Mahfudhotin, M.Si.

STATISTIKA

- **Statistika Deskriptif**
- **Distribusi Peluang**
- **Penaksiran, Uji Hipotesis, ANOVA**
- **Analisis Deret Waktu**

IAIN Kediri Press

STATISTIKA

Penulis: Mahfudhotin, M.Si.

Editor:
Layout: Epullah
Desain Cover: Audina
Cetakan: I Oktober 2022

Vi + 103 hlm. : 15,5 X 23 Cm

ISBN:978-623-7682-07-3

Diterbitkan oleh:
IAIN Kediri Press
Jl. Sunan Ampel 07 Ngronggo Kediri Jawa Timur
64127 Telp. (0354) 689282, Fax (0354) 686564

Percetakan:
Nadi Pustaka offset
Jl.Nakulo No.19A Pugeran
Maguwoharjo Depok Sleman Yogyakarta
Telp. 0274-4333626 / 081578626131

KATA PENGANTAR

بسم الله الرحمن الرحيم

Puji syukur penulis panjatkan kehadirat Tuhan yang Maha Esa karena atas kehendak-Nya makalah ini dapat terselesaikan tepat waktu. Shalawat serta salam tidak lupa penulis hanturkan kepada junjungan kita,Nabi Muhammad SAW yang telah membawa kita dari zaman jahiliyah menuju kezaman islamiyah, dari zaman kegelapan menujuzaman terang benerang ini.Karena beliaulah satu-satunya Nabi pembawa sekaligus pemberi syafaat kepadaseluruh umat kelak di yaumul qiayamah. Buku Statistika ini merupakan hasil dari kumpulan-kumpulan modul yang dikerjakan selama melakukan perkuliahan.

Secara umum, statistika berkaitan dengan pemodelan dan analisis data, merancang eksperimen, membuat kesimpulan, prediksi, dan keputusan dalam menghadapi ketidakpastian. Statistika adalah penerapan statistik; permasalahan terapan yang menjadi pemicu berkembangnya metode baru dan penerapan baru metodologi statistik lanjutan. Penelitian ini terletak pada berbagai bidang statistik, mulai dari kajian teoritis hingga kajian terapan.

Penulis menyadari buku ini masih banyak memiliki kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang positif agar karya ini menjadi lebih baik dan berdaya guna di masa yang akan datang.

Kediri 2022

Penulis

DAFTAR ISI

I. STATISTIKA DESKRIPTIF

A. TUJUAN

Menyarikan data atau mengambil informasi dari data (*data summary)* sebagai gambaran awal bagi peneliti/ pengamat, khususnya dalam mengeksplorasi data mentah yang ada. Satu hasil yang diharapkan dari menyarikan data adalah dapat menentukan bentuk distribusinya.

B. PENDAHULUAN

Hal yang terpenting dalam statistika atau analisis data adalah belajar melihat data secara **cermat**. Kita harus belajar melihat bukan hanya 'pada' suatu data akan tetapi juga 'ke dalam' data tersebut. Artinya kita harus dapat membaca data secermat kita membaca kata. Membaca angka (data) jauh lebih sederhana dan tidak kabur dibandingkan informasi suatu kata. Perhatikan data dibawah ini:

Tabel II-1Estates Production by Crops, Indonesia, 1990-2001 (Ton)

Year	Hevea Rubber	Palm Oil	Palm Kernel	Cocoa	Coffee	Tea	Quinine	Tobacco[1)]
1990	315,3	2.096.900	445,8	41,5	25,5	129,1	1,9	3,5
1991	330,1	1.843.600	406,2	30,6	26,4	125	2,1	4,9
1992	335	2.186.000	483,1	39,5	23,9	113	2,7	7,500
1993	335	21288.300	524,6	42,7	20,9	100	6,00	3,1
1994	326,4	1.930.300	472,1	43,7	19,700	98	3,00	5,1
1995	341	2.476.400	605,3	46,4	20,8	111,1	3,00	9,9
1996	334,6	2.569.500	626,6	48,8	28,5	132	4,00	7,1

1997	330,5	4.081.100	927,5	65,9	30,6	121	5,00	7,8
1998	332,6	4.013.100	912,068	60,926	28,5	132,7	4,00	7,7
1999	303,61	4.024.821	914,731	58,915	27,493	130,465	9,17	5,797
2000	336,2	4.094.073	930,603	60,572	29,5	127,902	9,34	6,312
2001[+)]	328,32	4.152.596	946,872	65,293	28,681	129,26	9,20	5,116

Note :

[1)]. *Including production which uses raw materials from smallholder.*

[+)]. *Estimation figures.*

Informasi apa yang Anda peroleh dari data pada tabel diatas?

Apabila Anda cermat maka Anda akan dapat melihat bahwa untuk produk *cocoa,* produksi terbesar adalah pada tahun 1997 dan terendah pada tahun 1994. Anda pun dapat mengatakan bahwa kenaikan produksi *Quinine* sangat drastis pada tahun 1999. Apabila Anda teruskan melihat dan membaca tabel tersebut maka Anda akan dapat informasi yang lebih banyak dan rinci.

Tapi yang jadi permasalahan adalah bagaimana cara Anda menyarikan informasi-informasi yang terkandung dalam data tersebut dengan efektif dan efisien, lalu menginterpretasikan maknanya kedalam suatu kalimat.

Untuk memahami data secara umum ada dua cara pendekatan yaitu: *eksplorasi* dan *konfirmasi*. Eksplorasi adalah suatu pendekatan untuk menemukan suatu informasi dari data. Cara ini dapat menghasilkan suatu taksiran, ide, atau hipotesis yang terkadang dipakai sebagai langkah awal kegiatan ilmiah. Sedangkan konfirmasi ditujukan untuk pengujian suatu dugaan atau hipotesis yang diperoleh dari eksplorasi sebelumnya.

