k adalah banyaknya variabel bebas

Dengan:

$$S_{b_k} = \frac{S_{Y.X_1.X_2...X_k}}{\sqrt{(\sum X_k^2 - n\bar{X}_k^2)(1 - r_{X_1X_2...X_k}^2)}}$$

Dan, $S_{Y.X_1X_2...X_k} =$

$$\sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum X_1 Y - b_2 \sum X_2 Y - \dots - b_k \sum X_k Y}{n - (k + 1)}}$$

Keterangan:

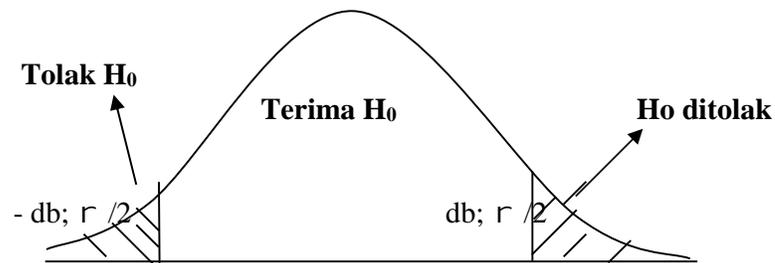
S_{b_k} : Kesalahan Baku Penduga b_k

$S_{Y.X_1X_2...X_k}$: Standar Error variabel Y berdasarkan variabel X yang diketahui

$r_{X_1X_2...X_k}$: Koefisien Korelasi Sederhana antar X_1 dan X_2 atau X_k

Menentukan daerah keputusan.

Daerah keputusan untuk menerima H_0 atau menolak H_0 dengan derajat bebas (db) dan taraf nyata α untuk uji 2 arah sebagai berikut:



4. Menentukan Keputusan

Nilai t – hitung lebih besar atau lebih kecil dari nilai t – tabel, maka H_0 ditolak atau Terima H_1 . Berarti Ada satu variabel bebas berpengaruh nyata terhadap variabel tidak bebas Y.

Catatan: Walaupun secara bersama-sama (simultan) variabel bebas berpengaruh nyata terhadap variabel tidak bebas Y, namun belum tentu secara parsial variabel bebas tersebut berpengaruh terhadap variabel tidak bebas Y.

5.2 Uji Signifikansi secara Simultan

Pengujian signifikansi terhadap model secara simultan (bersama-sama) dimaksudkan untuk melihat kemampuan seluruh variabel bebas mampu menjelaskan keragaman (memperkirakan) variabel tidak bebas Y.

Langkah-langkah pengujian:

1. Menyusun Hipotesis (Dugaan Sementara)

Hipotesis yang akan diuji adalah kemampuan variabel bebas menjelaskan tingkah laku variabel tidak bebas, apabila variabel bebas tidak dapat mempengaruhi variabel bebas dianggap nilai koefisiennya sama dengan nol, sehingga berapapun nilai variabel bebas tidak akan berpengaruh terhadap variabel bebas.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_k \neq 0$$

(minimal ada 1 variabel bebas (X) yang mempengaruhi Y)

2. Menentukan Daerah Kritis (daerah penolakan H_0)

Untuk uji ini digunakan Tabel F. Daerah kritis ditentukan oleh nilai Tabel F dengan derajat bebas pembilang pada kolom yaitu $(k - 1)$ dan derajat bebas penyebut pada baris yaitu $(n - k)$ dengan taraf nyata (tingkat kesalahan) α , jika diketahui $n =$ jumlah sampel dan $k =$ banyaknya variabel bebas dan tidak bebas. Sedangkan taraf nyata α untuk uji 1 arah ini biasa dipakai 1 %, 5 % atau lainnya.

3. Menentukan Nilai F hitung

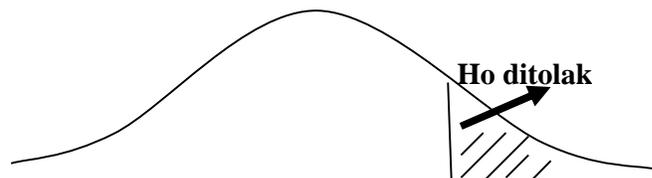
$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}$$

Keterangan : R^2 adalah koefisien determinasi

4. Menentukan Daerah Keputusan

Menentukan wilayah H_0 dan H_1 serta membandingkan dengan nilai F hitung untuk mengetahui apakah menerima H_0 atau menerima H_1

$$F_{\text{hit}} > F_{\alpha; (k-1); (n-k)}$$



$$F_{\alpha; (k-1); (n-k)}$$

5. Menentukan Keputusan

Jika nilai F - hitung lebih besar dari nilai F – Tabel, maka Tolak H_0 .

Kesimpulan: Variabel bebas secara simultan dapat menerangkan variabel tidak bebas, berarti model tersebut dapat digunakan.

Dalam pengujian koefisien regresi (model) secara simultan (bersamaan) dapat menggunakan cara lain yaitu: Uji F dengan Tabel Anova. Analisis Varians atau ANOVA adalah merupakan alat atau piranti yang dapat menggambarkan hubungan antara koefisien korelasi, koefisien determinasi dan kesalahan baku pendugaan. Untuk mengukur kesalahan baku kita menghitung *error* atau selisih Y dengan \hat{Y} atau dapat dinyatakan dalam bentuk persamaan $e = Y - \hat{Y}$ atau dalam bentuk lain $Y = \hat{Y} + e$. Dimana Y adalah nilai sebenarnya, \hat{Y} adalah nilai hasil regresi dan e adalah *error* atau kesalahan. Apabila kita membicarakan keragaman total terhadap Y yaitu $(Y - \bar{Y})^2$, maka keragaman tersebut terdiri dari dua unsur komponen yaitu :

- Keragaman dari komponen yang dapat dijelaskan yaitu komponen regresi Y dan nilai keragamannya dilambangkan dengan $(\hat{Y} - \bar{Y})^2$
- Keragaman dari komponen yang tidak dapat dijelaskan yaitu komponen *error* yang dilambangkan dengan $(Y - \hat{Y})^2$.

Keragaman total terhadap Y biasa disebut dengan SST/JKT (*sum square total*/jumlah kuadrat total), untuk keragaman regresi dengan SSR/JKR (*sum square regression*/jumlah kuadrat regresi) dan keragaman error dengan SSE/JKS (*sum square error*/jumlah kuadrat sisaan).

Secara umum rumus untuk menghitung JK Reg dan JKS adalah :

$$JK \text{ Reg} = b_1 \sum X_1 Y + b_2 \sum X_2 Y + \dots + b_k \sum X_k Y, \text{ sedangkan}$$

$$JKS = \sum (Y - \hat{Y})^2$$

atau dapat dihitung dengan rumus :

$JKS = \sum y^2 - JK \text{ Reg}$, dimana $\sum y^2$ merupakan jumlah kuadrat-kuadrat total dikoreksi, yang besarnya adalah

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}.$$

JK Reg dan JKS masing-masing memiliki derajat bebas (db), maka untuk mencari kuadrat tengah (*mean square*) adalah dengan cara membagi JK Reg dan JKS dengan masing-masing derajat kebebasannya, dapat ditulis dengan rumus :

$$KT \text{ Reg} = \frac{JK \text{ Reg}}{db \text{ Reg}} \text{ dan } KTS = \frac{JK (\text{Sisaan})}{db (\text{Sisaan})}$$

Dari kuadrat tengah (KT) tersebut di atas dibentuk statistik F dengan rumus :

$$F_{hitung} = \frac{KT \text{ Reg}}{KT \text{ Sisaan}}$$

Maka format tabel ANOVA dapat disusun sebagai berikut :

Tabel ANOVA

SK	df/db	SS/JK	MS/KT	F _{hit}	F
Regresi ($x_1 \dots x_k$)	k	JKR $= \sum (\hat{Y} - \bar{Y})^2$	JK Reg/db	$\frac{KT \text{ Reg}}{KTS}$	F ; k;n-k-i
Error/Sisaan	n-k-1	JKS $= \sum (Y - \hat{Y})^2$	JK Sisaan/db		
Total	n - 1	JKT $= \sum (Y - \bar{Y})^2$			

Lihat Contoh 1 :

Diperoleh model regresi linier berganda:

$$Y = 15,086 - 1,015 X_1 - 0,41 X_2$$

$$R^2 = 0,939 ; S_{Y.X_1X_2} = 0,72 ; S_{b_1} = 0,580 ; S_{b_2} = 0,156$$

Lakukan pengujian:

- Apakah X_1 , berpengaruh terhadap Y pada taraf nyata = 5%.
- Apakah X_2 , berpengaruh terhadap Y pada taraf nyata = 5%.
- Apakah X_1, X_2 secara bersama berpengaruh terhadap Y pada taraf nyata = 5%.

Jawab : a dan b

Hipotesis

$$H_0 : \beta_1 = \beta_2 = 0$$

H_1 : ada yang tidak nol

Daerah Kritis:

Uji t – Tabel dengan $db = n - k = 10 - 3 = 7$

dan $\alpha = 5\%$. Dengan pengujian 2 arah didapat nilai t-tabel = 2,36.

Perhatikan uji 2 arah maka ekor distribusi t ada dua sehingga $5\%/2 = 2,5\% = 0,025$; dan $db = 7$, didapat nilai kritis = 2,36

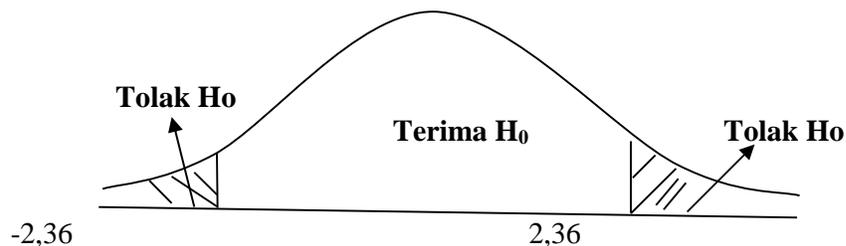
Nilai t hitung untuk β_1 ;

$$t \text{ hitung} = \frac{b_1 - \beta_1}{Sb_1} = \frac{-1,015 - 0}{0,58} = -1,75$$

Nilai t hitung untuk β_2 ;

$$t \text{ hitung} = \frac{b_2 - \beta_2}{Sb_2} = \frac{-0,41 - 0}{0,1558} = -2,637$$

4. Daerah Keputusan:



5. Menentukan Keputusan;

Nilai t hitung untuk koefisien Regresi $b_1 = -1,75$ lebih besar dari $-2,36$ atau berada pada daerah Terima H_0 . Jadi koefisien regresi tidak berbeda dengan nol, berarti variabel bebas X_1 tidak berpengaruh nyata terhadap variabel tidak bebas Y .

Kesimpulan: bahwa variabel harga minyak goreng tidak berpengaruh nyata terhadap permintaan minyak goreng

Nilai t hitung untuk koefisien regresi $b_2 = -2,637$ lebih kecil dari $-2,36$ atau berada pada daerah Tolak H_0 . Jadi koefisien regresi

berbeda dengan nol, berarti variabel bebas X_2 berpengaruh nyata terhadap variabel tidak bebas Y .

Kesimpulan: bahwa variabel pendapatan berpengaruh nyata terhadap variabel permintaan minyak goreng.

Untuk mempermudah penghitungan di atas, komputer dapat membantu mempercepat dengan bantuan salah satu *Software* yaitu *Excel*, *SPSS* dll. Dalam pengolahan data di atas menggunakan paket program *Excel*.

Dari hasil *output*, didapat nilai t - Statistik untuk Uji signifikansi model koefisien regresi $b_1 = -1,749$ dan $b_2 = -2,637$. Untuk menentukan daerah keputusan sama seperti cara di atas.

