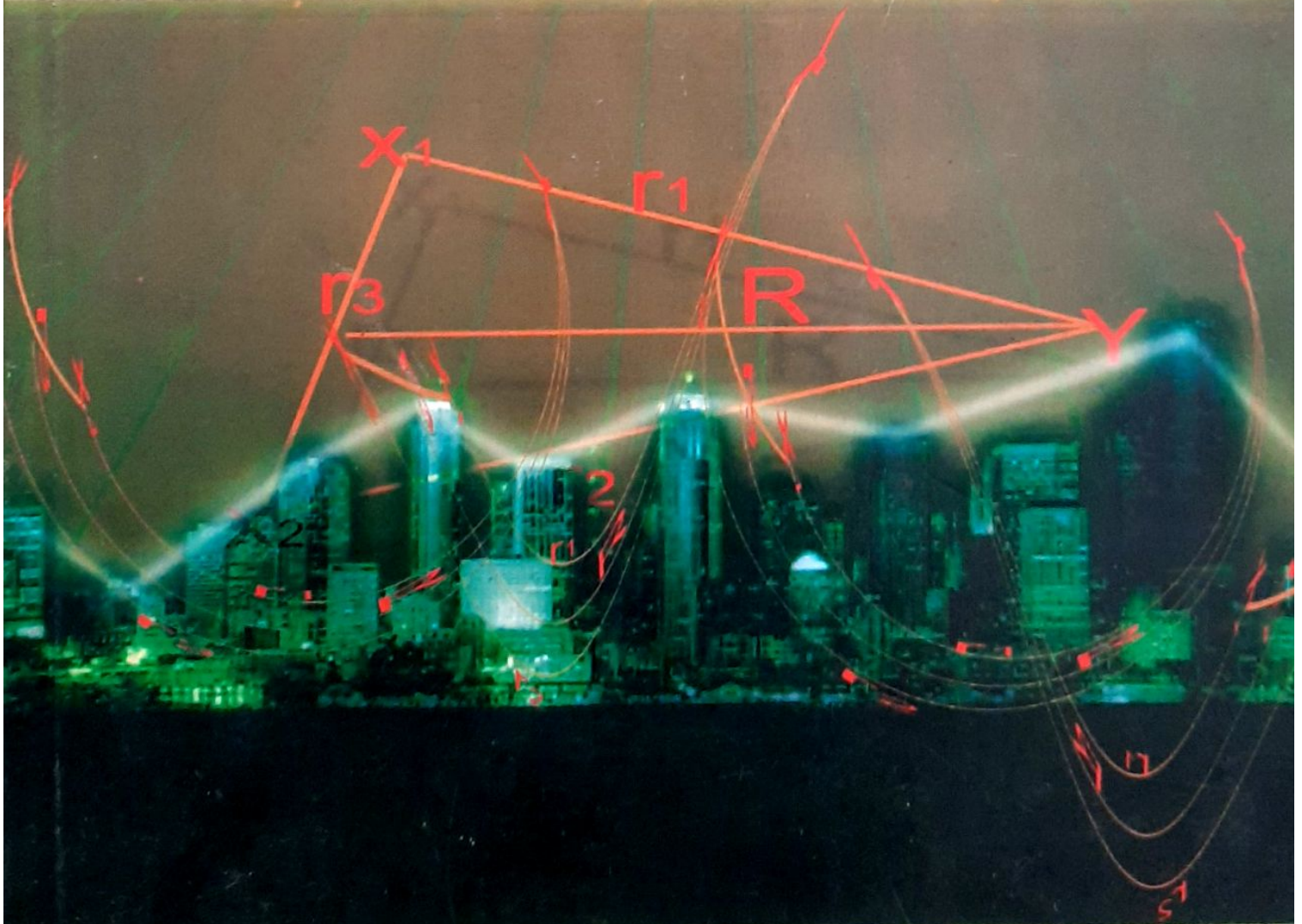


SALAFUDIN, S.Si, M.Si

STATISTIKA TERAPAN UNTUK PENELITIAN SOSIAL



STAIN PRESS
PEKALONGAN

STATISTIKA TERAPAN UNTUK PENELITIAN SOSIAL

Sanksi Pelanggaran Pasal 72:

Undang-Undang Nomor 19 Tahun 2002

tentang Perubahan atas Undang-Undang Nomor 12 Tahun 1997

tentang Hak Cipta

1. Barangsiapa dengan sengaja dan tanpa hak mengumumkan atau memperbanyak suatu ciptaan atau memberi izin untuk itu, dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp1.000.000,00 (satu juta rupiah) atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp5.000.000.000,00 (lima miliar rupiah).
2. Barangsiapa dengan sengaja menyebarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran Hak Cipta sebagaimana dimaksud dalam ayat (1), dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

STATISTIKA TERAPAN UNTUK PENELITIAN SOSIAL

ISBN : 979-3968-08-7

Penulis : Salafudin, S.Si, M.Si.

Editor : DR. H. Basukiyatno, M.Pd

Desain Cover : Wahyu Tri

Cetakan Pertama : 2005

KATA PENGANTAR

Statistik pada dasarnya adalah sebagai alat bantu untuk memberi gambaran atas suatu peristiwa melalui bentuk yang sederhana. Gambaran tersebut dapat berupa angka-angka atau berupa grafik dan diagram. Dalam penelitian yang menggunakan pendekatan kuantitatif, statistika berguna sebagai alat penentuan sampel penyajian data, dan untuk menganalisis data.

Buku ini merupakan statistika terapan (*applied statistics*), yaitu menyajikan bagaimana rumus-rumus statistika yang ada diterapkan, khususnya dalam penelitian pendidikan dan ilmu-ilmu sosial. Oleh karena itu dalam buku ini tidak dibahas bagaimana rumus-rumus statistika tersebut diperoleh. Sedapat mungkin digunakan bahasa dan diskripsi yang komunikatif.

Buku ini terdiri dari 18 Bab. Bab pertama berupa pendahuluan, yang mendeskripsikan tentang pengertian statistik, penggolongan, fungsi dan kegunaan statistik serta ciri khas statistik. Bab kedua tentang pengumpulan data statistik, populasi dan sampel, teknik sampling dan instrumen untuk mendapatkan data statistik.

Bab ketiga menguraikan tabel distribusi frekuensi, yang merupakan “jembatan” yang menghubungkan antara data mentah (*raw score*) dengan data yang akan dianalisis. Bab keempat diuraikan tentang pengolahan data statistik berisi tentang pengertian data, pengolongan dan sifat-sifat data. Juga diuraikan cara memilih teknik statistik. Selanjutnya bab kelima membahas tentang penyajian data statistik. Disini dibahas tentang grafik dan diagram yang merupakan visualisasi dari tabel distribusi frekuensi.

Statistik secara global di bedakan menjadi dua, yaitu statistik deskriptif dan Statistik Inferensia. Pada statistik deskriptif analisis dilakukan terhadap objek tanpa digeneralisasi terhadap objek yang lebih luas. Topik dalam statistik deskriptif, disamping masalah grafik dan diagram adalah ukuran tendensi sentral dan variabilitas data. Dua hal ini akan dibicarakan dalam bab keenam dan bab ketujuh.

Pada Statistik Inferensia, disamping mendeskripsikan suatu objek sampel, bisa dilakukan generalisasi terhadap populasi dari mana sampel diambil.

Hak cipta dilindungi undang-undang pada penerbit

STAIN Pekalongan Press

Jl. Kusumabangsa No 9 Pekalongan

Telp. 0285-412575, Faks. 0285-423418

E-mail: pipstainpk@yahoo.com

Inilah statistik yang banyak diaplikasikan dalam dunia penelitian, khususnya penelitian pendidikan dan ilmu-ilmu sosial.

Masalah hubungan (korelasi) antar variabel merupakan topik yang banyak menjadi perhatian dalam penelitian sosial. Teknik analisis korelasional ini diuraikan pada bab kedelapan sampai kelima belas. Pada bab kedelapan diuraikan secara global tentang analisis korelasional. Bab kesembilan sampai keempat belas mendeskripsikan secara detail Teknik Analisis Korelasional untuk Korelasi Bivariat. Bab kesembilan membahas teknik analisis korelasi produk momen. Bab kesepuluh menguraikan teknik analisis korelasi tata jenjang, bab kesebelas teknik analisis korelasi tau kendall, bab kedua belas teknik analisis korelasi phi, bab ketiga belas teknik analisis korelasi kontengensi dan bab keempat belas Teknik analisis korelasi poin biserial. Teknik Analisis Korelasi Multivariat dibahas dalam bab kelimabelas.

Berikutnya, pada bab keenam belas dibicarakan Analisis Varian (ANAVA). Selanjutnya, pada bab ketujuh belas dibahas analisis regresi.

Di samping masalah hubungan antarvariabel, topik lain yang lazim dalam penelitian pendidikan dan ilmu-ilmu sosial adalah masalah komparasi (perbandingan) antarvariabel. Teknik analisis komparasional ini dibahas dalam Bab kedelapan belas.

Selanjutnya, buku ini kami lengkapi dengan daftar tabel yang diperlukan dalam analisis data statistik, khususnya dalam pengujian hipotesis statistik.

Penulis mengharapkan saran dan kritik konstruktif pembaca. Semoga buku ini bisa melengkapi khasanah buku penelitian, terutama bidang sosial.

Penulis

DAFTAR ISI

Kata Pengantar — v

Daftar Isi — vii

BAB I PENDAHULUAN — 1

- 1.1. Pengertian dan Penggolongan Statistik — 1
- 1.2. Ciri Khas Statistik — 3
- 1.3. Fungsi dan Kegunaan Statistik — 4

BAB II PENGUMPULAN DATA STATISTIK — 11

- 2.1. Populasi (*Universe*) — 11
- 2.2. Sampel — 12

BAB III TABEL DISTRIBUSI FREKUENSI — 25

- 3.1. Analisis Data Berbasis Tabel Distribusi Frekuensi — 25
- 3.2. Pengertian Tabel Distribusi Frekuensi — 26
- 3.3. Macam-macam Tabel Distribusi Frekuensi — 27
- 3.4. Cara Membuat Tabel Distribusi Frekuensi — 33

BAB IV PENGOLAHAN DATA STATISTIK — 39

- 4.1. Pengertian Data Statistik — 39
- 4.2. Penggolongan dan Sifat-sifat Data Statistik — 39
- 4.3. Analisis Data Statistik — 44
- 4.4. Pedoman Memilih Teknik Statistik untuk Analisis Data — 47

BAB V PENYAJIAN DATA STATISTIK — 49

- 5.1. Pendahuluan — 49
- 5.2. Tabel Biasa — 49
- 5.3. Grafik — 50
- 5.4. Diagram — 51
- 5.5. Histogram dan Poligon — 53

BAB VI UKURAN TENDENSI SENTRAL — 55	
6.1. Pendahuluan — 55	
6.2. Rata-rata Hitung (Mean) — 55	
6.3. Median — 60	
6.4. Modus — 62	
6.5. Rata-rata Ukur — 64	
6.6. Rata-rata Harmonis — 64	
BAB VII UKURAN VARIABILITAS — 67	
7.1. Pendahuluan — 67	
7.2. Range — 68	
7.3. Deviasi — 69	
7.4. Deviasi Rata-rata — 70	
7.5. Variansi dan Standar Deviasi — 72	
7.6. Kegunaan Standar Deviasi — 75	
BAB VIII TEKNIK ANALISIS KORELASIONAL — 77	
8.1. Pendahuluan — 77	
8.2. Indeks Korelasi — 79	
8.3. Diagram Pencar — 80	
8.4. Macam-macam Teknik Analisis Korelasi — 81	
BAB IX TEKNIK ANALISIS KORELASI PRODUCT MOMENT — 83	
9.1. Pendahuluan — 83	
9.2. Indeks Korelasi dan Koefisien Determinasi — 83	
9.3. Rumusan Hipotesis — 84	
9.4. Interpretasi terhadap Nilai Indeks Korelasi — 84	
9.5. Interpretasi dan Pengujian Hipotesis dengan Menggunakan Uji t — 89	
BAB X TEKNIK ANALISIS KORELASI TATA JENJANG — 91	
10.1. Pendahuluan — 91	
10.2. Indeks Korelasi — 92	
10.3. Rumusan Hipotesis — 92	
10.4. Interpretasi terhadap Nilai Indeks Korelasi — 93	
10.5. Interpretasi dan Pengujian Hipotesis dengan Tabel z — 96	
10.6. Interpretasi dan Pengujian Hipotesis dengan Uji t — 97	

BAB XI TEKNIK ANALISIS KORELASI TAU KENDALL — 101	
11.1. Pendahuluan — 101	
11.2. Indeks Korelasi — 101	
11.3. Rumusan Hipotesis — 102	
11.4. Pengujian Hipotesis — 102	
BAB XII TEKNIK ANALISIS KORELASI PHI — 107	
12.1. Pendahuluan — 107	
12.2. Indeks Korelasi — 108	
12.3. Rumusan Hipotesis — 108	
12.4. Interpretasi terhadap Nilai Indeks Korelasi — 109	
12.5. Interpretasi dan Pengujian Hipotesis dengan Uji t — 112	
BAB XIII TEKNIK ANALISIS KORELASI KONTINGENSI — 115	
13.1. Pendahuluan — 115	
13.2. Indeks Korelasi — 116	
13.3. Rumusan Hipotesis — 116	
13.4. Interpretasi terhadap Nilai Indeks Korelasi — 117	
13.5. Interpretasi dan Pengujian Hipotesis dengan Chi Kuadrat — 119	
BAB XIV TEKNIK ANALISIS KORELASI POIN BISERIAL — 121	
14.1. Pendahuluan — 121	
14.2. Indeks Korelasi — 121	
14.3. Rumusan Hipotesis — 122	
14.4. Interpretasi terhadap Nilai Indeks Korelasi — 122	
BAB XV KORELASI MULTIVARIAT — 125	
15.1. Pendahuluan — 125	
15.2. Korelasi Ganda — 125	
15.3. Korelasi Parsial — 128	
BAB XVI ANALISIS VARIANSI — 131	
16.1. Pengertian — 131	
16.2. Analisis Variansi Satu Arah — 132	
16.3. Analisis Variansi Dua Arah — 136	
BAB XVII ANALISIS REGRESI — 145	
17.1. Pendahuluan — 145	
17.2. Regresi Linear Sederhana — 146	
17.3. Regresi Linear Berganda — 156	
17.4. Uji Regresi Linear Dengan Anava — 163	

BAB XVIII TEKNIK ANALISIS KOMPARASIONAL — 167

- 18.1. Pengertian — 167
- 18.2. Tes t — 168
- 18.3. Tes Chi Kuadrat — 174

LAMPIRAN:

- 1. Tabel I Luas di Bawah Lengkungan Kurva Normal — 178
- 2. Tabel II Nilai-Nilai Dalam Distribusi t — 180
- 3. Tabel III Nilai-Nilai r Product Moment — 182
- 4. Tabel IV Nilai-Nilai Chi Kuadrat — 183
- 5. Tabel V Tabel Nilai-Nilai Rho — 185
- 6. Tabel VI Harga-Harga Kritis Z Dalam Distribusi Normal — 186
- 7. Tabel VII Nilai-Nilai untuk Distribusi F — 188

DAFTAR PUSTAKA — 192

Kupersembahkan Kepada:
Ayah dan ibunda tercinta
Istri dan buah hatiku: shofwa, ilzam, izzuddin
Para sahabatku yang tengah terus berjuang di jalan-Nya

BAB I

PENDAHULUAN

1.1 Pengertian dan Penggolongan Statistika

Kata statistika berasal dari bahasa Italia, *statistita*, atau bahasa Latin *ratio status*. Kata ini mempunyai persamaan arti dengan kata *state* (bahasa Inggris) atau *staat* (bahasa Belanda), dan *statisteb* (bahasa Jerman). Kata-kata tersebut, yang di terjemahkan dalam bahasa Indonesia dengan negara, telah terdapat dalam karya William Shakespeare yang termashur, Hamlet (1602). Istilah statistika sendiri pertama kali di gunakan oleh Gottfried Achenwall (1719-1772) seorang profesor pada Marlborough dan Gottingen. Kemudian Zimmeman memperkenalkan istilah statistika di Inggris. Penggunaan statistika ini dipopulerkan oleh John Sinclair dalam pekerjaannya di Statistical Account of Scotland (1791-1799). Jadi jauh sebelum abad 18, umat manusia telah melakukan pencatatan dan penggunaan data.

Pada awalnya statistika diartikan sebagai kumpulan bahan keterangan (data), baik yang berwujud angka (data kuantitatif) maupun data yang tidak berwujud angka (data kualitatif), yang mempunyai arti bagi suatu negara. Metode statistika digunakan untuk mencatat sejarah, pelaksanaan sensus, pencatatan jumlah penduduk, pencacatan kekayaan teritorial dan pencatatan sumber daya yang dimiliki oleh suatu daerah kekuasaan teritorial (negara). Statistika juga di gunakan untuk mencatat pemilikan tanah, perkembangan penduduk (kelahiran dan kematian), perkawinan, penaksiran kematian yang disebabkan oleh suatu penyakit, dan sebagainya.

Dalam perkembangan selanjutnya, statistika hanya dibatasi pada keterangan yang berwujud angka (data kuantitatif) saja. Bahan keterangan yang tidak berwujud angka (data kualitatif) tidak dapat disebut sebagai statistika. Pada sisi lain, kata statistika tidak lagi digunakan dalam arti mengenai tugas-tugas kenegaraan, tetapi juga mempelajari segala keadaan yang terdiri dari data angka. Perkembangan statistika hingga sekarang ini begitu pesat dan dapat diterapkan hampir dalam semua aspek kehidupan.

Secara singkat, statistik secara terminologis dapat diartikan sebagai cara maupun aturan-aturan yang berkaitan dengan pengumpulan, pengolahan (analisis), dan penarikan kesimpulan, atas data yang berbentuk angka (data kuantitatif). Jadi, statistik merupakan suatu metode untuk mengumpulkan, mengolah, menganalisis dan menyimpulkan data kuantitatif. Berkaitan dengan hal ini statistik sering pula disebut sebagai metode statistik atau metode kuantitatif. Statistik dapat pula didefinisikan sebagai nilai yang didapat dari hasil pengolahan data kuantitatif. Dalam pengertian ini statistik dari sekumpulan data dapat berupa standar deviasi *mean*, *modus*, *median* atau lainnya. Di samping ada statistik, terdapat pula istilah *statistika* (bahasa Inggris: *statistics*) yang merupakan ilmu pengetahuan yang membicarakan tentang cara-cara mengumpulkan, mengolah, menganalisis dan menyimpulkan data kuantitatif.

Untuk mempelajari statistika ada dua jalan. Yang pertama melalui kajian statistika matematika atau statistika teoritis. Yang di bahas di sini adalah penurunan sifat-sifat, dalil-dalil, rumus-rumus dan segi-segi lainnya yang sifatnya teoritis matematis. Yang kedua adalah kajian statistika terapan atau statistika praktis (*applied statistics*) yang merupakan aplikasi statistika dalam berbagai aspek kehidupan dan ilmu pengetahuan. Jadi dalam statistika terapan, tidak dipersoalkan bagaimana didapatkannya rumus atau aturan, namun hanya di utamakan bagaimana mengaplikasikan aturan atau rumus tersebut dalam suatu bidang ilmu pengetahuan atau aspek kehidupan. Dengan demikian statistika praktis sifatnya instan. Buku ini merupakan statistik praktis.

Berdasarkan tingkat pekerjaannya statistik dibedakan menjadi dua yaitu:

1. Statistik deskriptif
2. Statistik inferensia

Statistik deskriptif disebut juga statistik sederhana, adalah statistik yang tingkat pekerjaannya mencakup cara-cara menghimpun, menyusun atau mengatur, mengolah dan menyajikan data angka agar dapat memberikan gambaran yang teratur, ringkas dan jelas mengenai suatu gejala, peristiwa atau keadaan. Statistik deskriptif disajikan dalam bentuk-bentuk tabel, diagram, histogram, poligon frekwensi, ogive, ukuran-ukuran penempatan (*median*, *kuartil*, *desil*, *persentil*), ukuran tendensi sentral (*rata-rata hitung*, *rata-rata ukur*, *rata-rata harmonik* dan *modus*), standar deviasi, angka baku.

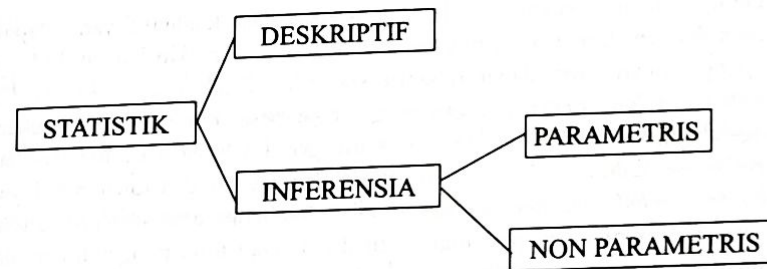
Sesuai dengan namanya, statistik deskriptif di gunakan untuk menggambarkan suatu gejala atau keadaan. Statistik deskriptif digunakan untuk menganalisis data penelitian, tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas (*generalisasi/inferensia*). Suatu penilaian yang tidak menggunakan sampel analisisnya akan menggunakan statistik deskriptif. Demikian juga penelitian yang menggunakan sampel, tetapi peneliti tidak

bermaksud membuat kesimpulan untuk populasi dari mana sampel diambil, maka statistik yang diambil adalah statistik deskriptif. Dalam hal ini teknik korelasi dan regresi juga dapat berperan sebagai statistik deskriptif.

Statistik inferensia itu disebut pula dengan statistik induktif atau statistik lanjut, adalah statistik yang menyediakan aturan-aturan penarikan kesimpulan (*conclusion*) yang bersifat umum dari sekumpulan data yang telah diolah. Statistik inferensia digunakan untuk menganalisis data sampel, dan hasilnya akan di generalisasikan (*diinferensikan*) untuk populasi dimana sampel diambil.

Ada dua macam statistik inferensia, yaitu: statistik parametrik dan statistik non parametrik. Statistik parametrik merupakan statistik yang digunakan untuk menganalisis data interval atau rasio dari populasi yang berdistribusi normal. Sedangkan statistik nonparametrik merupakan statistik yang di gunakan untuk menganalisis data ordinal atau nominal dari populasi yang distribusi, yaitu populasi yang berdistribusi apa saja, tidak harus berdistribusi normal.

Pembagian statistik dapat dilihat pada gambar 1.1 berikut:



Gambar 1.1 penggolongan statistik

1.2. Ciri Khas Statistika

Statistika merupakan salah satu bagian ilmu pengetahuan yang mempunyai ciri khas. Ciri tersebut ada 3, yaitu *kuantitatif*, *obyektif* dan *universal*.

1.2.1. Kuantitatif

Statistika selalu bekerja dengan angka (bilangan). Maka untuk melaksanakan tugasnya statistika selalu memerlukan keterangan yang bersifat kuantitatif. Jika statistik akan digunakan untuk menganalisis data kualitatif (data yang tidak berwujud angka/bilangan) maka data kualitatif tersebut lebih dahulu harus di kuantitatifkan. Proses transformasi dari data kualitatif menjadi data kuantitatif disebut proses *kuantifikasi*. Proses tersebut dilakukan dengan memberi skor kepada nilai kualitatif.

1.2.2 Obyektif

Statistika bersifat obyektif. Artinya, statistika bekerja menurut obyeknya, atau bekerja menurut apa adanya. Kesimpulan dan ramalan yang dihasilkan dari statistika, semata-mata berdasarkan pada angka yang diolah, bukan didasarkan pada subyektifitas atau pengaruh eksternal. Dengan kata lain statistika bekerja secara jujur, tanpa dipengaruhi oleh interes tertentu. Dengan sifat demikian, statistika dapat digunakan sebagai "alat penilaian kenyataan yang terjadi".