C. MENGHITUNG SARI NUMERIK

Sari Numerik merupakan informasi yang didapat dari data yang merupakan hasil dari eksplorasi tahap pertama. Sari numerik yang akan dibahas pada praktikum ini meliputi:

1. Ukuran Pemusatan Data / Parameter Lokasi

Taraf atau pusat atau parameter lokasi adalah salah satu yang terpenting dalam gambaran data. Taraf antara lain memberikan prediksi cepat yang terbaik mengenai besarnya harga yang mungkin dihasilkan oleh suatu proses tertentu. Sebagai contoh, rata – rata produksi *palm oil* dari tahun 1990-2001 merupakan taksiran terbaik bagi jumlah produksi tahun 2002. Ada berbagai jenis taraf, penggunaannya tergantung keperluan penelitian.

a. Rata-rata (Mean Sampel)

Rata-rata adalah taraf yang paling umum dan sering digunakan. Untuk menghitungnya dibutuhkan seluruh data yang ada. Rata-rata sangat berguna untuk tahap konfirmasi. Tetapi rata-rata mempunyai kelemahan, yaitu tidak tangguh (tidak *robust)* artinya apabila terdapat data ekstrim yang sangat mempengaruhi. Misalnya, dari 16 pelanggan sebuah minimarket, 15 orang pelangganberbelanja per minggu antara Rp 50.000 - Rp 70.000 dan satu orang lagi Rp 250.000. Akibatnya rata-rata belanja16 pelanggan minimarket tersebut tadi tidak akan menggambarkan keadaan sebenarnya.

Lambang untuk rata-rata sampel adalah $\bar{x}$. Formulanya adalah

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

dengan$n$ adalah banyaknya data, dan $x_i$adalah observasi ke-i, i = 1, 2, ..., n.

Catatan :

Penulisan x_i berarti data ke-i, untuk i = 1, 2, ... sedangkan untuk penulisan $x_{(i)}$ berarti data ke-i untuk i = 1,2, ... , setelah data tersebut diurutkan (sorting).

b. Median

Median adalah nilai tengah dari data. Untuk mendapatkan median secara manual Anda harus mengurutkan data terlebih dahulu (sorting).

Median termasuk taraf yang kokoh atau robust, artinya tidak terpengaruh oleh data ekstrim. Median menggunakan sedikit sekali data tetapi tidak kehilangan banyak informasi. Jika data tidak memiliki nilai ekstrim maka median akan mempunyai hasil yang hampir sama dengan rata-rata. Meskipun demikian, median memaksa kita menggantungkan diri pada satu atau dua observasi suatu kondisi yang kurang menyenangkan. Selain itu median masih mahal untuk digunakan dalam tahap konfirmasi karena membutuhkan teknik matematika yang rumit.Median disebut juga sebagai kuartil tengah (q_2).

c. Kuartil

Pusat dari data tidak harus selalu ditengah namun bisa di bagian seperempat terendah atau pun seperempat tertinggi tergantung dari kebutuhan peneliti dan maksud dari penelitian. Oleh karena itu Anda harus tahu bagaimana menghitung nilai kuartil.

i. Kuartil bawah (q_1) : nilai yang terletak pada observasi urutan ke-$\frac{(n+1)}{4}$.

$$q_1 = x_{\left(\frac{n+1}{4}\right)}$$

ii. Kuartil atas (q_3) : nilai yang terletak pada observasipada urutan ke- $\frac{3(n+1)}{4}$.

$$q_3 = x_{\left(\frac{3(n+1)}{4}\right)}$$

Kuartil juga digunakan dalam hal pemangkasan data, artinya kita memotong atau menghilangkan data seperempat terendah dan/atau seperempat tertinggi yang mungkin mengandung nilai ekstrim. Sehingga taraf yang dihasilkan dari data dibagian tengah ini akan lebih meyakinkan.

d. Trirata

Trirata disebut juga rata-rata tengah.

$$Trirata\ (TRI) = \frac{q_3 + q_1 + 2q_2}{4}$$

Trirata menggunakan lebih banyak data daripada median. Trirata memberikan bobot gAnda pada median karena jelas median dekat ke pusat yang merupakan tujuan penelitian.

e. Modus (mode)

Modus / Mode adalah nilai data yang paling sering muncul. Modus tidak memerlukan perhitungan yang rumit, cukup kejelian menghitung frekuensi dari data. Namun tidak semua data memiliki modus terutama data kontinu yang memiliki ketelitian lebih dari 2 angka desimal.

f. Rata-rata Geometri

Rata-rata geometri berbeda dengan rata-rata numerik (rata-rata biasa) karena menggunakan perkalian dalam menghitungnya. Rata-rata geometri cocok digunakan untuk data yang condong atau miring ke kiri artinya data yang bernilai kecil lebih banyak dari data yang bernilai besar. Jangan menggunakan rata-rata geometri jika ada data yang bernilai negatif.

$$GM = \sqrt[n]{x_1 x_2 \ldots x_n}$$

2. Ukuran Penyebaran Data / Parameter Dispersi

Mengetahui perkiraan pusat dari data sangat bermanfaat. Tetapi hal ini tidak cukup, kita perlu mengetahui bagaimana data menyebar disekitar pusat tersebut, *sebaran* data. Kalau sebarannya rendah, tarafdapat memberikan perkiraan yang paling baik, sedangkan kalau sebarannya tinggi taraf tidak terlalu membantu dalam peredeksi hasil tertentu.

a. Variansi dan Simpangan Baku (Standar Deviasi)

Variansi adalah salah satu statistik yang digunakan untuk mengukur sebaran data. Notasinya adalah s^2.

Rumus baku:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \qquad (1)$$

Namun secara praktis yang digunakan adalah:

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$ atau

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} \qquad (2) - (3)$$

Hal tersebut adalah karena ketelitian persamaan (2) dan (3) lebih baik daripada persamaan (1), khususnya untuk $\bar{x}$ yang dibulatkan.

b. Simpangan baku(*s*)

Simpangan baku(*s*) adalah akar positif dari variansi. Variansi tidak kokoh atau sangat dipengaruhi nilai ekstrim, tetapi variansi berguna dalam tahap eksplorasi atau inferensi.

c. Rentangan (*range*)

Range adalah selisih dari nilai maksimum dan nilai minimum. Rentangan merupakan ukuran sebaran yang paling sederhana. Namun *range* tidak tangguh karena didasarkan pada harga yang kemungkinan besar harga ekstrim.

d. Simpangan Kuartil

d_q = kuartil atas – kuartil bawah = $q_3 - q_1$. Ukuran ini merupakan solusi dari kelemahan rentangan dengan simpangan kuartil kita dapat menghindari nilai ekstrim.