Cara lain untuk menentukan daerah keputusan yaitu dengan melihat *P-value* dan dibandingkan dengan *taraf nyata* α (tanpa melihat Tabel-t). Dari hasil *output* diperoleh *P-Value* untuk uji signifikansi model koefisien regresi $b_1 = 0,1238$ dan $b_2 = 0,0336$ kemudian dibandingkan dengan $\alpha = 5\% = 0,05$. Ternyata *P-Value* untuk koefisien regresi b_1 lebih besar dari 0,05 maka Terima H_0 . Sedangkan *P-Value* untuk koefisien regresi b_2 lebih kecil dari 0,05 maka Tolak H_0 .

Jawaban: c

1. Hipotesis

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : \beta_1 \neq \beta_2 \neq 0$$

(minimal ada satu variabel X yang mempengaruhi variabel Y)

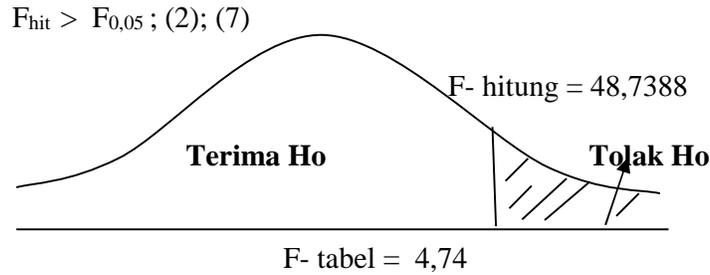
2. Daerah Kritis

Mencari nilai F -tabel dari Tabel F . Dari contoh diketahui jumlah sampel (n) = 10, dan jumlah variabel bebas ada 2 yaitu X_1, X_2 jadi $k = 2$. Jadi derajat bebas pembilang = $k = 2$, sedangkan derajat bebas penyebut = $n - k - 1 = 10 - 3 = 7$. Dengan taraf nyata $\alpha = 5\% = 0,05$ didapat nilai F -Tabel sebesar 4,74.

3. Nilai F - hitung

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} = \frac{0,933 / (3 - 1)}{(1 - 0,933) / (10 - 3)} = 48,7388$$

4. Daerah Keputusan:



5. Menentukan Keputusan;

Nilai F-hitung > nilai F-tabel dan berada di daerah Tolak Ho, berarti menunjukkan bahwa terdapat cukup bukti untuk menolak H_0 dan menerima H_1 .

Kesimpulan: Variabel bebas X_1 harga minyak goreng dan X_2 pendapatan secara nyata bersama-sama mempengaruhi variabel tidak bebas Y. Dalam hal ini variabel harga minyak goreng dan pendapatan secara nyata bersama-sama mempengaruhi variabel permintaan minyak goreng.

Cara yang lebih praktis untuk mengetahui signifikansi model regresi yaitu dengan bantuan paket program komputer, salah satunya Excel. Dari hasil *output*, diperoleh tabel ANOVA dengan melihat rincian nilai F-hitung sebesar 48,90635 dan dengan cara yang sama nilai F-hitung dibandingkan dengan F-tabel.

Atau dengan cara melihat nilai *Significance F* sebesar 7,69829E-05 dan dibandingkan dengan *P-Value* = 0,05. Berarti nilai *Significance F* lebih kecil dari *P-Value*, berarti Tolak Ho.

5.3 Uji Koefisien Korelasi Regresi Linier Berganda

Koefisien korelasi r merupakan sebuah nilai yang dihitung dari n pengamatan sampel. Sampel acak berukuran n yang lain tetapi diambil dari populasi yang sama biasanya akan menghasilkan nilai r yang berbeda pula. Dengan demikian r merupakan suatu nilai dugaan bagi koefisien korelasi linier sesungguhnya yang berlaku bagi seluruh anggota populasi. Kita lambangkan koefisien korelasi populasi yang sesungguhnya ini dengan ρ . Bila r dekat dengan 0, dapat disimpulkan bahwa $\rho = 0$. Bila r mendekati +1 atau -1 maka dapat disimpulkan bahwa $\rho \neq 0$.

Dengan demikian diperlukan suatu pengujian untuk mengetahui apakah r berada cukup jauh dari suatu nilai tertentu ρ_0 , agar kita mempunyai alasan untuk menolak hipotesis nol (H_0) bahwa $\rho = \rho_0$

dan menerima alternatifnya. Hipotesis alternatifnya (H_1) biasanya salah satu diantara $\rho < \rho_0$, $\rho > \rho_0$ atau $\rho \neq \rho_0$.

Langkah-langkah Pengujian :

1. Menyusun Hipotesis (dugaan sementara)

Tentukan pengujian hipotesis tentang ρ yang dapat dituliskan seperti di bawah ini :

a. Uji 2 Arah :

$H_0 : \rho = 0$ (Tidak ada korelasi linier)

$H_1 : \rho \neq 0$ (Ada korelasi linier)

b. Uji 1 Arah :

$H_0 : \rho = 0$

$H_1 : \rho > 0$ (Ada korelasi positif)

c. Uji 1 Arah :

$H_0 : \rho = 0$

$H_1 : \rho < 0$ (Ada korelasi negatif)

2. Menentukan daerah kritis

Daerah kritis ditentukan oleh nilai pada tabel t dengan derajat kebebasan(db): $n-k-1$ (n = banyaknya sampel/data) dengan tingkat signifikansi/ taraf nyata (untuk uji 1 arah) atau $/2$ (untuk uji dua arah). Nilai taraf nyata biasa dipakai 1%, 5% dan lainnya.

3. Menentukan nilai statistik pengujian

Nilai statistik pengujiannya digunakan uji t yaitu :

$$t = \frac{|R|\sqrt{n-k-1}}{\sqrt{1-R^2}},$$

$|R|$ = Harga mutlak dari koefisien korelasi linier berganda (selalu positif)

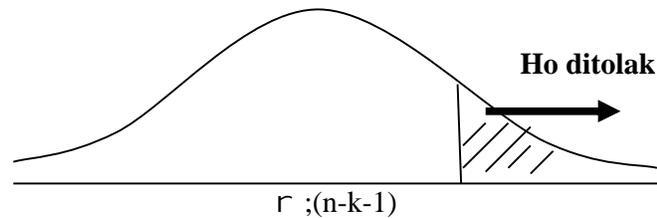
k = banyaknya variabel bebas

4. Menentukan daerah keputusan

Menentukan wilayah H_0 dan H_1 serta membandingkan dengan nilai t -hitung untuk mengetahui apakah menerima H_0 atau menerima H_1 .

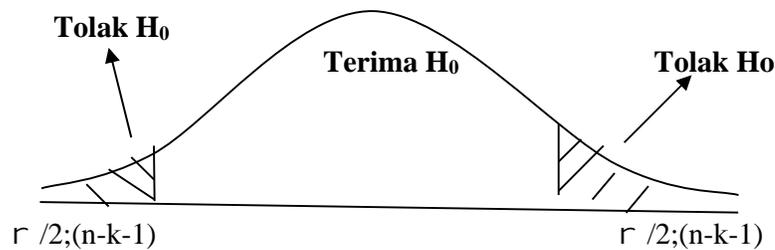
Uji 1 arah

$$t > t_{\alpha}; (n-k-1)$$



Uji 2 arah

$$t > t_{\alpha/2}; (n-k-1)$$



5. Menentukan Keputusan

Jika nilai t lebih besar dari nilai t Tabel, maka tolak H_0 . Berarti terdapat cukup bukti untuk menolak H_0 dan menerima H_1 .

Kesimpulan:

Jika H_0 ditolak pada tingkat signifikan α , maka X melalui persamaan $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$ merupakan prediktor bagi Y . Dengan perkataan lain ada korelasi linier yang nyata antara X_1, X_2, \dots, X_k dan Y .

Contoh 12:

Berdasarkan contoh 2 kita akan menguji apakah ada korelasi linier diantara ketiga variabel tersebut pada tingkat signifikansi $\alpha = 0,05$.

Diketahui $R^2 = 0,932$ $R = 0,965$

Jawab :

1. Menentukan Hipotesis

$H_0 : \rho = 0$

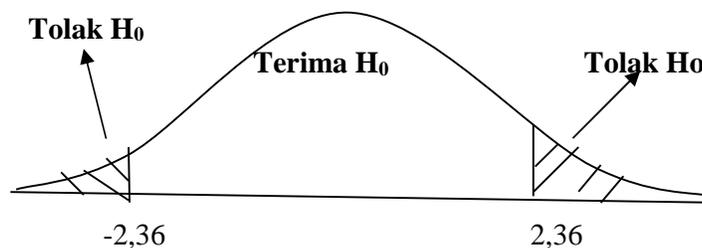
$H_1 : \rho \neq 0$

2. Menentukan daerah kritis

Uji t – Tabel dengan $db = n - k - 1 = 10 - 2 - 1 = 7$ dan $\alpha = 5\%$. Dengan pengujian 2 arah didapat nilai t -tabel = 2,36. Perhatikan uji 2 arah maka ekor distribusi t ada dua sehingga $5\% / 2 = 2,5\% = 0,025$; dan $db = 7$, didapat nilai kritis = 2,36

3. Menentukan nilai t hitung

$$t = \frac{|0,965| \sqrt{10 - 2 - 1}}{\sqrt{1 - 0,932}} = \frac{0,965 \cdot 2,65}{0,26} = \frac{2,55725}{0,26} = 9,84$$

4. Daerah keputusan**5. Menentukan keputusan**

Karena t hitung = 9,84 > 2,36 maka H_0 ditolak, artinya benar ada korelasi linier antara permintaan minyak goreng dengan harga dan pendapatan.

5.4 Rangkuman

1. Uji Signifikansi Secara Parsial

Pengujian ini dimaksudkan untuk menguji apakah suatu variabel bebas berpengaruh atau tidak secara nyata (signifikan) terhadap variabel tidak bebas. Untuk mengetahui apakah suatu variabel secara parsial berpengaruh nyata atau tidak digunakan statistik Uji t atau t – Student.

Langkah-langkah pengujian :

a. $H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

b. Tentukan

c. Daerah kritis: $|t_{hitung}| > t_{/2; n-k-1}$

d. Statistik uji:

$$t_{hitung} = \frac{b_k - \beta_k}{s(b_k)}$$

e. Keputusan:

- Terima H_0 : $|t_{hitung}| \leq t_{/2; n-k-1}$
- Tolak H_0 : $|t_{hitung}| > t_{/2; n-k-1}$

f. Kesimpulan

- Jika terima H_0 , maka parameter β_k tidak berguna dalam model
- Jika tolak H_0 , maka parameter β_k berguna dalam model

2. Uji Signifikansi Secara Simultan

Pengujian signifikansi terhadap model secara simultan (bersama-sama) dimaksudkan untuk melihat kemampuan seluruh variabel bebas mampu menjelaskan keragaman (memperkirakan) variabel tidak bebas Y.