1.2.3. Universal

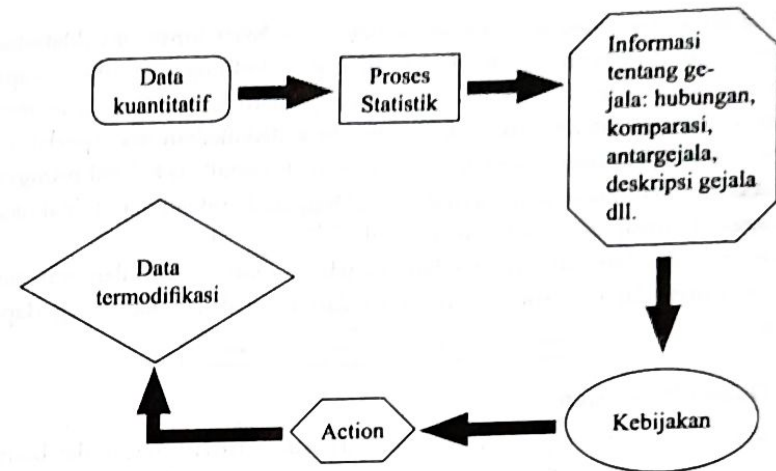
Statistika mempunyai ruang lingkup yang luas. Statistika dapat dipergunakan hampir semua cabang kehidupan. Beberapa yang dapat disebutkan misalnya: dalam bidang ekonomi terdapat statistik perdagangan, statistik ekspor-impor, statistik pertanian dan sebagainya. Dalam bidang kependudukan terdapat: statistik kelahiran dan kematian, statistik nikah-tolak-cerai-rujuk (NTCR), statistik kedatangan dan kepergian (imigrasi dan emigrasi). Dalam bidang hukum kita mengenal: statistik kriminalitas dan statistik kecelakaan lalu lintas. Dalam bidang pendidikan terdapat statistik pendidikan.

Dengan adanya kuantitatif, banyak gejala-gejala kualitatif yang dapat dianalisis dengan statistika. Seiring dengan itu, maka semakin banyak bidang ilmu pengetahuan yang dapat dibantu statistika. Sejak perang Dunia II, statistika mengalami perkembangan yang sangat pesat hingga memunculkan cabang-cabang baru dari statistika. Untuk mengoptimalkan suatu hasil usaha, dan resiko seminimal mungkin, dalam hal pengaturan distribusi misalnya, muncul *program linier* dan *operation reseach*. Dalam hal pengaturan antrian muncul *teori antrian*. sementara itu pertautan statistika dengan ilmu pengetahuan lain melahirkan cabang-cabang ilmu pengetahuan baru, antarlain: *Ekonometri* (pertautan antara psikologi dan statistika), *Biometri* (pertautan antara biologi dengan statistika) dan *Psikometri* (pertautan antara psikologi dengan statistika).

1.3. Fungsi dan Kegunaan Statistika

Pada hakikatnya statistik merupakan alat bantu yang dapat digunakan pada banyak bidang kehidupan. Statistik dapat digunakan pada dunia pendidikan, penelitian, bisnis dan perbankan, administrasi dan manajemen, dan dapat pula digunakan oleh para ilmuwan pada umumnya, bahkan masyarakat pun secara sadar atau tidak sadar menggunakan statistik.

Secara Diagramik, posisi proses statistik dapat dilihat pada diagram berikut:



Gambar 1.2. Posisi Statistik

Kegunaan statistik pada tiap aspek kehidupan akan diuraikan dalam uraian berikut:

1.3.1. Dalam Dunia Pendidikan

Dalam pendidikan statistik sebagai alat bantu untuk mengolah, menganalisis dan menyimpulkan hasil yang telah dicapai dalam kegiatan penilaian hasil proses belajar mengajar. Dalam penilaian tersebut, umumnya digunakan bilangan (data kuantitatif). Dengan data kuantitatif tersebut dapat dilihat:

- gambaran (deskripsi), baik gambaran secara khusus maupun umum tentang suatu gejala, keadaan atau peristiwa. Gambaran tersebut dapat berupa tabel, grafik diagram ataupun ringkasan data numerik.
- perkembangan atau pasang surut mengenai suatu gejala, keadaan atau peristiwa dari waktu ke waktu. Dengan demikian dapat dilihat *trend* (kecenderungan) yang terjadi.
- komparasi antara suatu gejala yang satu dengan gejala lainnya. Apakah suatu gejala dengan gejala lainnya terdapat perbedaan atau tidak. Bila terdapat perbedaan, apakah perbedaan tersebut cukup signifikan atau hanya kebetulan saja. Dengan melihat ada atau tidaknya perbedaan tersebut maka bisa diambil pilihan tindakan atau perlakuan (*treatment*) yang dapat sesuai dengan keadaan yang diharapkan.
- korelasi antara suatu gejala dengan gejala lainnya, apakah suatu gejala berhubungan dengan gejala lainnya. Jika berhubungan, apakah hubungan tersebut signifikan atau tidak, kuat atau lemah. Dengan melihat ada atau tidak adanya hubungan, dapat dilakukan suatu tindakan yang tepat.

- e. Pola hubungan antara suatu gejala dengan gejala lainnya, ini dilakukan dengan analisis regresi. Dengan melihat pola hubungan tersebut, dapat dilihat bobot suatu faktor (gejala) terhadap suatu faktor (gejala) lainnya. Dapat pula dilakukan perkiraan kejadian bila dilakukakan suatu perlakuan (*treatment*). Bila faktornya waktu, dapat pula di ramalkan hal-hal mungkin bakal terjadi di masa yang akan datang. Dengan demikian dapat dilakukan langkah antisipasi yang harus diambil.
- f. Kesimpulan dari data yang telah dianalisis. Dari kesimpulan tersebut, selanjutnya dapat diambil keputusan dan langkah-langkah yang dapat diambil.

1.3.2. Dalam Penelitian

Dalam konteks penelitian, kedudukan statistik adalah sebagai alat bantu. Urgensi penggunaan statistik tergantung pada jenis penelitian dan akurasi pemakaiannya, tergantung pada pemakainya. Maka anggapan yang menganggap bahwa, Statistik sebagai alat statistik yang paling tepat, maupun anggapan bahwa tanpa statistik maka penelitian tidak dapat dipertanggungjawabkan, harus dibuang jauh-jauh. Sebaliknya bahwa anggapan statistik tidak banyak memberi manfaat dalam pengembangan ilmu pengetahuan, tidak juga dapat diterima.

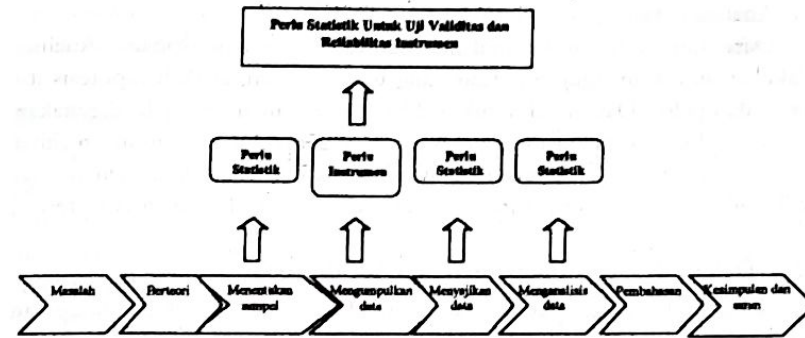
Penelitian merupakan upaya sistematis dalam menemukan, menganalisis, dan menafsirkan bukti-bukti empiris untuk memahami gejala atau untuk menemukan jawaban terhadap suatu permasalahan yang berkaitan dengan permasalahan tersebut. Penelitian ada yang bersifat kualitatif maupun kuantitatif. Penelitian kualitatif merupakan penelitian yang dalam analisisnya menggunakan metode kualitatif, bukan kuantitatif. Penelitian kuantitatif dalam analisisnya menggunakan metode kuantitatif. Pada metode kuantitatif, suatu gejala diukur dan dikuantitatifkan. Metode ini terbukti efektif untuk meneliti fenomena alam, yang pada perkembangan selanjutnya digunakan pula untuk meneliti gejala-gejala sosial.

Penelitian kuantitatif menggunakan teknik statistik. Statistik bekerja dengan angka-angka, oleh karenanya akan memaksa seorang pemakai statistik untuk terlibat dengan permainan angka-angka. Meskipun demikian anggapan bahwa statistik merupakan suatu yang sulit adalah tidak benar. Sebab pada dasarnya statistik adalah ilmu yang mudah dipelajari. Terlebih pada saat ini banyak alat bantu yang banyak menolong, mulai dari kalkulator, hingga *software* komputer yang siap pakai, antara lain SPSS, minitab, *microstat* bahkan *exell* pun menyimpan menu statistik yang cukup lengkap.

Penelitian dilakukan dengan cara sistematis. Suharsimi Arikunto (1998:16-17) menyatakan ada sebelas langkah dalam penelitian, yaitu: 1) memilih masalah, 2) studi pendahuluan, 3) merumuskan masalah, 4) merumuskan

hipotesis, 5) memilih pendekatan, 6) menentukan variabel dan sumber data, 7) menemukan dan menyusun instrument penelitian, 8) mengumpulkan data, 9) menarik kesimpulan, dan 10) menulis laporan.

Statistik digunakan dalam langkah-langkah penelitian menurut Sugiyono (1999, 13) kesertaan statistik dapat dinyatakan dalam gambar berikut:



Gambar 1.3. Proses penelitian dan statistik yang diperlukan

Dari diagram tersebut terlihat bahwa peranan statistik dalam penelitian adalah sebagai berikut:

1. alat untuk menentukan besarnya anggota sampel yang diambil dari populasi, agar jumlah sampel yang diperlukan dapat dipertanggungjawabkan.
2. alat-alat untuk menguji validitas dan reliabilitas instrument.
3. teknik-teknik untuk menyajikan data
4. alat untuk analisis data seperti menguji hipotesis penelitian yang diajukan.

Uraian masing-masing adalah sebagai berikut:

1. Penentuan sampel

Untuk menentukan sampel yang akan digunakan dalam penelitian terdapat berbagai teknik sampling. Secara garis besar ada dua teknik sampling, yaitu *probability sampling* dan *nonprobability sampling*. Peranan statistik adalah pada penentuan jumlah sampel harus dipenuhi agar dari penelitian terhadap sampel dapat dilakukan generalisasi terhadap populasi diambil darimana sampel diambil. Statistik juga mengatur bagaimana pengambilan sampel dilakukan agar dapat dipertanggungjawabkan secara ilmiah.

2. Pengujian validitas dan reliabilitas instrumen

Dalam penelitian kuantitatif, kualitas pengumpulan data sangat ditentukan oleh kualitas instrumen atau alat pengumpulan data yang digunakan. Instrumen tersebut dikatakan berkualitas dan dapat dipertanggungjawabkan bila sudah dibuktikan validitas dan reliabilitasnya.

3. Penyajian data

Agar mudah dipahami, data hasil penelitian perlu disajikan. Dengan penyajian data, data lebih komunikatif sehingga mudah dibaca dan difahami. Teknik-teknik penyajian itu antara lain dengan menggunakan table, grafik, diagram dengan berbagai bentuk.

4. Analisis data

Data yang telah terkumpul selanjutnya diolah dan dianalisis. Analisis dilakukan untuk menguji hipotesis yang telah disusun, apakah hipotesis itu benar atau palsu. Dalam hal untuk melihat perbedaan antar gejala, digunakan teknik analisis komparasional. Sedang bila dimaksudkan untuk melihat hubungan antara gejala, digunakan teknik analisis korelasional. Bila ingin melihat lebih jauh pola hubungan antara gejala tersebut, digunakan analisis regresi.

1.3.3. Dalam Bidang Manajemen dan Administrasi

Statistik digunakan para manajer dan administrator untuk pengumpulan data, penyajian data, pengolahan data, bahan evaluasi dan sebagai dasar pengambilan kebijakan maupun untuk memecahkan masalah (*problem solving*). Bagi manajer profesional, pengambilan keputusan dengan berbasis data merupakan suatu keniscayaan. Dengan demikian kebijakan dan pemecahan masalah dapat dilakukan secara akurat dan mantap.

1.3.4. Dalam Bidang Bisnis dan Perbankan

Dalam dunia bisnis dan perbankan kegunaan statistik sangat banyak. Bidang bisnis dan perbankan antara lain mencakup permasalahan *accounting, economic, finance, managemen dan marketing*. Bagaimana statistik dapat digunakan dalam persoalan bisnis, dapat diilustrasikan dalam paparan berikut.

1) Akutansi

Banyak keputusan yang dibuat oleh akuntan tentang status keuangan, likuiditas dan persediaan yang didasarkan pada analisis rasio keuangan. Akuntan menggunakan metode statistik untuk menganalisis kondisi keuangan perusahaan. Analisis keuangan dapat menunjukkan sehat tidaknya rasio keuangan perusahaan. Manager, karyawan dan investor perusahaan sangat berkepentingan dengan hasil analisis ini, sebab perusahaan-perusahaan yang memiliki rasio keuangan yang tidak sehat kemungkinan besar akan mengalami kebangkrutan.

2) Ekonomi

Gejala-gejala ekonomi, baik mikro maupun makro, berupa produktifitas ekonomi, laju inflasi, tingkat bunga maupun pengangguran merupakan obyek

pengamatan, kajian dan analisis ekonomi (ahli ekonomi). Mereka menggunakan metode statistika untuk mendapatkan berbagai angka indeks, seperti indeks harga konsumen atau indeks harga produsen. Mereka juga menggunakan metode statistika untuk mengukur laju inflasi dari waktu ke waktu. Untuk melakukan peramalan, ekonomi maka digunakan analisis runtun waktu (*time series analysis*), di samping analisis regresi untuk menganalisis dan menaksir perekonomian pada masa yang akan datang.

3) Keuangan

Investor (pemodal) dan manajer umumnya menginvestasikan dananya dengan membeli surat berharga. Mereka ingin mendapatkan penerimaan keuntungan yang maksimal dengan resiko yang seminimal mungkin. Untuk itu investor tidak membeli surat berharga sejenis saja. Mereka tidak ingin "meletakkan telur pada satu macam keranjang saja". Prinsip statistika dapat digunakan untuk mengukur penerimaan yang diharapkan (*expected return*), dan resiko yang memungkinkan harus ditanggung dapat diukur melalui ukuran statistika mengenai variabilitas atau dispersi.

4) Manajemen

Konsumen ingin mendapatkan barang dan jasa dengan kualitas setinggi mungkin, dengan harga yang serendah mungkin. Dalam era pasar global, tingkat persaingan ketat. Untuk dapat bersaing secara efektif, manajer harus selalu menggunakan inovasi untuk meningkatkan mutu dan produktifitasnya. Metode statistik dapat digunakan untuk melakukan pengendalian mutu. Metode pengendalian mutu terpadu (*total quality control*) secara statistik digunakan oleh perusahaan-perusahaan modern untuk mengelola dan secara terus-menerus mengembangkan proses perusahaan.

5) Pemasaran

Pemasaran merupakan suatu yang penting dalam bisnis. Produksi yang melimpah tanpa diikuti dengan pemasaran yang setimpal, akan mematikan bisnis. Untuk itu manager perlu mengetahui beberapa hal yang berkaitan dengan pemasaran produknya, terlebih bila akan meluncurkan produk baru. Manager harus mengetahui tentang hal-hal berikut: berapa besar pasar untuk suatu produk baru? Keuntungan apa yang ada di produk baru? Apa karakter, sikap, minat, dan opini konsumen yang tertarik terhadap suatu produk? Untuk menjawab pertanyaan-pertanyaan tersebut perlu dilakukan penelitian pasar, dan untuk itu perlu statistika.

1.3.5. Bagi Pembimbing Penelitian, Penguji Skripsi/Tesis dan Disertasi.

Bagi seorang pembimbing penelitian, bila penelitiannya menggunakan metode kuantitatif, pemahaman dan penguasaan statistika merupakan suatu tuntutan, agar ia dapat memberi bimbingan secara baik dan benar. Baik artinya ia bisa memberi arahan secara tepat, dan benar maksudnya arahan yang diberikan sesuai dengan kaidah dan dapat dipertanggungjawabkan secara ilmiah.

Tuntutan untuk memahami dan menguasai statistika juga berlaku bagi penguji skripsi/tesis dan disertasi. Hal ini dimaksudkan agar dapat melakukan penilaian secara benar dan obyektif. Dengan demikian akan dihasilkan alumnus yang berkualitas.

1.3.6. Bagi Intelektual/Pembaca/Pengguna Informasi

Statistika berguna untuk membantu memahami diagram, grafik atau tabel yang sering disajikan dalam buku, artikel atau pun sajian informasi di surat kabar, majalah, dan televisi. Dengan adanya pemahaman dan penguasaan statistika, minimal dasar-dasarnya, para intelektual, pembaca, pengguna informasi dapat mengerti maksud dari grafik, diagram atau tabel-tabel yang dijumpai. Dengan demikian pemahaman tentang informasi yang dibaca dapat utuh.

BAB II PENGUMPULAN DATA STATISTIK

Ada dua teknik yang digunakan untuk mengumpulkan data statistik, dengan sensus (*population research*) atau survei (*sample research*). Sensus adalah pengumpulan data dimana setiap unit anggota populasi diteliti. Jadi dilakukan pencacahan lengkap. Sedangkan survei (*sampling*) adalah cara pengumpulan data yang hanya meneliti sebagian saja dari unit anggota populasi. Pencacahan dilakukan terhadap sampel yang terpilih.

2.1. Populasi (*Universe*)

Populasi adalah wilayah generalisasi yang terdiri atas obyek yang akan diteliti yang didefinisikan dengan jelas, dengan karakteristik dan kuantitas tertentu. Populasi bukan hanya orang, tetap juga benda-benda hidup maupun mati yang ada. Populasi juga bukan hanya jumlah yang ada dalam obyek pengamatan, melainkan juga meliputi karakter/sifat yang ada pada obyek tersebut.

Misalnya akan dilakukan penelitian di SMA X, maka SMA X merupakan populasi. Sekolah tersebut mempunyai sejumlah orang, hal ini berarti SMA X merupakan populasi dalam arti jumlah. Tetapi SMA X juga mempunyai karakteristik tertentu pada orang-orangnya. Misalnya efektivitas mengajar gurugurunya, kedisiplinan siswa maupun gurunya (obyek), gaya kepemimpinan kepala sekolahnya, dan lain sebagainya. Yang demikian merupakan populasi dalam arti karakteristik.

Satu orang pun dapat digunakan sebagai populasi, karena satu orang itu mempunyai berbagai karakteristik. Misalnya, misalnya gaya bicara, hobi, cara bergaul, kepemimpinan dan lain-lain. Atau seorang siswa mempunyai catatan prestasi belajar dalam suatu mata pelajaran tertentu, dari semester I sampai VI. Catatan prestasi tersebut merupakan populasi.

Proses pengambilan data dari seluruh obyek pada populasi disebut sensus atau penelitian populasi (*population research*). Penelitian yang demikian biasanya sangat kompleks dan membutuhkan waktu, tenaga, dan biaya yang sangat besar. Di samping itu tidak dapat dilakukan pengamatan secara mendalam. Namun sensus mempunyai kelebihan, antara lain: dapat diketahui gambaran yang sebenarnya dari suatu populasi serta tidak mempunyai *sampling error*. Secara lengkap, kelebihan dan kekurangan sensus ditampilkan pada tabel berikut.

Tabel 2.1.
Kelebihan dan Kekurangan Sensus

No	KELEBIHAN	KEKURANGAN
1.	Dapat diketahui gambaran sebenarnya dari suatu populasi	Biaya, waktu dan tenaga yang dibutuhkan sangat besar
2.	Dapat diperoleh kerangka sampel (<i>sample frame</i>) yang berguna untuk survei	Kesalahan dari petugas (<i>nonsampling error</i>) sulit diperkirakan
3.	Tidak mempunyai <i>sampling error</i> (kesalahan karena pengambilan sampel)	Jenis data yang diperoleh terbatas dan sifatnya sederhana (tidak mendalam)

2.2. Sampel

Sampel merupakan sebagian dari jumlah karakteristik yang dimiliki populasi. Dengan kata lain sampel merupakan himpunan bagian dari populasi. Apa yang dipelajari dalam sampel, kesimpulannya dapat diberlakukan untuk populasi. Dengan kata lain sifat-sifat sampel dapat digeneralisasi untuk populasi.

Penelitian yang dilakukan terhadap sampel disebut penelitian sampel (*sample research*) atau survei. Penelitian sampel dilakukan disebabkan adanya kendala, misalnya adanya populasi yang sangat kompleks sehingga sulit didefinisikan, adanya kendala biaya, waktu serta tenaga. Dengan alasan inilah, penelitian sampel sering digunakan. Di samping itu, penelitian sampel sering dipilih karena terhadap obyek yang kecil dapat dilakukan pengamatan yang lebih teliti dan mendalam. Tetapi penelitian sampel mempunyai kekurangan, antara lain: gambaran tentang populasinya hanya merupakan taksiran, bukan merupakan gambaran yang sebenarnya, di samping itu diperlukan kerangka sampel (*sample frame*), dan sering kali metode pengambilan sampel tidak tepat, sehingga tidak dapat dihindarkan terjadinya kesalahan. Tabel 2.2. berikut menampilkan kelebihan dan kekurangan penelitian sampel.

Tabel 2.3
Kelebihan dan Kekurangan Penelitian Sampel (Survei)

No	Kelebihan	Kekurangan
1.	Biaya, waktu dan tenaga yang dibutuhkan jauh lebih sedikit	Gambaran tentang populasinya hanya merupakan taksiran
2.	Kesalahan dari <i>sampling error</i> dapat diukur	Memerlukan kerangka sampel
3.	Karakteristik/jenis data yang tercakup lebih banyak dan terinci	Metode pengambilan sampel kurang tepat

2.2.1. Teknik Sampling

Secara teoritis, hasil penelitian sampel dapat digeneralisasi untuk populasi sepanjang telah ditempuh prosedur yang benar. Salah satu prosedur yang harus dilakukan dalam penelitian sampel adalah teknik sampel, yakni kaidah-kaidah dalam menentukan besar sampel dan obyek yang menjadi sampel. Teknik pengambilan sampel dalam suatu penelitian perlu menggunakan kaidah yang bisa dipertanggungjawabkan secara ilmiah, atau lebih khusus secara statistik. Apabila salah dalam pengambilan sampel, maka ada kemungkinan hasil generalisasi atau penarikan kesimpulan untuk populasi menjadi tidak akurat.¹ Teknik Sampling meliputi dua hal, yaitu teknik penentuan ukuran sampel dan teknik pengambilan sampel.

2.2.1.1. Teknik Penentuan Ukuran Sampel

Ukuran sampel adalah jumlah sampel minimal yang harus diambil dari populasi agar sampel representatif. Penentuan ukuran sampel tergantung pada populasinya, yakni dilihat dari homogenitas populasi dan besar populasi. Bila populasi sangat homogen dan ukurannya kecil, maka dapat diambil sampel dengan ukuran kecil. Sebaliknya bila populasi sangat heterogen atau ukurannya besar, maka diperlukan populasi dalam ukuran besar.

Para ahli mengemukakan bermacam-macam cara dalam menentukan ukuran sampel, berikut paparnya.

1. Sebagai contoh, pada tahun 2002 kita dikejutkan oleh penelitian di Yogyakarta yang menyatakan 96% mahasiswa Yogyakarta telah melakukan hubungan seksual. Setelah ditelusuri, ternyata peneliti dalam mengambil sampel tidak menggunakan teknik pengambilan sampel yang benar. Ia tidak obyektif dalam mengambil sampel. Pengambilan sampel diarahkan pada obyek yang mendukung hipotesisnya.

a. Pendapat Slovin

Menurut Slovin, jumlah minimal yang dapat diambil agar sampel representatif terhadap populasi adalah:

$$n = \frac{N}{1+Nc^2}$$

dimana n = ukuran sampel (jumlah sampel)

N = ukuran populasi

c = persen kelonggaran ketidakteelitian karena kesalahan pengambilan sampel yang masih dapat ditolerir, misalnya 2%.

Pemakaian rumus di atas mempunyai asumsi bahwa populasi berdistribusi normal.

Contoh 2.1:

Akan dilakukan penelitian tentang minat siswa SMA/MA/SMK di wilayah eks Karesidenan Pekalongan terhadap STAIN Pekalongan. Berdasarkan informasi, jumlah siswa SMA/MA/SMK di seluruh wilayah eks Karesidenan Pekalongan adalah 10.000 siswa. Agar penelitian tersebut menghemat biaya, waktu dan tenaga, maka dilakukan penelitian sampel. Kesalahan yang diinginkan sebesar 3%. Maka sampel yang diambil berjumlah:

$$n = 10.000 / (1 + 10.000 \times 0,03^2) = 1.000 \text{ siswa}$$

Jadi jumlah sampel minimal yang harus diambil untuk penelitian tersebut adalah 1.000 siswa.

b. Pendapat Gay

Menurut Gay, ukuran minimum sampel yang dapat diterima berdasarkan desain penelitian yang digunakan, yaitu:

- a. Untuk penelitian deskriptif, minimal 10 % populasi. Untuk populasi relatif kecil minimal 20 % populasi.
- b. Untuk penelitian menggunakan teknik analisis korelasional minimal 30 obyek.

