

MILIK PERPUSTAKAAN  
UNIV. NEGERI PADANG

## MAKALAH



## TEKNIK ANALISIS DAN DATA



Oleh : Prof. Dr. H. Agus Irianto

MILIK PERPUSTAKAAN UNP PADANG	
DITERIMA TGL.	: 9-6-2003
SUMBER / HARGA	: Hadiah
KOLEKSI	: K1
NO. INVENTARIS	: 135 / K / 2003 - t. (2)
KLASIFIKASI	: 001.4225 IRI - 4

Disampaikan Pada :

SEMILOKA METODOLOGI PENELITIAN KUANTITATIF  
BAGI STAF PENGAJAR STAIN BATUSANGKAR

JAGA DAN PERGUNAKANLAH KOLEKSI  
INI DENGAN BAIK

**KERJASAMA**

SUATU SAAT ANAK DAN CUCU ANDA  
SANGAT MEMBUTUKANNYA

**SEKOLAH TINGGI AGAMA ISLAM NEGERI  
BATU SANGKAR**

**DENGAN**

**LEMBAGA PENELITIAN  
UNIVERSITAS NEGERI PADANG  
2003**

## TEKNIK ANALISIS DAN DATA<sup>(1)</sup>

*Prof. Dr. H. Agus Irianto<sup>(1)</sup>*

### A. Pengertian Statistik & Fungsinya

Statistik pada dasarnya merupakan alat bantu untuk memberi gambaran atas suatu kejadian melalui bentuk yang sederhana baik berupa angka-angka maupun grafik-grafik,. Mengingat perannya sebagai alat bantu, maka perlu disadari bahwa kunci keberhasilan analisis statistik masih terletak pada pemakaiannya. Anggapan yang mengatakan bahwa statistik merupakan alat yang paling canggih, maupun anggapan bahwa tanpa statistik, maka penelitian yang dilakukan kurang bisa dipertanggungjawabkan, harus dibuang jauh-jauh.

Semua istilah statistik hanya merupakan sekumpulan angka-angka yang menggambarkan keadaan penduduk, pendapatan masyarakat, tingkat produksi pertanian pada suatu waktu tertentu. Dalam pengertian disini statistik hanya memberi gambaran masa lalu sampai dibuat gambaran tersebut. Dewasa ini statistik tidak hanya merupakan sekumpulan angka-angka masa lalu saja, tetapi dengan statistik angka-angka yang terkumpul dapat digunakan untuk meramalkan kondisi di masa yang akan datang. Di samping itu kondisi yang bisa digambarkan dengan statistik berkembang pula, termasuk di dalamnya kondisi pendidikan, sosiologi, antropologi dan lain-lain. Perilaku manusia pun bisa dianalisis dengan menggunakan alat bantu statistik. Statistik juga mengandung unsur penyederhanaan atas data yang besar, sehingga mempermudah membaca data tersebut. Misalnya rata-rata penduduk Indonesia mempunyai anak sebanyak 4 orang, 60% penduduk Indonesia berdomisili di pulau Jawa.

Secara ringkas statistik dapat diartikan sebagai sekumpulan metoda dengan aturan-aturan untuk mengatur interpretasi hasil observasi (Grevetter dan Wollwam, 1985). Irianto (1988) memberi pengertian statistik sebagai sekumpulan cara maupun aturan-aturan yang berkaitan dengan pengumpulan, pengolahan (analisis), penarikan kesimpulan, atas data-data yang berbentuk angka, dengan menggunakan suatu asumsi tertentu. Definisi di atas ditujukan untuk penyeragaman bahwa dalam pembahasan

---

<sup>(1)</sup> *Disampaikan pada Semiloka Metodologi Penelitian bagi Tenaga Edukatif STAIN Batu Sangkar  
28 Mei 2003*

materi statistik yang akan dibahas lebih lanjut. Banyak definisi statistik yang bisa ditemui, walaupun berbeda bahasanya tetapi makna yang terkandung di dalamnya adalah sama.

Fungsi statistik ada dua yaitu deskripsi dan inferensial. Statistik deskriptif merupakan langkah awal dalam pembahasan statistik. Dalam hal ini pemakai statistik deskriptif tidak dapat mengambil kesimpulan yang bersifat umum (generalisasi), karena statistik di sini memang terbatas pada hal yang ada saja. Hasil analisisnya sederhana, bahkan kebanyakan perhitungannya bersifat penyederhanaan atas data yang terkumpul. Berdasarkan hasil penyederhanaan inilah pemakai diharapkan dapat mengambil arti dari data tersebut. Oleh karena sifatnya yang sederhana, maka dimungkinkan bisa muncul keraguan dan rasa tidak percaya atas hasil deskriptif ini. Anggapan ini sebenarnya tidaklah selalu benar, karena ada kemungkinan sifat data, ada yang hanya bisa dianalisis secara deskriptif. Jika kondisinya seperti itu, maka pemaksaan data untuk dianalisis dengan teknik analisis yang lebih tinggi justru akan menjerumuskan. Statistik inferensial merupakan pengembangan fungsi statistik. Di sini pemakai diberi lebih banyak informasi daripada informasi yang diberikan oleh statistik deskriptif. Dengan demikian, maka pemakai statistik inferensial akan bisa bicara lebih banyak tentang data yang dianalisisnya. Oleh karena sifatnya yang lebih baik daripada statistik deskriptif, tentunya langkah analisisnyapun lebih kompleks daripada analisis statistik deskriptif.

Contoh : Seorang guru matematika ingin melihat perbedaan hasil belajar siswanya antara siswa yang belajar di pagi hari (jam 1 dan 2) dengan siswa yang belajar siang hari (jam 6 dan 7). Sampel diambil secara random sebanyak 20 siswa masing-masing 10 siswa dan hasil pengumpulan nilai matematika sebagai berikut :

Siswa	Nilai kelas pagi	Nilai kelas siang
1	50	70
2	60	75
3	55	65
4	63	50
5	67	88
6	60	90
7	85	55
8	70	60
9	65	80
10	56	75

Data di atas dapat di susun sedemikian rupa sehingga mudah dipelajari/dibaca. Misalnya dihitung masing-masing kelompok, dalam hal ini jumlah nilai matematika :

- a. Kelompok siswa yang masuk jam 1 dan 2 adalah 631
- b. Kelompok siswa yang masuk jam 6 dan 7 adalah 708

Pada dasarnya statistik dapat digunakan untuk :

- a. Membantu peneliti untuk menentukan sampel, sehingga peneliti dapat bekerja efisien, tetapi hasilnya sesuai dengan objek yang ingin diteliti.
- b. Membantu peneliti untuk membaca data yang telah terkumpul, sehingga peneliti dapat mengambil keputusan yang tepat.
- c. Membantu peneliti untuk melihat ada tidaknya perbedaan antara kelompok satu dengan yang lainnya atas objek yang diteliti.
- d. Membantu peneliti untuk melihat ada tidaknya hubungan antara variabel yang satu dengan yang lainnya.
- e. Membantu peneliti dalam melakukan prediksi untuk waktu yang akan datang maupun masa lalu.
- f. Membantu peneliti untuk melakukan interpretasi atas data yang terkumpul.

## **B. Skala Pengukuran**

Statistik bekerja dengan angka-angka, sedangkan angka-angka tersebut berasal dari perhitungan kuantitas atas suatu objek maupun penelitian yang bersifat kuantitatif atas suatu objek. Dengan demikian maka data yang akan dianalisis dengan statistik harus berbentuk angka-angka. Apabila data yang ditemui belum berbentuk angka (kuantitatif), langkah awal yang harus dilakukan peneliti adalah melakukan perubahan data agar berbentuk angka. Angka-angka yang digunakan dalam analisis statistik pada dasarnya dapat dikategorikan menjadi 4 (empat) yaitu:

1. Skala nominal yaitu angka yang tidak mempunyai arti hitung. Angka yang diterapkan hanya merupakan simbol/tanda dari objek yang akan dianalisis.