Langkah-langkah pengujian :

$$SST = y^T y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

$$SSR = b^T X^T y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

$$SSE = SST - SSR$$

$$MSR = \frac{SSR}{k}$$

$$MSE = \frac{SSE}{n - k - 1}$$

a. $H_0: \beta_0 = \beta_1 = \dots = \beta_k = 0$

H_1 : tidak semua $\beta_j = 0$; $j = 0, 1, \dots, k$

b. Tentukan

c. Daerah kritis: $F_{hitung} > F_{(k, n-k-1)}$

d. Statistik uji:

$$F_{hitung} = \frac{MSR}{MSE}$$

e. Keputusan:

- Terima H_0 : $F_{hitung} \leq F_{(k, n-k-1)}$
- Tolak H_0 : $F_{hitung} > F_{(k, n-k-1)}$

f. Kesimpulan

- Jika terima H_0 , maka semua parameter tidak berguna dalam model
- Jika tolak H_0 , maka semua atau salah satu parameter berguna dalam model

Tabel Anova

SOV	SS	df	MS	F_{hitung}
Regression	$SSR = b^T X^T y - \frac{(\sum_{i=1}^n y_i)^2}{n}$	k	SSR/k	$\frac{MSR}{MSE}$
Error	$SSE = SST - SSR$	n-k-1	SSE/(n-k-1)	
Total	$SST = y^T y - \frac{(\sum_{i=1}^n y_i)^2}{n}$	n - 1		

3. Uji Koefisien Korelasi Regresi Linier Berganda

Langkah-langkah pengujian :

➤ Hipotesis

a. Uji 2 arah

H_0 : $\rho = 0$ (Tidak ada korelasi linier)

H_1 : $\rho \neq 0$ (Ada korelasi linier)

- b. Uji 1 arah
 $H_0 : \rho = 0$
 $H_1 : \rho > 0$ (Ada korelasi positif)
- c. Uji 1 arah
 $H_0 : \rho = 0$
 $H_1 : \rho < 0$ (Ada korelasi negatif)
- Tentukan
- Daerah kritis:
- a. Untuk $H_1 : \rho \neq 0$ (Ada korelasi linier)
 $|t_{hitung}| > t_{tabel} (/2; n-k-1)$
- b. Untuk $H_1 : \rho > 0$ (Ada korelasi positif)
 $t_{hitung} > t_{tabel} (; n-k-1)$
- c. Untuk $H_1 : \rho < 0$ (Ada korelasi negatif)
 $t_{hitung} < -t_{tabel} (; n-k-1)$
- Statistik uji:
- $$t_{hitung} = \frac{|R|\sqrt{n-k-1}}{\sqrt{1-R^2}}$$
- $|R|$ = Harga mutlak dari koefisien korelasi linier berganda (selalu positif)
 k = banyaknya variabel bebas
- Keputusan:
- a. Untuk $H_1 : \rho \neq 0$ (Ada korelasi linier)
- Tolak H_0 : $|t_{hitung}| > t_{tabel} (/2; n-k-1)$
 - Terima H_0 : $|t_{hitung}| < t_{tabel} (/2; n-k-1)$
- b. Untuk $H_1 : \rho > 0$ (Ada korelasi positif)
- Tolak H_0 : $t_{hitung} > t_{tabel} (; n-k-1)$
 - Terima H_0 : $t_{hitung} < t_{tabel} (; n-k-1)$
- c. Untuk $H_1 : \rho < 0$ (Ada korelasi negatif)
- Tolak H_0 : $t_{hitung} < -t_{tabel} (; n-k-1)$
 - Terima H_0 : $t_{hitung} > -t_{tabel} (; n-k-1)$
- Kesimpulan:
- Jika H_0 ditolak pada tingkat signifikan α , maka X melalui persamaan $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$ merupakan prediktor bagi Y . Dengan perkataan lain ada korelasi linier yang nyata antara X_1, X_2, \dots, X_k dan Y .

5.5 Soal-Soal

1. Lihat kembali *soal nomor 2 pada Bab I*. Pada tingkat kepercayaan 95 persen,
 - a. apakah penggunaan pupuk, secara parsial, berpengaruh nyata terhadap hasil panen jagung?
 - b. apakah penggunaan insektisida, secara parsial, berpengaruh nyata terhadap hasil panen jagung?
 - c. apakah penggunaan pupuk dan insektisida secara bersama-sama berpengaruh nyata terhadap hasil panen jagung?

2. Dilakukan sebuah penelitian di perusahaan penjualan peralatan perahu motor, dilakukan pengamatan biaya promosi dan biaya after sales terhadap jumlah omzet penjualan dan selama 6 tahun mulai dari tahun 2000 sd tahun 2005. Data sebagaimana pada tabel di bawah ini. Gunakan metode regresi linier berganda, bagaimanakah pengaruh variabel bebas terhadap variabel terikat secara simultan dengan $\alpha = 5\%$?

Tahun	2000	2001	2002	2003	2004	2005
Biaya promosi	7	4	16	3	21	8
Biaya after sales	33	41	7	49	5	31
Penjualan	42	33	75	28	91	55

Bab VI. Asumsi Dan Pelanggaran Asumsi Regresi

Berganda

Pada bab ini akan dibahas mengenai hal-hal yang harus dipenuhi (asumsi-asumsi pokok) sebelum melakukan regresi berganda, sehingga nilai koefisien regresi yang dihasilkan tidak bias.

Beberapa asumsi dalam regresi berganda adalah sebagai berikut :

1. Variabel tidak bebas dan variabel bebas memiliki hubungan yang linier. Jadi hubungan Y dengan X harus linier, bagaimana jika tidak linier? (terlihat di *scatter diagram*) Maka datanya ditransformasi terlebih dahulu menjadi linier. Jika *scatter diagram* terlihat :

- parabola maka diturunkan (*derivative*).
- bergelombang (*oscillatory*) maka dibuat *inverse trigonometric*.
- *exponential* maka ditarik logaritma.

2. Variabel tidak bebas harus bersifat kontinu atau berskala selang/interval. Variabel kontinu ini adalah variabel yang dapat menempati pada semua titik dan biasanya merupakan data dari proses pengukuran.

3. Nilai keragaman atau residu yaitu selisih antara data pengamatan dan data dugaan hasil regresi $(Y - \hat{Y})$ harus sama untuk semua nilai Y . Asumsi ini menyatakan bahwa nilai residu bersifat konstan untuk semua data Y , $(Y - \hat{Y}) = \text{„}$. Asumsi ini memperlihatkan kondisi HOMOSKEDASTISITAS yaitu nilai residu $(Y - \hat{Y})$ yang sama untuk semua nilai Y , menyebar normal dan mempunyai rata-rata 0.

4. Pengamatan-pengamatan untuk variabel tidak bebas dari suatu pengamatan ke pengamatan lain harus bebas atau tidak berkorelasi. Hal ini penting untuk data yang bersifat deret berkala.

Asumsi-asumsi tersebut haruslah dipenuhi dalam menyusun regresi berganda agar diperoleh hasil yang tidak bias. Untuk itu perlu dilakukan beberapa tes yang mungkin untuk mendeteksi pelanggaran terhadap asumsi-asumsi tersebut.

Beberapa pelanggaran asumsi regresi berganda sebagai berikut:

6.1 Multikolinieritas

Terdapat hubungan yang linier sempurna antara variabel tidak bebas, yang berarti melanggar asumsi 1 dalam regresi berganda. Salah satu akibat dari adanya hubungan antar variabel bebas, menurut Frish apabila terjadi multikolinier apalagi kolinier yang sempurna (koefisien korelasi antar variabel bebas = 1) maka koefisien regresi dari variabel bebas tidak dapat ditentukan dan standard errornya tidak terhingga.

$$\text{Contoh: } \hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$$

Dimana: \hat{Y} = Konsumsi
 X_1 = Total Pendapatan
 X_2 = Pendapatan dari Upah
 X_3 = Pendapatan dari bukan Upah

Disini total pendapatan (X_1) adalah penjumlahan dari pendapatan dari upah (X_2) dan pendapatan bukan dari upah (X_3), sehingga terjadi *perfect multicollinearity*.

Beberapa teknik untuk mengenali multikolinieritas :

- 1) Variabel bebas secara bersama-sama pengaruhnya nyata, atau uji F-nya nyata, namun ternyata setiap variabel bebasnya secara parsial pengaruhnya tidak nyata (uji t-nya tidak nyata)
- 2) Nilai koefisien determinasi R^2 sangat besar, namun ternyata variabel bebasnya berpengaruh tidak nyata (uji t-nya tidak nyata)
- 3) Nilai koefisien korelasi parsial yaitu r_{YX_1, X_2} , r_{YX_2, X_1} dan $r_{X_1 X_2, Y}$ ada yang lebih besar dari koefisien determinasinya.

Pada kasus contoh soal 1 terlihat bahwa dari uji F nyata, uji-t untuk variabel X_2 nyata, nilai $R^2 = 0,933$ relatif besar dan kuat namun variabel X_2 nyata, serta nilai koefisien korelasi parsial yaitu -0,55; -0,71 dan 0,143, nilai koefisien ini lebih kecil dari 0,933, sehingga dapat disimpulkan tidak terjadi multikolinier.

Ada beberapa cara yang dapat dilakukan, jika terjadi Multikolinieritas:

- a. Membuang variabel bebas yang diperkirakan sebagai penyebab multikolinier. Hal ini terlihat dari nilai korelasi parsial antar variabel bebas yang tinggi

b. Atau cara lain dengan menambah observasi atau data lagi.

Contoh penghitungan dengan

Klein dan Goldberger mencoba mencocokkan model regresi berikut ini

terhadap ekonomi Amerika Serikat :

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U_i$$

Dimana Y = Konsumsi (dalam milyar dolar)

X1 = Pendapatan upah (dalam milyar dolar)

X2 = Pendapatan non upah dan non pertanian (dalam milyar dolar)

X3 = Pendapatan pertanian (dalam milyar dolar)

Berikut ini adalah data pengamatan untuk kasus di atas :

Tabel 1

Data Mengenai Konsumsi (Y), Pendapatan Upah (X₁), Pendapatan Non Upah dan Non Pertanian (X₂), dan Pendapatan Pertanian (X₃) di Amerika Serikat
Dalam Milyar dollar

Tahun	Y	X ₁	X ₂	X ₃
1936	62,8	43,41	17,1	3,96
1937	65	46,44	18,65	5,48
1938	63,9	44,35	17,09	4,37
1939	67,5	47,82	19,28	4,51
1940	71,3	51,02	23,24	4,88
1941	76,6	58,71	28,11	6,37
1945*	86,3	87,69	30,29	8,96
1946	95,7	76,73	28,26	9,76
1947	98,3	75,91	27,91	9,31
1948	100,3	77,62	32,3	9,85
1949	103,2	78,01	31,39	7,21
1950	108,9	83,57	35,61	7,39
1951	108,5	90,59	37,58	7,98
1952	111,4	95,47	35,17	7,42

Sumber : L.R. Klein dan A.S. Goldberger, *An Economic Model of The United States, 1929 - 1952*, North Holland Publishing Company, Amsterdam, 1964, Hal. 131

* Data-data untuk tahun-tahun perang 1942 - 1944 hilang. Data untuk tahun-tahun lain dalam Milyar Dolar tahun 1939

Berdasarkan data di atas, kita ingin mengetahui ada atau tidaknya pelanggaran salah satu asumsi klasik multikolinearitas dalam data tersebut, sekaligus cara penanggulangannya jika pada data tersebut memang terdapat pelanggaran multikolinearitas, sehingga kita dapat memperoleh model regresi yang cocok untuk menggambarkan keadaan ekonomi di Amerika Serikat. Untuk menyelesaikannya, kita harus melakukan langkah-langkah sebagai berikut :

Sumber : L.R. Klein dan A.S. Goldberger, *An Economic Model of The United States, 1929 - 1952*, North Holland Publishing Company, Amsterdam, 1964, Hal. 131

* Data-data untuk tahun-tahun perang 1942 - 1944 hilang. Data untuk tahun-tahun lain dalam Milyar Dolar tahun 1939

Untuk menguji ada tidaknya multikolinearitas, kita dapat menggunakan nilai Toleransi atau VIF (Variance Inflation Factor), dengan rumus sebagai berikut :

$$VIF = \frac{1}{1-r_{12}^2} \qquad \text{Tolerance} = \frac{1}{VIF} = (1-r_{12}^2)$$

Jika nilai Toleransi kurang dari 0,1 atau nilai VIF melebihi 10 maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar variabel bebas.