e. Standar Error

Standar error dari mean sampel $\bar{x}$ adalah ukuran kedekatan antara rata-rata sample dengan rata-rata populasinya. Semakin dekat rata-rata sampel dengan rata-rata populasi maka semakin kecil standar errornya artinya semakin banyak sample yang kita ambil maka semakin baik rata-rata yang kita dapatkan.

$$\frac{simpangan\ \ baku\ \ Populasi}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}\ .$$

3. Ukuran Kemencengan dan Kelancipan

a. Kemencengan (*Skewness*)

Skewness adalah derajat kemencengan (derajat simetris) dari suatu distribusi. Kemencengan dari suatu distribusi didefinisikan sebagai :

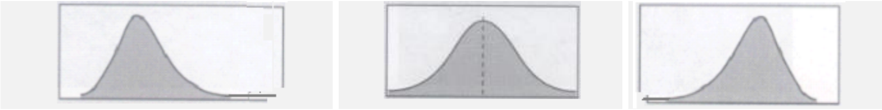
$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3}\ ,$$

dan ditaksir oleh:

$$\hat{\gamma} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}\ .$$

Suatu distribusi yang simetris, seperti distribusi normal dan *t*, memiliki nilai skewness = 0, sedangkan jika skewness nilainya positif menyatakan distribusi "menceng ke kanan" sehingga bagian kanan data berisi lebih banyak datum dibandingkan sebelah kiri. Artinya bahwa data menumpuk lebih banyak di nilai yang lebih kecil. Sedangkan "menceng ke kiri" diberikan oleh nilai skewness negatif, yaitu data menumpuk lenih banyak di nilai yang lebih besar.

Skewness Positif (mean > median)	Skewness, skew = 0 (mean = median)	Skewness Negatif (mean < median)



Gambar II. 1 Bentuk-bentuk distribusi berdasarkan skewness: positif (kiri), simetris (tengah), dan negatif (kanan).Skewness ini diperoleh dari membandingkan nilai rataan (mean) dan median.

b. Kelancipan (Kurtosis)

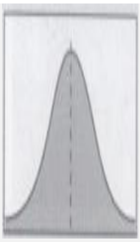
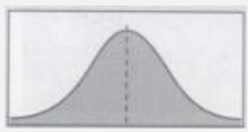
Kurtosis adalah derajat kelancipan dari suatu distribusi dibandingkan distribusi normal. Distribusi normal memiliki kurtosis = 0. Kurtosis didefinisikan sebagai berikut:

$$K(x) = \frac{E[(X-\mu)^4]}{\sigma^4},$$

dan ditaksir oleh:

$$\hat{K}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2}.$$

Bila kurtosis positif maka distribusi memiliki puncak yang lebih lancip dari distribusi normal. Sedangkan bila negatif maka distribusi memiliki puncak yang lebih landai dari distribusi normal.

Kurtosis Positif (leptokurtic dist.)	Kurtosis = 0 (normal distribution)	Kurtosis Negatif (platykurtic dist.)
		

Gambar II. 2 Bentuk-bentuk distribusi berdasarkan kurtosis relatif terhadap distribusi normal: leptokurtic (kiri), normal (tengah), dan platykurtic (kanan).

Pada umumnya, nilai kurtosis untuk distribusi normal adalah tiga (kurtosis = 3) seperti yang digunakan pada software Mic. Excel, Minitab, dan Matlab. Namun pada SPSS, nilai kurtosis untuk distribusi normal adalah nol (kurtosis = 0).

Contoh II-1:Data berikut menyatakan waktu penyembuhan penyakit *Typus* (dalam hari) dari 50 orang pasien :

21	20	31	24	15	21	24	18	33	8
26	17	27	29	24	14	29	41	15	11
13	28	22	16	12	15	11	16	18	17
29	16	24	21	19	7	16	12	45	24
21	12	10	13	20	35	32	22	12	10

Untuk menampilkan sari-sari numerik tersebut di atas.

Masukkan data dalam satu kolom padaData Editor, taskbar Data View. Beri nama variabel sebagai TYPUS, yang dapat diubah pada taskbar Variable View.

Pilih menu Analyze→Descriptive Statistics → Frequencies. Masukkan variabel yang akan dicari sari numeriknya kemudian klik statistics dan check sari numerik yang diinginkan, klik continue → klik OK.

Maka pada window report akan ditampilkan sebagai berikut :

Tabel II-2Output sari numerik data waktu penyembuhan penyakit *Typus* pada 50 orang pasien.

Statistics

TYPUS

N	Valid	50
	Missing	0
Mean		20,3200
Std. Error of Mean		1,18406
Median		19,5000
Mode		24,00
Std. Deviation		8,37255
Variance		70,100
Skewness		,825
Std. Error of Skewness		,337
Kurtosis		,587
Std. Error of Kurtosis		,662
Range		38,00
Minimum		7,00
Maximum		45,00
Sum		1016,00

Percentile s	25	13,7500
	50	19,5000
	75	24,5000

D. MENYAJIKAN DATA DALAM BENTUK GRAFIK

Pada tahap eksplorasi, selain menyusun informasi yang berbentuk angka diperlukan juga untuk melihat informasi data dari sajian data yang dapat menggambarkan data secarakeseluruhan. Ada beberapa cara penyajian data dalam bentuk grafik, antara lain:

- *Scatterplots* (diagram pencar)
- *Dotplot*
- Histogram
- Diagram Batang-Daun (*Stem-Leaf*)
- *Boxplot*

Dengan melihat sebuah grafik yang merepresentasikan data, dapat dilihat kecenderungan data secara umum, observasi yang yang tidak mengikuti kecenderungan data tersebut, dan hubungan yang penting antara variabel-variabel yang menyusun data tersebut.Dengan meng-gabungkan sari numerik dan grafik diharapkan Anda dapat lebih mencerna informasi dari data.