Misalnya: Seorang peneliti menghadapi data yang berkaitan dengan jenis kelamin (perempuan dan laki-laki). Agar peneliti dapat menggunakan statistik dalam analisisnya, dituntut untuk melakukan perubahan data tersebut menjadi bentuk angka. Jika peneliti menggunakan angka 1 sebagai simbol siswa perempuan dan 2 sebagai simbol siswa laki-laki, maka angka 1 dan angka 2 merupakan initial dari

jenis kelamin perempuan dan laki-laki. Untuk selanjutnya peneliti akan selalu berhadapan dengan angka 1 dan angka 2. Dalam hal ini angka 2 tidak berarti lebih besar daripada angka 1, karena angka-angka tersebut hanya sebagai simbol atau kode saja. Sepanjang angka-angka yang digunakan oleh peneliti hanya sebagai simbol, maka angka tersebut dimasukkan sebagai kelompok data yang berskala nominal.

2. Skala ordinal adalah suatu skala yang sudah mempunyai daya pembeda, tetapi perbedaan antara angka yang satu dengan yang lainnya tidak konstan (tidak mempunyai interval yang tetap).

Misalnya : Hasil ujian akhir suatu SMA menyatakan bahwa:

Siswa A sebagai juara 1

Siswa B sebagai juara 2

Siswa C sebagai juara 3

Dst.

Dalam hal ini angka satu mempunyai nilai lebih tinggi daripada angka 2 maupun 3, tetapi skala ini tidak bisa menunjukkan perbedaan kemampuan antara A, B, dan C secara pasti. Juara 1 tidak berarti mempunyai kemampuan dua kali lipat juara dua maupun mempunyai kemampuan tiga kali lipat dari kemampuan juara tiga. Di samping itu perbedaan kemampuan antara siswa juara 1 dengan siswa juara 2, juga berkemungkinan besar tidak sama dengan perbedaan kemampuan siswa juara 2 dengan siswa juara 3. Dengan demikian maka rentangan kemampuan siswa untuk masing-masing juara tidak selalu sama (tetap). Walaupun angka yang dipakai sebagai pengganti mempunyai rentangan yang sama. Penggunaan angka-angka tidak selamanya berpedoman pada angka-angka kecil adalah yang lebih baik, sehingga peneliti dapat menggunakan dasar bahwa angka yang besar adalah yang lebih baik. Mengingat posisi angka sebagai pengganti baik buruk, besar kecilnya suatu data, maka dalam melakukan deskripsi atas hasil analisis statistik harus hati-hati. Sifat konsisten harus dijalankan mulai dari pemberian kode sampai deskripsi.

3. Skala interval yaitu suatu skala yang mempunyai rentangan konstan antara tingkat satu dengan yang lainnya, tetapi tak mempunyai angka 0 mutlak.

Misalnya : Nilai siswa mempunyai rentangan 0 sampai dengan 10. Temperatur mempunyai rentangan dari 0 sampai dengan 100 celcius.

Dalam kasus ini siswa yang memperoleh nilai 8 mempunyai kemampuan 2 kali siswa yang memperoleh nilai 4, panas udara 15 derajat celsius merupakan 0,5 panas udara 30 derajat celsius. Tetapi siswa yang memperoleh nilai 0 bukan berarti tidak mempunyai pengetahuan sama sekali tentang yang diujikan, atau suhu udara berderajat 0 celsius bukan berarti udara tidak bersuhu. Rentangan nilai dari jenjang yang satu ke jenjang yang lainnya bersifat konstan (tetap). Sehingga skala ini dapat memberi gambaran tentang objek yang dinilai secara konsisten.

Dalam penelitian pendidikan sering dijumpai pengukuran terhadap objek-objek yang bersifat kejiwaan (psikologis). Untuk mengungkapkan objek-objek ini peneliti biasanya menggunakan alat ukur yang berskala (misalnya skala Likert). Sepanjang analisis skala tersebut didasarkan pada penjumlahan skor untuk setiap item, maka skor yang terkumpul dapat dikategorikan berskala interval.

4. Skala ratio adalah suatu skala yang mempunyai rentangan konstan dan mempunyai angka 0 mutlak.

Misalnya : Ukuran berat, panjang/tinggi, umur, dll. Seseorang mempunyai berat badan 100 kg. adalah 2 kali beratnya dari orang yang mempunyai berat badan 50 kg. Jika berat sesuatu benda adalah 0. Maka benda tersebut benar-benar tidak mempunyai berat. Hal ini menunjukkan kepada kita bahwa angka 0 mempunyai arti tersendiri (0 adalah mutlak adanya). Siswa yang mempunyai tinggi badannya 160 cm. Adalah  $\frac{3}{4}$  tinggi siswa yang tinggi badannya 120 cm. Jika ada orang yang tinggi 0 cm, maka orang tersebut nyata-nyata tidak bertinggi badan (walaupun dalam kenyataan tak ada orang yang mempunyai tinggi badan 0).

Mengingat masing-masing skala mempunyai arti yang sangat berbeda, maka teknik analisis statistik yang dipakainya (yang sesuai penggunaannya) juga berbeda-beda. Dari keempat skala nilai tersebut, skala ratio mempunyai nilai yang lebih tinggi dari skala nilai yang lainnya. Skala interval menduduki posisi kedua, mengingat kondisi skala interval tidak jauh berbeda dengan skala ratio, maka teknik analisis yang bisa ditetapkan pada kedua skala ini adalah sama. Skala nominal mempunyai nilai yang paling lemah. Skala ordinal mempunyai nilai lebih tinggi daripada skala nominal, tetapi skala ini tidak dapat disamakan dengan skala interval. Jika data yang kita hadapi berskala nominal

maupun ordinal, maka analisis statistik yang dapat digunakan adalah analisis non parametrik.

Analisis statistik inferensial mensyaratkan data berskala interval atau ratio.

### C. Variabilitas

Variabilitas merupakan kondisi dimana sekumpulan skor sama atau tidak. Jika sekumpulan skor itu sama, maka distribusi tersebut tidak mempunyai variabilitas. Besar kecilnya variabilitas merupakan gambaran tentang penyebaran distribusi.

Kegunaan perhitungan variabilitas adalah:

- a. Variabilitas memberikan indikasi bagaimana tingkat akurasi rata-rata dalam menjelaskan distribusi. Jika variabilitas kecil kemudian seluruh skor mengumpul dan setiap skor mendekati harga rata-ratanya, maka rata-rata sample representatif untuk seluruh distribusi skor. Sebaliknya jika variabilitas besar, maka skor tidak mendekati harga rata-ratanya, sehingga rata-rata sampel tidak representatif untuk seluruh distribusi skor.
- b. Variabilitas memberikan indikasi seberapa tepatnya suatu skor atau kelompok skor menggambarkan keseluruhan distribusi. Mengingat rata-rata populasi sering tidak diketahui maka peneliti lebih banyak menggunakan rata-rata yang berasal dari sampel. Jika variabilitas kecil, maka setiap skor akan akurat dalam menggambarkan keseluruhan distribusi. Sebaliknya jika variabilitas sampel distribusi besar, maka setiap skor atau sekumpulan skor tidak akurat dalam menggambarkan keseluruhan distribusi.

Ada 3 (tiga) macam perhitungan variabilitas:

1. **Range** adalah perbedaan antara skor terbesar dan skor terkecil. Range dapat dihitung dengan rumus

$$\text{Range} = \text{Skor tertinggi} - \text{Skor terkecil} + 1$$

Kelemahan pengukuran dengan rumus ini adalah tidak dapat menggambarkan bagaimana variasi skor di antara skor terkecil dengan skor terbesar.

Contoh: Dua buah distribusi sbb.

A.	1	8	9	9	10	10
B.	1	2	4	6	8	10

Kedua distribusi di atas mempunyai range yang sama yaitu:  $10 - 1 + 1 = 10$ , walaupun distribusi kedua kelompok skornya jelas berbeda.