Dengan bantuan *software SPSS 13*, kita dapat memperoleh nilai Toleransi atau VIF untuk data di atas pada tabel berikut ini.

Tabel 2
Nilai Toleransi dan VIF

Model		Collinearity Statistics	
		Tolerance	VIF
1	Pendapatan Upah	,081	12,297
	Pendapatan Non Upah dan Non Pertanian	,108	9,230
	Pendapatan Pertanian	,336	2,977

a. Dependent Variable: Konsumsi

1

Berdasarkan tabel 2 di atas, dapat dilihat bahwa terdapat nilai VIF (12,297) yang melebihi 10 dan nilai Tolerance (0,081) yang kurang dari 0,1. Oleh karena itu, kita menduga adanya pelanggaran asumsi klasik multikolinearitas pada data tersebut. Selain itu, kita dapat mendeteksi multikolinearitas melalui koefisien korelasi sederhana di antara variabel bebas. Jika koefisien korelasi sederhana mencapai atau melebihi 0,8 maka hal tersebut menunjukkan terjadinya masalah multikolinearitas dalam regresi.

Melalui SPSS 13 berdasarkan table 1, diperoleh nilai-nilai korelasi (korelasi parsial) antara Pendapatan upah, Pendapatan non upah dan Pendapatan pertanian

Tabel 3. Koefisien korelasi

		Correlations		
		Pendapatan Upah	Pendapatan Non Upah dan Non Pertanian	Pendapatan Pertanian
Pendapatan Upah	Pearson Correlation	1	,943**	,811**
	Sig. (2-tailed)		,000	,000
	N	14	14	14
Pendapatan Non Upah dan Non Pertanian	Pearson Correlation	,943**	1	,737**
	Sig. (2-tailed)	,000		,003
	N	14	14	14
Pendapatan Pertanian	Pearson Correlation	,811**	,737**	1
	Sig. (2-tailed)	,000	,003	
	N	14	14	14

** . Correlation is significant at the 0.01 level (2-tailed).

Berdasarkan tabel 3 di atas, dapat dilihat bahwa terdapat korelasi yang tinggi antara variabel bebas pendapatan upah (X1) dengan variabel pendapatan non upah dan non pertanian (X2) yaitu sebesar 0,943, dan antara variabel bebas pendapatan upah (X1) dengan variabel pendapatan pertanian (X3) yaitu sebesar 0,811. Dengan adanya nilai korelasi yang tinggi tersebut (melebihi 0,8), maka kita dapat menduga adanya pelanggaran asumsi klasik multikolinearitas pada data tersebut.

6.2 Heteroskedastisitas

Heteroskedastisitas untuk menunjukkan nilai varians residu $(Y - \hat{Y})$ antar nilai Y tidaklah sama atau hetero. Hal demikian sering terjadi pada data yang bersifat *cross section*, yaitu data yang dihasilkan pada suatu waktu dengan responden yang banyak. Kenapa terjadi perbedaan varian antar data pengamatan? Ada tiga kemungkinan :

Dalam data yang bersifat *cross section* memungkinkan banyak variasi seperti pendapatan ada yang hanya ratusan ribu per bulan, namun ada yang sampai miliaran. Perbedaan yang sangat besar bisa memungkinkan adanya varian yang berbeda antar data pengamatan.

Proses belajar, pada saat pertama bisa terjadi varian besar, dengan falsafah belajar, makin lama makin mengerti, maka dengan bertambahnya ilmu tersebut dapat membuat varian yang lebih kecil.

Teknik pengumpulan data, apabila jumlah data sedikit, cenderung akan bervariasi dibandingkan dengan jumlah datanya banyak.

Apa dampak dari terjadinya heteroskedastisitas?

Walaupun terjadi heteroskedastisitas, koefisien penduga yaitu b_1 dan b_2 tetap efisien, namun variannya atau kesalahan baku penduganya menjadi lebar atau tidak efisien.

Interval keyakinan untuk koefisien regresi menjadi semakin lebar dan uji signifikansi kurang kuat.

Apabila kita menggunakan OLS maka uji-t dan F tidak berfungsi sebagaimana mestinya, sehingga diperlukan perubahan-perubahan.

Bagaimana mendeteksi heteroskedastisitas?

Metode grafik yaitu menghubungkan antara Y dan $e^2 = (Y - \hat{Y})^2$, apabila hubungan antara Y dan e^2 tidak sistematis seperti makin membesar atau mengecil seiring bertambahnya Y , maka tidak terjadi heteroskedastisitas.

Uji korelasi *Rank Spearman* digunakan untuk menguji heteroskedastisitas.

6.3 Koefisien *rank* korelasi dari *Spearman*:

$$R_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Dimana

- d_i = perbedaan dalam *rank* yang ditempatkan untuk dua karakteristik yang berbeda dari sampel yang ke- i
- n = banyaknya sampel

	A	B	C	D	E	F	G	H
1								
2	Y	X	ej	R dr X	R dr ej	d	d ²	
3	12,4	12,1	1,017	5	9	-4	16	
4	14,4	21,4	1,26	7	10	-3	9	
5	14,6	18,7	0,181	8	4	4	16	
6	16	21,7	0,202	9	5	4	16	
7	11,3	12,4	0,221	3,5	6	-2,5	6,25	
8	10	10,4	0,602	1	7	-6	36	
9	16,2	20,6	0,908	10	8	2	4	
10	10,4	10,2	0,11	2	3	-1	1	
11	13,1	16	0,077	6	2	4	16	
12	11,3	12	0,037	3,5	1	2,5	6,25	
13							126,5	
14								

Gambar 5.

Tentukan hipotesis:

Ho: Ada heteroskedastisitas

H1: Tidak ada heteroskedastisitas

Koefisien korelasi *Rank Spearman* sebesar:

$$r_s = 1 - 6 \frac{126,5}{10(10^2 - 1)} = 0,2333$$

Mencari nilai t - observasi:

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

Daerah Kritis:

Lihat tabel – t dengan df $n-2 = 8$ dan tingkat kepercayaan sebesar 10 persen, didapat nilai kritis sebesar = 1,397

Daerah keputusan:

Jika nilai t – hitung melebihi nilai t - tabel, maka hipotesis nol diterima berarti menerima adanya heteroskedastisitas. Dari perhitungan di atas didapat nilai t- hitung sebesar 0,6786 lebih kecil dari nilai t- tabel sebesar 1,397, berarti menolak hipotesis nol. Kesimpulan; Tidak ada bukti adanya hubungan yang sistematis antara variabel yang menjelaskan dan nilai mutlak dari residual, maka tidak terdapat heteroskedastisitas.

Bagaimana cara mengatasi heteroskedastisitas?

Heteroskedastisitas tidak merusak sifat kebiasaan dan konsistensi dari penaksir. Oleh karena itu untuk mengatasinya dapat dilakukan dengan metode kuadrat terkecil tertimbang, nilai tertimbang dapat dilakukan berdasarkan apriori atau observasi.

6.4 Autokorelasi

Autokorelasi dikenalkan oleh Maurice G. Kendall dan William R. Buckland. Autokorelasi merupakan korelasi antara anggota observasi yang disusun menurut urutan waktu. Ada beberapa penyebab autokorelasi yaitu :

Kelembaman, biasanya terjadi dalam fenomena ekonomi dimana sesuatu akan mempengaruhi sesuatu mengikuti siklus bisnis atau saling kait mengkait.

Terjadi bias dalam spesifikasi yaitu ada beberapa variabel yang tidak termasuk dalam model.

Bentuk fungsi yang dipergunakan tidak tepat, seperti semestinya bentuk non-linier digunakan linier atau sebaliknya.

Bagaimana mendeteksi Autokorelasi?

Metode grafik yang menghubungkan antara error (e) atau residu dengan waktu, apabila terdapat hubungan yang sistematis baik meningkat atau menurun menunjukkan adanya autokorelasi.

Cara lain untuk mendeteksi adanya autokorelasi dengan melakukan Uji *durbin watson*.

6.5 Uji Durbin Watson

Di dalam model regresi linear klasik mengasumsikan bahwa tidak ada autokorelasi dalam disturbansi atau gangguan u_i .

Dengan lambang: $E(u_i u_j) = 0 \quad i \neq j$

Unsur gangguan yang berhubungan dengan observasi tidak dipengaruhi oleh unsur disturbansi atau gangguan yang berhubungan dengan pengamatan lain.

Misalkan:

Dalam data deret waktu kuartalan atas regresi *output* atau hasil atau masukan tenaga kerja dan modal. Jika ada pemogokan buruh yang mempengaruhi hasil dalam satu kuartal, maka produksi kuartal berikutnya tidak harus terpengaruh. Jika hasilnya lebih rendah dalam kuartal ini, maka tidak alasan untuk mengharapkan akan rendah pula untuk kuartal beikutnya.

Untuk data *cross section* atas regresi konsumsi terhadap pendapatan keluarga, pengaruh peningkatan pendapatan satu keluarga atas konsumsinya tidak diharapkan untuk mempengaruhi konsumsi keluarga lainnya.

Tetapi jika terdapat ketergantungan, maka dikatakan mempunyai **Autokorelasi**.

Gangguan yang disebabkan oleh pemogokan pada kuartal ini sangat mungkin mempengaruhi hasil atau *output* kuartal berikutnya. Peningkatan dalam konsumsi suatu keluarga, mungkin menyebabkan keluarga lain meningkatkan konsumsinya.

Dengan lambang: $E(u_i u_j) \neq 0 \quad i \neq j$

Statistik *d-Durbin Watson*: merupakan rasio dari jumlah kuadrat perbedaan dalam residual yang berturut-turut terhadap jumlah kuadrat regresi.

$$d = \frac{\sum_{t=2}^{t=n} (e_t - e_{t-1})^2}{\sum_{t=1}^{t=n} e_t^2}$$

Prosedur pengujian:

1. Susun persamaan regresi dan dapatkan e_i
2. Hitung nilai d
3. Tentukan nilai kritis dari d , untuk batas bawah d_L dan batas atas d_U
4. Tentukan hipotesis:

H_0 : Tidak ada autokorelasi

H1: Ada autokorelasi

5. Daerah keputusan:

Terima hipotesis nol, jika $d_u < d < 4 - d_u$

Tolak hipotesis nol jika $d < d_L$ dan $d > 4 - d_L$

6.6 Rangkuman

1. Asumsi-asumsi regresi linier berganda haruslah dipenuhi dalam menyusun regresi linier berganda, perlu dilakukan beberapa tes untuk mendeteksi pelanggaran terhadap asumsi-asumsi tersebut.
2. Cara yang dilakukan untuk mengatasi multikolinier yaitu dengan membuang variabel bebas yang diperkirakan sebagai penyebab multikolinier atau dengan menambah observasi atau data lagi.
3. Uji korelasi *rank Spearman* digunakan untuk menguji heteroskedastisitas.
4. Metode grafik yang menghubungkan antara error (e) atau residu dengan waktu, apabila terdapat hubungan yang sistematis baik meningkat atau menurun menunjukkan adanya autokorelasi. Cara lain untuk mendeteksi adanya autokorelasi dengan melakukan uji *durbin watson*.

6.7 Soal-Soal

1. Multikolinieritas berarti:
 - a. Terjadinya korelasi linear yang tinggi diantara variabel eksplanatori.
 - b. Terjadinya korelasi linear yang tinggi diantara variabel respon.
 - c. Terjadinya korelasi linear yang tinggi diantara variabel-variabel respon dengan variabel eksplanatori.
 - d. Jawaban a, b, dan c benar.