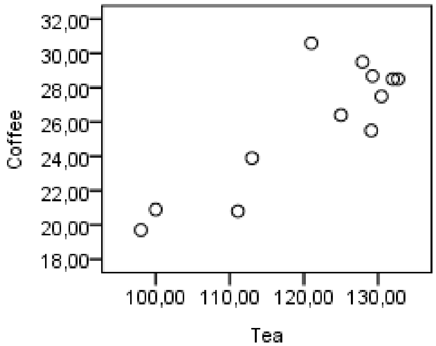
1. Scatterplot

Nama lain dari *scatterplot* adalah diagram pencar. *Scatterplot* adalah plot yang menAndai setiap observasi dengan titik dalam bidang koordinat segiempat *XY*. Setiap observasi terdiri dari dua nilai: setiap nilai merupakan titik absis (pada sumbu *X*) dan nilai kedua merupakan titik ordinat (pada sumbu *Y*). Sumbu *X* dan *Y* dapat dinamai sesuai dengan dua nama variabel yang sedang ingin diketahui hubungan antara kedua variabel tersebut.

Scatter plot dapat dimanfaatkan sebagai acuan awal untuk melihat hubungan yang mungkin dari dua variabel yang diamati.

Untuk menampilkan Scatter plots pada SPSS pilih menu Graphs →Legacy Dialog →Scatter/Dot, pilih simple scatter, klik Define, kemudian pilih variabel Y dan variabel X yang akan dibuat Scatter plot nya.Anda dapat memberi judul atau catatan kaki pada grafik dengan meng-klik kotak *titles.*

Contoh II-2: Perhatikan data produksi kopi dan teh pada Tabel II-1. Akan dilihat hubungan yang mungkin di antara dua variabel tersebut dengan menggunakan scatter plot. Diperoleh grafik sebagai berikut.



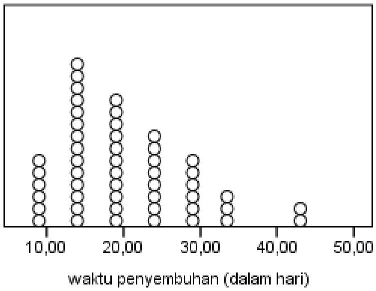
Gambar II. 3 Scatter plot dari produksi kopi dan teh pada Tabel II-1. Terlihat adanya hubungan yang cukup linier antara produksi kopi dan teh pada kurun waktu tersebut.

2. Dot Plot

Dot plot adalah suatu jenis plot dimana data yang ada diinterpretasikan dalam bentuk titik. Dot plot juga menggambarkan frekuensi dari data. Dalam hal ini data yang sama akan ditumpuk pada satu titik data yang bersangkutan.

Untuk menampilkan Dot plots pada SPSS pilih menu Graphs → Legacy Dialog→Scatter/Dot, pilih simple Dot kemudian pilih variabel yang akan dibuat Dot plot nya. Anda dapat memberi judul atau catatan kaki pada grafik dengan meng-klik kotak *titles.*Anda dapat mengedit tampilan grafik dengan meng-klik dua kali grafiknya.

Contoh II-3: Berikut adalah dotplot untuk data waktu penyembuhan penyakit Typus pada Contoh II-1.

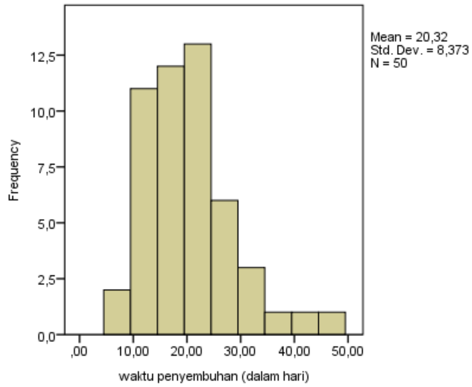


Gambar II.4Dotplot data waktu penyembuhan penyakit Typus.Terlihatbahwa data menumpuk di nilai yang lebih kecil.

3. Histogram

Nama lain dari histogram adalah diagram batang. Dengan histogram dapat ditentukan distribusi dari suatu sampel secara visual. Histogram dapat digunakan untuk menggambarkan frekuensi, frekuensi kumulatif, persentase dan persentase kumulatif.

Untuk menampilkan histogram pada SPSS pilih menu Graphs → Legacy Dialog →Histogram, kemudian pilih variabel yang akan dibuat histogramnya, dan klik OK.

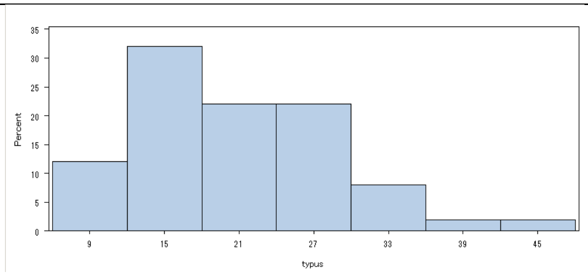


Gambar II.5 Histogram data waktu penyembuhan penyakit Typus. Karena histogram merupakan bentuk lain daripada dotplot yaitu sumbu horizontal berupa selang maka bentuk distribusi yang diperlihatkan mirip.

Untuk menampilkan histogram pada tambahkan perintah

```
procunivariatedata=work.typus;
histogram typus; /*perintah menampilkan histogram*/
run;
```

Maka Histogram akan muncul seperti gambar dibawah ini



4. Diagram batang-daun

Data statistik, yang dikumpulkan dalam jumlah amat banyak, akan sangat membantu dalam menelaah bentuk distribusi bila disajikan dalam bentuk gabungan tabel dan grafik yang dinamakan **diagram batang – daun (Stem-leaf)**.

Selain dapat memberikan gambaran bentuk distribusi data, diagram ini mampu mempertahankan tampilan data asli yang disusun terurut. Lebih jauh, tampilan diagram ini seperti bentuk histogram yang dirotasi 90° searah jarum jam.

Untuk menampilkan diagram batang daun :

Pilih menu Analyze → Descriptive Statistics → Explore. Kemudian masukkan variabel yang akan ditampilkan diagram batang daunnya pada Dependent List. Kemudian klik plots dan check Steam and Leaf, lalu klik Continue → OK.