- Interquartile range merupakan perbedaan antara quartile pertama dengan quartile ketiga. Interquartile range dapat dihitung dengan rumus

$$\text{Interquartile} = Q_3 - Q_1$$

Ketika interquartile range digunakan untuk menjelaskan variabilitas, biasanya ditransformasikan ke dalam semi interquartile range. Karena semi interquartile range terfokus pada pertengahan (50% dari distribusi), maka kondisi itu kurang dipengaruhi oleh skor yang ekstrim (terendah dan tertinggi). Oleh karenanya pemakaian rumus ini lebih baik daripada pemakaian rumus range untuk mengetahui variabilitas. Walaupun demikian semi interquartile range tidak dapat dihitung dari skor individual, sehingga dia tidak dapat memberikan gambaran lengkap bagaimana penyebaran skor.

Kedua cara diatas merupakan perhitungan variabilitas yang masih kasar, untuk itu tidak disarankan untuk sering dipakai secara mandiri.

- Simpangan baku (standar deviation) merupakan rata-rata penyimpangan setiap skor dengan rata-rata (mean) skornya. Langkah-langkah dalam perhitungan Simpangan baku sampel (Sd):
  - pertama-tama dihitung dulu rata-rata skornya.
  - Hitung perbedaan masing-masing skor dengan rata-rata skor.
  - Selisih masing-masing skor dengan rata-rata dikuadratkan dan jumlahkan.
  - Hasil penjumlahan dibagi dengan  $n-1$ , hasil perhitungan di sini disebut dengan variance.
  - Akar dari variance merupakan Sd.

Untuk menentukan besarnya simpangan baku dari sekumpulan skor perlu dihitung lebih dulu variance dari skor tersebut. Sedangkan variance atas skor yang terkumpul (sampel dapat dicari dengan rumus :



$$Sd^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Simpangan baku merupakan akar dari variance, sehingga simpangan baku dapat dihitung dengan rumus

$$Sd = \sqrt{Sd^2}$$

Contoh : Suatu distribusi nilai sbb.: 1, 2, 3, 4, 5.

Hitunglah simpangan baku dari data tersebut !

Untuk mempermudah dalam menghitung Sd, maka perlu disusun suatu tabel yang mengandung simpangan setiap skor dengan rata-ratanya dan kuadrat simpangan setiap skor dengan rata-rata. Data contoh di atas, dengan mudah dapat dihitung rata-ratanya yaitu membagi jumlah skor dengan jumlah sampel  $15/5 = 3$ .

Berdasarkan pada nilai rata-rata yang telah di peroleh, dapat ditentukan masing-masing simpangan skor dengan rata-rata.

Selanjutnya dapat disusun daftar kuadrat simpangan masing-masing skor dengan rata-ratanya. Hasil perhitungan diatas dapat dilihat pada tabel di berikut:

X	$X - \bar{X}$	$(X - \bar{X})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4
15	0	10

Dari tabel di atas, dengan mudah dapat dihitung simpangan bakunya, karena jumlah kuadrat simpangan masing-masing skor dengan rata-rata telah diketahui yaitu 10. Dengan menggunakan rumus di atas iperoleh variance, dan dengan mengakar variance diperoleh simpangan baku. Dengan demikian maka besarnya simpangan baku adalah  $Sd = \sqrt{10/4} = 1,58$

Melalui contoh sederhana di atas dapat diambil suatu kesimpulan bahwa variance merupakan rata-rata kuadrat simpangan, sedangkan simpangan baku merupakan rata-rata simpangan skor individual dengan meannya.

Jumlah simpangan skor individual dengan mean selalu 0, jika tidak 0 berarti ada kesalahan perhitungan, oleh karena untuk menghitung simpangan baku perlu dilakukan usaha untuk menghilangkan tanda - (negatif) dengan jalan mengkuadratkan setiap simpangan, baru kemudian dikembalikan dengan jalan mengakarnya.

Untuk populasi pembagi pada rumus bukan  $n-1$  tetapi  $N$ , dalam hal ini  $N$  merupakan jumlah populasi dan  $n$  merupakan jumlah sampel. Hal ini dilakukan untuk menghindari suatu bias dalam variabilitas sampel. Pembagi  $(n - 1)$  merupakan degrees of freedom (derajat kebebasan) untuk sampel.

Semakin kecil simpangan baku, berarti semakin terkumpul distribusi skornya, demikian pula sebaliknya. Dengan demikian maka semakin kecil simpangan baku, maka semakin baik prediksi rata-rata sampel terhadap rata-rata populasinya. Dengan kata lain sekumpulan skor sampel maupun skor individual dapat menggambarkan keseluruhan skor (skor populasi).

Jika kita telah dapat menghitung simpangan baku (sd), tugas selanjutnya adalah mendeskripsikan hasil perhitungan tersebut. Untuk itu mari kita ambil suatu contoh singkat : Seandainya hasil perhitungan rata-rata adalah 70 dan Sd sebesar 10, dan skor yang terkumpul mempunyai distribusi normal, maka nilai di antara  $(70 - 10)$  dan  $(70 + 10)$  sebanyak 68,27% dari seluruh data. Sedangkan nilai diantara  $| 70 - (2 \times 10) |$  dan  $| 70 + (2 \times 10) |$  sebanyak 95,45% dari seluruh data. Dengan demikian maka 68,27% siswa akan memperoleh nilai di antara 60 dan 80, atau 95,45% siswa akan memperoleh nilai di antara 50 dan 90.

Pada uraian rata-rata (mean) di muka, kita telah mencoba untuk melakukan pergeseran/perubahan skor dengan jalan menambah, mengalikan, mengurangi dan membagi dengan suatu bilangan konstan. Hasil perhitungan rata-rata (mean) akan berubah sesuai dengan perubahan skor. Hal ini juga akan terjadi pada perhitungan Sd, namun tidak seluruh Sd berubah sesuai dengan perubahan skor.

Contoh : Distribusi Nilai sbb:

X	X <sup>1</sup>	X <sup>2</sup>	X <sup>3</sup>	X <sup>4</sup>
0	2	-2	0	0
4	6	2	8	2
5	7	3	10	2,5
2	4	0	4	1
4	6	2	8	2
15	25	5	30	7,5

Keterangan : X adalah skor asli

X<sub>1</sub> adalah X + 2

X<sub>2</sub> adalah X - 2

X<sub>3</sub> adalah X x 2

X<sub>4</sub> adalah X : 2

Berdasarkan angka dalam tabel di atas dapat dihitung mean untuk masing-masing X dan Sd untuk masing-masing X dengan menggunakan rumus-rumus rata-rata (mean) dan Sd

X	=	3	Sd	=	2
X <sub>1</sub>	=	5	Sd <sub>1</sub>	=	2
X <sub>2</sub>	=	1	Sd <sub>2</sub>	=	2
X <sub>3</sub>	=	6	Sd <sub>3</sub>	=	4
X <sub>4</sub>	=	1,5	Sd <sub>4</sub>	=	1

Dari hasil perhitungan di atas tampak bahwa penambahan maupun pengurangan seluruh skor dengan suatu bilangan konstan tidak akan mempengaruhi Sd. Namun perubahan skor dengan cara mengalikan maupun membagi dengan bilangan konstan akan mengakibatkan perubahan Sd searah dengan perubahan skor (artinya jika skor dikalikan dengan 2 maka Sd yang baru menjadi 2 kali Sd semula, apabila skor dibagi dengan 2 maka Sd yang baru menjadi 1/2 Sd semula).

Perhatikan hasil perhitungan Sd di atas terutama untuk X, X<sub>1</sub>, X<sub>2</sub>. Ketiganya mempunyai harga yang sama, sedangkan distribusinya berbeda dan meannya juga berbeda. Kondisi ini memberikan petunjuk kepada kita bahwa dalam mencari perbedaan

distribusi tidak cukup hanya membandingkan Sdnya saja. Hal ini berlaku untuk semua jenis pengukuran variabilitas.