Contoh 2.2:

Seorang peneliti ingin mengetahui tingkat kesadaran menjalankan ibadah pada penduduk suatu desa. Untuk tujuan tersebut, ia melakukan pengamatan terhadap penduduk desa tersebut. Bila jumlah penduduk desa 1.000, maka sampel yang harus diambil sebanyak $1.000 \times 10\% = 100$ penduduk.

c. Pendapat Kracjie

Pendapat Kracjie hampir sama dengan Slovin, hanya Kracjie dalam melakukan perhitungan ukuran sampel didasarkan atas kesalahan 5%. Jadi sampel yang diperoleh itu mempunyai kepercayaan 95% terhadap populasi. Kracjie selanjutnya membuat tabel seperti ditunjukkan dalam tabel 2.3, sehingga kita tidak perlu melakukan perhitungan yang rumit.

Dari tabel tersebut, bila jumlah populasi 100, maka sampelnya 80. Bila populasi 1.000 maka ukuran sampelnya 278, dan bila populasinya 10.000 maka ukuran sampelnya 370. Dengan demikian makin besar populasi semakin kecil persentasi sampel. Oleh karena itu kurang tepat bila ukuran populasinya berbeda dengan persentasi yang sama, misalnya selalu 10%.

Tabel 2.3

Tabel Kracjie

N	S	N	S	N	S	N	S	N	S
10	10	100	80	280	162	800	260	2.800	338
15	14	110	86	290	165	850	265	3.000	341
20	19	120	92	300	169	900	269	3.500	346
25	24	130	97	320	175	950	274	4.000	351
30	28	140	103	340	181	1.000	278	4.500	354
35	32	150	108	360	186	1.100	285	5.000	357
40	36	160	113	380	191	1.200	291	6.000	361
45	40	170	118	400	196	1.300	297	7.000	364
50	44	180	123	420	201	1.400	302	8.000	367
55	48	190	127	440	205	1.500	306	9.000	368
60	52	200	132	460	210	1.600	310	10.000	370
65	56	210	135	480	214	1.700	313	15.000	375
70	59	220	140	500	217	1.800	317	20.000	377

75	63	230	144	550	226	1.900	320	30.000	379
80	66	240	148	600	234	2.000	322	40.000	380
85	70	250	152	650	242	2.200	327	50.000	381
90	73	260	155	700	248	2.400	331	75.000	382
95	76	270	159	750	254	2.600	335	100.000	384

2.2.1.2. Teknik Pengambilan Sampel

Teknik Pengambilan sampel pada dasarnya dikelompokkan menjadi dua yaitu *Probability sampling* dan *Non Probability sampling*. *Probability sampling* meliputi *simple random*, *proportionate stratified random*, *disproportionate stratified random* dan *area random*. *Non probability sampling* meliputi *sampling sistematis*, *sampling kuota*, *sampling aksidental*, *purposive sampling*, *sampling jenuh* dan *snowball sampling*. (Sugiyono, 1999: 57).

1. *Probability Sampling (Random Sampling)*

Probability sampling adalah teknik pengambilan sampel yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel (Saleh, 1988: 15). Metode ini dianggap merupakan metode yang terbaik, karena peneliti terbebas dari subyektifitas. Generalisasi hasil penelitian sampel terhadap populasi bisa lebih dipertanggungjawabkan. *Probability sampling* meliputi:

a. *Simple Random Sampling*

Dikatakan *simple* (sederhana) karena pengambilan sampel anggota populasi sepenuhnya dilakukan *random*, tanpa memperhatikan strata yang ada dalam populasi itu. Cara demikian bila anggota populasi homogen atau dianggap homogen. Teknik ini dapat digambarkan dalam gambar 2.1 berikut:



Gambar 2.1
Teknik *Simple Random Sampling*

Pengambilan sampel dengan cara ini dapat dilakukan dengan cara pengundian sebagai berikut:

Seluruh anggota pada populasi diberi nomor. Lalu secara acak dipilih nomor-nomor yang sesuai dengan banyaknya sampel yang telah ditentukan, dengan cara diundi. Pengundian bisa dilakukan dengan menggunakan "*kelintangan*", atau dengan menggunakan *mesin pembangkit angka acak*.¹ Pengundian dapat pula menggunakan tabel bilangan random yang sering dilampirkan pada buku-buku teks statistika.

Tabel bilangan random merupakan suatu tabel yang terdiri dari bilangan-bilangan yang disajikan dengan sangat tidak beraturan. Prinsip pemakaian tabel bilangan random adalah pertama-tama memberi nomor pada setiap anggota populasi. Daftar ini disebut kerangka pengambilan sampel (*sample frame*). Seandainya terdapat 50 anggota populasi, maka setiap anggota diberi nomor mulai dari 01 sampai 50. lalu gunakan jumlah digit pada tabel acak digit populasi.² Cara pemakaian pada tabel, pilih salah satu nomor dengan angka acak, gunakan digit terakhirnya. Anggota populasi pada *sample frame* yang mempunyai nomor yang cocok dengan angka di atas, diambil sebagai anggota sampel. Lalu lihat nomor yang berikutnya pada tabel, ambil dua digit terakhir. Anggota populasi pada *sample frame* yang mempunyai nomor sama terpilih sebagai anggota sampel. Demikian seterusnya sampai jumlah sampel yang diharapkan terpenuhi. Angka pada tabel yang nilainya di atas 50, atau angka yang berulang harus dibuang.

b. *Proportionate Stratified Random Sampling*

Teknik ini digunakan bila populasi mempunyai anggota/unsur yang tidak homogen dan berstrata secara proporsional. Sebagai contoh, suatu lembaga mempunyai pegawai dari latar belakang pendidikan yang beragam. Maka populasi pegawai itu berstrata. Misalnya jumlah pegawai yang lulus $S_2=10$, $S_1=40$, $D_3=200$, SMK Teknik = 400, SMK Ekonomi = 200, SMP = 150. Jumlah sampel yang harus diambil mencakup seluruh strata pendidikan tersebut yang diambil secara proporsional jumlah sampel.

Bila diambil kesalahan 5%, dengan menggunakan Tabel Krcajie, maka jumlah sampelnya 278. Karena sampel berstrata dan stratanya menurut tingkat/latar belakang pendidikan, maka sampel harus

1. Mesin pembangkit angka acak dapat berupa dadu bermata banyak, atau kini dapat menggunakan komputer.
2. Karena jumlah populasi 50 (dua digit), maka digit pada tabel angka acak adalah 2.

proporsional menurut latar belakang/tingkat pendidikan. Jumlah sampel untuk tiap tingkat/latar belakang pendidikan adalah:

$$S_2 = \frac{10}{1000} \times 278 = 2,78 = 3$$

$$S_1 = \frac{40}{1000} \times 278 = 11,12 = 11$$

$$D_3 = \frac{200}{1000} \times 278 = 54,60 = 55$$

$$\text{SMK Teknik} = \frac{400}{1000} \times 278 = 109,20 = 110$$

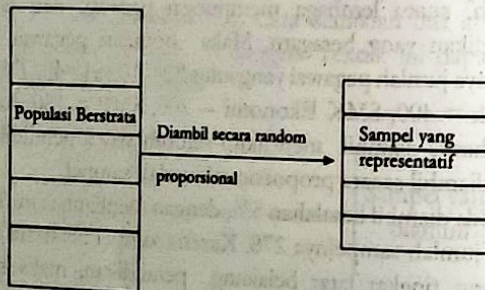
$$\text{SMK Ekonomi} = \frac{200}{1000} \times 278 = 54,60 = 55$$

$$\text{SMP} = \frac{150}{1000} \times 278 = 40,70 = 41$$

Terlihat bahwa pada teknik *proportionate stratified random sampling*, untuk jumlah tiap strata (subpopulasi) berbeda, maka jumlah sampel pada tiap strata (subsampel) pun berbeda.

Jika jumlah anggota tiap strata (subpopulasi) sama, maka jumlah sampel pada tiap strata (subsampel) akan sama.

Teknik *Proportionate Stratified Random Sampling* dapat digambarkan pada gambar 2.2 sebagai berikut:



Gambar 2.2
Teknik *Stratified Random Sampling*

c. *Disproportionate Stratified Random Sampling*

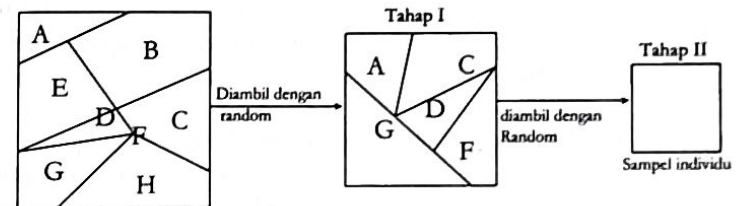
Teknik ini digunakan untuk menentukan jumlah sampel bila populasi berstrata tetapi kurang proporsional. Misalnya pegawai dari suatu PT tertentu mempunyai 3 orang lulusan S24 orang lulusan S1 90 orang lulusan D3 800 orang lulusan SMU, dan 700 orang lulusan SMP. Maka 3 orang lulusan S2 dan 4 orang lulusan itu diambil semuanya sebagai sampel, karena dua kelompok tersebut terlalu kecil bila dibandingkan kelompok, SMU dan SMP.

b. *Cluster Sampling (Area Sampling)*

Pengambilan sampel dengan kluster ini kadang-kadang dikaitkan dengan pengambilan sampel wilayah, sebab dalam pelaksanaannya sering dikaitkan dengan letak geografis. Namun teknik sampling ini bisa pula digunakan pada pengambilan sampel yang lebih umum, yakni pada suatu populasi yang berstruktur.

Dalam pemakaian teknik ini, sering digunakan melalui lebih dari satu tahap, yaitu tahap pertama menentukan sampel daerah, dan tahap berikutnya menentukan individu-individu yang ada pada daerah itu secara *sampling* juga. Teknik ini dapat digambarkan seperti pada gambar 2.3 berikut ini:

Populasi Daerah



Gambar 2.3
Teknik *Cluster Random Sampling*

Pengambilan sampel dengan cara ini mirip dengan cara *Proportionate Stratified Random Sampling*. Bedanya jika cara stratifikasi dipandang dari adanya subpopulasi yang homogen, sedang pada cara *cluster* unsur-unsurnya heterogen. Selanjutnya dari masing-masing kluster dipilih sampel secara random sebanyak yang dibutuhkan.

Misalnya di Indonesia terdapat 32 propinsi, dan sampelnya akan menggunakan 10 propinsi, maka pengambilan 10 propinsi itu dilakukan secara random, tetapi karena propinsi-propinsi di Indonesia berstrata maka pengambilan sampelnya perlu menggunakan *stratified random sampling*.

2. Non Probability Sampling

Non Probability sampling adalah teknik yang tidak memberi peluang/kesempatan yang sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel (Sugiyono, 1999: 60). Dengan cara demikian semua elemen populasi belum tentu mempunyai kesempatan untuk dipilih menjadi anggota sampel. Hal ini misalnya karena ada bagian tertentu yang secara sengaja tidak dimasukkan dalam pemilihan untuk mewakili populasi. Cara ini juga sering disebut sebagai pengambilan sampel berdasarkan pertimbangan tertentu oleh peneliti.

Ada 6 cara pengambilan sampel. Cara ini yaitu:

a. Sampling Sistematis

Sampling sistematis adalah teknik penentuan sampling berdasarkan urutan dari anggota populasi yang telah diberi nomor urut. Misalnya anggota populasi terdiri 100. Maka setiap anggota populasi diberi nomor dari 001 sampai dengan nomor 100. Misalnya peneliti menentukan hanya mengambil sampel dari populasi yang bernomor kelipatan 3. Maka anggota populasi yang terpilih menjadi anggota sampel adalah anggota populasi nomor 003, 006, 009 dan seterusnya sampai 099.

Bila peneliti memutuskan mengambil sampel dari anggota populasi yang bernomor genap, maka anggota sampel yang terpilih sebagai sampel adalah 002, 004, 006 dan seterusnya sampai 100. Jelas terlihat bahwa setiap anggota populasi tidak mempunyai kesempatan yang sama untuk terpilih sebagai sampel.

b. Sampling Kuota (Quota Sampling)

Teknik ini digunakan jika peneliti ingin mengkaji suatu fenomena dari beberapa sisi. Peneliti menentukan responden yang akan dipilih adalah orang-orang yang diperkirakan akan menjawab semua sisi tersebut. Misalnya akan diteliti perihal prestasi akademik mahasiswa dari mahasiswa aktif belajar di kelas, rajin membaca di perpustakaan dan turut serta dalam organisasi kemahasiswaan, maka sasaran kuesioner diarahkan pada mahasiswa yang aktif kuliah, rajin ke perpustakaan dan mahasiswa yang terlibat organisasi kemahasiswaan. Jadi mahasiswa-mahasiswa seperti itu jika dijadikan sampel akan menjadi wakil dari populasi seluruh mahasiswa.

c. Sampling Aksidental

Sampling aksidental adalah teknik penentuan sampel berdasarkan kebetulan, yaitu siapa saja yang secara kebetulan bertemu dengan

peneliti dapat digunakan sebagai sampel, bila kebetulan orang yang ditemui itu cocok sebagai sumber data.

Teknik ini disebut pula cara dipermudah (*convenience sampling*). Sampel ini nyaris tidak dapat diandalkan, tetapi biasanya paling mudah dan cepat dilakukan karena peneliti memiliki kebebasan untuk memilih siapa saja yang mereka temui.

Meskipun mempunyai ketidakhandalan yang tinggi, cara ini masih bermanfaat, misalnya pada tahap awal penelitian eksploratif saat studi penelitian pendahuluan untuk mencari petunjuk-petunjuk penelitian. Selanjutnya dari hasil tersebut dapat dilakukan langkah-langkah yang lebih terarah.

d. Sampling Purposive

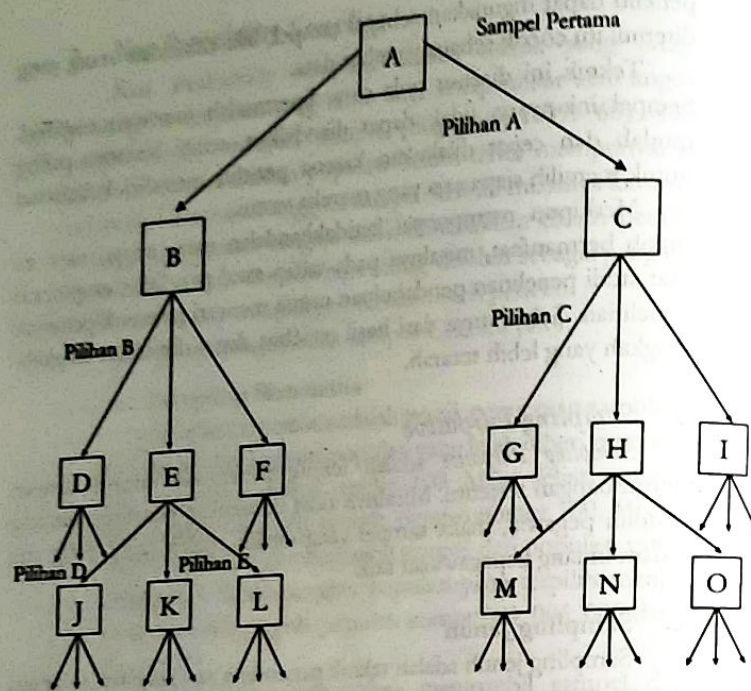
Sampling Purposive adalah teknik penentuan sampel dengan pertimbangan tertentu. Misalnya akan melakukan penelitian tentang disiplin pegawai, maka sampel yang dipilih adalah orang yang ahli dalam bidang kepegawaian saja.

e. Sampling Jenuh

Sampling jenuh adalah teknik penentuan sampel buka, sehingga semua anggota populasi digunakan sebagai sampel. Hal ini sering digunakan bila jumlah populasi relatif kecil, misalnya kurang dari 30 orang. Istilah lain dari sampel jenuh adalah sensus, dimana semua anggota populasi dijadikan sampel.

f. Snowball Sampling

Snowball sampling adalah teknik penentuan sampel yang mula-mula jumlahnya kecil. Kemudian sampel ini disuruh memilih teman-temannya untuk dijadikan sampel. Begitu seterusnya, sehingga jumlah sampel semakin banyak. Ibarat bola salju yang menggelinding, makin lama semakin besar. Pada penelitian kualitatif banyak menggunakan sampel *purposive* dan *snowball*. Teknik *snowball sampling*, ditunjukkan pada gambar 2.4 berikut:



Gambar 2.4. Snowball Sampling

2.2.1.3. Kekeliruan Sampling

Suatu penelitian sampel memang tidak dapat lepas dari kekeliruan, karena penelitian sampel hanya dilakukan terhadap sebagian populasi. Generalisasi yang dilakukan tentu saja akan memunculkan galat/error. Hal yang harus diupayakan adalah meminimalisasi kesalahan tersebut. Teknik sampling merupakan salah satu cara memperkecil kesalahan tersebut.

Namun demikian, kekeliruan tetap saja bisa terjadi, baik pada saat pengumpulan data, saat melakukan pengolahan data, atau pun pada saat penyajian informasi hasil penelitian. Pada tahap pengumpulan data sampel terdapat medan-medan dimana kekeliruan *sampling*¹ dapat terjadi. Oleh karena itu, perlu dilakukan antisipasi secara dini dengan melakukan penentuan sampel secara cermat, pengambilan data yang valid, pengolahan data yang akurat, dan penyajian data informasi yang baik.

¹ Maksudnya adalah terjadinya kekeliruan pada saat menelaah sampel, misalnya dalam menentukan jumlah sampel yang harus diambil.

2.2.2. Kegiatan Pengumpulan Data

Kegiatan pelaksanaan pengumpulan data dapat berbentuk :

- Pengamatan mendalam (*systematic observation*), yaitu pengamatan terhadap obyek yang akan dicatat datanya. Sebelumnya telah dilakukan persiapan matang, tentang obyek apa yang akan diamati secara mendalam. Tidak lupa, instrumen yang dibutuhkan juga dipersiapkan.
- Wawancara mendalam (*systematic interview*), yaitu pengumpulan data berbentuk pengajuan pertanyaan lisan. *Out line* dan *draft* pertanyaan telah dipersiapkan secara matang, demikian pula dengan instrumen yang dibutuhkan.
- Angket, yaitu pengumpulan data berbentuk pengajuan pertanyaan tertulis melalui sebuah daftar pertanyaan yang sudah dipersiapkan sebelumnya.
- Tes, misalnya tes hasil belajar, tes kecerdasan, tes kepribadian, tes minat dan bakat, dan lain sebagainya.

Dalam pengumpulan data statistik, beberapa alat (instrumen) yang biasa dipergunakan antara lain:

- Daftar atau daftar check (*check list*)
- Skala bertingkat (*rating scale*)
- Pedoman wawancara (*interview guide*)
- Questionnaire* (daftar pertanyaan yang pertanyaannya sudah disediakan jawabannya untuk dipilih, atau disediakan tempat untuk mengisi jawabannya).

BAB III

TABEL DISTRIBUSI FREKUENSI

3.1. Analisis Data Berbasis Tabel Distribusi Frekuensi

Masalah awal yang dihadapi peneliti setelah data terkumpul adalah bagaimana membuat data tersebut mudah dibaca dan komunikatif. Permasalahan demikian muncul karena hampir dalam setiap penelitian kuantitatif, akan didapatkan kumpulan data kuantitatif yang biasanya tidak teratur, berserak, dan masih merupakan bahan keterangan yang masih kasar dan mentah. Dengan kata lain kegiatan penelitian umumnya mendapatkan data yang belum bisa memberikan informasi secara ringkas, padat, jelas dan komunikatif. Untuk itu agar data angka yang telah berhimpun tersebut “dapat berbicara” dan dapat memberikan informasi yang berarti, diperlukan adanya tindak lanjut atau langkah tertentu.

Langkah tertentu tersebut dapat berupa mengurutkan data dari yang paling kecil ke yang paling besar, atau sebaliknya dari yang paling besar ke yang paling kecil. Pengurutan demikian untuk jumlah data sedikit akan cukup mudah, dilakukan dan dapat memudahkan untuk memahami data. Namun, bila jumlah data banyak, maka barisan data yang tersusun akan panjang. Hal demikian masih menyulitkan peneliti dalam mengamati data dan memberikan pemaknaan terhadap data tersebut. Maka perlu dilakukan pengelompokan data pada interval/selang tertentu, dan selanjutnya dihitung jumlah data pada kelompok tersebut. Data yang telah dikelompokkan, beserta frekuensi pada tiap kelompok data tersebut dinamakan distribusi frekuensi. Karena data tersebut disajikan dalam tabel, maka lazim disebut sebagai tabel distribusi frekuensi.

Dengan tabel distribusi frekuensi, data yang semula tidak teratur dan berserak, dapat diringkas dan disederhanakan sehingga dapat memberi informasi yang jelas dan sistematis mengenai ciri atau sifat yang dimilikinya. Dapat dikatakan distribusi frekuensi merupakan “pintu gerbang” yang harus

dilewati untuk dapat melakukan analisis statistik yang lebih dalam. Dengan kata lain distribusi frekuensi merupakan bagian yang sangat penting dalam analisis data statistik. Mengingat hal demikian, pembuatan tabel distribusi frekuensi sedapat mungkin diupayakan untuk dikuasai.

3.2. Pengertian Tabel Distribusi Frekuensi

Sebelum mendefinisikan tabel distribusi frekuensi, terlebih dahulu perlu difahami beberapa pengertian tentang variabel, frekuensi, distribusi dan distribusi frekuensi.

Variabel yang dalam bahasa Inggrisnya *variable*, secara etimologis berarti "peubah", "faktor tak tetap", atau "gejala yang dapat diubah-ubah atau berubah-ubah". Dalam statistik, variabel merupakan suatu gejala yang menjadi fokus untuk diamati. Variabel merupakan atribut dari suatu obyek yang mempunyai variasi antara satu dengan yang lainnya. Tinggi, berat badan, usia, sikap, motivasi belajar, kedisiplinan, merupakan atribut dari seseorang, dan akan menjadi variabel bila menjadi fokus yang diamati. Variabel umumnya bersifat kualitatif, namun dilambangkan dengan angka.

Contoh 3.1:

- Nilai test, pada dasarnya adalah gejala kualitatif, tetapi dilambangkan dengan angka pada interval 0 – 100.
- Usia juga merupakan gejala kualitatif, tetapi lazim dilambangkan dengan angka, misalnya 5 tahun, 17 tahun, 30 tahun, 40 tahun, 50 tahun, 60 tahun, 70 tahun dan sebagainya.

Kata frekuensi (Ing. *frequency*), mempunyai arti "kekerapan", "keseringan", atau "jarang-kerapnya". Dalam statistik, yang dimaksud dengan frekuensi adalah : angka (*bilangan*) yang menunjukkan berapa kali suatu nilai dari variabel tertentu muncul.

Contoh 3.2:

Dari 10 mahasiswa didapat data IPK masing-masing sebagai berikut:
3,50 3,50 3,00 3,00 2,60 3,10 2,70 3,50 3,20 2,60

Pada data tersebut terlihat mahasiswa yang mempunyai IPK 2,60 ada 2, yang mempunyai IPK 2,70 ada 1, yang mempunyai IPK 3,00 ada 2, yang mempunyai IPK 3,10 ada 1, dan yang mempunyai IPK 3,20 ada 1, sedang yang mempunyai IPK 3,50 ada 3. Maka dapat dikatakan IPK 2,60 berfrekuensi 2, IPK 3,00 berfrekuensi 2, IPK 3,50 berfrekuensi 3, Frekuensi IPK 2,70 dan 3,10 serta 3,20 adalah 1 karena ada 1 yang mempunyai IPK 2,70, 3,10 dan 3,20. Frekuensi IPK 2,80, adalah 0 karena tidak ada yang mempunyai IPK 2,80 demikian juga untuk IPK 2,90 dan 3,30 dan 3,40.

Kata distribusi (*distribution*) berarti "penyaluran", "pembagian", atau "pencaran". Jadi distribusi frekuensi berarti pembagian frekuensi, atau pencaran

frekuensi. Dalam statistik, distribusi frekuensi merupakan suatu penggambaran pencaran variabel/gejala yang dilambangkan dengan angka, dimana semua data angka yang mewakili variabel/gejala tersebut terdistribusikan.

Umumnya distribusi frekuensi disajikan dalam bentuk tabel, yakni berbentuk kolom dan lajur. Kolom berupa variabel dan frekuensi, sedang lajur memuat nilai variabel dan besar frekuensi. Tabel yang menyajikan distribusi frekuensi tersebut dinamakan tabel distribusi frekuensi. Seringkali tabel distribusi frekuensi disebut tabel frekuensi saja.