2. Beberapa cara mendeteksi adanya multikolinearitas:
 - a. Regresi komponen utama, indeks kondisi, VIF.
 - b. Tol, indeks kondisi, VIF.
 - c. Regresi komponen utama, regresi ridge, regresi kuadrat terkecil parsial.
 - d. Tidak ada jawaban yang benar.
3. Heteroskedastisitas berarti:
 - a. Adanya korelasi yang tinggi diantara variabel eksplanatory.
 - b. Adanya korelasi antara pengamatan yang satu dengan pengamatan yang lainnya.
 - c. Adanya *varians error* yang tidak sama.
 - d. Adanya *varians Y* yang sama.
4. Beberapa cara mengatasi jika terdapat heteroskedastisitas adalah
 - a. Uji korelasi *Rank Spearman*, Uji *Park*, Uji *Goldfeld-Quadrat*.
 - b. Uji *Park*, Metode *WLS*, Uji *White*.
 - c. Plot antara X dan Y, plot antara eror dan Y.
 - d. Transformasi variabel, metode *WLS*.
5. Pada model ekonometrika sering terjadi kasus autokorelasi, karena...
 - a. Variabel-variabel ekonomi saling terikat.
 - b. Kasalahan penggunaan model.
 - c. Pada umumnya model ekonometrika menggunakan data *time series*, dimana anantara pengamatan ke-t dengan pengamatan sebelumnya ada keterkaitan.
 - d. Jawaban a, b, dan c benar.

Bab VII. Cara Memilih Independen Variabel

Bagaimana jika calon independen variabel yang akan digunakan dalam analisis regresi ditentukan oleh teori dalam bidang yang bersangkutan (*subject matter*), bukan ditentukan oleh ilmu statistik atau *computer software*.

Ukuran yang sering dipakai untuk memilih calon independen variabel dalam pembentukan persamaan regresi yang terbaik yaitu *adjusted R²* yang merupakan perbaikan dari *R²*.

Beberapa cara atau metode pemilihan independen variabel :

Sebagai contoh:

Kita memiliki calon independen variabel X_1, X_2, X_3, X_4 maka untuk memilih independen variabel yang ikut dalam persamaan regresi yang terbaik, maka :

Masukan semua calon independen variabel X_1, X_2, X_3, X_4 kemudian mulai keluarkan satu calon independen variabel. Didapat persamaan regresi tanpa X_1 dan hitung *adjusted R²*.

Lalu masukan kembali X_1 buat persamaan regresi tanpa X_2 dan hitung *adjusted R²*.

Kemudian masukan kembali X_2 dan buat persamaan regresi tanpa X_3 , hitung *adjusted R²* dan seterusnya sampai X_4 .

Selanjutnya keluarkan dua calon independen variabel, dimulai dengan X_1 dan X_2 tidak ikut dalam persamaan regresi kemudian hitung *adjusted R²* nya.

Lalu masukan kembali X_2 sedangkan X_1 dan X_3 tidak masuk dalam persamaan regresi, hitung *adjusted R²*.

Demikian seterusnya dicatat *adjusted R²* setiap persamaan, cari persamaan regresi yang memberikan *adjusted R²* yang tertinggi.

7.1 Backward Elimination

Cara lain memilih adalah masukkan X_1, X_2, X_3, X_4 , kemudian keluarkan dari persamaan regresi calon independen variabel yang memiliki korelasi parsial yang terendah dengan dependen variabel dan hitung $adjusted R^2$. Demikian seterusnya dicatat $adjusted R^2$ nya, cari persamaan regresi yang mempunyai $adjusted R^2$ tertinggi.

Prosedur pembentukan model menggunakan *backward elimination*:

- 1) Lakukan penghitungan dengan persamaan regresi yang mengandung semua variabel.
- 2) Hitung F-uji parsial untuk setiap variabel prediktor dimana variabel prediktor tersebut dianggap sebagai variabel paling akhir yang dimasukkan ke dalam persamaan regresi.
- 3) Nilai F-uji parsial terkecil, misal F_L , dibandingkan dengan tingkat signifikansi yang telah ditentukan, misal F_0 , dengan ketentuan sebagai berikut:
 - Jika $F_L < F_0$, maka hilangkan variabel prediktor terkait, Z_L , dari persamaan regresi dan ulangi langkah 1 dengan persamaan regresi yang baru.
 - Jika $F_L > F_0$, maka terima persamaan regresi tersebut sebagai model terbaik.

Contoh :

- Ilustrasi *backward elimination* menggunakan data *Hald*:

- 1) Penghitungan dengan persamaan regresi yang mengandung semua variabel, diperoleh:

$$\hat{Y} = 62,41 + 1,551 X_1 + 0,510 X_2 + 0,102 X_3 - 0,144 X_4$$

- 2) Hitung nilai F-uji parsial, diperoleh hasil sebagai berikut:

$$F_{1|234} = 4,34 \qquad F_{2|134} = 0,50$$

$$F_{3|124} = 0,02 \qquad F_{4|123} = 0,04$$

- 3) F-uji parsial terkecil adalah $F_{3|124} = 0,02$ dan titik kritis untuk $\alpha = 0,05$ adalah $F_0(1,8;0,95) = 5,32$. Karena $F_{3|124} < F_0$ maka variabel X_3 dikeluarkan dari persamaan regresi.

- 4) Penghitungan dengan persamaan regresi baru (setelah X_3 dikeluarkan) diperoleh persamaan sebagai berikut:

$$\hat{Y} = 71,65 + 1,452 X_1 + 0,416 X_2 - 0,237 X_4$$

Model di atas memiliki nilai $R^2 = 0,982$ dan nilai overall F = 166,83

- 5) Hitung nilai F-uji parsial untuk mencari kemungkinan variabel lain yang akan dieliminasi, diperoleh hasil sebagai berikut:

$$F_{1/24} = 154,01 \quad F_{2/14} = 5,03$$

$$F_{4/12} = 1,86$$

- 6) F-uji parsial terkecil adalah $F_{4/12} = 1,86$ dan titik kritis untuk $\alpha = 0,05$ adalah $F_0(1,9;0,95) = 5,12$. Karena $F_{4/12} < F_0$ maka variabel X_4 dikeluarkan dari persamaan regresi.
- 7) Penghitungan dengan persamaan regresi baru (setelah X_4 dikeluarkan) diperoleh persamaan sebagai berikut:

$$\hat{Y} = 52,58 + 1,468 X_1 + 0,662 X_2$$

Model di atas memiliki nilai $R^2 = 0,979$ dan nilai overall F = 229,5

- 8) Hitung nilai F-uji parsial, diperoleh hasil sebagai berikut:

$$F_{1/2} = 146,52 \quad F_{2/1} = 208,58$$

- 9) Terlihat bahwa kedua F-uji parsial sangat signifikan, jauh di atas $F_0(1,8;0,95) = 4,96$, sehingga hasil akhir diperoleh model persamaan regresi:

$$\hat{Y} = 52,58 + 1,47 X_1 + 0,66 X_2$$

7.2 Forward Inclusion

Cari satu dari X_1, X_2, X_3, X_4 yang memiliki korelasi parsial yang tertinggi dengan dependen variabel, masukan dalam persamaan regresi dan hitung *adjusted R²*. Demikian seterusnya tambahkan ke dalam persamaan regresi catat *adjusted R²*, cari persamaan regresi yang mempunyai *adjusted R²* tertinggi.

- Metode *forward selection* adalah pemodelan dimulai dari nol variabel (*empty model*), kemudian satu persatu variabel prediktor dimasukan sampai kriteria tertentu dipenuhi.
- Dalam metode ini, variabel prediktor yang telah dimasukkan dalam model tidak dapat dikeluarkan lagi.
- Langkah-langkah metode forward adalah sebagai berikut :
 - a) Membuat model dengan meregresikan variabel respon Y dengan setiap variabel prediktor. Kemudian dipilih model yang mempunyai nilai R^2 tertinggi. Misal model tersebut adalah yang memuat prediktor X_a , yaitu $Y = \alpha_0 + \alpha_1 X_a$.
 - b) Meregresikan variabel respon Y , dengan prediktor X_a , ditambah dengan setiap prediktor selain X_a dan prediktor lain. Kemudian semua model diperiksa dan dilihat

peningkatan pada R^2 dan nilai F parsial untuk semua model diperiksa dan dipilih yang terbaik sesuai kriteria. Misal model terpilih mengandung tambahan prediktor X_b , yaitu model

$$Y = o + aX_a + bX_b + \dots$$

- c) Proses diulang sampai tidak ada lagi variabel di luar variabel prediktor yang telah terpilih yang memenuhi kriteria.

- Ilustrasi *forward selection* dengan data Hald:

- a) Variabel pertama yang dimasukkan adalah variabel prediktor yang memiliki korelasi paling tinggi dengan variabel respon. Dari matriks korelasi berikut ini:

Matrix M1

1.00000	0.22858	-0.82413	-0.24545	0.73072
0.22858	1.00000	-0.13924	-0.97295	0.81625
-0.82413	-0.13924	1.00000	0.02954	-0.53467
-0.24545	-0.97295	0.02954	1.00000	-0.82131
0.73072	0.81625	-0.53467	-0.82131	1.00000

dimana urutan dari kiri ke kanan dan dari atas ke bawah adalah variabel X_1, X_2, X_3, X_4 , dan Y , terlihat variabel X_4 memiliki korelasi paling tinggi terhadap Y yaitu sebesar -0,8213. Dan dari hasil uji F-parsial diperoleh sebagai berikut:

Partial F-Values of Variables Not in Regression

1	2	3	4
12.60	21.96	4.40	22.80

Terlihat X_4 memenuhi kriteria, sehingga X_4 terpilih dan diperoleh model:

$Y = o + 4X_4 + \dots$, atau hasil penghitungan diperoleh:

$$Y = 118 - 0,738 X_4$$

Model di atas memiliki nilai $R^2 = 0,675$ dan nilai overall F = 176,63

- b) Tambahkan setiap variabel lainnya ke dalam model yang telah mengandung X_4 dan hitung nilai F-parsial untuk setiap model, diperoleh:

Subscripts of Variables in Regression	Partial F-Values of Variables Not in Regression			
	1	2	3	4
4	108.22	0.17	40.29	—

Terlihat variabel X_1 memiliki nilai F-parsial paling tinggi, sehingga X_1 dipilih dan model persamaan regresi menjadi seperti berikut ini:

$Y = o + 1X_1 + 4X_4 + \dots$, atau hasil penghitungan diperoleh:

$$Y = 103 + 1,44 X_1 - 0,614 X_4$$

Model di atas memiliki nilai $R^2 = 0,972$ dan nilai overall $F = 176,63$, sudah sangat signifikan.

- c) Tambahkan setiap variabel lainnya ke dalam model yang telah mengandung X_1 dan X_4 dan hitung nilai F-parsial untuk setiap model, diperoleh:

Subscripts of Variables in Regression	Partial F-Values of Variables Not in Regression			
	1	2	3	4
14	—	5.03	4.24	—

Terlihat bahwa $F_{2|1,4} = 5,03$ dan $F_{3|1,4} = 4,24$ keduanya lebih kecil dari $F_{(1,9;0,95)} = 5,12$ sehingga penambahan variabel-variabel tersebut ke dalam model tidak akan memberikan perubahan yang signifikan.

7.3 Stepwise

Adalah gabungan dari *backward elimination* dan *forward inclusion*, independen variabel lain dapat masuk dalam persamaan regresi lalu keluar dari persamaan regresi bergantung *adjusted R^2* .