Untuk membandingkan variabilitas suatu distribusi yang satu dengan yang lain perlu dicari nilai variabilitas relatifnya, yaitu dengan jalan membagi nilai variabilitas mutlak (Sd, Range, Semi Interquartile Range) dengan rata-rata skornya. Ukuran variabilitas mutlak yang sering dipakai adalah Sd.

Variabilitas relatif sering disebut dengan coefisien of variation (CV) sedangkan CV dapat dihitung dengan rumus :

$$CV = Sd / \bar{X}$$

Contoh: Bandingkan variabilitas distribusi pada soal di depan

- Untuk  $X_1$      $CV = 2/3 = 0,67$
- Untuk  $X_2$      $CV = 2/5 = 0,4$
- Untuk  $X_3$      $CV = 2/1 = 2$
- Untuk  $X_4$      $CV = 4/6 = 0,67$
- Untuk  $X_5$      $CV = 1/1,5 = 0,67$

Berdasarkan hasil perhitungan coefisien of variation kita dapat melakukan perbandingan distribusi. Tampak kini bahwa distribusi  $X_2$  merupakan distribusi yang paling baik di antara sekumpulan distribusi, karena mempunyai nilai CV paling kecil. Walaupun Sd distribusi  $X_1, X_2, X_3$  adalah sama, bukan berarti bahwa distribusi tersebut sama. Hal ini terbukti melalui perhitungan CV. Sd yang terkecil bukan petunjuk mutlak untuk mengatakan distribusinya terbaik. Untuk lebih tepatnya dalam melakukan perbandingan distribusi, kita pakai dua dasar yaitu Sd dan CV. Apabila suatu distribusi mempunyai Sd kecil dan CVnya juga rendah, maka jelas distribusi tersebut adalah yang terbaik.

#### D. Alat-Alat Statistik

##### 1. Deskriptif Univariat

Berikut ini adalah ikhtisar alat statistik yang bisa dipakai untuk menggambarkan profil statistik satu variabel.

Jenis Data	Alat Statistik	Penjelasan
Nominal	<ul style="list-style-type: none"> <li>• Modus (Mo)</li> <li>• Frekwensi relatif (f relatif)</li> <li>• Prosentase (%)</li> <li>• Angka mutlak</li> </ul>	<ul style="list-style-type: none"> <li>• Untuk menjelaskan kecenderungan terbanyak</li> <li>• Untuk menjelaskan penyebaran data menurut frekwensinya</li> <li>• Untuk mengetahui proporsi data dalam persen (%)</li> <li>• Untuk mengetahui jumlah mutlak dari data</li> </ul>
Ordinal	<ul style="list-style-type: none"> <li>• Modus (Mo)</li> <li>• Median (Md)</li> <li>• Simpangan</li> <li>• Kuartil</li> <li>• Prosentase (%)</li> <li>• Angka mutlak</li> </ul>	<ul style="list-style-type: none"> <li>• Untuk menjelaskan kecenderungan terbanyak</li> <li>• Untuk menjelaskan kecenderungan tengah</li> <li>• Untuk menjelaskan penyebaran</li> <li>• Untuk menjelaskan penyebaran</li> <li>• Untuk menjelaskan proporsi data dalam persen (%)</li> <li>• Untuk menjelaskan (menggambarkan) jumlah mutlak dari data</li> </ul>
Interval/Rasio	<ul style="list-style-type: none"> <li>• Rerata (<math>\bar{X}</math>)</li> <li>• Median (Md)</li> <li>• Modus (Mo)</li> <li>• Rentangan</li> <li>• Simpangan baku (SD)</li> <li>• Coefiseen of Variation (CV)</li> </ul>	<ul style="list-style-type: none"> <li>• Untuk menjelaskan kecenderungan tengah</li> <li>• Untuk menjelaskan kecenderungan tengah</li> <li>• Untuk menjelaskan kecenderungan terbanyak</li> <li>• Untuk menggambarkan pola penyebaran (minimum-maksimum)</li> <li>• Untuk menjelaskan pola penyebaran atau homogenitas data</li> <li>• Untuk menjelaskan pola penyebaran Skor dengan standarisasi</li> </ul>

## 2. Untuk Mengukur Asosiasi/Korelasi Antara Dua Variabel

Berikut ini adalah ikhtisar beberapa rumus atau alat statistik yang bisa digunakan di penelitian ilmu sosial.

Jenis Data	Alat Statistik	Penjelasan
Data nominal (X) dan data nominal (Y)	<ul style="list-style-type: none"> <li>• Lambda (<math>\lambda</math>)</li> <li>• Tau (<math>\tau</math>) Goodman &amp; Kruskal dan Kendall</li> <li>• Phi Koefisien (<math>\phi</math>)</li> <li>• Chi-Square (<math>\chi^2</math>)</li> <li>• Koefisien Kontingensi</li> </ul>	<ul style="list-style-type: none"> <li>• Untuk mengukur tingkat asosiasi antara 2 variabel yang semuanya nominal. Misalnya asosiasi antara <i>jenis kelamin dan tingkat pendidikan</i> pegawai di suatu organisasi.</li> <li>• Sama seperti Lambda tetapi lebih teliti/rinci</li> <li>• Tujuan sama, rumus berbeda. Terutama untuk data dikotomi</li> <li>• Tujuan sama, rumus berbeda. Data nominal apa saja, tidak hanya dikotomi</li> <li>• Untuk data kategorikal</li> <li>• Untuk mengukur kekuatan asosiasi/korelasi</li> </ul>
Data ordinal (X) dan data ordinal (Y)	<ul style="list-style-type: none"> <li>• Rho Spearman</li> <li>• Gamma (<math>\gamma</math>)</li> <li>• Point Biserial</li> </ul>	<ul style="list-style-type: none"> <li>• Untuk mengukur asosiasi dua variabel yang datanya di ranking.</li> <li>• Sama seperti Rho Spearman, tetapi rumus ini bisa digunakan untuk memprediksi kecenderungan ranking variabel Y (dengan membaca ranking variabel X)</li> <li>• Simplifikasi dari analisis product-moment. Data bersifat dikotomi, misalnya laki-laki – perempuan; betul – salah, bekerja – tidak bekerja, dan sebagainya.</li> <li>• Bisa juga antara dikotomi dan data kontinyus.</li> </ul>
Data interval/Rasio (X) dan data Interval/Rasio (Y)	<ul style="list-style-type: none"> <li>• Product-moment (Pearson)</li> </ul>	<ul style="list-style-type: none"> <li>• Kedua variabel harus berbentuk kurva normal.</li> </ul>

### 3. Alat Statistik Untuk Mengukur Korelasi Lebih Dari Dua Variabel.

Berikut ini adalah ikhtisar alat statistik untuk mengukur lebih dari dua variabel.