Contoh 3.3:

Data IPK 10 mahasiswa pada contoh di atas dapat disajikan tabel distribusi frekuensi sebagai berikut:

Tabel 3.1
Distribusi Frekuensi IPK 10 Mahasiswa

IPK Mahasiswa	Frekuensi
2,60	2
2,70	1
2,80	0
2,90	0
3,00	2
3,10	1
3,20	1
3,30	0
3,40	0
3,50	3
Total	10

3.3. Macam-macam Tabel Distribusi Frekuensi

Tabel distribusi frekuensi dibedakan menjadi dua yaitu: Tabel Distribusi Frekuensi Data Tunggal dan Tabel Distribusi Frekuensi Data Kelompok. Masing-masing bentuk tabel dapat dibuat Tabel Distribusi Frekuensi Kumulatif dan Tabel Distribusi Frekuensi Relatif. Selanjutnya, dapat pula dibentuk Tabel Distribusi Frekuensi Relatif Kumulatif.

3.3.1. Tabel Distribusi Frekuensi Data Tunggal

Tabel Distribusi Frekuensi Data Tunggal adalah tabel distribusi frekuensi yang menyajikan frekuensi dari data tunggal (data yang tidak dikelompokkan).

Contoh 3.4:

Data nilai rapor PAI 100 siswa SMA Ulul Albab disajikan dalam tabel distribusi sebagai berikut:

Tabel 3.1
Tabel Distribusi Frekuensi Nilai Rapor PAI Siswa SMA Ulul Albab

Nilai	Frekuensi
10	2
9	5
8	30
7	50
6	10
5	3
Total	100

Pada tabel di atas variabel nilai dinyatakan secara tunggal (bukan kelompok). Terlihat Nilai 10 mempunyai frekuensi 2, artinya yang mempunyai nilai 10 ada 2. Nilai 6 frekuensinya 10, yang berarti siswa yang mendapat nilai 6 ada 10.

Pada tabel distribusi frekuensi data tunggal setiap unit data hanya berisi satu angka. Dengan demikian secara tepat dapat dilihat berapa frekuensi suatu nilai data tertentu. Hal ini merupakan kelebihan tabel distribusi frekuensi data tunggal. Namun tabel distribusi frekuensi data tunggal mempunyai kelemahan. Bila jumlah data banyak, dengan rentang yang lebar, tabel akan panjang. Karena setiap unit data harus mewakili satu angka, maka dengan rentang yang lebar harus disediakan pula unit angka yang panjang. Misal datanya mencakup nilai 46 sampai 100, maka harus didaftar unit data sejumlah 55 yaitu 46, 47, 48, ..., 98, 99, 100. Hal demikian menjadikan tabel "kurang indah", tidak praktis dan tidak sederhana serta kurang komunikatif. Oleh karena itu untuk data besar dan rentang lebar, biasanya tabel distribusi frekuensi data tunggal tidak digunakan.

3.3.2. Tabel Distribusi Frekuensi Data Kelompok

Tabel Distribusi Frekuensi Data Kelompok adalah tabel distribusi frekuensi yang menyajikan data yang dikelompokkan (bukan data tunggal). Pada Tabel Distribusi Frekuensi Data Kelompok, data masuk dalam kelas interval. Kelas interval memuat data yang mempunyai nilai pada rentang interval tersebut.

Kelas interval mempunyai batas bawah dan batas atas. Batas bawah kelas interval merupakan nilai minimum pada kelas interval, sedang batas atas merupakan nilai maksimum pada kelas interval.

Contoh 3.5:

Nilai hasil ujian Statistika Pendidikan dari sejumlah 80 Mahasiswa STAIN Pekalongan disajikan dalam tabel 3.2 berikut:

Tabel 3.2
Tabel Distribusi Frekuensi Nilai Hasil Ujian Statistika Pendidikan Mahasiswa STAIN Pekalongan

Nilai	Frekuensi
96-100	2
91-95	5
86-90	6
81-85	7
76-80	11
71-75	18
66-70	11
61-65	7
56-60	6
51-55	5
46-50	2
Total	80

Pada tabel distribusi frekuensi data kelompok dalam satu unit data terdapat lebih dari satu angka. Frekuensi menunjukkan banyaknya kemunculan data pada suatu interval nilai, bukan suatu nilai data tertentu. Pada tabel tersebut variabel nilai berbentuk interval. Interval nilai 96 – 100 menunjukkan nilai-nilai pada rentang 96 sampai 100. Pada interval tersebut batas bawahnya adalah 96 sedang batas atasnya 100. Frekuensi pada interval tersebut 2, artinya mahasiswa yang mendapatkan nilai pada rentang 96 sampai 100 ada 2. Dua orang tersebut mungkin mendapat nilai 96 dan 100, atau 96,5 dan 98, atau 97 dan 99 atau bisa juga keduanya 96 atau 100. Jadi pada Tabel Distribusi Frekuensi Data Kelompok tidak dapat ditentukan secara pasti berapa frekuensi suatu nilai. Frekuensi hanya menunjuk nilai pada interval tertentu. Itulah kelemahan Tabel Distribusi Frekuensi Data Kelompok. Namun tabel ini mempunyai kelebihan, yaitu lebih praktis, sederhana dan "lebih indah", khususnya untuk data yang banyak dan rentangnya lebar. Misalnya, pada data yang mencakup nilai 46 sampai 100 tidak harus disediakan 55 unit data, melainkan cukup 11.

3.3.3. Tabel Distribusi Frekuensi Kumulatif

Tabel Distribusi Frekuensi Kumulatif adalah tabel distribusi yang menyajikan frekuensi yang dijumlahkan secara kumulatif, dari atas kebawah atau sebaliknya. Bila penjumlahan dilakukan dari atas kebawah, maka disebut tabel Distribusi Frekuensi Kumulatif Atas disimbolkan dengan f_{ka} , bila penjumlahan dilakukan dari bawah ke atas disebut Distribusi frekuensi Kumulatif Bawah disimbolkan dengan f_{kb} .

Contoh 3.6:

Data nilai rapor PAI 100 siswa SMA Ulul Albab yang disajikan dalam tabel 3.1 dapat dibentuk Tabel Distribusi Frekuensi Kumulatif Atas sebagai berikut:

Tabel 3.3
Tabel Distribusi Frekuensi Kumulatif Atas Nilai Rapor PAI
Siswa SMA Ulul Albab

Nilai	Frekuensi (f)	f_{ka}
10	2	2
9	5	7
8	30	37
7	50	87
6	10	97
5	3	100
Total	100	-

Pada tabel tersebut frekuensi kumulatif atas 10 adalah 2, artinya yang mendapat nilai kurang atau sama dengan 10 ada 2. Frekuensi kumulatif atas 9 adalah 7, yakni jumlahan dari 2 dengan 5. Nilai f_{ka} 7 ini mengandung arti, yang mendapat nilai kurang atau sama dengan 9 ada 7. Frekuensi kumulatif atas 5 adalah 100 yang merupakan jumlahan dari 2, 5, 30, 50, 10 dan 3. Berarti yang mendapat nilai kurang atau sama dengan 5 ada 100.

Contoh 3.7:

Data nilai rapor PAI 100 siswa SMA Ulul Albab yang disajikan dalam tabel 3.1 dapat dibentuk Tabel Distribusi Frekuensi Kumulatif Bawah sebagai berikut:

Tabel 3.4

Tabel Distribusi Frekuensi Kumulatif Bawah Nilai Rapor PAI
Siswa SMA Ulul Albab

Nilai	Frekuensi (f)	f_{kb}
10	2	100
9	5	98
8	30	93
7	50	63
6	10	13
5	3	3
Total	100	-

Pada tabel di atas, Frekuensi Kumulatif Bawah (f_{kb}) dari 5 adalah 3, berarti yang mendapat nilai kurang atau sama dengan 5 ada 3. Frekuensi kumulatif bawah (f_{kb}) dari 6 adalah 13, yakni penjumlahan dari 3 dan 10. Nilai f_{kb} ini bermakna yang mendapat nilai kurang atau sama dengan 6 ada 13. Frekuensi kumulatif bawah (f_{kb}) dari 10 adalah 100, yakni penjumlahan dari 3, 10, 50, 30, 5 dan 2. Nilai f_{kb} ini bermakna yang mendapat nilai kurang atau sama dengan 10 ada 100.

3.3.4. Tabel Distribusi Frekuensi Persentasi

Tabel distribusi frekuensi persentasi merupakan tabel distribusi yang frekuensinya dinyatakan dalam bentuk persen disimbolkan dengan p. Tabel ini disebut pula Tabel Relatif. Pada tabel ini frekuensi tiap unit dibagi jumlah seluruh data dikalikan 100%. Tabel ini dapat menunjukkan berapa jumlah relatif atau proporsi suatu unit dibanding jumlah keseluruhan data.

Contoh 3.8:

Data nilai rapor PAI 100 siswa SMA Ulul Albab yang disajikan dalam tabel 3.1 dapat dibentuk Tabel Distribusi Frekuensi persentasi sebagai berikut:

Tabel 3.5
Tabel Distribusi Frekuensi Persentasi Nilai Rapor PAI
Siswa SMA Ulul Albab

Nilai	Frekuensi (f)	P
10	2	2
9	5	5
8	30	30
7	50	50
6	10	10
5	3	3
Total	100	100

Pada tabel tersebut terlihat frekuensi relatif 5 adalah 3%, yang didapat dari $3/100 \times 100\%$. Hal ini berarti yang mendapat nilai 5 ada 3% dari seluruh data. Frekuensi relatif 7 adalah 50, didapat dari $50/100 \times 100\%$, yang berarti yang mendapat nilai 7 ada 50%. Total frekuensi relatif adalah 100%.

3.3.5. Tabel Distribusi Frekuensi Persentasi Kumulatif

Tabel Distribusi Frekuensi Persentasi Kumulatif adalah tabel distribusi frekuensi persentasi dimana persentasinya dijumlahkan terus meningkat dari atas ke bawah atau dari bawah ke atas. Bila penjumlahan dari atas disebut Tabel Persentasi Kumulatif Atas (pka), sedang bila dari bawah disebut Tabel Persentasi Kumulatif Bawah (pkb).

Contoh 3.9:

Data nilai rapor PAI 100 siswa SMA Ulul Albab yang disajikan dalam tabel 3.1 dapat dibentuk Tabel Distribusi Frekuensi Persentasi Kumulatif Atas dan Kumulatif Bawah sebagai berikut:

Tabel 3.6
Tabel Distribusi Frekuensi Persentasi Kumulatif Atas dan Kumulatif Bawah
Nilai Rapor PAI Siswa SMA Ulul Albab

Nilai	Frekuensi (f)	P(%)	Pka(%)	Pkb(%)
10	2	2	2	100
9	5	5	7	98
8	30	30	37	93
7	50	50	87	63
6	10	10	97	13
5	3	3	100	3
Total	100	100	-	-

3.4. Cara Membuat Tabel Distribusi Frekuensi

Tabel distribusi frekuensi dibuat dengan tujuan agar data observasi dapat lebih mudah dipahami oleh pihak yang berkepentingan. Pada distribusi frekuensi terdapat beberapa kelas yang masing-masing kelas tersebut memuat data observasi. Pada distribusi frekuensi, masing-masing kelas mempunyai interval (selang) yang besarnya sama untuk semua kelas.

Sebetulnya tidak ada ketentuan khusus yang harus dipenuhi dalam pembuatan tabel distribusi frekuensi. Yang penting tabel distribusi frekuensi tersebut dapat memberikan informasi yang baik tentang data pengamatan. Namun demikian, agar distribusi frekuensi yang dibuat cukup baik dalam penyajian, maka penyusunan tabel distribusi frekuensi hendaknya memperhatikan jumlah kelas (k), panjang interval kelas (i), dan titik tengah kelas (*mid point*).

Ada beberapa hal yang perlu diperhatikan agar tabel distribusi frekuensi cukup baik, yaitu:

1. Jumlah kelas pada suatu tabel distribusi frekuensi jangan terlalu banyak, namun juga jangan terlalu sedikit.
2. Hindari adanya kelas pada tabel distribusi frekuensi yang tidak mengandung data pengamatan sehingga frekuensinya nol.
3. Semua data harus termuat dalam tabel distribusi frekuensi tersebut.

Seperti sudah diuraikan di depan, secara garis besar tabel distribusi frekuensi dibedakan menjadi Tabel Distribusi Frekuensi Data Tunggal dan Tabel Distribusi Frekuensi Data Kelompok. Berikut akan diuraikan cara membuat tabel distribusi frekuensi untuk keduanya.

3.4.1. Cara Membuat Tabel Distribusi Frekuensi Data Tunggal
Pembuatan tabel distribusi frekuensi data tunggal relatif mudah. Langkahnya adalah:

1. Tentukan nilai tertinggi (X_{max}) dan nilai terendah (X_{min})
2. Siapkan tabel yang terdiri 3 kolom, masing untuk Nilai, Turus/Tally, dan Frekuensi (f).
3. Urutkan data dari nilai tertinggi ke nilai terendah, letakkan pada kolom Nilai.
4. Hitung frekuensi masing-masing nilai dengan bantuan Turus/Jari-jari, letakkan pada kolom Turus.
5. Ubah turus/jari-jari menjadi frekuensi.
6. Terakhir, jumlahkan semua frekuensi.

Contoh 3.10:

Data nilai ulangan Matematika 40 siswa kelas II SMP Al Ummah adalah sebagai berikut:

6	9	7	5	7	8	10	7	5	6
4	6	9	7	6	5	7	8	8	3
5	7	6	8	9	10	4	6	7	9
4	5	10	6	4	7	9	7	8	7

Untuk menyajikan data tersebut ke dalam Tabel Distribusi Frekuensi Data Tunggal langkahnya adalah sebagai berikut:

1. Nilai tertinggi (X_{max}) = 10, nilai terendah (X_{min}) = 3.
2. Tabel distribusi frekuensi dengan 3 kolom:

Tabel 3.7
Tabel Distribusi Frekuensi Nilai Matematika
Siswa SMP Al Ummah

Nilai	Turus	Frekuensi
Jumlah		

3. Mengurutkan data dari nilai tertinggi ke nilai terendah, dimasukkan pada kolom Nilai.
4. Penghitungan Turus.

5. Mengubah Turus menjadi frekuensi.
 6. Menjumlah Frekuensi
- Dengan langkah 3 sampai 6 terbentuk tabel Distribusi Frekuensi Nilai Matematika seperti dinyatakan dalam tabel 3.8

Tabel 3.8
Tabel Distribusi Frekuensi Nilai Matematika
Siswa SMP Al Ummah

Nilai	Turus	Frekuensi
10	///	3
9	////	5
8	////	5
7	//// //	10
6	//// //	7
5	////	5
4	////	4
3	/	1
Jumlah	-	40

3.4.2. Cara Membuat Tabel Distribusi Frekuensi Data Kelompok

3.4.2.1. Menentukan Jumlah Kelas Interval

Langkah awal dalam membuat tabel distribusi frekuensi adalah menentukan kelas interval. Sebenarnya tidak ada ketentuan baku tentang jumlah kelas interval, tergantung dari opini atau kesukaan pembuat tabel tersebut. Akan tetapi harus diingat bahwa jumlah kelas yang baik adalah yang tidak terlalu banyak atau pun terlalu sedikit. Berkaitan dengan itu untuk menentukan jumlah kelas interval terdapat beberapa pedoman. Dalam hal ini akan diuraikan 2 pedoman.

a. Menggunakan Rumus Sturges

Menurut Sturges jumlah kelas interval ditentukan dengan rumus:

$$K = 1 + 3,3 \log n$$

Rumus 3.1

- K = Jumlah interval kelas
n = Jumlah data pengamatan
log = Logaritma

Misalkan jumlah data 100, maka jumlah interval kelas adalah

$$\begin{aligned} K &= 1 + 3,3 \log 100 \\ &= 1 + 3,3 \cdot 2 \\ &= 1 + 6,6 \\ &= 7,6 \text{ dapat dibulatkan menjadi } 7 \text{ atau } 8 \end{aligned}$$

b. Menggunakan Ancar-ancar

Menurut (Anas Sudijono, 2003: 50) jumlah klas interval sebaiknya antara 10 sampai 20. Ancar-ancar ini bisa dipakai, tetapi tidak harus dipenuhi.

3.4.2.2. Menentukan Rentang Data (R)

Rentang data merupakan selisih antara nilai tertinggi (X_{max}) dengan nilai terendah (X_{min}), atau diformulasikan:

$$R = X_{max} - X_{min}$$

Rumus 3.2

3.4.2.3. Menghitung Panjang Kelas Interval (i)

Panjang interval kelas pada tabel distribusi frekuensi dirumuskan dengan:

$$i = R/K$$

Rumus 3.3

- i = Panjang interval kelas
- R = Rentang data
- K = Banyaknya interval kelas

Hasil perhitungan yang seringkali didapat berupa nilai pecahan. Untuk itu dilakukan pembulatan. Pembulatan dapat ke atas atau pun ke bawah.

3.4.2.4. Menentukan Batas-batas Kelas

Penentuan batas-batas kelas interval dilakukan dengan menjamin seluruh data observasi masuk dalam tabel distribusi frekuensi, jumlah kelas interval dan panjang interval sesuai dengan yang ditentukan. Untuk itu terlebih dahulu ditentukan bilangan-bilangan dasar. Bilangan dasar terkecil umumnya adalah data terendah, bilangan berikutnya adalah bilangan dasar terkecil ditambah panjang interval, begitu seterusnya. Namun, dalam prakteknya, bilangan-bilangan ini dapat digeser (ditambah atau dikurangi).

3.4.2.5. Menyusun Kelas-Kelas Interval

Bilangan-bilangan dasar selanjutnya menjadi batas bawah kelas interval, batas atas kelas interval ditentukan dengan formula:

$$Ba = Bb + i - 1$$

Rumus 3.4

- Ba = Batas atas kelas interval
- Bb = Batas bawah kelas interval
- i = Panjang interval kelas

Kelas-kelas interval ini selanjutnya dimasukkan dalam tabel yang terdiri 3 kolom, masing-masing untuk Interval Nilai, Turus, dan Frekuensi.

3.4.2.6. Menghitung Frekuensi

Pada langkah ini, frekuensi tiap kelas interval dihitung dengan menggunakan bantuan turus. Selanjutnya turus diubah menjadi frekuensi, kemudian dijumlahkan.

Contoh 3.11:

Data nilai ulangan Biologi 30 siswa kelas II SMA Al Wathoniyah adalah sebagai berikut:

59	54	60	71	58	48
56	64	77	65	40	51
41	46	49	64	73	67
87	67	89	62	78	55
86	64	94	80	68	84

Data tersebut dapat dibuat menjadi tabel distribusi frekuensi dengan langkah-langkah sebagai berikut:

1. **Menentukan Banyak Kelas Interval**
Banyaknya data 30 dengan Rumus Sturges $K = 1 + 3,3 \log n$
Maka $K = 1 + 3,3 \log 30 = 1 + 4,875 = 5,8745$ dibulatkan 6
2. **Menentukan Rentang Data (R)**
Nilai tertinggi (X_{max}) = 94, nilai terendah (X_{min}) = 40
 $R = X_{max} - X_{min} = 94 - 40 = 54$
3. **Menghitung Panjang Kelas Interval (i)**
 $i = R/K = 54/5,8745 = 9,19$ dibulatkan 10
4. **Menentukan Batas-batas Kelas**
Nilai terendah (X_{min}) = 40, panjang kelas interval 10, maka bilangan-bilangan dasarnya adalah 40, 50, 60, 70, 80, 90.
5. **Menyusun Kelas-kelas Interval**
Bilangan-bilangan dasar tersebut selanjutnya menjadi Batas-batas bawah Kelas Interval. Dengan rumus 3.4 Batas-batas Atas Kelas Intervalnya berturut-

turut adalah 49, 59, 69, 79, 89, 99. Dengan demikian Kelas-kelas Intervalnya adalah: 40-49, 50-59, 60-69, 70-79, 80-89 dan 90-99.

Kelas-kelas Interval ini selanjutnya dimasukkan ke dalam tabel berikut:

Tabel 3.9

Tabel Distribusi Frekuensi Nilai Ulangan Biologi Siswa SMA Al Wathoniyah

Nilai (X)	Turus	Frekuensi (f)
90-99		
80-89		
70-79		
60-69		
50-59		
40-49		
Jumlah		

6. Menghitung Frekuensi

Pada langkah ini dilakukan penghitungan frekuensi dengan bantuan turus. Selanjutnya turus diubah menjadi frekuensi dan dijumlah.

Tabel 3.10

Tabel Distribusi Frekuensi Nilai Ulangan Biologi Siswa SMA Al Wathoniyah

Nilai (X)	Turus	Frekuensi (f)
90-99	/	1
80-89	###	5
70-79	////	4
60-69	### ////	9
50-59	### /	6
40-49	###	5
Jumlah	-	30

BAB IV PENGOLAHAN DATA STATISTIK

4.1. Pengertian Data Statistik

Data (jamak dari *datum*) adalah sesuatu yang diketahui atau dianggap diketahui. Ia adalah bahan mentah, yang bila diolah/dianalisis dengan baik dapat menjadi informasi. Data statistik merupakan sekumpulan nilai kuantitatif dari suatu obyek.

Contoh 4.1:

- Data volume penjualan suatu produk setiap bulan selama dua tahun. Data di sini adalah kumpulan besarnya volume penjualan setiap bulan selama dua tahun.
- Data perkembangan harga beras/kilogram setiap hari di Pasar Surogenen Pekalongan dalam bulan September 2004. Data di sini adalah kumpulan harga beras/kilogram setiap hari dalam bulan September 2004.
- Data kelahiran dan kematian di Kelurahan Sugihwaras dalam tahun 2004. Data di sini adalah catatan tentang jumlah kematian dan kelahiran penduduk Kelurahan Sugihwaras dalam tahun 2004.

4.2. Penggolongan dan Sifat-sifat Data Statistik

Data statistik dibedakan beberapa macam, tergantung dari cara menggolongkannya. Penggolongannya dapat berdasarkan cakupan pengambilannya, sumber data, waktu pengumpulannya, pengelompokan angka, dan sifat nilai/skore.

4.2.1. Berdasarkan Cakupan Pengambilan Data

Bila dilihat dari cakupan pengambilannya, data statistik dibedakan menjadi dua, yaitu data populasi dan data sampel. Apabila kumpulan data tersebut melibatkan semua anggota dari obyek yang diambil datanya, maka data

10=N	7	5	6	8	5	4	7	6	6	6	60	396
p	0,7	0,5	0,6	0,8	0,5	0,4	0,7	0,6	0,6	0,6		
q	0,3	0,5	0,4	0,2	0,5	0,6	0,3	0,4	0,4	0,4		

$$\sum X_i = 60, \sum X_i^2 = 396, N = 10$$

$$M_t = \frac{\sum X_i}{N} = \frac{60}{10} = 6$$

$$SD_t = \sqrt{\frac{\sum X_i^2}{N} - \left(\frac{\sum X_i}{N}\right)^2}$$

$$= \sqrt{\frac{396}{10} - \left(\frac{60}{10}\right)^2}$$

$$= \sqrt{39,6 - 36} = \sqrt{3,6} = 1,897$$

Untuk butir soal nomor 2

$$M_p = \frac{6+4+9+8+8+6+3}{7} = \frac{44}{7} = 6,286$$

$$p = 0,7, q = 0,3$$

$$r_{pb} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

$$r_{pb} = \frac{6,286 - 6}{1,897} \sqrt{\frac{0,7}{0,3}}$$

$$= \frac{0,286}{1,897} \sqrt{\frac{0,7}{0,3}} = 0,15076 \times 1,5275 = 0,230296 = 0,230$$

Banyak data = N = 10

$$db = 10 - 2 = 8$$

Nilai r_t pada tingkat kesalahan 5% adalah 0.632. Sedang pada tingkat kesalahan 1% adalah 0.765. Jadi $|r_{pb}| < r_t$ maka H_0 diterima, H_a ditolak. Berarti, butir 2 tidak valid.

Soal Bab IV:

Selidikilah validitas butir-butir soal nomor 2, 3, 4, 5, 6, 7, 8, 9, dan 10!

BAB XV KORELASI MULTIVARIAT

15.1. Pendahuluan

Korelasi multivariat merupakan korelasi antarvariabel, dimana terdapat lebih dari satu variabel independen (bebas) dan satu variabel dependen (terikat). Korelasi multivariat ada dua macam, yaitu: korelasi ganda (*multiple correlation*) dan korelasi parsial.