Metode *stepwise regression* pada prinsipnya hampir sama dengan *forward selection* dimana pemodelan dimulai dari nol variabel (*empty model*), kemudian satu persatu variabel prediktor dimasukkan sampai kriteria tertentu dipenuhi.

Perbedaannya pada *stepwise regression* dimungkinkan untuk membatalkan/menghilangkan kembali variabel yang sebelumnya telah dimasukkan dalam model. Setiap ada variabel baru yang dimasukkan ke dalam model, signifikansi dari setiap variabel yang sudah ada di dalam model diperiksa kembali.

Dengan kata lain pada tiap tahap dalam prosedur *forward selection*, lakukan kembali uji signifikansi (uji F-parsial) untuk setiap variabel dalam model, dan hilangkan dari model untuk variabel yang tidak signifikan (variabel dengan nilai F-parsial kurang dari nilai F-uji).

7.4 Rangkuman

1. Backward Elimination

Cara lain memilih adalah masukkan X_1, X_2, X_3, X_4 , kemudian keluarkan dari persamaan regresi calon independen

variabel yang memiliki korelasi parsial yang terendah dengan dependen variabel dan hitung $adjusted R^2$. Demikian seterusnya dicatat $adjusted R^2$ nya, cari persamaan regresi yang mempunyai $adjusted R^2$ tertinggi.

2. Forward Selection

Cari satu dari X_1, X_2, X_3, X_4 yang memiliki korelasi parsial yang tertinggi dengan dependen variabel, masukan dalam persamaan regresi dan hitung $adjusted R^2$. Demikian seterusnya tambahkan ke dalam persamaan regresi catat $adjusted R^2$, cari persamaan regresi yang mempunyai $adjusted R^2$ tertinggi.

3. Stepwise

Adalah gabungan dari *backward elimination* dan *forward inclusion*, independen variabel lain dapat masuk dalam persamaan regresi lalu keluar dari persamaan regresi bergantung $adjusted R^2$.

Tabel. Data Hald

X_1	X_2	X_3	X_4	Y
7	26	6	60	78,5
1	29	15	52	74,3
11	56	8	20	104,3
11	31	8	47	87,6
7	52	6	33	95,9
11	55	9	22	109,2
3	71	17	6	102,7
1	31	22	44	72,5
2	54	18	22	93,1
21	47	4	26	115,9
1	40	23	34	83,8
11	66	9	12	113,3
10	68	8	12	109,4

Bab VIII. Penutup

8.1 Kesimpulan

Regresi dan korelasi merupakan analisis untuk melihat hubungan variabel bebas dan tidak bebas. Dalam melihat pola hubungan ini, langkah awal yang dilakukan yaitu dengan menggunakan scatter plot (plot data). Kemudian tentukan analisis regresi yang tepat pada pola hubungan yang terdapat dalam plot datanya.

Regresi linier merupakan persamaan garis lurus yang menunjukkan perubahan variabel bebas berpengaruh pada perubahan variabel tidak bebasnya. Data baik digunakan untuk persamaan regresi linier jika nilai koefisien korelasi antara variabel tidak bebas dan bebasnya mendekati 1 (satu).

8.2 Implikasi

Sebagai tambahan pengetahuan peserta Diklat Fungsional Statistisi Tingkat Ahli untuk diterapkan pada unit kerjanya masing-masing dengan kasus-kasus yang riil.

Dengan menghitung regresi dan korelasi peserta dapat melihat fenomena yang terjadi pada data di unit kerjanya.

8.3 Tindak Lanjut

Peserta diklat fungsional statistisi tingkat ahli dapat mengembangkan regresi dan korelasi ini pada kasus yang lebih dari 3 (tiga) variabel, yaitu 1 (satu) variabel tidak bebas dan lebih dari 2 (dua) variabel bebasnya serta pada kasus regresi non linier.

Aplikasi Komputer

Untuk penghitungan persamaan regresi lebih dari 2 variabel bebas dan nilai-nilai statistiknya dapat menggunakan bantuan paket komputer MS. Excel atau SPSS

Di bawah ini adalah hasil *output* penghitungan regresi pada contoh 2 dengan menggunakan paket komputer Microsoft Excel.

	A	B	C	D	E	F	G	H
1	SUMMARY OUTPUT							
2								
3	<i>Regression Statistics</i>							
4	Multiple R	0,996030126						
5	R Square	0,993214204						
6	Adjusted R Square	0,914132549						
7	Standard Error	0,715114241						
8	Observations	10						
9								
10	<i>ANOVA</i>							
11		df	SS	MS	F	Significance F		
12	Regression	2	50,02028136	25,01014068	48,90536336	7,69829E-05		
13	Residual	7	3,57971864	0,511388377				
14	Total	9	53,6					
15								
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%
17	Intercept	15,08616647	3,035108769	4,970552168	0,001618165	7,909279804	22,26305314	7,909279804
18	X Variable 1	-1,015240328	1,580455635	-1,749040203	0,123766216	-2,387798935	0,357318279	-2,387798935
19	X Variable 2	-0,410902695	0,155321874	-2,637002864	0,033569316	-0,779362616	-0,04244278	-0,779362616
20								

Gambar 12.

Berdasarkan *output* tersebut maka persamaan regresi linier berganda untuk contoh 2 adalah:

$$Y = 15,086 - 1,015X_1 - 0,41X_2$$

Dengan Kesalahan Baku / *Standard Error* $SE =$

$$\sqrt{\frac{SSE}{n-(k+1)}} = \sqrt{\frac{3,57971864}{10-(2+1)}} = 0,71511424$$

Kesalahan Baku Penduga adalah sebagai berikut:

$$Sb_0 = 3,035108769; \quad Sb_1 = 0,580455685; \quad Sb_2 = 0,155821874$$

$$\text{Koefisien Determinasi Linier Berganda } R^2 = \frac{SSR}{SST} = \frac{50,02028136}{53,6} = 0,933214204$$

Dan Korelasi Linier Berganda adalah $R = 0,966030126$

Pengujian :

▪ Uji F

Untuk pengujian apakah variabel bebas secara simultan dapat menerangkan variabel tidak bebas pada tingkat signifikansi α , dengan kata lain apakah model tersebut dapat digunakan, maka dilakukan pengujian hipotesis yang telah dijelaskan sebelumnya dengan menggunakan uji F. Dari hasil *print out* terlihat bahwa $F_{hit} = 48,906$ dan signifikan $F = 0,00007698$, berarti pada tingkat signifikansi $\alpha = 0,05$ hipotesis nol ditolak, karena $\alpha > \text{Signifikan } F$. Kesimpulan : model tersebut dapat digunakan.

▪ Uji T

Untuk menguji apakah ada pengaruh nyata secara parsial masing-masing variabel bebas X_k pada variabel tidak bebas Y , kita gunakan uji t pada tingkat signifikansi α . Dari hasil *print-out* terlihat :

t hitung untuk $X_1 = -1,749$ dan $P\text{-Value} = 0,1237$. Ternyata $\alpha = 0,05 < \text{nilai } P\text{-Value}$, berarti hipotesis nol diterima dengan kata lain X_1 tidak berpengaruh pada Y .

t hitung untuk $X_2 = -2,637$ dan $P\text{-Value} = 0,0335$. Ternyata $\alpha = 0,05 > \text{nilai } P\text{-Value}$, berarti hipotesis nol ditolak dengan kata lain X_2 berpengaruh pada Y .

▪ Interval Kepercayaan 95% untuk parameter B_0 adalah :

$$7,909 \leq B_0 \leq 22,263$$

▪ Interval Kepercayaan 95% untuk parameter B_1 adalah:

$$-2,388 \leq B_1 \leq 0,357$$

▪ Interval Kepercayaan 95% untuk parameter B_2 adalah:

$$-0,779 \leq B_2 \leq -0,042$$

Latihan

1. Dalam rangka meningkatkan keuntungan usaha, maka para pengusaha eceran menambah tenaga kerjanya dan memperluas areal parkir sebagai pelayanan yang baik kepada konsumen. Dilakukan survei terhadap 15 pengusaha eceran.

Hasil survei tersebut adalah sebagai berikut :

Tabel 1.

No	Penerimaan keuntungan (Rp.000/bln)	Jumlah Tenaga Kerja	Luas area parkir mobil (M ²)
1	280	14	18
2	130	7	3
3	123	13	8
4	160	12	7
5	450	39	36
6	570	67	84
7	315	20	5
8	64	4	8
9	340	48	55
10	670	81	98
11	370	12	65
12	644	59	9
13	128	44	12
14	415	28	9
15	387	65	12

- a. Buatlah Persamaan Regresinya
- b. Hitunglah Koefisien Determinasi dan Koefisien Non Determinasi
- c. Ujilah dengan uji F, uji t dan uji koefisien korelasinya!
- d. Apa yang lebih mempengaruhi penerimaan keuntungan, apakah harus menambah tenaga kerja saja, luas parkir saja atau keduanya?

2. Dibawah ini adalah data mengenai jumlah rata-rata penerimaan, rata-rata biaya pelayanan dan rata-rata biaya promosi dari 20 perusahaan parawisata.

Tabel 2

Perusahaan	Rata-rata penerimaan dalam milyar (y)	Rata-rata biaya pelayanan dalam milyar (x_1)	Rata-rata biaya promosi dalam milyar (x_2)
1	96	6	20
2	83	22	30
3	126	18	40
4	61	8	10
5	59	12	10
6	90	10	10
7	82	17	30
8	88	11	20
9	86	16	20
10	76	23	30
11	102	7	20
12	108	12	30
13	96	24	40
14	70	16	20
15	80	9	10
16	113	11	30
17	76	22	20
18	74	11	10
19	98	16	20
20	80	8	20

- Tentukan persamaan regresi berganda dengan menggunakan ketiga variabel di atas.
- Hitunglah koefisien korelasi, koefisien determinasi dan Koefisien Parsial !
- Hitunglah Kesalahan Baku/*Standar Error* dan *Standar Error Estimation*-nya !
- Menurut anda, variabel independen mana yang memiliki dampak yang lebih menentukan terhadap variabel dependen? Beri evaluasi !

3. Hasil pengolahan dengan komputer terhadap data yang telah dikumpulkan dari sampel berukuran 15 yang menganalisis hubungan antara variabel tak bebas (Y) dengan variabel bebas X_1 dan X_2 diperoleh sebagai berikut :

ANOVA

	df	SS	MS	F	Significance F	
Regression	2	228014,6263	114007,3	168,4712	1,65411E-09	
Residual	12	8120,603016	676,7169			
Total	14	236136,2293				
	Coefficient	Standard Error	t Stat	P-Value	Lower 95%	Upper 95%
Intercept	662,151	21,093110433	26,65094	4,78E-12	516,1930837	608,10893
X1	-5,4365	0,336216167	-16,1699	1,64E-09	-6,16913267	-4,704029
X2	-20,0123	2,34505227	-8,54313	1,91E-06	-25,116201	-14,90844

- Berdasarkan program komputer di atas, buatlah persamaan regresinya!
- Hitung Koefisien Determinasi Regresi Linear Berganda dan jelaskan artinya!
- Berdasarkan model, buatlah nilai ramalan bagi y untuk $X_1 = 3$ dan $X_2 = 5$
- Tentukan *Standar Error*/ Kesalahan baku dan *Standar Error Estimation*-nya, lalu apa artinya!
- Berapakah nilai taksiran /penduga untuk parameter B_0, B_1 dan B_2 ?

4. Di bawah ini adalah *print out* hasil 60 data yang digunakan untuk menyelidiki hubungan antara motivasi baca, tingkat pendidikan ibu dan tingkat pendidikan ayah dengan kemahiran membaca literatur bahasa Inggris mahasiswa tingkat I pada suatu universitas. ($\alpha = 0,05$).