Alat Statistik	Penjelasan
<ul style="list-style-type: none"> <li>• Korelasi ganda (multiple correlation)</li> <li>• Korelasi parsial (partial correlation)</li> <li>• Analisis faktor (factor analysis)</li> <li>• Regresi linear ganda (multiple linear regression)</li> </ul>	<ul style="list-style-type: none"> <li>• Digunakan untuk mengukur korelasi antara lebih dari variabel (bebas atau terikat)</li> <li>• Digunakan untuk mengukur korelasi antara beberapa variabel, dengan satu atau lebih variabel bebas dikontrol</li> <li>• Digunakan untuk menentukan apakah sekelompok variabel bisa direduksi menjadi beberapa faktor saja.</li> <li>• Digunakan untuk mengukur kekuatan hubungan antara satu variabel terikat dengan beberapa variabel bebas.</li> <li>• Untuk memprediksi variabel terikat berdasarkan beberapa variabel bebas.</li> </ul>

### 4. Alat Statistik Untuk Uji Perbedaan.

Berikut ini adalah ikhtisar alat statistik untuk melakukan uji beda antara satu variabel dengan variabel lainnya, dengan satu sampel atau lebih dari satu sampel

Alat Statistik	Penjelasan
<ul style="list-style-type: none"> <li>• Tes Z</li> <li>• Tes T</li> <li>• Tes F</li> <li>• Anova</li> <li>• Anacova</li> <li>• Mann-whitney U-test</li> <li>• Kruskal-Wallis test</li> <li>• Chi-Square test</li> </ul>	<ul style="list-style-type: none"> <li>• Dipergunakan untuk menguji signifikansi perbedaan antara sampel (satu sampel) dengan populasi. Variasi populasi harus diketahui</li> <li>• Juga digunakan untuk uji beda dua sampel</li> <li>• Sama dengan tes Z. Sampel <math>\leq 30</math>. Distribusi populasi normal.</li> <li>• Juga digunakan untuk uji beda antara dua sampel</li> <li>• Sama dengan test, tetapi "mean" lebih dari dua.</li> <li>• Dipergunakan bila sampel yang dibandingkan lebih dari dua.</li> <li>• Sama seperti Anova, tetapi ada variabel bebas (X) yang dikontrol.</li> <li>• Untuk menguji (non parametrik) apa dua sampel (X) yang tidak berhubungan berbeda secara signifikan</li> <li>• Untuk menguji (non parametrik) apakah dua sampel (X) yang berhubungan berbeda secara signifikan</li> <li>• Untuk menguji (non parametrik) apakah dua set distribusi frekwensi berbeda secara signifikan.</li> </ul>

## E. Korelasi

### 1. Pengertian Korelasi

Korelasi merupakan suatu hubungan antara satu variabel dengan variabel lainnya. Hubungan antara variabel tersebut bisa secara korelasional dan juga secara kausal. Jika hubungan tersebut tidak menunjukkan sifat sebab akibat, maka korelasi tersebut dikatakan korelasional artinya sifat hubungan variabel satu dengan variabel lainnya tidak jelas mana variabel sebab dan mana variabel akibat. Sebaliknya, jika hubungan tersebut menunjukkan sifat sebab akibat, maka korelasinya dikatakan kausal, artinya jika variabel yang satu merupakan sebab, maka variabel yang lainnya merupakan akibat.

Pembahasan korelasi minimal menyangkut dua kelompok nilai atau dua variabel. Variabel-variabel tersebut bisa berasal dari subjek penelitian yang sama, tetapi bisa juga terjadi pada atau berasal dari subjek penelitian yang berbeda. Misalnya pada penelitian mahasiswa di FIK UNP, khususnya pengukuran atas tinggi badan dan tinggi lompatan. Setiap satu subjek akan memberikan dua macam nilai yaitu tinggi badannya dan tinggi lompatannya.

Contoh dua nilai yang dapat dicari hubungannya tetapi subjeknya berbeda adalah pengukuran antara tinggi badan ayah dan tinggi badan anaknya setelah dewasa. Apakah ada hubungan antara tinggi badan anak dan ayah?

Beberapa penyebaran yang mungkin terjadi sebagai berikut:

1. Memanjang tegak (mendekati sejajar dengan sumbu vertikal)
2. Memanjang rebah (mendekati sejajar dengan sumbu horisontal)
3. Memanjang ke kanan atas
4. Memanjang ke kanan bawah
5. Bulat tidak menunjukkan arah pasti.

### 2. Korelasi Pearson

Korelasi yang sering digunakan oleh peneliti (terutama peneliti yang mempunyai data-data interval) adalah korelasi Pearson atau Product moment correlation. Sebelum kita menggunakan korelasi ini terlebih dulu kita harus memperhatikan data yang terkumpul, apakah memenuhi persyaratan yang diminta oleh rumus korelasi ini.



Adapun beberapa persyaratan yang harus dipenuhi apabila kita menggunakan rumus ini adalah:

1. **Pengambilan** sampel dari populasi harus random (acak)
2. Data yang dicari korelasinya harus berskala interval atau ratio.
3. Variasi skor kedua variabel yang akan dicari korelasinya harus sama.
4. Distribusi skor kedua variabel yang dicari korelasinya hendaknya merupakan distribusi unimodal.
5. Hubungan antara variabel X dan Y hendaknya linear.

Korelasi Person (Pearson Product moment) dapat dihitung dengan rumus :

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

Selain itu korelasi Pearson dapat juga dihitung dengan menggunakan rumus

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

Hasil perhitungan korelasi pada dasarnya dapat dikelompokkan menjadi 3 (tiga) kelompok besar :

1. **Korelasi positif kuat**, apabila hasil perhitungan korelasi mendekati +1 atau sama dengan +1. Ini berarti bahwa setiap kenaikan skor/nilai pada variabel X akan diikuti dengan kenaikan skor/nilai variabel Y. Sebaliknya jika variabel X mengalami penurunan, maka akan diikuti penurunan variabel Y.
2. **Korelasi negatif kuat**, apabila hasil perhitungan korelasi mendekati -1 atau sama dengan -1. Ini berarti bahwa setiap kenaikan skor/nilai pada variabel X akan diikuti dengan penurunan skor/nilai variabel Y., Sebaliknya apabila skor/nilai dari variabel X turun, maka skor/nilai dari variabel Y akan naik.
3. **Tidak ada korelasi**, apabila hasil perhitungan korelasi mendekati atau sama dengan 0. Hal ini berarti bahwa naik turunnya skor/nilai satu variabel tidak mempunyai kaitan dengan naik turunnya skor/nilai variabel yang lainnya. Apabila skor/nilai variabel X naik tidak selalu diikuti dengan naik atau turunnya skor/nilai variabel Y, demikian juga sebaliknya.

135/K (2003 - t<sub>1</sub> (2) .  
001.422 5 iri t<sub>1</sub>

### 3. Korelasi Spearman

Apabila data yang kita hadapi mempunyai skala ordinal, maka korelasi produk moment tidak dapat digunakan. Untuk itu telah dikemukakan suatu rumus sederhana tetapi akurat yaitu Spearman Correlation. Korelasi Pearson (produk moment correlation) didasarkan pada hubungan linier, sedangkan Spearman correlation justru tidak memperhatikan sifat hubungan linier antara kedua variabel yang akan dicari korelasinya.

Korelasi Spearman dapat dicari dengan rumus :

$$r_s = (\rho) = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

Keterangan :

- D merupakan selisih antara X dan Y
- 6 merupakan angka konstan

### 4. Pengujian Signifikansi Korelasi

Langkah awal dalam pengujian di sini juga menyusun hipotesis nol dan hipotesis alternatif. Baru kemudian hasil kita hitung t untuk sampel kecil atau Z untuk sampel besar. Nilai untuk korelasi Pearson dapat dicari dengan rumus

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

Sedangkan nilai t untuk korelasi Spearman dapat dihitung dengan rumus :

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

Derajat kebesarannya adalah  $n - 2$

Jika sampelnya besar maka kita akan menggunakan Z, sedangkan nilai Z untuk korelasi Pearson dihitung dengan rumus

$$Z = r \sqrt{n-1}$$

Nilai Z untuk korelasi Spearman dihitung dengan rumus 6.10:

$$Z = r_s \sqrt{n-1}$$

Dalam hal ini kita menggunakan asumsi bahwa sampling berdistribusi mendekati norma; (dengan rata-rata = 0 dan standard deviasi =  $1/\sqrt{n-1}$ ).