15.2. Korelasi Ganda

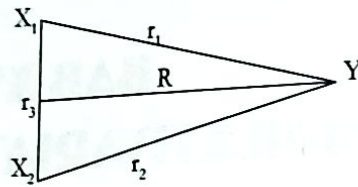
Korelasi ganda (*multiple correlation*) merupakan korelasi antara dua variabel atau lebih secara bersama-sama. Fungsi korelasi ganda adalah sebagai alat bantu untuk menganalisis data dalam penelitian, yaitu untuk melihat hubungan antara suatu variabel dependen atau variabel yang diprediksi (sering disebut dengan *criterion variable*) dengan dua atau lebih variabel independen yang berfungsi sebagai *predictor*. Sesuai dengan fungsinya, maka variabel independen itu disebut dengan istilah variabel prediktor (*predictor variable*).

Guilford dan Fruchter mengemukakan pengertian *multiple correlation* sebagai berikut:

"The coefficient of multiple correlation incates the strength of relationship between one variable and two or more others combined with optimal weights. The multiple correlation is related to the intercorrelations between independent variables as well as to their correlations with the independent variable".

Korelasi ganda dapat terdiri dari dua variabel independen dan satu variabel dependen. Dapat pula terdiri lebih dari 2 variabel independen dan satu variabel dependen. Pembahasan di sini dibatasi hanya untuk 2 variabel independen dan satu variabel dependen. Untuk kasus lebih dari 2 variabel independen dibahas dalam analisis regresi.

Ilustrasi korelasi ganda untuk 2 variabel independen dan satu variabel dependen digambarkan sebagai berikut:



Gambar 15.1

Koefisien korelasi ganda disimbolkan dengan **R**. Persyaratan dalam menggunakan rumusan *multiple correlation* ini sama dengan menggunakan rumus korelasi bivariat. Berikut ini adalah bentuk dasar formulasi rumusan *multiple correlation* dengan tiga variabel (satu variabel yang diprediksi/variabel dependen dan dua lainnya adalah variabel *predictor*/variabel independen).

$$R_{Yx_1x_2} = \sqrt{\frac{r_1^2 + r_2^2 - 2r_1r_2r_3}{1 - r_3^2}}$$

Rumus 15.1.

Keterangan:

$R_{Yx_1x_2}$: koefisien *multiple correlation*

r_1 : korelasi antara variabel independen 1/*predictor* 1 (x_1) dengan variabel dependen/kriterium (Y)

r_2 : korelasi antara variabel independen 2/*predictor* 2 (x_2) dengan variabel dependen/kriterium (Y)

r_3 : korelasi antara variabel independen 1/*predictor* 1 (x_1) dengan variabel independen 2/*predictor* 2 (x_2)

Contoh 15.1:

Seorang peneliti hendak menganalisis hubungan antara efektivitas organisasi (Y) dengan semangat kerja karyawan (x_1) dan kelancaran komunikasi (x_2). Jumlah responden yang diteliti sebanyak 60 karyawan, sedangkan hasil perhitungan dasar mengenai koefisien korelasi produk momennya adalah sebagai berikut ; a) koefisien korelasi antara variabel efektivitas organisasi dan variabel semangat kerja (r_1) = 0.46; b) koefisien korelasi antara variabel efektivitas organisasi dan variabel kelancaran komunikasi (r_2) = 0.52; sedangkan c) koefisien korelasi antara variabel semangat kerja dan variabel kelancaran

komunikasi (r_3) = 0.17. Maka besarnya koefisien korelasi gandanya dapat diselesaikan sebagai berikut:

$$R = \sqrt{\frac{0.46^2 + 0.52^2 - 2(0.46)(0.52)(0.17)}{1 - (0.17)^2}}$$

$$R = \sqrt{\frac{0.21 + 0.27 - (0.04)}{1 - (0.03)}}$$

$$R = \sqrt{0.4124}$$

$$R = 0.64$$

Tes signifikansi untuk *multiple correlation* ini dapat memakai formula rumusan F sebagai berikut:

$$F = \frac{[R^2]}{1-R^2} \frac{[N - k - 1]}{k}$$

Rumus 15.2

Keterangan:

R : koefisien korelasi ganda (*multiple correlation*)

N : jumlah kasus atau responden

k : jumlah prediktor

Contoh 15.2:

Pada perhitungan di atas diperoleh koefisien korelasi ganda $R = 0.64$ dan jumlah responden $N = 60$, sedangkan jumlah prediktor adalah 2, sehingga dapat dihitung besarnya tes signifikansi sebagai berikut:

$$F = \frac{[0.64^2]}{1 - 0.64^2} \frac{[60 - 2 - 1]}{2}$$

$$F = \frac{[0.4096]}{0.5904} \frac{[57]}{2}$$

$$F = [0.694] [28.5]$$

$$F = 19.779$$

Derajat kebebasan (*degree of freedom*) dicari sebagai berikut:

- *degree of freedom satu* (df_1) = k

- derajat kebebasan dua (df_2) = (N - k - 1).

Berdasarkan harga kritik pada tabel F, yaitu sebesar 3.17 untuk taraf kepercayaan 95% dan 5.01 untuk taraf kepercayaan 99%.

Rumusan hipotesis nihil (H_0) dan hipotesis alternatif (H_a) yang diajukan adalah sebagai berikut:

H_0 : Tidak ada korelasi yang signifikan antara efektivitas organisasi dengan intensitas semangat kerja dan kelancaran komunikasi

H_a : Terdapat korelasi yang signifikan antara efektivitas organisasi dengan intensitas semangat kerja dan kelancaran komunikasi

Untuk pengujian hipotesis, dengan mengkonsultasikan besarnya F perhitungan dengan F dalam tabel.

Terlihat bahwa F perhitungan lebih besar dari F harga kritiknya (F perhitungan > F tabel), maka hipotesis nihil yang diajukan ditolak. Tertolakny hipotesis nihil tersebut, berarti hipotesis kerja atau hipotesis alternatifnya (H_a) diterima, sehingga dapat disimpulkan bahwa, efektivitas organisasi berkaitan secara signifikan dan sangat meyakinkan dengan intensitas semangat kerja serta kelancaran komunikasi. Atau dapat pula dikatakan bahwa, efektivitas organisasi dapat diprediksi dari intensitas semangat kerja para karyawannya dan tingkat kelancaran komunikasi dalam organisasi tersebut.

Pemahaman lebih lanjut mengenai pemaknaan korelasi ganda ini, adalah bahwa semakin kecil koefisien korelasi antarvariabel prediktornya (seperti dalam contoh di atas koefisien korelasi prediktornya adalah 0.17), hal ini akan diikuti dengan semakin kuatnya koefisien korelasi gandanya. Atau dengan arti lain bahwa, semakin efektif suatu organisasi dalam mencapai tujuan, hal ini akan diikuti dengan semakin kecilnya kadar korelasi antarvariabel prediktornya, yaitu variabel semangat kerja dan kelancaran komunikasi.

15.3. Korelasi Parsial

Korelasi parsial digunakan untuk menganalisis data penelitian yang mengetahui pengaruh atau mengetahui hubungan antara variabel independen dan dependen, dimana salah satu variabel independennya dibuat tetap/dikendalikan. Jadi koefisien korelasi parsial merupakan angka yang menunjukkan arah dan kuatnya hubungan antara dua variabel atau lebih setelah satu variabel yang diduga dapat mempengaruhi hubungan variabel tersebut dikendalikan untuk dibuat tetap keberadaannya.

Persyaratan yang harus dipenuhi jika hendak menggunakan analisis korelasi parsial adalah:

1. Data semuanya berskala interval
2. Desain penelitiannya adalah korelasional
3. Variabel penelitiannya lebih dari dua (2)

Rumus korelasi parsial antara X_1 dengan Y, bila variabel X_2 dikendalikan, atau korelasi antara X_1 dengan Y, bila variabel X_2 tetap adalah sebagai berikut:

$$R_{Y.X_1.X_2} = \frac{r_{YX_1} - r_{YX_2}r_{X_1X_2}}{\sqrt{(1-r_{X_1X_2}^2)(1-r_{YX_2}^2)}}$$

Rumus 15.3

Bila X_1 yang dikendalikan, maka rumusnya adalah:

$$R_{Y.X_1.X_2} = \frac{r_{YX_2} - r_{YX_1}r_{X_1X_2}}{\sqrt{(1-r_{X_1X_2}^2)(1-r_{YX_1}^2)}}$$

Rumus 15.4

Untuk pengujian hipotesis dilakukan transformasi ke t dengan rumus:

$$t = \frac{r\sqrt{n-3}}{\sqrt{1-r^2}}$$

Rumus 15.5

Contoh 15.3:

Jika diketahui koefisien korelasi antara IQ dengan nilai ujian statistik (r_{YX_1}) = 0,70, koefisien korelasi antara nilai ujian statistik dengan lamanya belajar (r_{YX_2}) = 0,40, dan koefisien korelasi antara IQ dengan lamanya belajar ($r_{X_1X_2}$) = -0,10. Tentukan: berapa koefisien korelasi antara IQ dengan nilai ujian statistik, bila waktu belajar diabaikan (diparsialkan).

Jawab:

$$\begin{aligned} R_{Y.X_1.X_2} &= \frac{r_{YX_1} - r_{YX_2}r_{X_1X_2}}{\sqrt{(1-r_{X_1X_2}^2)(1-r_{YX_2}^2)}} \\ &= \frac{0,70 - (-0,10).0,40}{\sqrt{(1-0,40^2)(1-(-0,10)^2)}} \\ &= \frac{0,70 - (-0,10).0,40}{\sqrt{(1-0,40^2)(1-(-0,10)^2)}} \\ &= \frac{0,70 - 0,040}{\sqrt{(1-0,16)(1-0,10)}} \end{aligned}$$

$$= \frac{0,70 + 0,040}{\sqrt{(0,84)(0,99)}}$$

$$= \frac{0,70 + 0,040}{\sqrt{0,8316}} = \frac{0,74}{0,911921} = 0,811$$

Jadi bila lama belajar diabaikan, maka korelasi antara IQ dengan nilai ujian statistik 0.81.

Untuk mengetahui apakah koefisien parsial tersebut signifikan atau tidak, maka diuji dengan rumus 15.5. Misalkan jumlah sampel yang diambil 25, maka:

$$t = \frac{r\sqrt{n-3}}{\sqrt{1-r^2}}$$

$$= \frac{0,811 \sqrt{25-3}}{\sqrt{1-(0,811)^2}}$$

$$= \frac{0,811 \sqrt{22}}{\sqrt{1-0,657721}}$$

$$= \frac{0,811 \cdot 4,6904158}{0,58504615}$$

$$= 6,5019$$

Selanjutnya t hitung ini dikonsultasikan/dibandingkan dengan nilai t tabel, dengan $db = n - 1 = 25 - 1$. Pada taraf kesalahan 5%, untuk uji dua pihak maka harga t tabel = 2,064. Maka t hitung lebih besar dari t tabel. Dengan demikian koefisien korelasi yang dihitung tersebut adalah signifikan. Dengan demikian dapat digeneralisasikan untuk seluruh populasi dari mana sampel tersebut diambil.

BAB XVI ANALISIS VARIANSI (ANAVA)

16.1. Pengertian

Analisis Variansi (ANAVA) merupakan salah satu metode yang digunakan dalam statistik. Ide dasar dari analisis variansi adalah kumpulan hasil pengamatan sesuatu hal yang nilai datanya bervariasi. Pada bab VII sudah dibicarakan tentang ukuran variabilitas, khususnya untuk satu populasi.

Analisis variansi digunakan untuk mendapatkan pemecahan terhadap masalah pada suatu eksperimen yang terdiri dari 2 atau lebih populasi ($k \leq 2$). Selain itu analisis variansi juga dapat digunakan untuk mengukur besarnya variasi-variasi yang terjadi di dalam eksperimen tersebut, dimana variasi-variasi yang terjadi sangat ditentukan oleh jenis pengamatan yang dilakukan dalam eksperimen tersebut.

Contoh 16.1:

Seseorang ingin melakukan penelitian terhadap 3 jenis varietas padi, yaitu IR 32, IR 36 dan VUTW yang masing-masing ditanam pada setiap Ha sawah. Kemudian hasil produksi ketiga jenis padi tadi dicatat satu persatu. Selanjutnya diajukan hipotesis yang menyatakan bahwa ketiga jenis variasi tersebut memiliki rata-rata produksi yang sama untuk setiap hektarnya. Untuk menguji ada tidaknya perbedaan rata-rata produksi tiap hektar dari ketiga jenis varitas padi tersebut secara serempak dapat kita gunakan analisis variansi.

Peranan analisis variansi pada prinsipnya ada dua, yaitu:

- Mengambil dan menghitung komponen variasi
- Pengujian apakah signifikan atau tidak

Analisis variansi ditampilkan dalam bentuk daftar ringkasan yang disebut Tabel Analisis Variasi (Tabel Anava). Tabel ini memuat kolom-kolom sebagai berikut:

1. Sumber Keragaman /SK= sumber variasi = *source of variation*
2. Derjat Bebas/db = *degree of freedom*
3. Jumlah Kuadrat/JK = *sum of square/SS*
4. Kuadrat Rerata/KR = *Mean Squares/MSS*
5. Nilai F perhitungan/ $F_{hit} = F_{Calculated} = F_{observed}$
6. Nilai F dari tabel $F_{ab} = F_{table}$

Untuk kebutuhan analisis variansi dipergunakan distribusi F, yang tabelnya tersedia dalam buku-buku statistik.

16.2. Analisis Variansi Satu Arah

Analisis satu arah atau disebut juga anava sederhana. Pada anava ini hasil penelitian disajikan dalam bentuk tabel sederhana, sebagai berikut:

Tabel 16.1

	Treatment				
	1	2	3	k
	X11	X12	X13	X1k
	X21	X22	X23	...	X2k

	Xn1	Xn2	Xn3	...	Xnk
Jumlah	T1	T2	T3	Tk

Tabel anavanya disajikan sebagai berikut:

Tabel 16.2

Sumber variasi	df	SS	MSS	F_{hit}	F_{ab}	
					5%	1%
Antara Treatment	k-1	SSTr	$MSTr = \frac{SSTr}{k-1}$	$MSSTr$	$F_{0,05;k-1,k(n-1)}$	$F_{0,01;k-1,k(n-1)}$
Dalam Treatment (Error)	k(n-1)	SSE	$MSE = \frac{SSE}{k(n-1)}$			
Total	nk-1	SST	-	-	-	-

Dimana:

$$SSTr = \frac{1}{n} \sum_{j=1}^k T_j^2 - \frac{G^2}{nk} \dots \dots \dots (16.1)$$

$$SST = \sum_{j=1}^k \sum_{i=1}^n X_{ij}^2 - \frac{G^2}{nk} \dots \dots \dots (16.2)$$

$$SSE = SST - SSTr \dots \dots \dots (16.3)$$

G = Grand Total menunjukkan total n X k pengamatan yang ada = $\sum_{j=1}^k \sum_{i=1}^n X_{ij}$

16.2.1. Pengujian Hipotesis

Pengujian hipotesis diawali dengan merumuskan hipotesis. Rumusan hipotesisnya adalah:

H_0 : sebanyak k mean adalah sama, artinya tidak terdapat perbedaan antara treatment

H_1 : paling sedikit 2 dari k mean populasi yang tidak sama

Pengujian hipotesis dilakukan dengan membandingkan nilai F_{hit} dengan

F_{ab} . Jika $F_{hit} > F_{ab}$, maka H_0 ditolak, berarti paling sedikit 2 dari k mean populasi tidak sama. Artinya ada efek dalam perbedaan treatment.

Jika $F_{hit} \leq F_{ab}$, maka H_0 diterima, dan berarti sebanyak k mean adalah sama, artinya tidak ada efek dalam perbedaan treatment.

Contoh 16.2:

Pemerintah Kota Pekalongan ingin mengetahui apakah ada perbedaan rata-rata penghasilan para pedagang kaki lima untuk empat lokasi di kota Pekalongan. Untuk keperluan tersebut, masing-masing lokasi diambil lima orang dan ditanyakan penghasilan sebulannya. (dalam puluhan ribu rupiah). Didapat hasil sebagai berikut:

Tabel 16.3

	Lokasi				Total
	A	B	C	D	
	22	25	26	26	
	26	27	29	28	
	25	28	33	27	
	25	26	30	30	
	31	29	33	30	
Jumlah	129	135	151	141	G= 556

Selanjutnya dibuat tabel anava sebagai berikut:

Tabel 16.4

Sumber variasi	df	SS	MSS	F _{hit}	F _{tab}	
					5%	1%
Lokasi (Treatment)	3	52,8	$\frac{52,8}{3} = 17,6$	$\frac{17,6}{6,28} = 2,80$	3,24	5,29
Error	16	100,4	$\frac{100,4}{16} = 6,28$			
Total	19	153,2	-	-	-	-

Dimana:

$$n = 5, k = 4$$

$$\begin{aligned} SSTr &= \frac{1}{5}(129^2 + 135^2 + 151^2 + 141^2) - \frac{556^2}{5,4} \\ &= \frac{77548}{5} - \frac{556^2}{5,4} \\ &= 15509,6 - 15456,8 \\ &= 52,8 \end{aligned}$$

$$\begin{aligned} SST &= \sum_{j=1}^k \sum_{i=1}^n X_{ij}^2 - \frac{G^2}{nk} \\ &= (22^2 + 26^2 + \dots + 30^2) - \frac{556^2}{5,4} \\ &= 153,2 \end{aligned}$$

$$\begin{aligned} SSE &= SST - SSTr \\ &= 153,2 - 52,8 \\ &= 100,4 \end{aligned}$$

$$MSTr = \frac{SSTr}{k-1} = \frac{52,8}{3} = 17,6$$

$$MSE = \frac{SSE}{k(n-1)} = \frac{100,4}{16} = 6,28$$

$$F_{hit} = \frac{MSTr}{MSE} = \frac{17,6}{6,28} = 2,80$$

H₀ : tidak terdapat perbedaan penghasilan pedagang antara satu lokasi dengan lokasi lainnya

H_a : paling sedikit 2 dari lokasi mempunyai penghasilan yang berbeda

F_{hit} : 2,80

Pada taraf signifikan 5% $F_{tab} = F_{0,05;k-1(n-1)} = F_{0,05;3,16} = 3,24$. Dengan demikian $F_{hit} \leq F_{tab}$, maka H₀ diterima, dan berarti tidak ada perbedaan rata-rata penghasilan pedagang kaki lima dari empat lokasi.

Pada taraf signifikan 1% $F_{tab} = F_{0,01;k-1(n-1)} = F_{0,01;3,16} = 5,29$. Dengan demikian $F_{hit} \leq F_{tab}$, maka H₀ diterima, dan berarti tidak ada perbedaan rata-rata penghasilan pedagang kaki lima dari empat lokasi.

Contoh 16.3:

Seorang ahli pendidikan ingin mengetahui IQ dari anak-anak asal desa petani, desa bukan petani dan kota pada suatu SD. Anak yang berasal dari desa petani diambil 5 orang, dari desa bukan petani 4 orang dan kota 6 orang yang setelah dites hasilnya sebagai berikut:

Tabel 16.5

Desa Petani	Desa Bukan Petani	Kota
128	80	130
90	90	120
130	110	80
125	100	90
130		122
		112
603	380	654

Penyelesaian:

Masing-masing kolom tidak sama, maka rumusnya mengalami perubahan sebagai berikut:

$$SSTr = \sum_{j=1}^k \frac{T_j^2}{n_j} - \frac{G^2}{N} \dots \dots \dots (16.4)$$

Dimana $N = \sum_{j=1}^k n_j$

$$SST = \sum_{j=1}^k \sum_{i=1}^n X_{ij}^2 - \frac{G^2}{N} \dots \dots \dots (16.5)$$

$$SSE = SST - SSTr$$

Nilai df yang bersesuaian untuk SSTr, SSE dan SST adalah k-1, N-k dan N-1. Untuk data di atas maka:

$$N = n_1 + n_2 + n_3 = 5 + 4 + 6 = 15$$

$$SSTr = \frac{1}{5}(603)^2 + \frac{1}{4}(380)^2 + \frac{1}{6}(654)^2 - \frac{1637^2}{15}$$

$$= 72721,8 + 36100 + 71286 - 178651,27$$

$$= 1456,53$$

$$SST = 128^2 + 90^2 + 130^2 + \dots + 112^2 - \frac{1637^2}{15} = 5085,73$$

$$SSE = 5085,73 - 1456,533629,2$$

Tabel anavanya adalah:

Tabel 16.6

Sumber variasi	df	SS	MSS	F _{hit}	F _{tab}	
					5%	1%
Antar Asal (antar treatment)	2	1456,53	$\frac{1456,53}{2} = 728,26$	$\frac{728,26}{302,43} = 2,41$	3,89	6,93
Error	12	3629,2	$\frac{3629,2}{12} = 302,43$			
Total	14	5085,73	-	-	-	-

H₀ : Tidak terdapat perbedaan rata-rata IQ antara anak-anak kota, desa petani dan desa bukan petani

H₁ : IQ anak-anak kota, desa petani dan desa bukan petani berbeda

$$F_{hi} = 2,41$$

Pada taraf signifikan 5% F_{tab} = F_{0,05,2,12} = 3,89. Dengan demikian F_{hit} ≤ F_{tab}, maka H₀ diterima, dan berarti tidak ada perbedaan rata-rata IQ antara anak-anak kota, anak desa petani dan anak desa bukan petani.

Pada taraf signifikan 1% F_{tab} = F_{0,01,2,12} = 6,93. Dengan demikian F_{hit} ≤ F_{tab}, maka H₀ diterima, dan berarti tidak ada perbedaan rata-rata IQ antara anak-anak kota, anak desa petani dan anak desa bukan petani.

16.3. Analisis Variansi Dua Arah

Dalam anava dua arah akan dilihat 2 variasi, yaitu variasi yang timbul karena treatment dan variasi yang timbul karena kategorisasi/blok. Dengan kata lain dalam analisis ini data observasi dibagi menjadi dua komponen yaitu komponen treatment dan komponen blok. Misalnya dalam kasus penanaman

padi, treatment yang digunakan adalah jenis padi (IR 32, IR 36 dan VUTW), sedang bloknnya misalnya penggunaan pupuk yang terdiri beberapa macam. Analisis variansi dua arah dibedakan menjadi dua, yaitu data yang tidak ada interaksi antara treatment dan blok dan data yang ada interaksi antara treatment dengan blok. Dalam kasus penanaman padi di atas, terdapat interaksi antara jenis padi dengan pupuk yang digunakan, sehingga anava akan mencakup tiga hipotesis, ialah pengaruh antarjenis padi adalah nol, pengaruh antarjenis pupuk adalah nol dan pengaruh antarinteraksi adalah nol.

16.3.1. Anova Tanpa Interaksi

Contoh 16.4:

Suatu penelitian dilakukan terhadap siswa yang mempunyai orang tua dengan latar belakang ekonomi berbeda. Penelitian dimaksudkan untuk melihat skor pemahaman terhadap suatu materi pelajaran yang diberikan dengan menggunakan 3 metode yang berlainan pada 3 kelompok siswa yang masing-masing terdiri dari 4 siswa. Hasilnya skor pemahaman masing-masing siswa adalah seperti pada tabel 16.7. Selidikilah apakah terdapat perbedaan/pengaruh latar belakang pekerjaan orang tua, juga apakah ada perbedaan/pengaruh metode pembelajaran terhadap pemahaman siswa.