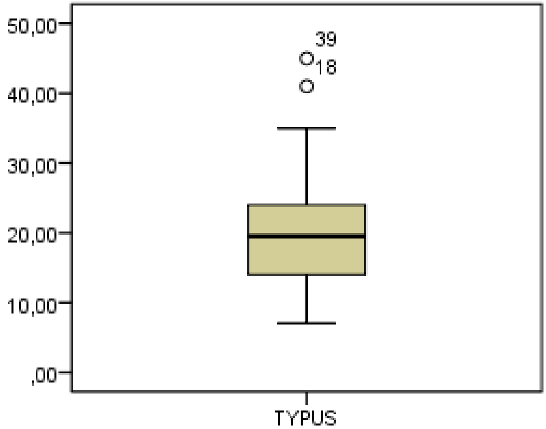
```
TYPUS Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2,00        0 .  78
    11,00        1 .  00112222334
    12,00        1 .  555666677889
    13,00        2 .  0011112244444
     6,00        2 .  678999
     3,00        3 .  123
     1,00        3 .  5
     2,00 Extremes    (>=41)

 Stem width:      10,00
 Each leaf:        1 case(s)
```

Gambar II.6Diagram batang dan daun data waktu penyembuhan penyakit Typus. Satuan untuk batang adalah puluhan (10,00) sedangkan untuk daun adalah satuan (1 case).

5. Box Plot

Box Plot merupakan salah satu grafik yang paling sering dipakai. Box plot dapat memperlihatkan data yang memiliki nilai ekstrim selain Anda dapat langsung melihat nilai kuartil 1,2 dan 3.

Untuk menampilkan boxplot, pilih menu Analyze → Descriptive Statistics → Explore, kemudian masukan variabel yang akan dubuat boxplotnya pada Dependent List, kemudian klikplots dan pada boxplot pilih Dependent together, lalu klik Continue → OK.

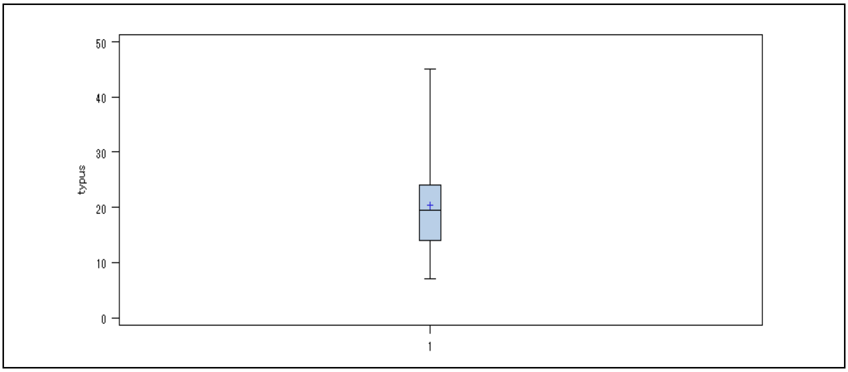


Gambar II.7Boxplot data waktu penyembuhan penyakit Typus. Terdapat dua datum yang terpisah dari yang lain, yang diduga merupakan pencilan atas.

Untuk menampilkan boxplot, ketik perintah

```
procboxplotdata=work.typus;
plot typus*data1;
run;
```

Maka Result Viewer akan menampilkan



Box plot dapat memperlihatkan data outliers atau pencilan. Ada dua jenis pencilan yaitu:

i. Data pencilan yaitu data yang berada 1.5 d_q dari kuartil atas atau 1.5 d_q dari kuartil bawah.
ii. Data pencilan jauh yaitu data yang berada 3 d_q dari kuartil atas atau 3 d_q dari kuartil bawah.

6. PP Plot dan QQ plot

Distribusi data dapat dikatakan normal atau tidak dapat dilihat dengan menggunakan grafik P – P Plots. Misalkan pada data tinggi siswa akan dicari tahu apakah distribusi datanya normal atau tidak. Analisis P-P dan Q-Q plot merupakan analisis plot grafik probabilitas secara umum yang digunakan untuk menetapkan apakah distribusi suatu variabel tertentu sesuai dengan variabel yang telah di tetapkan.

P-P plot menganalisis plot grafik antara variabel proporsi komulatif dengan proporsi setiap anggota / case- proporsi setiap anggota / case-nya. Q-Q plot menganalisis plot grafik antara variabel quantile (quantile merupakan nilai yang akan membagi case dalam jumlah tertentu yang besarnya sama pada setiap kelompoknya) dengan quantile setiap anggota / casenya.

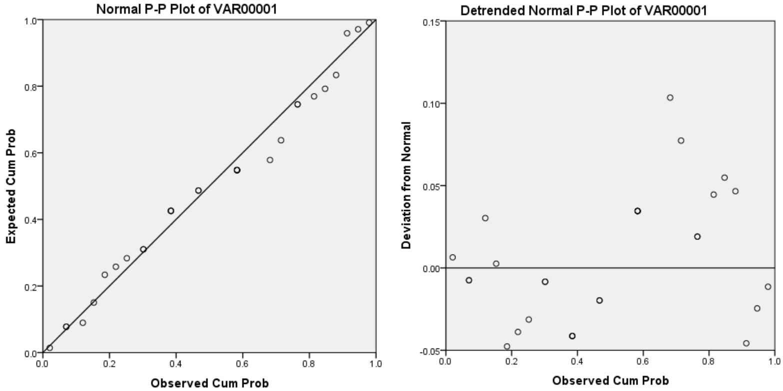
Contoh 1: Data berikut menyatakan tinggi badan kelas X dari 30 siswa:

165	178	137	168	163
190	167	159	167	158
147	148	152	167	176
147	156	163	196	188
159	175	174	174	167
165	170	167	157	163

Langkah – langkahnya adalah sebagai berikut :

1. Klik menu Analyze > Descriptive Statistics > P-P Plots
2. Masukkan variabel tinggi ke kotak Variables. Pastikan pilihanTest Distribution pada pilihan Normal. Lalu klik OK.

Normal P-P Plots. Grafik P-P plots dan Detrended Normal P-P Plots terlihat seperti di bawah ini:



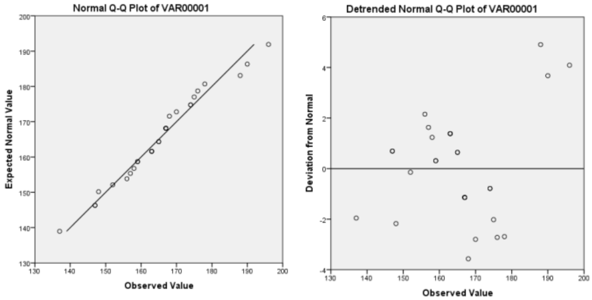
Gambar 1P-P plot dan Detrended Normal P-P plot

Garis diagonal dalam grafik ini menggambarkan keadaan ideal dari data yang mengikuti distribusi normal. Titik-titik di sekitar garis adalah keadaan data yang kita uji. Jika kebanyakan titik-titik berada sangat dekat dengan garis atau bahkan menempel pada garis, maka dapat kita simpulkan jika data kita mengikuti distribusi normal.