ANOVA

	df	SS	MS	F	Significance F
Regression	4	21,4725	5,3681	41,8147	0,000E+00
Residual	55	7,0608	0,1284		
Total	59	28,5333			
	Coefficient	Standard Error	t Stat		
Intercept	2,5026				
Motivasi	0,0050	0,0013	3,8462		
TP Ibu	0,1750	0,0195	8,9744		
TP Ayah	0,0411	0,181	0,2271		

Adjusted R Squared = 0,7393

R Squared =

Multiple R =

Standard Error =

- a. Lengkapi isian titik-titik pada hasil *print-out* di atas!
- b. Buatlah model dan persamaan regresi linier bergandanya!
- c. Apakah ada pengaruh dari ketiga unsur yang diselidiki terhadap kemahiran membaca literatur bahasa Inggris?
- d. Apakah ada korelasi linier yang nyata antara motivasi membaca, tingkat pendidikan ibu, ayah dan kemahiran membaca pada tingkat signifikansi $\alpha = 0,05$?

Jawaban

1. a. Persamaan umum regresi berganda dengan 2 variabel bebas:

$$Y = b_0 + b_1X_1 + b_2X_2 + \epsilon_i$$

Menggunakan cara dihalaman

no	y	x1	x2	yx1	yx2	x1x1	x2x2	x1x2
1	280	14	18	3920	5040	196	324	252
2	130	7	3	910	390	49	9	21
3	123	13	8	1599	984	169	64	104
4	160	12	7	1920	1120	144	49	84
5	450	39	36	17550	16200	1521	1296	1404
6	570	67	84	38190	47880	4489	7056	5628
7	315	20	5	6300	1575	400	25	100
8	64	4	8	256	512	16	64	32
9	340	48	55	16320	18700	2304	3025	2640
10	670	81	98	54270	65660	6561	9604	7938
11	370	12	65	4440	24050	144	4225	780
12	644	59	9	37996	5796	3481	81	531
13	128	44	12	5632	1536	1936	144	528
14	415	28	9	11620	3735	784	81	252
15	387	65	12	25155	4644	4225	144	780
jumlah	5046	513	429	226078	197822	26419	26191	21074

$$A = n \sum X_1 Y - \sum X_1 \sum Y$$

$$B = n \sum (X_2)^2 - (\sum X_2)^2$$

$$C = n \sum X_1 X_2 - \sum X_1 \sum X_2$$

$$D = n \sum X_2 Y - \sum X_2 \sum Y$$

$$E = n \sum (X_1)^2 - (\sum X_1)^2$$

$$F = EB - C^2$$

Dari persamaan-persamaan tersebut dicari nilai koefisien b_0 , b_1 dan b_2 sebagai berikut:

$$b_1 = \frac{AB - CD}{F}$$

$$b_2 = \frac{DE - AC}{F}$$

$$b_0 = \frac{\sum Y - b_1 \sum X_1 - b_2 \sum X_2}{n}$$

Dengan memasukkan nilai-nilai pada tabel maka diperoleh:

$$\begin{array}{lll} A=802572 & B=208824 & C=96033 \\ D=802596 & E=133116 & F=18575478495 \end{array}$$

$$b_1 = 4,8731 \quad b_2 = 1,6024 \quad b_0 = 123,9113$$

b. Koefisien determinasi:

$$\begin{aligned} R^2 &= \frac{b_1 \sum x_{1i} y_i + b_2 \sum x_{2i} y_i}{\sum y_i^2} \\ &= \frac{(4,8731 \times 226078) + (1,6024 \times 197882)}{2221264} = 0.6387 \end{aligned}$$

$$\text{Koefisien Non-determinasi: } 1 - (0,6387) = 0,3613$$

c. Uji F (Uji Simultan):

1. Menyusun Hipotesis (dugaan sementara).

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k = 0$$

$H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_k \neq 0$ (minimal ada 1 X yang mempengaruhi Y)

2. Menentukan Daerah Kritis (daerah penolakan H_0)

Untuk uji ini digunakan Tabel – F. Daerah kritis ditentukan oleh nilai Tabel – F dengan derajat bebas pembilang pada kolom yaitu ($k - 1$) dan derajat bebas penyebut pada baris yaitu ($n - k$) dengan taraf nyata (tingkat kesalahan) α , jika diketahui n = jumlah sampel dan k = banyaknya variabel bebas dan tidak bebas. Sedangkan taraf nyata α untuk uji 1 arah ini biasa dipakai 1 %, 5 % atau lainnya.

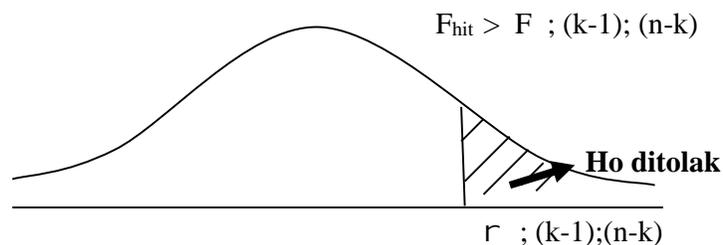
3. Menentukan Nilai F– hitung

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} = 10,6067$$

Keterangan : R^2 adalah koefisien Determinasi

4. Menentukan Daerah Keputusan

Menentukan wilayah H_0 dan H_1 serta membandingkan dengan nilai F–hitung untuk mengetahui apakah menerima H_0 atau menolak H_0



5. Menentukan Keputusan

Jika nilai F - hitung lebih besar dari nilai F – Tabel, maka terima H1 atau Tolak Ho. Berarti terdapat cukup bukti untuk menolak Ho dan menerima H1.

Kesimpulan: Variabel bebas secara simultan dapat menerangkan variabel tidak bebas, berarti model tersebut dapat digunakan.

Uji T (Uji Parsial):

1. Menentukan dugaan sementara (Hipotesis).

$$\begin{array}{ll} H_0 : \beta_k = 0 & H_0 : \beta_k = 0 \\ H_1 : \beta_k \neq 0 & H_1 : \beta_k \neq 0 \end{array}$$

2. Menentukan daerah kritis (wilayah ditolaknya H_0).

Daerah kritis ditentukan oleh nilai Tabel – t dengan derajat bebas yaitu (n-k), dan taraf nyata α , dimana n= jumlah sampel dan k = banyaknya variabel bebas dan tidak bebas. Sedangkan taraf nyata α yang biasa dipakai 1 %, 5 % atau lainnya. Dengan melakukan Uji 2 Arah, maka nilai t-hitung dapat dicari dengan menggunakan taraf nyata $\alpha / 2$.

3. Menentukan Nilai t – hitung.

$$\text{Nilai t hitung} = \frac{b_k - \beta_k}{Sb_k}$$

$$= \frac{b_k - 0}{Sb_k}$$

k adalah banyaknya variabel bebas

Dengan:

$$Sb_k = \frac{S_{Y.X_1.X_2...X_k}}{\sqrt{\left(\sum X_k^2 - n\bar{X}_k^2\right)\left(1 - r_{X_1X_2...X_k}^2\right)}}$$

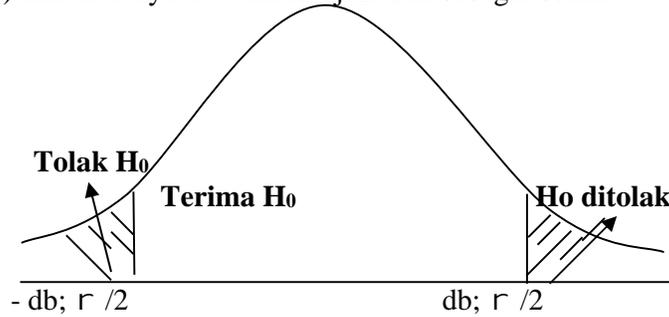
$$\text{Dan, } S_{Y.X_1X_2...X_k} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum X_1Y - b_2 \sum X_2Y - \dots - b_k \sum X_kY}{n - (k + 1)}}$$

Keterangan:

- Sb_k : Kesalahan Baku Penduga β_k
 $S_{Y.X_1X_2...X_k}$: Kesalahan Baku variabel Y berdasarkan variabel X yang diketahui
 $r_{X_1X_2...X_k}$: Koefisien Korelasi Sederhana antar X_1 dan X_2 atau X_k

4. Menentukan daerah keputusan.

Daerah keputusan untuk menerima H_0 atau menolak H_0 dengan derajat bebas (db) dan taraf nyata α untuk uji 2 arah sebagai berikut:



5. Menentukan Keputusan

Nilai t – hitung lebih besar atau lebih kecil dari nilai t – tabel, maka H_0 ditolak atau Terima H_1 . Berarti Ada satu variabel bebas berpengaruh nyata terhadap variabel tidak bebas Y .

Dari nilai koefisien parsial dapat diketahui bahwa variabel pertama (tenaga kerja) berpengaruh lebih signifikan dibandingkan variabel kedua (luas parkir mobil).

2. a. Persamaan umum regresi berganda dengan 2 variabel bebas:

$$Y = b_0 + b_1X_1 + b_2X_2 + e_i$$

NO	Y	X1	X2	YX1	YX2	YY	X1X1	X2X2	X1X2
1	96	6	20	576	1920	9216	36	400	120
2	83	22	30	1826	2490	6889	484	900	660
3	126	18	40	2268	5040	15876	324	1600	720
4	61	8	10	488	610	3721	64	100	80
5	59	12	10	708	590	3481	144	100	120
6	90	10	10	900	900	8100	100	100	100
7	82	17	30	1394	2460	6724	289	900	510
8	88	11	20	968	1760	7744	121	400	220
9	86	16	20	1376	1720	7396	256	400	320
10	76	23	30	1748	2280	5776	529	900	690
11	102	7	20	714	2040	10404	49	400	140
12	108	12	30	1296	3240	11664	144	900	360
13	96	24	40	2304	3840	9216	576	1600	960
14	70	16	20	1120	1400	4900	256	400	320
15	80	9	10	720	800	6400	81	100	90
16	113	11	30	1243	3390	12769	121	900	330
17	76	22	20	1672	1520	5776	484	400	440
18	74	11	10	814	740	5476	121	100	110
19	98	16	20	1568	1960	9604	256	400	320
20	80	8	20	640	1600	6400	64	400	160
jumlah	1744	279	440	24343	40300	157532	4499	11400	6770

$$\begin{aligned}
A &= n \sum X_1 Y - \sum X_1 \sum Y \\
B &= n \sum (X_2)^2 - (\sum X_2)^2 \\
C &= n \sum X_1 X_2 - \sum X_1 \sum X_2 \\
D &= n \sum X_2 Y - \sum X_2 \sum Y \\
E &= n \sum (X_1)^2 - (\sum X_1)^2 \\
F &= EB - C^2
\end{aligned}$$

Dari persamaan-persamaan tersebut dicari nilai koefisien b_0 , b_1 dan b_2 sebagai berikut:

$$\begin{aligned}
b_1 &= \frac{AB - CD}{F} \\
b_2 &= \frac{DE - AC}{F} \\
b_0 &= \frac{\sum Y - b_1 \sum X_1 - b_2 \sum X_2}{n}
\end{aligned}$$

Dengan memasukkan nilai-nilai pada tabel maka diperoleh:

$$\begin{aligned}
A &= -31108 \quad B = 34400 \quad C = 4720 \quad D = 38640 \\
E &= 12139 \quad F = 395303200
\end{aligned}$$

$$b_1 = -3,1684 \quad b_2 = 1,5580 \quad b_0 = 97,12318$$

b. Hitunglah koefisien korelasi

Koefisien Parsial

Koefisien korelasi parsial Y dan X_2 , jika X_1 konstan dapat dihitung dengan rumus :

$$r_{YX_2 \cdot X_1} = \frac{r_{YX_2} - r_{YX_1} \cdot r_{X_1X_2}}{\sqrt{(1 - r_{YX_1}^2)(1 - r_{X_1X_2}^2)}}$$