Tabel 16.7

Pekerjaan Orang Tua (Blok)	Metode Pembelajaran			Jumlah Blok
	Ceramah	Ceramah dan Diskusi	Diskusi	
Petani	7,0	14,0	8,5	29,5
Pedagang	16,5	15,5	16,5	48,0
Buruh	10,5	15,0	9,5	35,0
PNS	13,5	21,0	13,5	48,0
Jumlah Treatment (T)	47,0	65,5	48,0	G=160,5

Penyelesaian:

Kita lakukan perhitungan-perhitungan dengan menggunakan rumus-rumus sebagai berikut;

$$SST = \sum_{j=1}^k \sum_{i=1}^n X_{ij}^2 - \frac{G^2}{nk} \dots \dots \dots (16.6)$$

$$SSTr = \frac{1}{n} \sum_{j=1}^k T_j^2 - \frac{G^2}{nk} \dots \dots \dots (16.7)$$

$$SSB = \frac{1}{n} \sum_{j=1}^k B_j^2 - \frac{G^2}{nk} \dots (16.8)$$

$$\text{Error} = SST - SSTR - SSB \dots (16.9)$$

Maka:
 $SST = 7^2 + 14^2 + (8,5)^2 + \dots + (13,5)^2 - \frac{(160,5)^2}{4,3}$

$$= 2316,75 - 2146,69 = 170,06$$

$$SSTR = \frac{1}{4} (47^2 + 65,5^2 + 48^2) - 2146,69 = 54,12$$

$$SSB = \frac{1}{3} (29,5^2 + 48^2 + 35^2 + 48^2) - 2146,69 = 87,73$$

$$\text{Error/Residu} = SST - SSTR - SSB = 170,06 - 54,12 - 87,73 = 28,21$$

Selanjutnya dibuat tabel anava sebagai berikut:

Tabel 16.8

Sumber Variasi	df	SS	MSS	Fhit	Ftab	
					5%	1%
Antar Metode Pembelajaran (Treatment)	k-1 = 3-1 = 2	54,12	$\frac{54,12}{2} = 5,75$	$\frac{27,06}{4,702} = 5,75$	5,74	10,92
Antar Pekerjaan Orang tua (Blok)	n-1 = 4-1 = 3	87,73	$\frac{87,73}{3} = 29,24$	$\frac{29,24}{4,702} = 6,22$	4,76	9,76
Error/Residu	(k-1)(n-1) = 6	28,21	$\frac{28,21}{6} = 4,702$	-	-	-
TOTAL	11	170,06	-	-	-	-

H_{01} : tidak terdapat perbedaan skor pemahaman karena perbedaan metode pembelajaran

H_{a1} : terdapat perbedaan skor pemahaman karena perbedaan metode pembelajaran

H_{02} : tidak terdapat perbedaan skor pemahaman karena perbedaan pekerjaan orang tua

H_{a2} : terdapat perbedaan skor pemahaman karena perbedaan pekerjaan orang tua

Untuk taraf signifikan 5% ternyata $F_{Treatment} = FT$ dan untuk $F_{Blok} = FB$ masing-masing lebih besar dari F_{tab} sehingga hipotesis nol ditolak. Artinya terdapat perbedaan karena pengaruh metode belajar dan pekerjaan orang tua.

16.3.2. Anava Dengan Interaksi

Untuk melihat adanya interaksi antara *treatment* dengan blok, dilakukan replikasi/pengulangan sebanyak r kali dalam setiap kombinasi dan blok.

Contoh 16.5:

Suatu penelitian tentang berat badan (dalam kg) bagi anak umur 5 tahun menurut suku bangsa dan wilayah dibesarkan. Hasilnya adalah sebagaimana tercantum dalam tabel 16.9. Masing-masing dari suku bangsa dan wilayah dibesarkan ditimbang 3 anak. Berdasarkan data tersebut, buatlah anava tentang pengaruh antarwilayah dibesarkan, antarsuku bangsa dan interaksi antarwilayah dan suku bangsa.

Tabel 16.9

Dibesarkan di	Suku Bangsa			
	Jawa	Batak	Sunda	Bali
Kota	16,13,19	18,17,21	14,16,13	13,14,16
Desa	9,15,11	15,13,12	7,12,9	3,1,9
Desa Lalu Kota	22,25,17	14,16,12	11,14,12	13,17,14
Kota Lalu Desa	14,16,12	20,16,17	8,11,10	6,8,12

Penyelesaian:

Terlebih dahulu dibuat tabel anava sebagai berikut:

Tabel 16.10

Sumber Variasi	df	SS	MSS	F _{hit}	F _{tab}	
					5%	1%
Antar Treatment	k-1	SSTR	MSTR	$\frac{MSTR}{MSE}$		
Antar Blok	n-1	SSB	MSB	$\frac{MSB}{MSE}$		
Interaksi	(k-1)(n-1)	SSI	MSI	$\frac{MSI}{MSE}$		
Error/Residu	nk(r-1)	SSE	MSE	-		
Total	rnk-1	SST	-	-		

Dimana:

$$SST = \sum_i \sum_k \sum_r X^2 - \frac{G^2}{mk} \dots \dots \dots (16.10)$$

$$SSI = \frac{1}{n} \sum_i \sum_k \left(\sum_r X \right)^2 - \frac{G^2}{mk} - SSTr - SSB \dots \dots \dots (16.11)$$

$$SSTr = \frac{1}{nr} \sum_k T^2 - \frac{G^2}{mk} \dots \dots \dots (16.12)$$

$$SSB = \frac{1}{kr} \sum_i B^2 - \frac{G^2}{mk} \dots \dots \dots (16.13)$$

$$G = \sum_i \sum_k \sum_r X \dots \dots \dots (16.14)$$

Dari tabel 16.9 maka k = 4, n = 4 dan r = 3. Selanjutnya dibuat tabel sebagai berikut:

Tabel 16.11

Dibesarkan di	Suku Bangsa / T				Jumlah B
	Jawa	Batak	Sunda	Bali	
Kota	16,13,19	18,17,21	14,16,13	13,14,16	190
Desa	9,15,11	15,13,12	7,12,9	3,1,9	116
Desa Lalu Kota	22,25,17	14,16,12	11,14,12	13,17,14	187
Kota Lalu Desa	14,16,12	20,16,17	8,11,10	6,8,12	150
Jumlah T	189	191	137	126	643=G

$$SST = 16^2 + 13^2 + 19^2 + 18^2 + 17^2 + 21^2 + \dots + 6^2 + 8^2 + 12^2 - \frac{(643)^2}{3.4.4} = 977,5$$

$$SSTr = \frac{1}{4.3} (189^2 + 191^2 + 137^2 + 126^2) - \frac{643^2}{3.4.4} = 290,4$$

$$SSB = \frac{1}{4.3} (190^2 + 116^2 + 187^2 + 150^2) - \frac{643^2}{3.4.4} = 305,3$$

$$SSI = \frac{1}{3} ((16 + 13 + 19)^2 + (18 + 17 + 21)^2 + \dots + (3 + 1 + 9)^2 + (22 + 25 + 17)^2 + \dots + (6 + 8 + 12)^2) - \frac{643^2}{3.4.4} - 290,4 - 305,3 = 181,1$$

$$SSE = 977,5 - 290,4 - 305,3 - 181,1 = 200,7$$

$$MST = SSTr / (k-1) = 290,4 / (4-1) = 96,8$$

$$MSB = SSB / (n-1) = 305,3 / (4-1) = 101,8$$

$$MSI = SSI / (k-1)(n-1) = 181,1 / (4-1)(4-1) = 20,1$$

$$\begin{aligned} MSE &= SSE / nk(r-1) = 200,7 / 4.4(3-1) = 6,27 \\ F_{hitT} &= MSTr / MSE = 96,8 / 6,27 = 15,4 \\ F_{hitB} &= MSB / MSE = 101,8 / 6,27 = 16,2 \\ F_{hitI} &= MSI / MSE = 20,1 / 6,27 = 3,2 \end{aligned}$$

Dari hasil perhitungan tersebut akan dihasilkan tabel anava sebagai berikut

Tabel 16.11

Sumber Variasi	df	SS	MSS	F _{hit}	F _{tab}	
					5%	1%
Antar Suku Bangsa (Treatment)	3	290,4	96,8	15,4	2,90	4,46
Antara Wilayah (Blok)	3	305,3	101,8	16,2	22,90	4,46
Interaksi	9	181,1	20,1	3,2	2,19	3,01
Error/ Residu	32	200,7	6,27	-		
Total	47	977,5	-	-		

- H_{0T} : tidak terdapat perbedaan berat badan karena perbedaan suku bangsa
- H_{1T} : terdapat perbedaan berat badan karena perbedaan suku bangsa
- H_{0B} : tidak terdapat perbedaan berat badan karena perbedaan wilayah
- H_{1B} : terdapat perbedaan berat badan karena perbedaan wilayah
- H_{0T} : tidak terdapat perbedaan berat badan karena adanya interaksi antara suku bangsa dan wilayah
- H_{1T} : terdapat perbedaan berat badan karena adanya interaksi antara suku bangsa dan wilayah

Karena F_{hitT}, F_{hitB} dan F_{hitI} masing-masing lebih besar dari nilai-nilai F_{tab}. Baik pada taraf signifikan 5% maupun 1%, maka H₀ tidak dapat diterima. Jadi masing-masing suku bangsa, wilayah tempat membesarkan dan interaksi adalah signifikan pada berat badan anak.

Latihan Bab XVI:

- Empat macam campuran makanan diberikan kepada kambing dalam rangka percobaan untuk meningkatkan pertambahan berat dagingnya. Setelah percobaan selesai, pertambahan berat dagingnya dicatat dan hasilnya sebagai berikut

Tabel 16.12

	Pertambahan Berat Karena Makanan Ke			
	1	2	3	4
Data Hasil Pengamatan	12 20 23 10 17	14 15 10 19 22	6 16 16 20	9 14 18 19
Jumlah	82	80	58	60
Rata-rata	16,4	16,0	14,5	15,0

Selidikilah apakah perbedaan makanan menyebabkan perbedaan pertambahan berat badan?

- Di desa Sugihwaras diadakan wawancara tentang KB, yaitu berkaitan dengan besar keluarga yang diinginkan menurut pendidikannya. Didapat data sebagai berikut:

Tabel 16.13

SD	SMP	SMA
5	3	1
3	2	1
4	4	2
3	1	2
3	3	1
2	2	2

Ujilah apakah ada pengaruh antara pendidikan terhadap. Keinginan besar keluarga?

- Pada suatu percobaan terhadap 20 siswa, dilakukan pembagian menjadi 5 kelompok belajar. Masing-masing kelompok terdiri dari 4 mahasiswa. Terhadap setiap kelompok diajukan materi yang sama dan diberlakukan metode yang berbeda. Setelah dilakukan tes, didapat hasil skor sebagai berikut:

Tabel 16.14

Kelompok Belajar	Metode			
	1	2	3	4
1	7	1	6	15
2	8	4	4	11
3	6	5	9	16
4	3	2	6	15
5	4	3	10	13

Buatlah anava dan ambillah kesimpulan mengenai pengaruh antarkelompok belajar dan pengaruh antarmetode mengajar terhadap hasil tes tersebut!

BAB XVII

ANALISIS REGRESI

17.1 Pendahuluan

Seringkali kita ingin melihat hubungan antargejala atau variabel, seperti hubungan antara panjang dan bobot bayi, tingkatan pendidikan ibu dan gizi balita, dan pendapatan dengan pengeluaran seseorang. Untuk mengetahui ada tidaknya hubungan antarvariabel digunakan teknik analisis korelasional yang telah dibahas dalam bab-bab sebelumnya.

Bila antarvariabel terdapat hubungan, maka suatu variabel bersifat mempengaruhi variabel lainnya. Variabel yang mempengaruhi disebut variabel bebas (*independent variable*), sedangkan variabel yang dipengaruhi disebut variabel tak bebas (*dependent variable*). Secara kuantitatif hubungan antara variabel bebas dan variabel tak bebas tersebut dapat dimodelkan dalam suatu bentuk matematik, sehingga kita dapat menduga nilai suatu variabel tak bebas bila diketahui nilai variabel bebas. Persamaan matematik yang menggambarkan hubungan antara variabel bebas dan tak bebas sering dengan persamaan regresi.

Regresi dan korelasi mempunyai hubungan yang sangat erat. Setiap regresi pasti ada korelasinya, tetapi korelasi belum tentu dilanjutkan dengan regresi. Korelasi yang bisa dilanjutkan dengan regresi adalah bila hubungan antarvariabel mempunyai hubungan kausalitas/sebab akibat.

Regresi pertama kali dipergunakan sebagai konsep statistik pada tahun 1877 oleh Sir Francis Galton. Dia telah melakukan studi tentang kecenderungan tinggi badan anak yang lahir terhadap orang tuanya adalah menurun (regress) mengarah pada tinggi badan rata-rata penduduk.

Istilah regresi pada mulanya bertujuan untuk membuat perkiraan nilai satu variabel (tinggi badan anak) terhadap satu variabel yang lain (tinggi badan orang tua). Pada perkembangan selanjutnya, analisis regresi dapat digunakan

sebagai alat untuk membuat perkiraan nilai suatu variabel dengan menggunakan beberapa variabel lain yang berhubungan dengan variabel tersebut.

Analisis regresi digunakan untuk membuat perkiraan nilai suatu variabel dengan menggunakan beberapa variabel lain yang berhubungan dengan variabel tersebut. Dalam analisis regresi, untuk mengetahui hubungan antara variabel yang satu dengan variabel lain menggunakan suatu persamaan estimasi atau persamaan regresi. Persamaan estimasi adalah suatu formula matematis yang menunjukkan hubungan keterkaitan antara satu atau beberapa variabel yang nilainya sudah diketahui (*known variabel*) dengan variabel yang nilainya belum diketahui (*unknown variabel*). Persamaan estimasi yang diperoleh dari regresi menunjukkan pola hubungan (*the pattern of relationship*) antara variabel independen dengan variabel dependen. Selanjutnya untuk mengetahui derajat hubungan antarvariabel dapat digunakan analisis korelasi. Analisis korelasi dapat digunakan untuk mengetahui seberapa baik suatu persamaan estimasi dapat digunakan untuk menjelaskan hubungan antarvariabel tersebut.

Dilihat dari banyaknya variabel (bebas) independen, persamaan regresi dapat terdiri dari satu variabel bebas dan satu variabel tak bebas, atau beberapa variabel bebas dengan satu variabel tak bebas. Persamaan yang *pertama* disebut regresi sederhana misalnya persamaan yang menggambarkan hubungan antara tinggi bapak dengan tinggi anak laki-lakinya. Dalam contoh ini tinggi bapak sebagai variabel bebas dan tinggi anak sebagai variabel tak bebas.

Persamaan *kedua* disebut regresi berganda, misalnya hubungan antara variabel bebas tingkat pendidikan, pendapatan, dan jumlah anak terhadap variabel tak bebas pengeluaran konsumsi keluarga.

Regresi sederhana dan regresi berganda dapat berbentuk linear atau pun non linear. Regresi sederhana yang linear disebut regresi linear sederhana. Pada regresi ini, kurva persamaan regresinya berbentuk garis lurus. Model regresi sederhana yang non linear dapat berupa kuadrat, logaritmik, eksponensial atau pun lainnya. Regresi berganda yang linear disebut regresi linear berganda.

17.2 Regresi Linear Sederhana

Regresi linear sederhana adalah persamaan regresi yang menggambarkan hubungan antara satu variabel bebas (X) dan satu variabel tak bebas (Y), dimana hubungan keduanya dapat digambarkan sebagai suatu garis lurus. Sehingga hubungan kedua variabel tersebut dapat dituliskan dalam bentuk persamaan linear.

Bentuk umum persamaan linear sederhana yang menunjukkan hubungan antara dua variabel, yaitu variabel X dan Y adalah:

$$Y = a + bX \dots\dots\dots (17.1)$$

Dimana:

Y adalah variabel dependen

a adalah intersep (titik potong kurva terhadap sumbu Y)

b adalah gradien/kemiringan kurva linear, disebut juga sebagai koefisien regresi sederhana

X adalah variabel independen

Pada analisis regresi, hubungan antara variabel dependen (Y) dengan variabel independen (X) dinyatakan sebagai persamaan regresi yang merupakan persamaan estimasi linear. Persamaan tersebut dinyatakan sebagai berikut:

$$\hat{Y} = a + bX \dots\dots\dots (17.2)$$

dimana:

\hat{Y} (Y hat/ Y topi) adalah nilai estimasi Y

a = intersep kurva estimasi/konstanta

b = gradien/kemiringan kurva estimasi disebut juga sebagai koefisien regresi

X = nilai X

Perlu diingat, hubungan antara X dan Y hanya merupakan estimasi linear, bukan hubungan yang sebenarnya. Karena dalam kenyataannya hubungan tersebut tidak sepenuhnya linear, melainkan ada factor error (e). Secara pasti hubungan antara Y dan X dinyatakan:

$$Y = a + bX + e \dots\dots\dots (17.3)$$

Dimana e adalah error

Nilai a dan b persamaan (16.2) ditentukan dengan metode kuadrat terkecil (*least square methode*). Nilai a dan b didapat dengan rumus:

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \dots\dots\dots (17.4)$$

$$a = \bar{Y} - b\bar{X} \dots\dots\dots (17.5)$$

dimana:

\bar{Y} = nilai rata-rata Y

\bar{X} = nilai rata-rata X

n = jumlah data yang digunakan sebagai sampel

Contoh 17.1:

Seorang pengusaha ingin mengetahui hubungan antara nilai penjualan (Y) dengan biaya promosi (X). Data nilai penjualan (Y) dan biaya promosi (X) terangkum dalam tabel 17.1. Buatlah persamaan regresi linear yang menunjukkan hubungan antara nilai penjualan (Y) dengan biaya promosi (X).

Tabel 17.1

Nilai Penjualan (Y)	Biaya Promosi (X)
64	20
61	16
84	34
70	23
88	27
92	32
72	18
77	22

Penyelesaian:

Persamaan regresi linearnya adalah:

$$\hat{Y} = a + bX, \text{ dimana:}$$

\hat{Y} = estimasi/taksiran nilai penjualan pada biaya promosi (X) tertentu

a = Nilai penjualan bila biaya promosi nol

b = gradien/kemiringan kurva estimasi = perubahan nilai penjualan apabila terjadi perubahan satu satuan biaya promosi

X = Biaya promosi

Untuk menentukan nilai a dan b digunakan rumus 17.4 dan 17.5 dan dibuat table perhitungan sebagai berikut:

Tabel 17.2

Tabel Perhitungan

Nilai Penjualan (Y)	Biaya Promosi (X)	XY	X ²
64	20	1280	400
61	16	976	256
84	34	2856	1156
70	23	1610	529
88	27	2376	729
92	32	2944	1024

72	18	1296	324
77	22	1694	484
$\sum Y = 608$	$\sum X = 192$	$\sum XY = 15032$	$\sum X^2 = 4902$

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$= \frac{8(15032) - (192)(608)}{8(4902) - (192)^2}$$

$$= \frac{3520}{2352}$$

$$= 1,4965 \approx 1,5$$

$$a = \bar{Y} - b\bar{X}$$

$$= \frac{\sum Y}{n} - b \frac{\sum X}{n}$$

$$= \frac{608}{8} - 1,5 \frac{192}{8}$$

$$= 76 - 1,5(24)$$

$$= 40$$

Untuk mendapatkan nilai a dan b dapat pula digunakan *software* SPSS.

Dengan SPSS didapat hasil sebagai berikut:

Independent Variable: X

Coefficients

Model	Unstandardized Coefficients		Sig.
	B		
1 (Constant)	40.082		.004
X	1.497		.006

a Dependent Variable: Y

Dari perhitungan tersebut didapat nilai a = constant = 40,082 = 40, dan nilai koefisien X = b = 1,497 = 1,5.

Jadi persamaan regresinya adalah

$$\hat{Y} = 40 + 1,5X$$

17.2.1. Pengujian terhadap Koefisien regresi

Persamaan regresi yang diperoleh dalam suatu proses perhitungan tidak selalu tepat. Untuk itu perlu dilakukan analisis persamaan regresi. Dalam analisis

ini dilakukan pengujian terhadap koefisien regresi. Pengujian dilakukan untuk mengetahui apakah variabel independen berpengaruh secara signifikan terhadap variabel dependen. Pengujian dilakukan dengan langkah-langkah sebagai berikut:

a. Menghitung Kesalahan Standar Estimasi

Kesalahan standar estimasi (Se) menunjukkan ketepatan persamaan estimasi untuk menjelaskan nilai variabel dependen yang sesungguhnya. Semakin kecil nilai kesalahan standar estimasi, semakin tinggi ketepatan persamaan estimasi yang dihasilkan untuk menjelaskan nilai variabel dependen yang sesungguhnya. Sebaliknya semakin besar nilai kesalahan standar estimasi, semakin rendah ketepatan persamaan estimasi yang dihasilkan untuk menjelaskan nilai variabel dependen yang sesungguhnya.

Kesalahan standar estimasi didapat dengan rumus:

$$Se = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{N - 2}} \dots\dots\dots(17.6)$$

Contoh 17.2:

Tentukan kesalahan standar estimasi pada contoh 17.1

Jawab:

Dibuat tabel perhitungan sebagai berikut:

Tabel 17.3
Tabel Perhitungan

Nilai Penjualan (Y)	Biaya Promosi (X)	XY	X ²	Y ²
64	20	1280	400	4096.00
61	16	976	256	3721.00
84	34	2856	1156	7056.00
70	23	1610	529	4900.00
88	27	2376	729	7744.00
92	32	2944	1024	8464.00
72	18	1296	324	5184.00
77	22	1694	484	5929.00
$\sum Y = 608$	$\sum X = 192$	$\sum XY = 15032$	$\sum X^2 = 4902$	$\sum Y^2 = 47094,00$

$$Se = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{N - 2}}$$

$$= \sqrt{\frac{\sum 47094 - (40)(608) - (1,5)(15032)}{8 - 2}}$$

$$= \sqrt{\frac{226}{6}} = 6.14$$

b. Merumuskan Hipotesis

Rumusan hipotesisnya adalah:

H₀ : β = 0 : variabel independen (X) tidak berpengaruh terhadap variabel dependen (Y)

H_a : β ≠ 0 : variabel independen (X) berpengaruh secara signifikan terhadap variabel dependen (Y)

c. Menentukan Nilai t Test (t Hitung)

Nilai t test ditentukan dengan rumus sebagai berikut:

$$t_{hit} = \frac{b - \beta}{S_b} \dots\dots\dots(17.7)$$

Dimana:

b = koefisien regresi

β = 0, karena pada perumusan hipotesis nol (H₀), β = 0

S_b adalah kesalahan standar koefisien regresi, ditentukan dengan rumus:

$$S_b = \frac{S_e}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}} \dots\dots\dots(17.8)$$

d. Menentukan Nilai t Table

Nilai t table ditentukan dengan derajat kebebasan (*degree of freedom*) dan tingkat signifikansi tertentu. Derajat kebebasan ditentukan dengan rumus:

$$db = N - 2 \dots\dots\dots(17.9)$$

Tingkat signifikansi dapat 1% atau pun 5%.

e. Membandingkan Nilai t Test dengan t Table

Jika, |t_{hit}| ≥ t_{table} maka H₀ ditolak, H_a diterima. Maka disimpulkan variabel independen berpengaruh secara signifikan terhadap variabel dependen.

Jika $|t_{\text{hit}}| < t_{\text{tabel}}$, maka H_0 diterima, H_1 ditolak. Maka disimpulkan variabel independen tidak berpengaruh secara signifikan terhadap variabel dependen

Contoh 17.3.

Pada contoh di atas, selidikilah, apakah variabel independen (X) berpengaruh secara signifikan terhadap variabel dependen (Y)?

Jawab:

- Dari hasil perhitungan, $S_e = 6,14$
- Rumusan Hipotesis

Rumusan hipotesisnya adalah:

$H_0: \beta = 0$: variabel independen (X) tidak berpengaruh terhadap variabel dependen (Y)

$H_1: \beta \neq 0$: variabel independen (X) berpengaruh secara signifikan terhadap variabel dependen (Y)

c. Menghitung t_{hit}

$$S_b = \frac{S_e}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}}$$

$$S_b = \frac{6,14}{\sqrt{4902 - \frac{(192)^2}{8}}}$$

$$= \frac{6,14}{17,15}$$

$$= 0,359$$

$$t_{\text{hit}} = \frac{1,5 - 0}{0,359} = 4,18$$

d. Menentukan Nilai t Table

$$db = N - 2 = 8 - 2 = 6$$

Pada tingkat signifikan 1% nilai $t_{\text{tabel}} = t_{\left(\frac{0,01}{2}\right)} = t_{(6,0,005)} = 3,707$

Pada tingkat signifikan 5% nilai $t_{\text{tabel}} = t_{\left(\frac{0,05}{2}\right)} = t_{(6,0,025)} = 2,447$

e. Membandingkan t_{hit} dengan t_{tabel}

Dari hasil perhitungan telah didapat $t_{\text{hit}} = 4,18$

Pada tingkat signifikan 1%

Nilai $t_{\text{tabel}} = t_{(6,0,005)} = 3,707$, maka:
 $t_{\text{hit}} = 4,18 > t_{\text{tabel}} = 3,707$, sehingga H_0 ditolak, H_1 diterima. Maka disimpulkan variabel independen berpengaruh secara signifikan terhadap variabel dependen.

Pada tingkat signifikan 5%

Nilai $t_{\text{tabel}} = t_{(6,0,025)} = 2,477$, maka:
 $t_{\text{hit}} = 4,18 > t_{\text{tabel}} = 2,477$, sehingga H_0 ditolak, H_1 diterima. Maka disimpulkan variabel independen berpengaruh secara signifikan terhadap variabel dependen.

17.2.2. Penaksiran Nilai Variabel Dependen

Bila persamaan regresi sudah didapat, maka kita dapat menaksir nilai variabel dependen untuk suatu nilai variabel independen tertentu.