## F. Regresi

### 1. Pengertian Regresi

Sering kali peneliti ingin melihat kondisi di waktu yang akan datang dengan suatu dasar keadaan sekarang atau ingin melihat kondisi di waktu yang lalu dengan dasar keadaan sekarang. Sifat ingin melakukan prediksi atau taksiran mulai berkembang dalam dunia ekonomi, tetapi sekarang banyak dilakukan di dunia pemerintahan. Bahkan dewasa ini, melakukan prediksi keadaan karyawan untuk waktu yang akan datang merupakan kondisi yang sangat dibutuhkan dalam dunia pemerintahan. Melalui prediksi yang baik, perencanaan baik yang menyangkut promosi, metode pengembangan SDM, fasilitas ruang, dan nilai-nilainya, akan dapat direalisasikan seefisien mungkin.

Sebagaimana layaknya arti kata prediksi, prediksi di sinipun bukanlah merupakan hal yang pasti, tetapi merupakan suatu keadaan yang mendekati kebenaran. Jika kita membandingkan nilai asli variabel yang kita predik dengan nilai prediksinya berkemungkinan besar akan terdapat perbedaan. Perbedaan tersebut bisa terlalu besar maupun terlalu kecil. Sepanjang perbedaan tersebut tidak besar maka prediksi yang kita lakukan merupakan hasil kerja yang luar biasa. Penyimpangan-penyimpangan nilai asli dan nilai prediksi ini sering terjadi karena dalam melakukan prediksi kita berdasarkan nilai rata-rata, dan menggunakan suatu persamaan yang menggambarkan suatu garis tertentu. Sifat yang menggambarkan garis itu bermacam-macam, ada yang lurus, hiperbola, parabola dan lain-lainnya. Untuk menentukan rumus mana yang akan dipakai tergantung pada teori yang dipakai dan kondisi data yang diperoleh, karena masing-masing rumus dikembangkan melalui beberapa asumsi.

Salah satu syarat untuk dapat melakukan prediksi atas variabel terikat di waktu yang akan datang, maupun di dalam populasinya, dengan dasar beberapa skor variabel bebas dan variabel terikat (sebagai sampel) adalah adanya hubungan signifikan antara variabel bebas dan variabel terikat. Sebenarnya statistik tidak dapat membedakan antara data hasil analisis statistik bisa menyesatkan jika dasar teoritisnya tidak kuat. Suatu misal: kita mempredik antara banyaknya panen cengkeh dengan banyaknya janda, mungkin hasil analisis statistik menyatakan bahwa panen cengkeh dapat dipredik berdasarkan banyaknya janda, tetapi secara nyata hasil analisis ini tidak dapat dibuktikan.

Kalau toh dapat dibuktikan hubungan kausal antara variabel banyak janda dan banyak penen cengkeh terdapat banyak variabel antaranya. Dengan banyaknya janda akan banyak kucing di rumah-rumah, karena tikus akan lari dari rumah ke kebun. Di kebun, tikus akan memakan serangga kecil yang merupakan hama cengkeh, sehingga pohon-pohon cengkeh dapat berbunga dengan baik. Kami rasa hubungan seperti ini tidaklah tepat, kecuali seluruh variabel antara memang dimasukkan dalam analisis statistik.

Ada beberapa pola persamaan regresi dengan satu variabel bebas yang dapat digunakan untuk melakukan prediksi, diantaranya:

1. Linier dengan persamaan  $\hat{Y} = a + bX$
2. Parabola dengan persamaan  $\hat{Y} = a + bX + cX^2$
3. Hiperbola dengan persamaan  $\hat{Y} = 1/(a + bX)$ .
4. Fungsi pangkat tiga dengan persamaan  $\hat{Y} = a + bx + cx^2 + dx^3$
5. Dan lain-lain

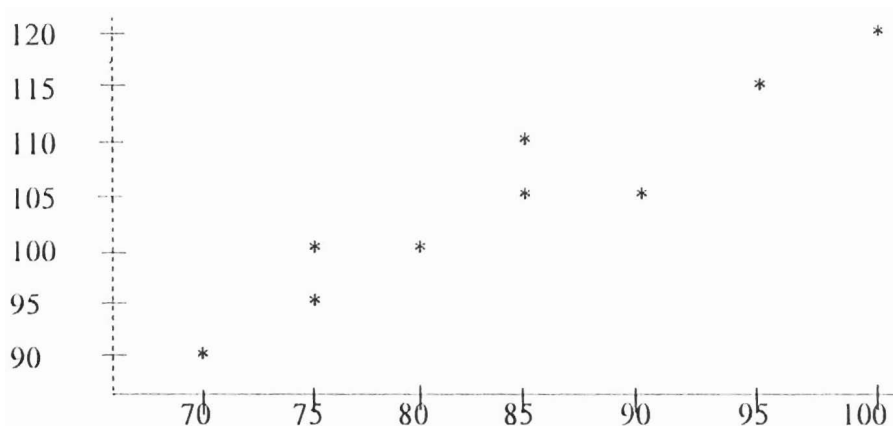
Sedangkan regresi dengan variabel bebas lebih dari satu sampai saat ini baru dikembangkan dengan model linier yang rumus persamaannya:  $\hat{Y} = a + b_1X_1 + b_2X_2 + \dots + b_iX_i$ , dimana jumlah variabel bebas sebanyak  $i$ .

Untuk mempermudah pemahaman regresi perlu kita kembali pada pola penyebaran skor (titik penyebaran skor yaitu titik-titik perpotongan antara nilai  $X$  dan nilai  $Y$ ).

Contoh: Misalnya kita mempunyai data dari dua buah variabel yaitu variabel intelegensi ( $X$ ) dan variabel hasil belajar ( $Y$ ) yang penyebarannya sbb.

X	90	100	100	95	105	110	105	105	115	120
Y	70	75	80	75	85	85	85	80	95	100

Berdasarkan data tersebut di atas kita buat gambar diagramnya (scatter-diagram) sbb:



Jika antara titik satu dengan titik yang lainnya dihubungkan dengan suatu garis, maka akan diperoleh garis yang tidak lurus. Tetapi jika diambil suatu garis yang mewakili rata-rata dari seluruh titik-titik tersebut, maka akan diperoleh garis lurus. Garis lurus itulah yang merupakan garis regresi linier.

Melalui persamaan garis lurus itulah kita dapat melakukan prediksi rata-rata nilai variabel terikat. Jadi dengan mengetahui nilai variabel bebas kita dapat mengetahui rata-rata nilai variabel terikatnya. Tentu dengan kondisi dan situasi yang tidak berbeda dengan sampel, atau dengan kata lain yang dipredik terbatas pada populasi yang diambil sampel. Apabila kita ingin mencrapkan hasil prediksi tersebut pada populasi yang diambil sampelnya, nilai prediksi belum tentu sama dengan nilai aslinya. Hal ini disebabkan oleh karena diprediksi adalah nilai rata-ratanya. Dengan demikian, nilai prediksi dapat dikatakan baik apabila nilai prediksi tidak jauh menyimpang (kalau mungkin sama) dari nilai rata-ratanya tidak terlalu besar (kalau mungkin nol).

Rumus persamaan regresi linier  $\hat{Y} = \alpha + \beta X$  bukan merupakan persamaan yang tepat, artinya persamaan tersebut merupakan pendekatan dari persamaan  $\hat{Y} = \alpha + \beta X$ . Persamaan yang sebenarnya terlalu sukar untuk dihitung (walaupun sebenarnya dapat dihitung dengan bantuan komputer yang canggih) sehingga hasil perhitungan  $\alpha$  dan  $\beta$  yang merupakan pendekatan  $\alpha$  dan  $\beta$  perlu diuji kecocokannya. Jika ternyata  $\alpha$  berfungsi sebagai pengganti  $\alpha$  dan  $\beta$  berfungsi sebagai pengganti  $\beta$ , maka persamaan di atas dapat digunakan sebagai pengganti persamaan sebenarnya, yang fungsinya untuk melakukan prediksi

Nilai a dapat dihitung dengan rumus:

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

Sedangkan nilai b dapat dihitung dengan rumus :

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

### 3. Pengujian Signifikansi Koefisien Regresi

Sebelum kita beranjak lebih lanjut perlu kita melihat apakah koefisien regresi kita itu cukup signifikan? Untuk itu perlu pengujian signifikansi koefisien regresi. Mengingat koefisien regresi yang sangat mempengaruhi nilai taksiran adalah b, maka pengujian di sini cukup hanya menguji signifikansi koefisien regresi b. Walaupun demikian masih dimungkinkan untuk melakukan pengujian signifikansi koefisien regresi a, sebagai kepuasan peneliti. Pengujian kedua koefisien regresi ini mempunyai langkah yang sama.