Koefisien korelasi parsial Y dan X_1 , jika X_2 konstan, dihitung dengan rumus :

$$r_{YX_1 \cdot X_2} = \frac{r_{YX_1} - r_{YX_2} \cdot r_{X_1X_2}}{\sqrt{(1 - r_{YX_2}^2)(1 - r_{X_1X_2}^2)}}$$

Sedangkan koefisien korelasi parsial X_1 dan X_2 , jika Y konstan dihitung dengan rumus :

$$r_{X_1X_2.Y} = \frac{r_{X_1X_2} - r_{YX_1}r_{YX_2}}{\sqrt{(1-r_{YX_1}^2)(1-r_{YX_2}^2)}}$$

keterangan:

r_{YX_1} = koefisien korelasi linier sederhana antara 2 variabel yaitu Y dan X_1

r_{YX_2} = koefisien korelasi linier sederhana antara 2 variabel yaitu Y dan X_2

$r_{X_1X_2}$ = koefisien korelasi linier sederhana antara 2 variabel yaitu X_1 dan X_2

Koefisien determinasi:

$$\begin{aligned} R^2 &= \frac{b_1 \sum x_{1i}y_i + b_2 \sum x_{2i}y_i}{\sum y_i^2} \\ &= \frac{(-3,1684 \times 24343) + (1,5580 \times 40300)}{157532} = -0,091 \end{aligned}$$

Koefisien Non-determinasi: $1 - (-0,091) = 1,091$

c. Kesalahan Baku/Standar Error dan Standar Error Estimationnya

$$\begin{aligned} S_{Y.X_1X_2,\dots,X_k} &= \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum X_1Y - b_2 \sum X_2Y - \dots - b_k \sum X_kY}{n - (k + 1)}} \\ &= 126,9635 \end{aligned}$$

d. Variabel yang paling berpengaruh signifikan adalah variabel rata-rata biaya pelayanan karena nilai mutlak koefisiennya paling besar.

3. a. buatlah persamaan regresinya:

$$Y = 662,151 - 5,4365X_1 - 20,0123X_2 + i$$

b. Hitung Koefisien Determinasi Regresi Linear Berganda dan jelaskan artinya!

$$R^2 = \frac{SSR}{SST} = 0,9656$$

Keterangan : $SSR = 228014,6263$

$SST = 236136,2293$

c. Berdasarkan model, buatlah nilai ramalan bagi y untuk $X_1 = 3$ dan $X_2 = 5$

$$\begin{aligned} Y &= 662,151 - 5,4365X_1 - 20,0123X_2 + \epsilon_i \\ &= 662,151 - 5,4365(3) - 20,0123(5) + \epsilon_i \\ \bar{Y} &= 545,78 \end{aligned}$$

d. Tentukan Standar Error/ Kesalahan baku dan Standar Error Estimationnya, lalu apa artinya!

$$\sqrt{\frac{SSE}{(n-(k+1))}} = \sqrt{\frac{81206030}{12}} = 26,0138$$

e. Berapakah nilai taksiran /penduga untuk parameter B_0, B_1 dan B_2 ?

$$\begin{aligned} b_0 &= 662,151 \\ b_1 &= -5,4365 \\ b_2 &= -20,0123 \end{aligned}$$

4. Adjusted R Squared = 0,7393

R Squared =

Multiple R =

Standard Error =

Lengkapilah isian titik-titik pada hasil print-out di atas!

	df	SS	MS	F	Signif icanc e F
Regression	<u>4</u>	21,4725	<u>5,3681</u>	<u>41,8147</u>	0,000 E+00
Residual	<u>55</u>	7,0608	<u>0,1284</u>		
Total	59	28,5333			
	Coefficient	Standar Error	t Stat		
Intercept	2,5026				
Motivasi	0,0050	0,0013	<u>3,8462</u>		
TP Ibu	0,1750	0,0195	<u>8,9744</u>		
TP Ayah	0,0411	0,181	<u>0,2271</u>		

a. Buatlah model dan persamaan regresi linier bergandanya!

$$Y = 2,5026 + 0,0050X_1 + 0,1750X_2 + 0,0411X_3 + \epsilon_i$$

b. Apakah ada pengaruh dari ketiga unsur yang diselidiki terhadap kemahiran membaca literatur bahasa Inggris?

Ada, karena signifikansi lebih kecil dari α .

g. Apakah ada korelasi linier yang nyata antara motivasi membaca, tingkat pendidikan ibu, ayah dan kemahiran membaca pada tingkat signifikansi $\alpha = 0,05$?

Daftar Pustaka

Anto Dayan (1990), Pengantar Metode Statistik Jilid II, Jakarta, LP3S.

Analisis Regresi, RK. Sembiring, Penerbit ITB 1995, Bandung.

J. Supranto, MA (2008), Statistik Teori & Aplikasi edisi ketujuh, Erlangga, Jakarta.

Pengantar Statistika, Ronald E. Walpole, Gramedia 1992, Jakarta.

Statistika untuk Ekonomi dan Keuangan Modern, Jilid 2, Suharyadi, Purwanto SK, Penerbit Salemba 4, Tahun 2004, Jakarta.

Lampiran.

Jawaban:

Bab 3

$$Y = -68,3185 + 1,450517X_1 + 9,344678X_2$$

Bab 4

1. a. $r_{yx_1x_2} = 0,8070601923$, $R^2 = 0,65134$

b. Interpretasi :

Dalam kasus ini, didapat korelasi antara pendapatan per bulan dan jumlah anggota keluarga terhadap pengeluaran per bulan adalah **0,8070601923**. Hal ini menunjukkan bahwa terjadi hubungan yang sangat kuat. Sedangkan arah hubungan adalah positif dilihat dari nilai r yang positif, artinya semakin tinggi pendapatan per bulan dan jumlah anggota keluarga maka semakin tinggi pengeluaran per bulan, dan semakin rendah pendapatan per bulan dan jumlah anggota keluarga maka semakin rendah pengeluaran per bulan

Nilai koefisien $r_{yx_1x_2}^2 = 65,13461541\%$ memberi makna bahwa naik turunnya (variasi) pengeluaran (Y) disebabkan oleh pendapatan (X1) dan jumlah anggota keluarga (X2) sebesar **65,14%** sedangkan sisanya sebesar 34,86% disebabkan oleh faktor-faktor lainnya yang juga turut mempengaruhi pengeluaran (Y) tetapi tidak dimasukkan ke dalam persamaan regresi linear berganda.

2 $r_{yx_1x_2} = -0,2651550303$, interpretasi: Dalam kasus ini, didapat korelasi antara pengaruh curah hujan terhadap pertumbuhan bibit kelapa sawit dimana dosis pemupukan dikendalikan (dibuat tetap) adalah -0,265. Hal ini menunjukkan bahwa terjadi hubungan yang rendah. Sedangkan arah hubungan adalah negatif dilihat dari nilai r yang negatif, artinya semakin tinggi intensitas curah hujan maka semakin lambat pertumbuhan bibit kelapa sawit, dan semakin rendah intensitas curah hujan maka semakin cepat pertumbuhan bibit kelapa sawit.

Bab 5

$$Sb_1 = 1,015821$$

$$Sb_2 = 0,403722$$

Bab 6

$$0,018 \quad 1 \quad 2,773$$

artinya diduga bahwa rata-rata berat akhir binatang (Y) naik sekitar antara 0,018 sampai 2,773 kg untuk setiap kenaikan satu kilogram berat awal binatang (X1) bila jumlah makanan tetap (X2).

Bab 7

Penyelesaian:

1) Hipotesis

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \text{tidak semua } \beta_k = 0; \quad k = 1, 2$$

2) Taraf nyata: = 5%

3) Daerah penolakan

Tolak H_0 jika $F_{\text{hitung}} > F_{\text{tabel}}(0,05; 2, 6-2-1)$

dengan $F_{\text{tabel}} = F(0,05; 2, 3) = 9,55$

4) Statistik uji

Untuk data soal di atas dilakukan perhitungan sebagai berikut.

	X_{i1}	X_{i2}	Y_i	X_{i1}^2	X_{i2}^2	$X_{i1}X_{i2}$	$X_{i1}Y_i$	$X_{i2}Y_i$	Y_i^2
1	7	33	42	49	1089	231	294	1386	1764
2	4	41	33	16	1681	164	132	1353	1089
3	16	7	75	256	49	112	1200	525	5625
4	3	49	28	9	2401	147	84	1372	784
5	21	5	91	441	25	105	1911	455	8281
6	8	31	55	64	961	248	440	1705	3025
Σ	59	166	324	835	6206	1007	4061	6796	20568

$$X = \begin{bmatrix} 1 & 7 & 33 \\ 1 & 4 & 41 \\ 1 & 16 & 7 \\ 1 & 3 & 49 \\ 1 & 21 & 5 \\ 1 & 8 & 31 \end{bmatrix} \quad Y = \begin{bmatrix} 42 \\ 33 \\ 75 \\ 28 \\ 91 \\ 55 \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & \Sigma X_{i1} & \Sigma X_{i2} \\ \Sigma X_{i1} & \Sigma X_{i1}^2 & \Sigma X_{i2}X_{i1} \\ \Sigma X_{i2} & \Sigma X_{i2}X_{i1} & \Sigma X_{i2}^2 \end{bmatrix} = \begin{bmatrix} 6 & 59 & 166 \\ 59 & 835 & 1007 \\ 166 & 1007 & 6206 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 34,57856 & -1,55089 & -0,65704 \\ -1,65089 & 0,080308 & 0,031128 \\ -0,65704 & 0,031128 & 0,012685 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_{i1} Y_i \\ \Sigma X_{i2} Y_i \end{bmatrix} = \begin{bmatrix} 324 \\ 4061 \\ 6796 \end{bmatrix}$$

$$b = (X'X)^{-1} X'Y$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 34,57856 & -1,65089 & -0,65704 \\ -1,65089 & 0,080308 & 0,031128 \\ -0,65704 & 0,031128 & 0,012685 \end{bmatrix} \begin{bmatrix} 324 \\ 4061 \\ 6796 \end{bmatrix} = \begin{bmatrix} 51,67219 \\ 1,944315 \\ -0,60691 \end{bmatrix}$$

$$b'xY = [51,67219 \quad 1,944315 \quad -0,60691] \begin{bmatrix} 324 \\ 4061 \\ 5796 \end{bmatrix} = 20529,45$$

$$SST = Y'Y - n\bar{Y}^2 = 20568 - 6(324/6)^2 = 3072$$

$$SSR = b'X'Y - n\bar{Y}^2 = 20529,45 - 6(324/6)^2 = 3033,452$$

$$SSE = SST - SSR = 3072 - 3033,452 = 38,54755$$

Tabel ANOVA

SOV	db	SS	MS	F _{hitung}
Regresi	2	3033,452	1516,726	118,0407
Error	3	38,54755	12,84918	
Total	5	3072		

$$F_{hitung} = 1516,726 / 12,84918 = 118,0407$$

5) Keputusan

Tolak H_0 karena $F_{hitung} > 9,55$

6) Kesimpulan

Dengan tingkat kepercayaan 95%, cukup bukti untuk mengatakan bahwa model regresi tersebut signifikan untuk data biaya promosi dan biaya after sales terhadap jumlah omzet penjualan.

Bab 8.

1. A
2. B
3. C
4. D
5. C