Contoh 17.4:

Bila nilai biaya promosi (X) = 40 maka berapa estimasi nilai produksi (Y), pada contoh di atas?

Penyelesaian:

Persamaan regresinya adalah:

$$\hat{Y} = 40 + 1,5X$$

Untuk biaya promosi (X) = 40 maka:

$$\hat{Y} = 40 + 1,5(40)$$

$$= 40 + 60$$

$$= 100$$

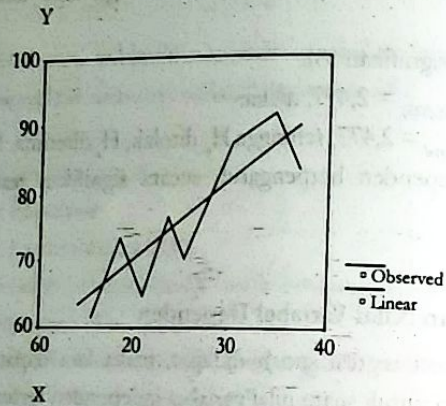
Jadi dengan biaya promosi 40, maka nilai produksinya diperkirakan 100.

17.2.3. Kurva Pendekatan Untuk Y / Garis Regresi

Pasangan (X, Y) dapat digambarkan pada bidang kartesius sebagai titik-titik. Bila titik-titik tersebut dihubungkan, akan membentuk kurva. Kurva tersebut merupakan kurva yang sebenarnya, dan belum membentuk garis linear. Garis linear dapat dibuat sebagai kurva pendekatan yang merupakan grafik dari persamaan regresi. Grafik tersebut disebut pula garis regresi.

Pada contoh di atas, kurva pendekatan untuk Y dan kurva sebenarnya dari sebaran data digambarkan pada gambar 17.1.

Gambar 17.1
Kurva Pendekatan untuk Y



Garis lurus menunjukkan kurva pendekatan/kurva regresi

17.2.4. Menghitung Koefisien Korelasi

Seperti diungkapkan di depan, setiap regresi pasti ada korelasinya. Indeks korelasi (r) ditentukan dengan rumus:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\} \{N\sum Y^2 - (\sum Y)^2\}}} \dots \dots \dots (17.10)$$

Contoh 17.5:

Pada contoh 17.1 hitung koefisien korelasinya!

Jawab:

Table Perhitungannya adalah sebagai berikut:

Tabel 17.4
Tabel Perhitungan

Nilai Penjualan (Y)	Biaya Promosi (X)	XY	Y ²	X ²
64	20	1280	4096	400
61	16	976	3721	256
84	34	2856	7056	1156
70	23	1610	4900	529

88	27	2376	7744	729
92	32	2944	8464	1024
72	18	1296	5184	324
77	22	1694	5929	484
$\sum Y = 608$	$\sum X = 192$	$\sum XY = 15032$	$\sum Y^2 = 47094$	$\sum X^2 = 4902$

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\} \{N\sum Y^2 - (\sum Y)^2\}}}$$

$$r = \frac{8(15032) - (192)(608)}{\sqrt{\{8(4902) - (192)^2\} \{8(47094) - (608)^2\}}}$$

$$r = \frac{120256 - 116736}{\sqrt{\{39216 - 36864\} \{376752 - 369664\}}}$$

$$r = \frac{3520}{\sqrt{\{2352\} \{7088\}}}$$

$$r = \frac{3520}{\sqrt{16670976}}$$

$$r = \frac{3520}{4083,01065}$$

$$= 0.862$$

Jadi koefisien korelasi antara biaya promosi dengan nilai penjualan adalah 0.862.

Koefisien determinasi = $R^2 = 0.743$, berarti biaya promosi mempunyai pengaruh terhadap nilai produksi sebesar 74,3 %. Sedangkan sisanya, sebesar 25,7% dipengaruhi oleh faktor-faktor lain.

Nilai koefisien korelasi dan koefisien determinasi dapat dicari dengan *software* SPSS, dengan hasil sebagai berikut:

Model	R	R Square	Std. Error of the Estimate
1	.862	.743	6.15761

a Predictors: (Constant), X

17.3. Regresi Linear Berganda

Model regresi linear sederhana yang telah dibahas, digunakan pada model yang menggunakan satu variabel independen (X) untuk mengestimasi nilai variabel dependen (Y). Bila terdapat lebih dari satu variabel independen, yaitu X1, X2, X3, ..., Xp maka digunakan model regresi linear berganda. Model ini banyak digunakan dalam kasus ekonomi, bisnis, pendidikan atau pun ilmu social secara umum.

Model linear berganda mengasumsikan adanya hubungan fungsional antara variabel dependen (Y) dengan variabel-variabel independen sebanyak N (X1, X2, X3, ..., Xp). Hubungan fungsional antara variabel independen (Y) dengan variabel-variabel dependen dinyatakan dengan:

$$Y = f(X_1, X_2, X_3, \dots, X_p) \dots\dots\dots (17.11)$$

Pada regresi linear berganda, persamaan regresinya adalah:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_p X_p \dots\dots\dots (17.12)$$

dimana:

- \hat{Y} = nilai estimasi Y
- a = konstanta = nilai \hat{Y} bila semua nilai variabel independen 0
- $b_1, b_2, b_3, \dots, b_p$ = koefisien-koefisien regresi
- $X_1, X_2, X_3, \dots, X_p$ = variabel-variabel independen

17.3.1. Regresi Linear dengan 2 Variabel Independen

Bila regresi linear dengan hanya dua variabel independen maka persamaan regresinya adalah:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 \dots\dots\dots (17.13)$$

dimana:

- \hat{Y} = nilai estimasi Y
- a = konstanta = nilai \hat{Y} bila semua nilai variabel independen 0
- b_1, b_2 = koefisien-koefisien regresi
- X_1, X_2 = variabel-variabel independen

Hubungan antara X dan Y hanya merupakan estimasi, bukan hubungan yang sebenarnya. Karena dalam kenyataannya hubungan tersebut tidak sepenuhnya linear, melainkan ada *factor error* (e). Secara pasti hubungan antara Y dan X dinyatakan:

$$Y = a + b_1 X_1 + b_2 X_2 + e \dots\dots\dots (17.14)$$

Dimana e adalah *error*

17.3.1. 1. Menentukan Persamaan Regresi Linear dengan 2 Variabel Independen

Persamaan regresi linear dengan 2 variabel independen mempunyai bentuk $= a + b_1 X_1 + b_2 X_2$.

Nilai a, b₁ dan b₂ persamaan tersebut ditentukan dengan sistem persamaan:

$$\sum Y = Na + b_1 \sum X_1 + b_2 \sum X_2 \dots\dots\dots (17.15.a)$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 \dots\dots\dots (17.15.b)$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \dots\dots\dots (17.15.c)$$

Besar nilai a, b₁ dan b₂ dapat dicari dengan menyelesaikan system persamaan 3 variabel, dengan tiga persamaan tersebut. Metode yang bisa digunakan adalah metode substitusi, eliminasi atau pun metode *cramer*.

Contoh 17.6:

Suatu penelitian bertujuan untuk melihat faktor-faktor yang mempengaruhi keikutsertaan KB (Y), yaitu penghasilan kurang (X₁) dan banyaknya anak (X₂). Penelitian dilakukan di 22 desa, dan didapat catatan persentasi penduduk yang mengikuti KB (Y), persentasi yang berpenghasilan kurang (X₁), dan rata-rata banyaknya anak (X₂) sebagai berikut:

Tabel 17.5

Data Persentasi Penduduk yang mengikuti KB (Y), Persentasi Penduduk yang Berpenghasilan Kurang (X₁), dan Rata-rata Banyak Anak Penduduk (X₂) di 22 Desa

Variabel No. Desa	Y	X1	X2
1	27.00	57.00	3.00
2	34.00	93.00	2.85
3	27.00	79.00	3.20
4	24.00	26.00	2.49
5	35.00	69.00	3.07
6	18.00	24.00	2.38
7	33.00	76.00	3.74
8	39.00	61.00	2.62
9	35.00	82.00	2.53
10	25.00	29.00	3.17
11	29.00	36.00	3.24

12	13.00	24.00	2.38
13	13.00	7.00	2.29
14	40.00	88.00	3.72
15	23.00	57.00	2.65
16	20.00	20.00	2.56
17	25.00	3.00	2.97
18	21.00	54.00	3.24
19	19.00	64.00	1.84
20	17.00	76.00	2.54
21	35.00	89.00	3.66
22	17.00	18.00	2.80

Dari data tersebut tentukan persamaan regresi linearnya.

Penyelesaian:

Terlebih dahulu dibuat tabel kerja/tabel perhitungan sebagai berikut:

Tabel 17.6
Tabel Perhitungan

No	Y	X ₁	X ₂	X ₁ Y	X ₂ Y	X ₁ X ₂	X ₁ ²	X ₂ ²
1	27.00	57.00	3.00	1539.00	81.00	171.00	3249.00	9.00
2	34.00	93.00	2.85	3162.00	96.90	265.05	8649.00	8.12
3	27.00	79.00	3.20	2133.00	86.40	252.80	6241.00	10.24
4	24.00	26.00	2.49	624.00	59.76	64.74	676.00	6.20
5	35.00	69.00	3.07	2415.00	107.45	211.83	4761.00	9.42
6	18.00	24.00	2.38	432.00	42.84	57.12	576.00	5.66
7	33.00	76.00	3.74	2508.00	123.42	284.24	5776.00	13.99
8	39.00	61.00	2.62	2379.00	102.18	159.82	3721.00	6.86
9	35.00	82.00	2.53	2870.00	88.55	207.46	6724.00	6.40
10	25.00	29.00	3.17	725.00	79.25	91.93	841.00	10.05
11	29.00	36.00	3.24	1044.00	93.96	116.64	1296.00	10.50
12	13.00	24.00	2.38	312.00	30.94	57.12	576.00	5.66
13	13.00	7.00	2.29	91.00	29.77	16.03	49.00	5.24
14	40.00	88.00	3.72	3520.00	148.80	327.36	7744.00	13.84
15	23.00	57.00	2.65	1311.00	60.95	151.05	3249.00	7.02
16	20.00	20.00	2.56	400.00	51.20	51.20	400.00	6.55
17	25.00	3.00	2.97	75.00	74.25	8.91	9.00	8.82
18	21.00	54.00	3.24	1134.00	68.04	174.96	2916.00	10.50

19	19.00	64.00	1.84	1216.00	34.96	117.76	4096.00	5.39
20	17.00	76.00	2.54	1292.00	43.18	193.04	5776.00	6.45
21	35.00	89.00	3.66	3115.00	128.10	325.74	7921.00	13.40
22	17.00	18.00	2.80	306.00	47.60	50.40	324.00	7.84
Jumlah	569.00	1132	62.94	32603	1679	3356.2	75570	185.1656

Dari hasil perhitungan di atas di dapat:

$$N = 22 \quad \sum X_1^2 = 75570 = 3356,20$$

$$\sum Y = 569 \quad \sum X_2^2 = 185,11656$$

$$\sum X_1 = 1132 \quad \sum X_1 Y = 32603$$

$$\sum X_2 = 62,94 \quad \sum X_2 Y = 1679$$

Nilai-nilai ini dimasukkan ke dalam persamaan:

$$\sum Y = Na + b_1 \sum X_1 + b_2 \sum X_2$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

Sehingga terbentuk persamaan:

$$569 = 22a + 1132 b_1 + 62,94 b_2$$

$$32603 = 1132a + 75570 b_1 + 3356,2 b_2$$

$$1679,5 = 62,94a + 3356,2 b_1 + 185,1656 b_2$$

Didapat nilai $a = -0,98171142$

$$b_1 = 0,14607215 \quad b_2 = 6,7563390$$

Sehingga persamaan regresinya adalah:

$$\hat{Y} = -0,98171142 + 0,14607215 X_1 + 6,7563390 X_2$$

Untuk menentukan nilai a dapat pula dilakukan dengan menggunakan komputer *software* SPSS:

Coefficients

Model	Unstandardized Coefficients	Sig.
	B	
1 (Constant)	-.982	.892
X1	.146	.005
X2	6.756	.021

a Dependent Variable: Y

Terlihat nilai konstant = $a = -0,982$

Nilai koefisien $X_1 = b_1 = 0,146$

Nilai koefisien $X_2 = b_2 = 6,756$

17.3.1.2. Penaksiran Nilai Variabel Dependen Regresi Linear 2 Variabel Independen

Dengan menggunakan persamaan regresi estimasi dapat dilakukan penaksiran nilai Variabel Dependen (Y) bila diketahui nilai variabel-variabel independennya. Penaksiran dilakukan dengan memasukkan nilai variabel-variabel independen ke persamaan regresi. Misalnya pada contoh di depan, bila nilai $X_1 = 10$ dan nilai $X_2 = 4$, yaitu persentasi penduduk yang berpenghasilan kurang 10, dan rata-rata banyaknya anak 4, maka nilai Y (persentasi banyaknya penduduk yang ikut KB) diestimasi sebagai berikut:

$$\begin{aligned} \hat{Y} &= -0,98171142 + 0,14607215(10) + 6,7563390(4) \\ &= -0,98171142 + 1,4607215 + 27,025356 \\ &= 27,50436608. \end{aligned}$$

Jadi diestimasi persentasi banyak penduduk yang ikut KB 27%.

17.3.2. Regresi Linear dengan 3 Variabel Independen

Dalam beberapa kasus kadang-kadang dijumpai suatu variabel dependen (Y) dipengaruhi oleh 3 variabel independen (X_1, X_2, X_3). Misalnya, suatu penelitian ingin melihat adanya hubungan antara kemampuan kerja (X_1), pemahaman terhadap tugas (X_2) dan motivasi kerja (X_3) dengan produktivitas kerja (Y).

Maka persamaan regresinya adalah:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3 \dots \dots \dots (16.16)$$

Persamaan tersebut merupakan persamaan estimasi, bukan merupakan hubungan yang sebenarnya antara X dengan Y, karena masih terdapatnya *factor error* (e). Secara pasti hubungan antara Y dan X dinyatakan:

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + e \dots \dots \dots (17.17)$$

Dimana e adalah *error*.

17.3.2.1. Menentukan Persamaan Regresi Linear dengan 3 Variabel Independen

Nilai-nilai a, b_1 , b_2 dan b_3 , dapat dicari dengan menggunakan sistem persamaan sebagai berikut:

$$\sum Y = Na + b_1 \sum X_1 + b_2 \sum X_2 + b_3 \sum X_3 \dots \dots \dots (17.18.a)$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + b_3 \sum X_1 X_3 \dots \dots \dots (17.18.b)$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 + b_3 \sum X_2 X_3 \dots \dots \dots (17.18.c)$$

$$\sum X_3 Y = a \sum X_3 + b_1 \sum X_1 X_3 + b_2 \sum X_2 X_3 + b_3 \sum X_3^2 \dots \dots \dots (17.18.d)$$

Metode yang digunakan untuk menyelesaikan sistem persamaan tersebut dengan metode substitusi, eliminasi atau dengan *Cramer*.

Contoh 17.7:

Suatu penelitian bertujuan untuk melihat faktor-faktor yang mempengaruhi besarnya sumbangan mahasiswa STAIN dalam ribuan rupiah terhadap korban bencana Tsunami NAD (Y), yaitu rata-rata banyaknya menghadiri ceramah agama dalam seminggu (X_1), banyaknya menghadiri aksi-aksi sosial (X_2) dan rata-rata simpanan tabanasnya setiap bulan (X_3) dalam ribuan. Penelitian dilakukan dengan sampel 10 mahasiswa dan didapat data sebagai berikut:

Tabel 17.7

Data Besar Sumbangan Mahasiswa (dalam ribuan rupiah), Kehadiran dalam Aksi Sosial dan Rata-rata Besar Simpanan Tabanas Setiap Bulan

No.	1	2	3	4	5	6	7	8	9	10
Y	7	6	3	4	8	1	9	7	5	3
X1	6	5	1	4	7	1	6	5	4	2
X2	8	10	5	6	5	3	9	8	5	8
X3	10	11	13	14	13	11	10	16	12	10

Tentukan persamaan regresinya.

Penyelesaian:

Terlebih dahulu dibuat tabel kerja/tabel perhitungan sebagai berikut:

Tabel 17.8

Tabel Perhitungan

	Y	X_1	X_2	X_3	$X_1 Y$	$X_2 Y$	$X_3 Y$	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	X_1^2	X_2^2	X_3^2
1	3	1	5	13	3	15	39	5	13	65	1	25	169
2	4	4	6	14	16	24	56	24	56	84	16	36	196
3	8	7	5	13	56	40	104	35	91	65	49	25	169
4	1	1	3	11	1	3	11	3	11	33	1	9	121
5	9	6	9	10	54	81	90	54	60	90	36	81	100
6	7	5	8	16	35	56	112	40	80	128	25	64	256
7	5	4	5	12	20	25	60	20	48	60	16	25	144
8	3	2	8	10	6	24	30	16	20	80	4	64	100
9	3	1	5	13	3	15	39	5	13	65	1	25	169
10	4	4	6	14	16	24	56	24	56	84	16	36	196
Jml	53	41	67	120	263	384	638	295	494	795	209	493	1476

$$\begin{array}{lll}
 N = 10 & \sum X_1^2 = 209 & \sum X_1 Y = 638 \\
 \sum Y = 53 & \sum X_2^2 = 493 & \sum X_1 X_2 = 295 \\
 \sum X_1 = 41 & \sum X_3^2 = 1476 & \sum X_1 X_3 = 494 \\
 \sum X_2 = 67 & \sum X_1 X_2 = 263 & \sum X_2 X_3 = 795 \\
 \sum X_3 = 120 & \sum X_2 X_3 = 384 &
 \end{array}$$

Sistem persamaan yang dibentuk adalah:

$$\begin{aligned}
 \sum Y &= Na + b_1 \sum X_1 + b_2 \sum X_2 + b_3 \sum X_3 \\
 \sum X_1 Y &= a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + b_3 \sum X_1 X_3 \\
 \sum X_2 Y &= a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 + b_3 \sum X_2 X_3 \\
 \sum X_3 Y &= a \sum X_3 + b_1 \sum X_1 X_3 + b_2 \sum X_2 X_3 + b_3 \sum X_3^2
 \end{aligned}$$

Jika dimasukkan nilai-nilai di atas maka terbentuk sistem persamaan:

$$\begin{aligned}
 53 &= 10a + 41b_1 + 67b_2 + 120b_3 \\
 263 &= 41a + 209b_1 + 295b_2 + 494b_3 \\
 384 &= 67a + 295b_1 + 493b_2 + 795b_3 \\
 638 &= 120a + 494b_1 + 795b_2 + 1476b_3
 \end{aligned}$$

Dari sistem persamaan tersebut didapatkan nilai-nilai a, b1, b2 dan b3 sebagai berikut:

$$a = -0,7694 \quad b_1 = 1,0173 \quad b_2 = 0,1969 \quad \text{dan} \quad b_3 = 0,0483$$

Sehingga persamaan regresi linearnya adalah:

$$\hat{Y} = -0,7694 + 1,0173 X_1 + 0,1969 X_2 + 0,0483 X_3$$

Dengan menggunakan SPSS didapat hasil yang sama:

Coefficients

Model	Unstandardized Coefficients		Sig.
	B		
1 (Constant)		-,769	,768
	X1	1,017	,001
	X2	,197	,307
	X3	4,827E-02	,788

a Dependent Variable: Y

Terlihat

$$\text{Konstant} = a = -0,769$$

$$b_1 = 1,017$$

$$b_2 = 0,197$$

$$b_3 = 0,04827$$

17.3.2.2. Penaksiran Nilai Variabel Dependen Regresi Linear Berganda

Dengan menggunakan persamaan regresi estimasi dapat dilakukan penaksiran nilai Variabel Dependen (Y) bila diketahui nilai variabel-variabel independennya. Penaksiran dilakukan dengan memasukkan nilai variabel-variabel independen ke persamaan regresi. Misalnya pada contoh di depan, bila seorang mahasiswa mempunyai nilai $X_1 = 6$, nilai $X_2 = 5$ dan $X_3 = 12$, yaitu banyaknya mengikuti ceramah agama dalam seminggu rata-rata 6, dan banyaknya menghadiri aksi-aksi sosial 5 kali, dan banyaknya tabungan tabanas Rp 12.000, maka nilai Y (besarnya sumbangan mahasiswa) diestimasi sebagai berikut:

$$\hat{Y} = -0,7694 + 1,0173 X_1 + 0,1969 X_2 + 0,0483 X_3$$

Jadi diestimasi persentasi banyak penduduk yang ikut KB 27 %.

17.4. Uji Regresi Linear Dengan ANAVA

Untuk menganalisis apakah suatu model regresi cocok atau tidak bisa dilakukan dengan Anava. Dalam hal ini hipotesis nolnya adalah Koefisien Korelasi Berganda (korelasi antara $X_1, X_2, X_3, \dots, X_n$ dengan Y) adalah 0. Artinya tidak ada korelasi antara $X_1, X_2, X_3, \dots, X_n$ dengan Y, dengan kata lain model regresi tidak cocok. Hipotesis alternatifnya adalah Koefisien Korelasi Berganda (korelasi antara $X_1, X_2, X_3, \dots, X_p$ dengan Y) tidak 0. Jadi ada korelasi antara $X_1, X_2, X_3, \dots, X_p$ dengan Y, dengan kata lain model regresi cocok.

Tabel anovanya adalah sebagai berikut:

Tabel 17.9

Tabel Anova

Sumber Variasi	df	SS	MSS	F_{hit}	F_{Tab}	
					$F_{0,05;p,p-1}$	$F_{0,01;p,p-1}$
Variabel Regresi X_1, X_2, \dots, X_p	p	SSR	MSR	$\frac{MSR}{MSE}$		
Error/Residu	n-p-1	SSE	MSE			
Total	n-1	SST	-			

Dimana:

$$\text{SSR} = a \sum Y + b_1 \sum X_1 Y + b_2 \sum X_2 Y + \dots + b_p \sum X_p Y - \frac{1}{n} (\sum Y)^2 \dots \dots \dots (17.19)$$

$$\text{SST} = \sum Y^2 - \frac{1}{n} (\sum Y)^2 \dots \dots \dots (17.20)$$

$$\text{SSE} = \text{SST} - \text{SSR} \dots \dots \dots (17.21)$$

$$\text{MSR} = \text{SSR}/p \dots \dots \dots (17.22)$$

$$MSE = SSE / (n-p-1) \dots \dots \dots (17.23)$$

Jika F_{hitung} kurang dari F_{tabel} pada taraf pada taraf signifikan tertentu, maka hipotesis nol diterima. Berarti tidak terdapat korelasi ganda anda $X_1, X_2, X_3, \dots, X_p$ dengan Y . Maka model regresi yang didapatkan tidak dapat dipakai.

Contoh 17.8:

Suatu penelitian tentang banyaknya angkatan kerja (Y) dan angkatan kerja (X) dilakukan di 10 desa. Hasil penelitian disajikan dalam tabel 17.10. Buatlah Anava, dan bagaimana kesimpulannya?

Tabel 17.10

No. Desa	1	2	3	4	5	6	7	8	9	10
Banyaknya Angkatan Kerja (Y)	160	200	140	150	130	160	150	150	140	220
Banyaknya Pengangguran (X)	70	72	68	66	67	69	68	71	70	69

Penyelesaian:

Dari data di atas dibuat tabel perhitungan sebagai berikut:

Tabel 17.11

Y	X	XY	Y ²	X ²
160.00	70.00	11200.00	25600.00	4900.00
200.00	72.00	14400.00	40000.00	5184.00
140.00	68.00	9520.00	19600.00	4624.00
150.00	66.00	9900.00	22500.00	4356.00
130.00	67.00	8710.00	16900.00	4489.00
160.00	69.00	11040.00	25600.00	4761.00
150.00	68.00	10200.00	22500.00	4624.00
150.00	71.00	10650.00	22500.00	5041.00
140.00	70.00	9800.00	19600.00	4900.00
220.00	69.00	15180.00	48400.00	4761.00
$\Sigma X = 1600.00$	$\Sigma Y = 690.00$	$\Sigma XY = 110600.00$	$\Sigma Y^2 = 263200.00$	$\Sigma X^2 = 47640.00$

Dari tabel di atas didapat:

$$\Sigma X = 1.600$$

$$\Sigma Y = 690$$

$$\Sigma XY = 110.600$$

$$\Sigma Y^2 = 263.200$$

$$\Sigma X^2 = 47.640$$

Persamaan regresinya adalah: $\hat{Y} = a + bX$

$$b = \frac{n \Sigma XY - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2} = \frac{10(110600) - (690)(1600)}{10(47640) - (690)^2}$$

$$= \frac{1106000 - 1104000}{476400 - 476100} = \frac{2000}{300} = 20/3$$

$$= 6,66667$$

$$a = \bar{Y} - b\bar{X}$$

$$= \frac{1600}{10} - 20/3 \cdot \frac{690}{10}$$

$$= 160 - 460$$

$$= -300$$

Jadi persamaan regresinya: $\hat{Y} = -300 + 6,66667X$

Selanjutnya:

$$SST = \Sigma Y^2 - (\Sigma Y)^2/n = 263200 - (1600)^2/10 = 7200$$

$$SSR = a\Sigma Y + b\Sigma XY - (\Sigma Y)^2/n = (-300)(1600) + (6,66667)(110600) - (1600)^2/10 = 1333,33333$$

$$SSE = SST - SSR = 7200 - 1333,33333 = 5866,66667$$

Tabel 17.12

Tabel Anova

Sumber Variasi	df	SS	MSS	F_{hitung}	F_{tabel}	
					$F_{0,05;p-1}$	$F_{0,01;p-1}$
Variabel Regresi X	1	1333,33333	1333,33333	1,82	5,32	11,26
Error/Residu	8	5866,66667	733,33333	-	-	-
Total	9	7200	-	-	-	-

Dari tabel ternyata $F_{hitung} < F_{tabel}$. Berarti model tersebut tidak cocok untuk dipasangkan pada data di atas.