Apabila kita menguji signifikansi koefisien regresi b, maka diperlukan perhitungan nilai t dari koefisien regresi b dengan rumus:

$$t = (b - \beta) / \sigma$$

Misal hasil perhitungan b dari 10 responden adalah 0,93 dan  $\sigma$  adalah 0,13, maka nilai t hitung adalah

$$\begin{aligned} t &= (0,93 - 0) / 0,13 \\ &= 7,153846154 \\ &= 7,15 \end{aligned}$$

Jika menggunakan  $\alpha = 0,05$  maka  $t_{0.025}$  dengan  $dk = 8$  adalah 2,306. Dengan demikian maka kita menolak hoipotesis nol, artinya koefisien regresi adalah signifikan, sehingga ada hubungan linier yang signifikan antara X dan Y, dan nilai b = 0,93 bukan semata-mata disebabkan oleh faktor random (kebutuhan) saja.

Secara singkat langkah-langkah dalam perhitungan regresi linier sederhana adalah:

1. Menghitung nilai a dan b untuk menentukan persamaan regresi linear sederhana.

2. Menguji signifikansi koefisien regresi (bila seluruh koefisien regresi diuji, tetapi yang terpenting adalah koefisien regresi b).
3. Menghitung variansi untuk selanjutnya digunakan untuk menentukan standard error penaksiran.
4. Menentukan confidence interval dari penaksiran.
5. Menghitung koefisien korelasi, untuk mengetahui koefisien determinasi.
6. Menguji signifikansi daripada koefisien korelasi.
7. Melakukan interpretasi.

Langkah lain untuk menguji hipotesis berkaitan dengan regresi linier adalah melalui analisis variance atau analisis variansi. Dalam hal ini kita akan berhubungan dengan jumlah kuadrat (sum of squares) dari masing-masing variabel. Di sini jumlah kuadrat variabel terikat merupakan jumlah dari: kuadrat jumlah Y dibagi dengan jumlah sampel, hasil kali b dengan jumlah hasil kali simpangan Y dengan  $\hat{Y}$

1. regresi a, dengan derajat kebebasan 1.
2. Regresi (b/a), dengan derajat kebebasan 1.
3. Sisa, dengan derajat kebebasan n-2

Sum of squares yang berkaitan dengan regresi a dapat di hitung dengan rumus:

$$SS_a = (\sum Y)^2 / n$$

Sum of squares yang berkaitan dengan regresi b/a dapat dihitung dengan rumus :

$$SS_{b/a} = b \sum (X - \bar{X})(Y - \bar{Y})$$

Rumus lain :

Sum of squares sisa dapat dihitung dengan rumus :

$$SS_{sisa} = \sum Y^2 - SS_a - SS_{b/a}$$

Mean squares yang berkaitan dengan regresi a dapat dihitung dengan rumus:

$$MS_a = SS_a / dk SS_a$$

Mean squares yang berkaitan dengan regresi b/a dapat dihitung dengan rumus :

$$MS_{b/a} = SS_{b/a} / dk SS_{b/a}$$

Mean squares yang berkaitan dengan regresi sisa hitung dengan rumus :

$$MS_{sisa} = SS_{sisa} / dk SS_{sisa}$$

Tahap akhir dalam pengujian hipotesis signifikansi kontribusi atau sumbangan variabel X terhadap variabel Y adalah menghitung nilai F yang dapat diperoleh dengan rumus ;

$$F = MS_{b/n} : MS_{\text{sisas}}$$

Sumber Variansi	Dk	SS	MS	F
Regresi a	1	71402,5	71402,5	
Regresi b/a	1	625,425	625,425	
Sisa	8	97,075	12,134375	51,54
Total	10	72125		

Sedangkan syarat-syarat yang harus dipenuhi dalam perhitungan di atas adalah :

- Sampel diambil secara random (acak)
- Variabel X dan variabel Y mempunyai hubungan yang kausal, dimana X merupakan sebab dan Y merupakan akibat.
- Nilai Y mempunyai penyebaran yang berdistribusi normal
- Persamaan tersebut hendaknya benar-benar linear.

## F. Anova

### 1. Pengertian

Membandingkan satu rata-rata populasi dengan satu rata-rata populasi yang lain digunakan t test. Sering kali kita menghadapi banyak rata-rata (lebih dari satu rata-rata). Apabila kita mengambil langkah pengujian perbedaan rata-rata tersebut satu persatu (dengan t tes) akan memakan waktu, tenaga yang banyak. Di samping itu kita akan menghadapi resiko salah yang besar. Untuk itu telah ditemukan cara analisis yang mengandung kesalahan lebih kecil dan dapat menghemat waktu serta tenaga yaitu dengan ANOVA (Analysis of variances).

Pada saat kita menghadapi beberapa kelompok sampel perlu kita sadari dari awal kondisi sampel tersebut sebelum kita melakukan analisis lebih lanjut. Tuntutan ini disebabkan karena pola sampel akan berpengaruh terhadap pengujian hipotesis yang akhirnya berpengaruh terhadap kesimpulan yang diambil.

Pada dasarnya pola sampel dikelompokkan menjadi dua kelompok yaitu:



- a. Seluruh sampel, baik yang berada pada kelompok pertama sampai dengan yang ada kelompok lain, berasal dari populasi yang sama. Untuk kondisi hipotesis nol terbatas pada tidak ada efek dari treatment (perlakuan).
- b. Sampel yang ada di kelompok satu berasal dari populasi yang berbeda dengan populasi sampel yang ada dikelompok lainnya. Untuk kondisi ini hipotesis nol dapat berbunyi tidak ada perbedaan efek treatment antar kelompok.

Mengingat ANOVA berkaitan dengan pengujian hipotesis yang multipel (ganda), maka perhitungannya lebih kompleks dari pada tes. Pada saat melakukan pengujian hipotesis (perbedaan dua rata-rata) dengan menggunakan t tes selalu menanggung kesalahan tipe I sebesar alpha. Untuk ANOVA kesalahan tipe I disebut dengan experiment wise alpha level yang besarnya:

$$1 - (1 - \alpha)^N$$

N merupakan banyaknya tes jika menggunakan t tes (dilakukan satu persatu).

Misalnya untuk pengujian perbedaan rata-rata dari 5 kelompok sampel. Jika diambil alpha sebesar 0,05, maka dengan penggunaan t tes besarnya resiko kesalahan tipe I untuk sekali pengujian adalah 0,05, dan untuk 10 kali pengujian berarti menanggung kesalahan tipe I sebesar 0,50. Apabila kita menggunakan ANOVA kesalahan tipe I yang harus ditanggung adalah :

$$1 - (1 - 0,05)^{10} = 0,40$$

## 2. *Macam-Macam ANOVA*

Pada dasarnya ANOVA dapat dibagi menjadi dua kelompok besar yaitu :

- a. Beberapa kelompok yang dihadapi merupakan pembagian dari satu independent variable (variabel bebas). Kondisi ini yang sering disebut dengan single factor experiment (analisis variance satu arah), yang merupakan topik bahasan pada bab ini.
- b. Beberapa kelompok yang dihadapi merupakan pembagian dari beberapa independent variable (variabel bebas). Kondisi seperti ini sering tersebut dengan two factor experiment (analisis variance dua arah).

Untuk mempermudah memahami konsep pengelompokan ANOVA dalam dua bagian besar, perlu kiranya kita lihat suatu contoh model ANOVA yang mungkin ditemui.