Jika dalam grafik terdapat titik yang berada sangat jauh dari garis. Ini adalah titik yang sama yang kita lihat dalam stem and leaf plots. Keberadaan titik ini menjadi peringatan bagi kita untuk berhati-hati melakukan analisis berikutnya.

Gambar 2 ini menggambarkan selisih antara titik-titik dengan garis diagonal pada grafik sebelumnya. Jika data yang kita miliki mengikuti distribusi normal dengan sempurna, maka semua titik akan jatuh pada garis 0,0. Semakin banyak titik-titik yang tersebar jauh dari garis ini menunjukkan bahwa data kita semakin tidak normal.

Dengan cara yang sama lakukan langkah di atas menggunakan pilihan menu Q-Q Plots . Pada Q-Q Plot terdapat 2 macam menu. Berikut adalah penjelasannya. Output 1 :



Gambar 2QQ plot dan Detrended Normal Q-Q Plot

E. TRANSFORMASI DATA

Membakukan bentuk distribusi adalah melakukan transformasi data sehingga hasilnya memiliki bentuk distribusi yang baku, yaitu simetris, berpuncak tunggal, menyempit ke kiri dan ke kanan. Cara melakukan transformasi tersebut adalah sebagai berikut :

1. Tentukan sari numeriknya (mean, kuartil bawah, median, kuartil atas). Hal ini bisa dilakukan dengan mudah, cukup dengan membuat steam-leaf dari data tersebut.
2. Buatlah box plot-nya berdasarkan sari numerik tersebut.
3. Analisis apakah distribusinya skewness positif, simetris, atau skewness negatif dengan menghitung derajat kemencengannya.
4. Gunakan tangga transformasi Tukey untuk memilih transformasi yang sesuai dengan distribusi pada no. 3.

Kenapa memakai tangga transformasi Tukey?

Ini dilakukan karena tangga transformasi Tukey adalah rumusan matematika yang sederhana dan mudah dioperasionalkan. Tangga transformasi tersebut adalah sebagai berikut :

$-\frac{1}{x^2}$	$-\frac{1}{x}$	$\log x$ $\sqrt{x}$	x x^2	x^3	10^x
Kuat	Sedang	Lemah	Lemah	Sedang	Kuat
skewness positif		Simetri		skewness negatif	

Transformasi yang dianjurkan untuk data dengan skewness lebih besar dari nol adalah $\sqrt{x}$, log(x), dan seterusnya. Sedangkan transformasi yang dianjurkan untuk data dengan skewness kecil dari nol adalah x^2, x^3, dan seterusnya.

Perlu diperhatikan bahwa apabila melakukan transformasi terhadap data maka telah terjadi perubahan satuan data. Sehingga dalam melakukan analisis terhadap hasil pengolahan data, perlu dilakukan penyesuaian terhadap interpretasinya.

F. SOAL LATIHAN

1. Buatlah sari numerik dari data Estates Production by Crops, Indonesia, 1990-2001 (Ton)untuk masing-masing variabel (mean, median, quartil, variansi, maksimum-minimum, kurtosis, dan skewness).
2. Buatlah histogram, boxplot dan diagram batang daun dari data Estates Production by Crops, Indonesia, 1990-2001 (Ton) untuk masing-masing variabel.
3. Data berikut menyatakan panjang dalam sentimeter dari 30 kecambah sejenis :

2,0	3,0	0,3	3,3	1,3	0,4
0,2	6,0	5,5	6,5	0,2	2,3
1,5	4,0	5,9	1,8	4,7	0,7
4,5	0,3	1,5	0,5	2,5	5,0
1,0	6,0	5,6	6,0	1,2	0,2

Buatlah sari numerik dari data tersebut di atas yang meliputi :

Rata-rata	Simpangan
Median	Data terbesar dan data terkecil
Modus	Quartil bawah, tengah dan atas
Variansi	Jumlah (sum)
Standar deviasi	

Buat semua grafik yang telah diajarkan dari data diatas!

4. Skor berikut menyatakan nilai ujian akhir suatu mata kuliah Biostatistika :

23	60	79	32	57	74	52	70	82	36
80	79	81	95	41	65	92	85	55	76
52	10	64	75	78	25	80	98	81	67
41	71	83	54	64	72	88	62	74	43
60	78	89	76	84	48	84	90	15	79
34	67	17	82	69	74	63	80	85	61

   a. Buat diagram batang-daun untuk nilai diatas dengan batang 1, 2, ..., 9, dan jelaskan bentuk distribusinya!
   b. Berapa rata-rata, variansi, dan simpangan baku dari data di atas?

5. Data berikut menyatakan berat badan 50 siswa kelas XII:

51	64	56	61	57	62	57	62	64	69
59	49	56	67	55	65	61	67	50	59
54	75	64	67	55	58	59	65	59	64
50	58	56	52	63	61	60	53	66	62
59	60	59	51	59	56	55	54	57	50

   a. BuatlahBlok Spot, diagram batang daun, PP Plot dan QQ Plot dari data di atas, berikan penjelasan mengenai normalitas data tersebut

6. Data berikut menyatakan besar keuntungan perusahaan ABC dalam 50 hari:

13000000	13000000	18000000	18000000	21000000
24000000	20000000	29000000	27000000	13000000
14000000	36000000	32000000	26000000	16000000
15000000	25000000	16000000	25000000	21000000
18000000	19000000	19000000	22000000	18000000
13000000	21000000	28000000	14000000	20000000
12000000	17000000	31000000	24000000	21000000
15000000	16000000	19000000	18000000	22000000

16000000	14000000	30000000	13000000	14000000
18000000	14000000	14000000	16000000	20000000

a. Buatlah PP Plot dan QQ Plot dari dua data di atas, berikan penjelasan mengenai normalitas data tersebut.