Soal-soal:

Buatlah persamaan regresi linear dan ujliah untuk soal-soal pada Bab IX dan bab X.

BAB XVIII

TEKNIK ANALISIS

KOMPARASIONAL

18.1. Pengertian

Komparasi berarti perbandingan. Penelitian komparasional merupakan penelitian yang berusaha untuk menemukan persamaan atau pun perbedaan tentang benda, orang, prosedur kerja atau pun ide. Pada penelitian komparatif, hipotesis yang digunakan adalah hipotesis komparatif, yaitu hipotesis tentang adanya perbedaan atau tidak adanya perbedaan antarvariabel yang diteliti. Teknik yang digunakan disebut teknik analisis komparasional, yaitu suatu analisis statistik yang digunakan untuk menguji hipotesis komparatif. Teknik analisis komparatif ada beberapa macam. Penentuan teknik apa yang digunakan didasarkan pada bentuk komparasi sampel dan macam data yang dianalisis. Komparasi dapat 2 sampel atau pun lebih dari 2 sampel. Masing-masing dapat berpasangan/berkorelasi atau pun saling independen.

Sampel berpasangan adalah bila *variabel pertama* berkorelasi dengan *variabel kedua*. Misalnya, bila seorang peneliti ingin melihat apakah ada perbedaan pemahaman keagamaan responden antara sebelum dan sesudah mendapatkan ceramah agama. Dalam hal ini satu responden akan diambil dua data, yaitu sebelum mendapat ceramah (misal sebagai variabel X) dan sesudah mendapat ceramah (sebagai variabel Y).

Sampel independen adalah bila antarvariabel sampel tidak saling berkaitan. Misalnya bila seorang peneliti pendidikan ingin melihat ada/tidak adanya perbedaan antara 2 metode pengajaran. Dalam hal ini data bisa diambil dari dua kelompok sampel yang berlainan. Terhadap sampel pertama diperlakukan metode X, dan kelompok sampel kedua diperlakukan metode Y. Skor yang didapat dari variabel X tidak berkorelasi dengan skor variabel Y.

Berbagai teknik statistik yang digunakan untuk menguji hipotesis komparatif dapat dilihat pada tabel 18.1. berikut:

Tabel 18.1

Macam data	Bentuk Komparasi			
	Dua Sampel		Sampel	
	Korelasi	Independen	Korelasi	Independen
Interval & Rasio	test dua sampel	test dua sampel	Anova satu arah Anova Dua Arah	Anova satu arah Anova Dua Arah
Nominal	Mc Nemar	Chi Kuadrat dua sampel	Chi Kuadrat k sampel	Chi Kuadrat k sampel

Akan diuraikan Test t dan Test Chi kuadrat. Anava (Analisis Variansi) sudah diuraikan dalam bab XVI.

18.2. Tes t

Tes t merupakan salah satu tes statistik yang digunakan pada analisis komparasi. Tes digunakan untuk menguji kebenaran atau ketidakbenaran suatu hipotesis yang menyatakan bahwa di antara dua buah rata-rata sampel diambil secara random dari populasi yang sama tidak terdapat perbedaan yang signifikan. Tes ini didasarkan kepada distribusi t yang pertama kali dikenalkan oleh WS Gosset yang menggunakan nama samaran Student. Huruf t pada kata student itulah yang digunakan untuk tes ini.

18.2.1. Tes t untuk Sampel yang Berkorelasi

Pada kasus dimana sampel berkorelasi (berpasangan), maka langkahnya adalah sebagai berikut:

Merumuskan hipotesis

Rumusan hipotesisnya adalah:

H_0 : tidak terdapat perbedaan mean yang signifikan antarsampel / antara variabel I dengan variabel II

H_1 : terdapat perbedaan mean yang signifikan antarsampel/antara variabel I dengan variabel II

a. Menghitung nilai t_0 dengan rumus:

$$t_0 = \frac{M_D}{SE_{MD}} \dots \dots \dots (18.1)$$

Dimana:

$$D = X - Y \dots \dots \dots (18.2)$$

$$M_D = \frac{\sum D}{N} \dots \dots \dots (18.3)$$

$$SE_{MD} = \frac{SD_D}{\sqrt{N-1}} \dots \dots \dots (18.4)$$

$$SD_D = \sqrt{\frac{\sum D^2}{N} - \left(\frac{\sum D}{N}\right)^2} \dots \dots \dots (18.5)$$

D = beda antara skor variabel I dengan variabel II

M_D = nilai mean dari selisih antara skor variabel I dengan skor variabel II

SE_{MD} = standar Error dari Mean Beda antara skor variabel I dengan skor variabel II

SD_D = standar Deviasi dari beda

b. Uji Hipotesis, yaitu membandingkan besar nilai t_0 (t tabel) dengan t_c .

Nilai t_c dilihat pada tabel nilai t. Dengan derajat kebebasan

$$df = N - 1 \dots \dots \dots (18.6)$$

Pada uji hipotesis ini berlaku ketentuan:

Bila $|t_0| \geq t_c$ maka H_0 ditolak, H_1 diterima. Berarti disimpulkan terdapat perbedaan yang signifikan antara variabel I dengan variabel II.

Bila $|t_0| < t_c$ maka H_0 diterima, H_1 ditolak. Berarti disimpulkan tidak terdapat perbedaan yang signifikan antara variabel I dengan variabel II

Contoh 18.1:

Suatu penelitian bertujuan ingin melihat ada atau tidak adanya perbedaan perilaku keagamaan siswa SMA Ulul Albab antara sebelum dan sesudah mengikuti Pesantren Kilat RAIHAN (Ramadhan Intensif dengan Pelatihan). Dari sampel berjumlah 10 didapat data seperti tertulis pada tabel 18.2. Bagaimana kesimpulan yang didapat?

Tabel 18.2
Data Skor Perilaku Keagamaan

Nama Siswa	Skor Perilaku Keagamaan	
	Sesudah Mengikuti Peskil (X)	Sebelum Mengikuti Peskil (Y)
A	8	7
B	8	7
C	5	7
D	7	6
E	6	6
F	6	5
G	8	5
H	9	8
I	9	7
J	8	6

Penyelesaian:

Terlebih dari dibuat tabel kerja sebagai berikut:

Tabel 18.2

Nama Siswa	Skor Perilaku Keagamaan		D	D ²
	X	Y		
A	8	7	1	1
B	8	7	1	1
C	5	7	-2	4
D	7	6	1	1
E	6	6	0	0
F	6	5	1	1
G	8	5	3	9
H	9	8	1	1
I	9	7	2	4
J	8	6	0	0
Jumlah	-	-	8	22

Rumusan hipotesisnya adalah:

H_0 : tidak terdapat perbedaan yang signifikan dalam perilaku keagamaan antara sebelum dengan sesudah mengikuti Pesantren Kilat

H_1 : terdapat perbedaan yang signifikan dalam perilaku keagamaan antara sebelum dengan sesudah mengikuti Pesantren Kilat.

$$M_D = \frac{\sum D}{N} = \frac{8}{10} = 0,8$$

$$SD_D = \sqrt{\frac{\sum D^2}{N} - \left(\frac{\sum D}{N}\right)^2}$$

$$= \sqrt{\frac{22}{10} - (0,8)^2} = \sqrt{2,2 - 0,64} = \sqrt{1,56}$$

$$SE_{MD} = \frac{SD_D}{\sqrt{N-1}} = \frac{\sqrt{1,56}}{\sqrt{10-1}} = \sqrt{\frac{1,56}{9}} = 0,416$$

$$= \frac{0,8}{0,416} = 1,923$$

$N = 10$

Derajat kebebasan = $df = N - 1 = 10 - 1 = 9$

Pada taraf signifikan 5% Nilai $t_c = 2,262$

Pada taraf signifikan 1% Nilai $t_c = 3,250$

$t_o = 1,923$, maka $|t_o| = 1,923$

Dengan demikian:

Pada Taraf signifikan 5% $|t_o| \geq t_c$ maka H_0 ditolak, H_1 diterima. Berarti disimpulkan terdapat perbedaan yang signifikan dalam perilaku keagamaan antara sebelum dengan sesudah mengikuti Pesantren Kilat.

Pada Taraf signifikan 1% $|t_o| \geq t_c$ maka H_0 ditolak, H_1 diterima. Berarti disimpulkan terdapat perbedaan yang signifikan dalam perilaku keagamaan antara sebelum dengan sesudah mengikuti Pesantren Kilat.

18.2.2. Tes t untuk Sampel yang Independen

Pada kasus dimana sampel tidak berkorelasi (saling lepas/independen), maka langkahnya adalah sebagai berikut:

a. Merumuskan hipotesis

Rumusan hipotesisnya adalah:

H_0 : tidak terdapat perbedaan mean yang signifikan antarsampel/antara variabel I dengan Variabel II

H_1 : terdapat perbedaan mean yang signifikan antarsampel/antara variabel I dengan variabel II

b. Menghitung nilai t_0 dengan rumus:

$$t_0 = \frac{M_1 - M_2}{SE_{M_1 - M_2}} \quad (18.7)$$

dimana:

$$M_1 = \frac{\sum X}{N_1} \quad (18.8)$$

$$M_2 = \frac{\sum Y}{N_2} \quad (18.9)$$

$$SE_{M_1 - M_2} = \sqrt{SE_{M_1}^2 + SE_{M_2}^2} \quad (18.10)$$

$$SE_{M_1} = \frac{SD_1}{\sqrt{N_1 - 1}} \quad (18.11)$$

$$SE_{M_2} = \frac{SD_2}{\sqrt{N_2 - 1}} \quad (18.12)$$

$$SD_1 = \sqrt{\frac{\sum (X - M_1)^2}{N_1}} \quad (18.13)$$

$$SD_2 = \sqrt{\frac{\sum (Y - M_2)^2}{N_2}} \quad (18.14)$$

c. Uji hipotesis, yaitu membandingkan besar nilai t_1 (t tabel) dengan t_0 .

Nilai t_1 dilihat pada tabel nilai t .

Dengan derajat kebebasan

$$df = N_1 + N_2 - 2 \quad (18.15)$$

Pada uji hipotesis ini berlaku ketentuan:

Bila $|t_0| \geq t_1$ maka H_0 ditolak, H_1 diterima. Berarti disimpulkan terdapat perbedaan yang signifikan antara variabel I dengan variabel II.

Bila $|t_0| < t_1$ maka H_0 diterima, H_1 ditolak. Berarti disimpulkan tidak terdapat perbedaan yang signifikan antara variabel I dengan variabel II.

Contoh 18.2:

Seorang guru melakukan eksperimen tentang metode mengajar A dan metode B. Diambil dua kelas sebagai eksperimennya. Dari masing-masing kelas diambil beberapa sampel. Dari kelas A diambil sampel sebanyak 9 siswa dan kelas B diambil sampel sebanyak 13 siswa. Pengambilan sampel dilakukan secara random, dan didapat hasil pengumpulan data seperti pada tabel 18.3. Bagaimana kesimpulannya?

Tabel 18.3

Kelas A (Variabel X)	Kelas B (Variabel Y)
70	63
60	60
80	70
75	80
76	74
75	75
71	85
65	64
85	65
	60
	90
	75
	75

Penyelesaian:

Sebelumnya dibuat tabel kerja sebagai berikut:

Tabel 18.4

X	X - M_1	(X - M_1) ²	Y	Y - M_2	(Y - M_2) ²
85	12	144	90	18	324
80	7	49	85	13	169
76	3	9	80	8	64
76	3	9	80	8	64
75	2	4	75	3	9
75	2	4	75	3	9
71	-2	4	75	3	9
70	-3	9	74	2	4
65	-8	64	70	-2	4
60	-13	169	65	-7	49
			64	-8	64
			63	-9	81
			60	-12	144
			60	-12	144
657	0	456	936	0	1074

$$M_1 = \frac{\sum X}{N_1} = \frac{657}{9} = 73$$

$$M_2 = \frac{\sum Y}{N_2} = \frac{936}{13} = 72$$

$$SD_1 = \sqrt{\frac{\sum (X - M_1)^2}{N_1}} = \sqrt{\frac{456}{9}} = 7,118$$

$$SD_2 = \sqrt{\frac{\sum (X - M_1)^2}{N_1}} = \sqrt{\frac{1074}{13}} = 9,089$$

$$SE_{M1} = \frac{SD_1}{\sqrt{N_1 - 1}} = \frac{7,118}{\sqrt{9 - 1}} = 2,526$$

$$SE_{M2} = \frac{SD_2}{\sqrt{N_2 - 1}} = \frac{9,089}{\sqrt{13 - 1}} = 2,624$$

$$SE_{M1-M2} = \sqrt{SE_{M1}^2 + SE_{M2}^2} = 3,642$$

$$t_o = \frac{M_1 - M_2}{SE_{M1-M2}} = \frac{73 - 72}{3,642} = 0,275$$

$$df = N_1 + N_2 - 2 = 9 + 13 - 2 = 20$$

Rumusan hipotesisnya adalah:

H_0 : tidak terdapat perbedaan yang signifikan antara metode mengajar A dengan metode mengajar B

H_a : terdapat perbedaan yang signifikan antara metode mengajar A dengan metode mengajar B

Pada taraf signifikansi 5% nilai $t_t = 2,086$

Pada taraf signifikansi 1% nilai $t_t = 2,845$

Dengan demikian:

Pada taraf signifikan 5% $|t_o| < t_t$ maka H_0 diterima, H_a ditolak. Berarti disimpulkan tidak terdapat perbedaan yang signifikan antara metode mengajar metode A dengan metode mengajar B.

Pada taraf signifikan 1% $|t_o| < t_t$ maka H_0 diterima, H_a ditolak. Berarti disimpulkan tidak terdapat perbedaan yang signifikan antara metode mengajar metode A dengan metode mengajar B.

18.3. Test Chi Kuadrat

Test Chi Kuadrat merupakan suatu teknik analisis komparasional untuk mengetahui ada atau tidak adanya perbedaan yang signifikan dari data statistik yang menggunakan skala nominal. Skala nominal menunjukkan jumlah atau frekuensi suatu variabel. Dengan kata lain Test Chi Kuadrat digunakan untuk

menganalisis perbedaan antarvariabel yang mendasarkan pada frekuensi dari data yang diselidiki.

Langkah untuk melakukan Test Chi Kuadrat adalah:

a. Merumuskan hipotesis

Rumusan hipotesisnya adalah:

H_0 : tidak terdapat perbedaan yang signifikan antara variabel I dengan variabel II

H_a : terdapat perbedaan mean yang signifikan antara variabel I dengan variabel II

b. Menghitung nilai χ^2

Rumusnya adalah:

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{ft} \dots \dots \dots (18.15)$$

Dimana:

f_o = Frekuensi Observasi tiap sel

f_t = Frekuensi teoritis sel

Nilai ini selanjutnya disebut χ^2

Untuk keperluan itu perlu dibuat tabel perhitungan

c. Menentukan nilai derajat kebebasan (df)

$$df = (k-1)(n-1) \dots \dots \dots (18.16)$$

Dimana:

k = banyaknya kriteria variabel I

n = Banyak kriteria variabel II

d. Menentukan nilai χ^2 pada taraf signifikan 5% atau 1%

e. Membandingkan χ^2_o dengan χ^2_t

Bila $\chi^2_o \geq \chi^2_t$ maka H_0 ditolak, H_a diterima. Berarti disimpulkan terdapat perbedaan yang signifikan antara variabel I dengan variabel II.

Bila $\chi^2_o < \chi^2_t$ maka H_0 diterima, H_a ditolak. Berarti disimpulkan tidak terdapat perbedaan yang signifikan antara variabel I dengan variabel II.

Contoh 18.3:

Untuk mengetahui apakah ada perbedaan yang signifikan dalam prestasi mahasiswa bila kuliah diberikan pada pagi hari atau sore hari dilakukan penelitian dengan sampel sebanyak 100 mahasiswa. Dari hasil pengumpulan data didapat tabulasi data seperti pada tabel 18.5. Bagaimana kesimpulan yang didapat pada taraf signifikan 5%?

Tabel 18.5

Nilai Ujian	A	B	C	D	E	Total
Waktu Kuliah						
Pagi	12 = a	18 = b	13 = c	11 = d	6 = e	60
Sore	3 = f	7 = g	17 = h	4 = i	9 = j	40
Total	15	25	30	15	15	100

a,b,c,d,..., j adalah nama-nama sel

Tabel 18.6

Tabel Perhitungan

Sel	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
a	12	$\frac{60 \times 15}{100} = 9$	3	9	1
b	18	$\frac{60 \times 25}{100} = 15$	3	9	1,5
c	13	$\frac{60 \times 30}{100} = 18$	-5	25	1,39
d	11	$\frac{60 \times 15}{100} = 9$	2	4	0,44
e	6	$\frac{60 \times 15}{100} = 9$	-3	9	1
f	3	$\frac{40 \times 15}{100} = 6$	-3	9	1,5
g	7	$\frac{40 \times 25}{100} = 10$	-3	9	0,9
h	17	$\frac{40 \times 30}{100} = 12$	5	25	2,083
i	4	$\frac{40 \times 15}{100} = 6$	-2	4	0,667
j	9	$\frac{40 \times 15}{100} = 6$	3	9	1,5
Total	100	-	-	-	11,98

Dari tabel perhitungan di atas didapat: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 11,98 = \chi^2$

Rumusan hipotesisnya:

H_o : tidak terdapat perbedaan yang signifikan prestasi mahasiswa antara kuliah pagi dengan sore.

H_a : terdapat perbedaan yang signifikan prestasi mahasiswa antara kuliah pagi dengan sore.

$$df = (k-1)(n-1) = (5-1)(2-1) = 4$$

Pada taraf signifikansi 5% Nilai $\chi^2 = 9,488$

Dengan demikian:

Pada Taraf signifikansi 5% $\chi^2 > \chi^2_{table}$ maka H_o ditolak, H_a diterima. Berarti disimpulkan terdapat perbedaan yang signifikan prestasi mahasiswa antara kuliah pagi dengan sore.

Soal:

1. Pada suatu penataran yang dilakukan terhadap para guru, seorang penatar ingin mengetahui apakah penataran tersebut mempunyai dampak terhadap pengetahuan guru. Untuk itu dilakukan *pre test* dan *post test* yang hasilnya adalah seperti tertuang dalam tabel 18.7. Bagaimana kesimpulan yang didapat?

Tabel 18.7

Pre Test	4	5	5	6	5	4	8	4	5	6
Post Test	7	7	8	7	6	8	9	6	5	9

2. Dari hasil penelitian kemampuan berbahasa Indonesia pada Jurusan Syariah dan Jurusan Tarbiyah didapat data sebagai berikut.
Jurusan Syariah : 90, 90,85, 85, 92, 83, 84,92, 93, 94, 95, 85, 89, 87,80.
Jurusan Tarbiyah : 81, 83, 79, 78, 80, 90, 70, 80, 85, 86, 87,75, 88,92,82
Bagaimana kesimpulannya?
3. Suatu penelitian ingin mengetahui apakah terdapat perbedaan kebiasaan minum Coca Cola antara orang kaya dan orang miskin. Dari 100 responden didapat data seperti tertera pada tabel 18.8

Tabel 18.8

Kebiasaan Minum Cola	Tingkat Ekonomi		Total
	Kaya	Miskin	
Minum	30 = a	10 = b	40
TidakMinum	10 = c	50 = d	60
Total	40	60	100

Bagaimana kesimpulannya?

TABEL I
Luas Di Bawah Lengkungan Kurve Normal Dari 0 S/d Z

Z	0	1	2	3	4	5	6	7	8	9
0,0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0479	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2258	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2612	2642	2673	2703	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4419	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	4772	4778	4783	4788	4793	4808	4808	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4898	4896	4898	4901	4004	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4025	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4043	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974

2,8	4074	4975	4976	4977	4977	4987	4979	4979	4980	4981
2,9	4981	4982	4982	4083	4984	4984	4985	4985	4986	4986
3,0	4987	4987	4987	4988	4988	4989	4989	4989	4990	4990
3,1	4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3,2	4993	4993	4994	4994	4994	4994	4994	4994	4995	4995
3,3	4995	4995	4995	4986	4996	4996	4996	4996	4997	4997
3,4	4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3,5	4998	4998	4998	4998	4998	4998	4998	4998	4998	4998
3,6	4998	4998	4999	4999	4999	4999	4999	4999	4999	4999
3,7	4999	4999	4999	4999	4999	4999	4999	4999	4999	4999
3,8	4999	4999	4999	4999	4999	4999	4999	4999	4999	4999
3,9	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000

DAFTAR PUSTAKA

- Alghifari. 1997. *Analisis Statistik untuk Bisnis Dengan Regresi, Korelasi dan Non Parametrik*. Yogyakarta: BPFE.
- Alghifari. 1997. *Statistika Ekonomi*. Yogyakarta: Bagian Penerbitan YKPN.
- BAPEDA Jateng. 2001. *Statistik Pembangunan*
- Casella, George, Berger, Roger L.. 1990. *Statistical Inference*. Pacific Grove, California: Wadsworth & Brooks/Cole Advance Books & Software.
- Dudewicz, Edward J, Mishra, Satya N. 1995. *Statistika Matematika Modern*. Bandung: ITB.
- Fauzy, Akhmad. 2001. *Statistik Industri I*. Yogyakarta: UII Press.
- Greene, William H. 1993. *Economic Analysis*. New York: Macmillan Publishing Company.
- Irianto, Agus. 2004. *Statistik; Konsep dan Aplikasinya*. Jakarta: Kencana.
- IA, Suparman. 1995. *Statistik Sosial*. Jakarta: Rajagrafindo Persada.
- Saleh, Samsubar. 1988. *Statistik Induktif*. Yogyakarta: Liberty.
- Soejoeti, Zanzawi. 1990. *Peluang dan Statistika*. Yogyakarta: Fakultas MIPA UGM.
- Soepeno, Bambang. 1997. *Statistik Terapan Dalam Penelitian Ilmu-ilmu Sosial Dan Pendidikan*. Jakarta: Rineka Cipta.
- Sudiyono, Anas. 1997. *Pengantar Statistik Pendidikan*. Jakarta: Rajawali Press
- Sudjana. 1996. *Metoda Statistika*. Bandung: Taarsito.
- Sugiyono. 1999. *Statistika Untuk Penelitian*. Jakarta: CV Alfa Beta.
- Usman, Husaini dan Akbar, R. Purnama Setiady, 1995, *Pengantar Statistika*, Jakarta, Bumi Aksara
- Wahana Komputer. 2002. *10 Model Penelitian dan Pengolahannya dengan SPSS 10.01*. Semarang.

STATISTIKA TERAPAN
UNTUK
PENELITIAN SOSIAL



STAIN PRESS

Statistik bisa dijadikan sebagai alat bantu untuk memberikan gambaran atas suatu peristiwa melalui bentuk yang sederhana, baik berupa angka-angka atau grafik dan diagram.

Buku ini menyajikan rumus-rumus statistika terapan yang diperlukan dalam penelitian pendidikan dan ilmu-ilmu sosial dengan menggunakan bahasa dan deskripsi yang komunikatif.

ISBN 979-3968-08-7

STAIN PRESS
PEKALONGAN

Jl. Kusumabangsa No. 9 Pekalongan
Telp. (0285) 412575, Faks. (0285) 423418
E-mail : pipstainpkl@yahoo.com