### 1. Single factor experiment :



### Metode Mengajar

A	B	C	D
Sampel	Sampel	Sampel	Sampel

## 2. Two factor experiment

### Metode Mengajar

Metode Mengajar		A	B	C
Jenis	I.	Sampel	Sampel	Sampel
Kelamin	P	Sampel	Sampel	Sampel

Melalui pengamatan model di atas, jelas bahwa dalam two factor experiment pengelompokan dipengaruhi oleh lebih dari satu variabel. Untuk memperoleh data pada setiap sel, maka pengelompokan itu dilakukan pada saat pengambilan sampel. Oleh karena itu model ini lebih cocok untuk penelitian eksperimenal dimana dari awal sampel-sampel tersebut telah dikotak-kotakan. Melalui control yang baik maka perembesan informasi antar kelompok dapat ditanggulangi. Walaupun demikian kita menutup kemungkinan ANOVA diterapkan pada penelitian survey.

## 3. Asumsi dasar dalam ANOVA

### a. Kenormalan

Setiap harga dalam sampel berasal dari distribusi normal, sehingga distribusi skor sampel dalam kelompokpun hendaknya normal. Kenormalan dapat diatasi dengan memperbanyak sampel dalam kelompok, karena semakin banyak  $n$  maka distribusi akan mendekati normal. Apabila sampel tiap kelompok kecil dan tidak normal dapat pula diatasi dengan jalan melakukan transformasi.

### b. Kesamaan variansi

Masing-masing kelompok hendaknya berasal dari populasi yang mempunyai variansi yang sama. Untuk sampel yang sama pada setiap kelompok, kesamaan variansi dapat diabaikan. Tetapi jika banyaknya sampel pada masing-masing kelompok tidak sama, maka kesamaan variansi populasi memang sangat diperlukan. Kalau hal ini diabaikan bisa menyesatkan (terutama dalam pengambilan keputusan). Apabila

variansi berbeda dan banyaknya sampel tiap kelompok tidak sama, diperlukan langkah penyelamatan yaitu dengan jalan melakukan transformasi (misalnya dengan mengtransformasikan dengan logaritma).

c. Pengamatan Bebas.

Sampel hendaknya diambil secara acak (random), sehingga setiap pengamatan merupakan informasi yang bebas. Asumsi ini merupakan asumsi yang tak bisa ditawar lagi, dengan kata lain tak ada cara untuk mengatasi tidak terpenuhinya asumsi ini. Dengan demikian maka setiap peneliti harus merencanakan secara cermat dalam pengambilan sampel.

Asumsi-asumsi di atas hendaknya dipenuhi oleh data yang akan dianalisis dengan ANOVA. Ketidakterpenuhinya asumsi ini dapat menimbulkan kesimpulan yang salah. Hal ini mengandung arti bahwa kesimpulan penelitian yang dianalisis dengan ANOVA tidak memberi arti apa-apa. Walaupun ada asumsi yang sifatnya tidak kaku. Artinya dapat diatasi dengan jumlah sampel namun pengujian atas terpenuhinya asumsi merupakan tindakan yang disarankan.

#### 4. Analisis Sesudah ANOVA

Sesudah perhitungan  $F$  tes dan kita dapat membandingkannya dengan  $F$  tabel, analisis kita sebenarnya belumlah selesai. Hal ini disebabkan karena kesimpulan yang didasarkan pada perhitungan  $F$  tes dalam ANOVA hanyalah merupakan kesimpulan yang masih luas (kasar). Seandainya  $F$  signifikan (menolak  $H_0$ ), ini berarti ada perbedaan efek treatment terhadap output dari masing-masing kelompok. Namun informasi perbedaan efek tersebut masih bersifat umum, karena  $F$  tes sama sekali tidak menunjukkan efek treatment terhadap kelompok mana yang berbeda.

Untuk mempermudah dalam pemahaman konsep di atas marilah kita perhatikan sebuah ilustrasi yang berkenaan pengujian hipotesis. Misalnya kita menghadapi 4 (empat) kelompok, maka hipotesis non dan hipotesis alternatifnya sbb :

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H_1$  : paling sedikit 1 (satu)  $\mu$  tidak sama.

Jika  $F$  hitung  $>$  daripada  $F$  tabel, maka kita akan menolak  $H_0$ , sehingga yang kita hadapi ada beberapa kemungkinan yaitu :

$$\mu_1 = \mu_2 = \mu_3 \neq \mu_4$$

$$\mu_1 = \mu_2 \neq \mu_3 = \mu_4$$

$$\mu_1 \neq \mu_2 = \mu_3 = \mu_4$$

$$\mu_1 \neq \mu_2 = \mu_3 \neq \mu_4$$

$$\mu_1 = \mu_2 \neq \mu_3 \neq \mu_4$$

$$\mu_1 \neq \mu_2 = \mu_3 \neq \mu_4$$

$$\mu_1 \neq \mu_2 \neq \mu_3 = \mu_4$$

$$\mu_1 = \mu_3 \neq \mu_2 \neq \mu_4$$

$$\mu_1 = \mu_4 \neq \mu_2 \neq \mu_3$$

$$\mu_2 = \mu_4 \neq \mu_1 \neq \mu_3$$

$$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

Beberapa kemungkinan di atas merupakan informasi yang lebih teliti, sehingga deskripsi kita akan lebih tajam. Oleh karena itu ANOVA harus dilanjutkan lagi dengan analisis lain yang dapat memberikan informasi yang lebih teliti lagi. Analisis lanjutan ANOVA sering disebut dengan pasca ANOVA (post hoc).

Untuk memperoleh informasi lebih lanjut, telah ditemukan beberapa tes statistik. Yang umum dipakai dalam pendidikan adalah Tukey's HSD. Pembahasan lebih lanjut yang akan dibicarakan disini, jika kita menolak hipotesis nol.

Langkah analisis pasca ANOVA :

1. Hitung Tukey's HSD dengan rumus :

$$HSD = q \sqrt{\frac{MSw}{n}}$$

*Keterangan*

M adalah banyak sampel per kelompok.

Q adalah the studentized range statistic, yang dapat dilihat dalam tabel yang sudah disusun, dengan memakai dasar alpha (  $\alpha$  ), k dan dk.

K adalah banyaknya kelompok

$$Dk = (\sum n) - K$$

2. Cari perbedaan rata-rata antar kelompok, dan untuk mempermudah dalam menginterpretasikan perlu disusun dalam satu tabel khusus.
3. Interpretasikan nilai HSD yaitu dengan jalan membandingkan perbedaan rata-rata antar kelompok dengan hasil perhitungan HSD. Apabila perbedaan rata-rata antar kelompok itu lebih besar daripada nilai HSD, maka perbedaan tersebut dapat dikatakan signifikan.

MILIK PERPUSTAKAAN  
UNIV. NEGERI PADANG

## DAFTAR BACAAN

- Blalock, Herbert M. Jr., *Social Statistics*. Tokyo : McGraw-Hill, 1981.
- Gravetter, Frederick J., Larry B. Wallman. *Statistics For The Behavioral Sciences*. San Fransisco : West Publishing Company, 1985.
- Kennedy, John J., Andrew J. Bush. *An Introduction to The Design and analysis of Experiments Behavioral Research*. Lanham : University Press of America, 1985.
- Kerlinger, Fred N., Elazar J Pedhazur. *Multiple Regression in Behavioral Research*. New York : Holt Rinehart and Wiston, Inc. 1973.
- Ott, Lyman, Richard F. Larson, William Mendenhall. *Statistics, A Toll For The Sosial Sciences*. Boston : Duxbury, 1987.
- Sprinthall, Ricard C. *Basic Statistical Analysis*. New Jersey : Printice Hall, Inc, 1987.
- Sudjana. *Metode Statistika*. Bandung : Tarsito, 1984
- Tuckman, Bruce W. *Conducting Educational Research*. New York : Hartcourt Brace Javanovick, Inc, 1972.