II. DISTRIBUSI PELUANG

A. TUJUAN

1. Memperkenalkan Distribusi Peluang Diskrit dan Kontinu
2. Menghitung nilai peluang untuk peubah acak yang berdistribusi Binomial dan Poisson.
3. Dapat menghitung peluang dari peubah acak yang berdistribusi Normal dan Eksponensial

B. PENDAHULUAN

Dalam kehidupan sehari-hari, manusia selalu dihadapkan pada bermacam-macam permasalahan yang harus dapat dicari penyelesaiannya. Untuk itu, diperlukan pengamatan untuk memperoleh informasi sehingga dari semua kemungkinan yang akan terjadi dapat diambil keputusan yang objektif. Keseluruhan pengamatan yang ingin diteliti, berhingga ataupun tidak, membentuk *populasi*. Jadi populasi terdiri atas keseluruhan pengamatan yang menjadi perhatian, dan *sampel* adalah suatu *himpunan bagian* dari populasi.

Percobaan statistika(eksperimen acak) merupakan suatu istilah untuk menunjukkan setiap proses yang menghasilkan pengamatan yang berkemungkinan. Dengan kata lain, *percobaan statistika* adalah percobaan yang hasilnya tidak dapat ditentukan di muka.

Suatu percobaan dikatakan acak apabila :

1. Percobaan dapat diamati atau diukur.
2. Hasil percobaan yang akan terjadi tidak dapat diper-kirakan sebelumnya karen adanya galat (*error)*.
3. Semua hasil yang mungkin dari suatu eksperimen dapat dirumuskan dalam suatu ruang sampel.
4. Proporsi keberhasilan dapat diketahui dari hasil – hasil sebelumnya.
5. Dalam kondisi yang sama, percobaan tersebut dapat diulang oleh pengamat ataupun orang lain.

Himpunan seluruh hasil yang mungkin muncul dalam percobaan statistika dinamakan *ruang sampel (S)*. Ruang sampel terbagi atas dua jenis, yaitu :

a. Ruang sampel diskrit : banyaknya elemen pada ruang sampel yang dapat dihitung (*countable*).

   Contoh : Ruang sampel dari percobaan pengecekan sepatu pada pabrik AAA yang dapat dikategorikan dengan sepatu yang cacat atau tidak.

b. Ruang sampel kontinu: elemen – elemen dari ruang sampel nerupakan bagian dari suatu interval.

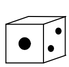
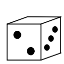
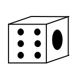
   Contoh : Ruang sampel dari percobaan pengukuran berat badan mahasiswa statistik IAIN Kediri. S= $\{x \mid 30 < x < 100\}$

Kejadian atau *event (E)* adalah himpunan bagian (subset) dari ruang sampel S. Kemungkinan terjadinya suatu kejadian sebagai hasil percobaan statistika yang dinilai dengan menggunkan kemungkinan sekumpulan bilangan real 0 sampai 1 disebut *peluang*. Jika suatu ruang sampel mempunyai $n(S)$ elemen dan suatu kejadian memiliki $n(E)$ elemen maka peluang kejadian E adalah

$$P(E) = \frac{n(E)}{n(S)}$$

Dalam suatu percobaan sering kali yang menjadi pusat perhatian kita bukan percobaan itu sendiri tapi akibat dari percobaan tersebut. Jadi rincian ruang sampel itu sendiri bukan menjadi perhatian tapi suatu pemetaan darinya. Yang menjadi perhatian utama seorang penjudi, misalnya, bukan bilangan berapa yang muncul dalam lantunan dadu tapi apakah dia menang atau kalah. Pemetaan ini yang dinamakan *peubah acak*. Jadi *peubah acak* ialah suatu *pemetaan* yang mengaitkan setiap unsur dalam ruang sampel *pada suatu bilangan riil*.

Contoh :

1. S = { , , ... ,  }

   $\boldsymbol{X}$= {1, 2, ... ,6}

b. S = banyaknya mahasiswa yang bolos kuliah statistik. (dengan catatan jumlah mahasiswa 60 orang)

   $\boldsymbol{X}$= {0,1,2, ... , 60}

Suatu percobaan yang diubah menjadi peubah acak memiliki beberapa keuntungan, yaitu:

- Merepresentasikan masalah ke dalam titik real
- Dapat dipetakan

- Lebih mudah dalam penulisan

Seperti ruang sampel, peubah acak juga terbagi atas 2 jenis :

i. Peubah Acak Diskrit : Himpunan terhitung $\{x_1, x_2, \dots,\}$ berhingga atau tak hingga

   Contoh : Banyak mata kuliah yang diambil semester ganjil oleh mahasiswa Matematika ditahun kedua

   X = 0, jika tidak mengambil satupun

   X = 1, jika mengambil 1 mata kuliah

   X = 2, jika mengambil 2 mata kuliah, dst

ii. Peubah Acak Kontinu

   Contoh : Nilai IP yang diperoleh mahasiswa Matematika di tahun kedua pada semester ganjil.

   X = (0,1], jika $0 < \text{IP} \leq 1$

   X = (1,2], jika $1 < \text{IP} \leq 2$

   X = (2,3], jika $2 < \text{IP} \leq 3$

   X = (3,4], jika $3 < \text{IP} \leq 4$

Pada peubah acak, peluang untuk suatu kejadian dibentuk menjadi suatu fungsi peluang dan fungsi kumulatif dari peluang tersebut direpresentasikan menjadi fungsi distribusi.

C. DISTRIBUSI PELUANG DISKRIT

Himpunan pasangan terurut (x, $f(x)$) merupakan suatu *distribusi peluang peubah acak diskrit* jika untuk setiap kemungkinan hasil x memenuhi :

1. $P(X = x) \geq 0$
2. $\sum_x P(X = x) = 1$

3. $P(X = x) = f(x)$

Distribusi kumulatif $F(x)$ suatu peubah acak diskrit X dengan fungsi kepadatan peluang $f(x)$ dinyatakan oleh:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \text{ untuk } -\infty \langle x \langle \infty$$

Distribusi kumulatif juga dikenal sebagai fungsi distribusi.

Beberapa jenis distribusi peluang diskrit diantaranya adalah *distribusi seragam diskrit, distribusi binomial, distribusi Poisson* dan *distribusi Hipergeometrik*. Namun yang akan dibahas pada modul ini hanyalah distribusi Binomial dan distribusi Poisson.

1. DISTRIBUSI BINOMIAL

Suatu percobaan sering kali terdiri dari beberapa usaha. Tiap usaha memiliki 2 kemungkinan hasil yaitu sukses atau gagal. Kita dapat menentukan atau memilih salah satu hasil sebagai sukses. Misalnya pada pengujian barang hasil produksi, dengan tiap pengujian atau usaha dapat menunjukkan apakah suatu barang cacat atau tidak cacat. Proses seperti ini disebut *proses Binomial*. Tiap usaha disebut *usaha Bernoulli*.

Percobaan binomial harus memenuhi persyaratan sebagai berikut :

1) Percobaan terdiri atas n usaha yang berulang.
2) Tiap usaha memberi hasil yang dapat dikelompokkan menjadi sukses atau gagal.
3) Peluang sukses, dinyatakan dengan p, tidak berubah dari usaha yang satu ke yang berikutnya.
4) Tiap usaha bebas dengan usaha lainnya.

Banyaknya *X* yang sukses dalam n usaha Bernoulli disebut *peubah acak binomial*. Distribusi peluang dari *X* disebut *distribusi binomial* dengan parameter *n* dan *p* (ditulis $\boldsymbol{X \sim B(n,p)}$), dengan *n* menyatakan banyaknya percobaan dan *p* menyatakan peluang sukses pada setiap percobaan.

Banyaknya cara menyusun *k* sukses dari *n* percobaan diberikan oleh koefisien binomial,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \text{ untuk k = 0,1, … , n}$$

Maka peluang sukses sebanyak *k* buah dari n percobaan adalah sebagai berikut :

$$B(k;n,p) = P\ (\ X = k\ ) = \binom{n}{k} p^k (1-p)^{n-k}$$

dengan mean dari *X* adalah *np* dan variansi dari *X* adalah *np(1-p)*.

Langkah-langkah menghitung nilai peluang suatu peubah acak yang berdistribusi Binomial melalui SPSS:

1. Pilih menu Transform→compute variable
2. Pada Target variable, ketikkan nama variabel (bebas) sebagai tempat keluaran hasil perhitungan.
3. Pilih fungsi cdf.binom (quant, n, p) untuk dimasukkan dalam Numeric Expression.
4. Ganti quant dengan nilai titik yang akan dihitung peluangnya (x), dengan n banyaknya percobaan, dan p peluang “sukses” atau “gagal”.
5. OK

2. DISTRIBUSI POISSON

Percobaan yang menghasilkan peubah acak *X*yang bernilai numerik, yaitu banyaknya hasil selama selang waktu tertentu atau dalam daerah tertentu, disebut *percobaan Poisson*.Panjang selang waktu tersebut dapat semenit, sehari, seminggu, sebulan, atau setahun.

Untuk selang waktu, banyaknya hubungan telepon per jam yang diterima suatu kantor, banyaknya hari di sekolah ditutup karena banjir, atau banyaknya pertandingan sepakbola yang diundurkan karena hujan selama musim hujan merupakan contoh percobaan poisson yang hasil pengamatannya dapat diubah menjadi peubah acak *X*. Sedangkan untuk daerah yang dimaksud dapat berupa sepotong garis, suatu luas daerah, sutau isi benda, atau pun barangkali sepotong benda. Dalam hal ini peubah acak *X* mungkin menyatakan banyaknya tikus sawah per hektar, banyaknya bakteri dalam suatu kultur, ataupun banyaknya salah tik per halaman.

Suatu *percobaan Poisson* mendapat namanya dari *proses Poisson* dan memiliki sifat berikut:

1) Banyaknya hasil yang terjadi dalam suatu selang waktu atau daerah tertentu tidak terpengaruh oleh (bebas dari) apa yang terjadi pada selang waktu atau daerah lain yang terpisah. Dalam hubungan ini proses Poisson dikatakan tidak punya ingatan.
2) Peluang terjadinya suatu hasil (tunggal) dalam selang waktu yang amat pendek atau dalam daerah yang kecil sebanding dengan panjang selang waktu atau besarnya

daerah dan tidak tergantung pada banyaknya hasil yang terjadi di lluar selang waktu atau daerah tersebut.

3) Peluang terjadinya lebih dari satu hasil dalam selang waktu yang pendek atau daerah yang sempit tersebut dapat diabaikan.

Banyaknya hasil X dalam suatu percobaan Poisson disebut suatu *peubah acak Poisson*. Distribusi dari peluangnya disebut *distribusi Poisson*. Suatu peubah acak berdistribusi Poisson ditulis $\boldsymbol{X \sim POI(\lambda)}$. Jika X berdistribusi Poisson yangmenyatakan banyaknya kejadian (sukses) yang terjadi dalam suatu selang waktu atau daerah tertentu dinyatakan dengan t, maka fungsi kepadatan peluangnya adalah sebagai berikut :

$$p(x, \beta t) = \frac{e^{-\beta t}(\beta t)^x}{x!}, \; x = 0,1,2, \ldots$$

dengan $\beta t = \lambda$ menyatakan rata-rata banyaknya kejadian (sukses) yang terjadi pada suatu interval waktu atau daerah tertentu. Rataan dan variansi dari suatu peubaacak yang berdistribusi Poisson adalah sama yakni λ .

Langkah-langkah menghitung nilai peluang suatu peubah acak yang berdistribusi Poisson melalui SPSS:

1. Pilih menu Transform→computevariable
2. Pada Target variabel, ketikkan nama variabel (bebas) sebagai tempat keluaran hasil perhitungan.
3. Pilih fungsi cdf.poisson (quant, mean) untuk dimasukkan dalam Numeric Expression.
4. Ganti quant dengan nilai titik yang akan dihitung peluangnya (x), mean dengan rataan (λ).

5. OK

3. HAMPIRAN DISTRIBUSI POISSON TERHADAP DISTRIBUSI BINOMIAL

Bila suatu eksperimen menghasilkan variabel berdistribusi Binomial dengan banyaknya pengamatan (n) cukup besar dan nilai peluang 'sukses' (p) mendekati nol, maka eksperimen tersebut dapat dihitung melalui distribusi Poisson dengan $\lambda = n\,p$.

D. DISTRIBUSI PELUANG KONTINU

Bila ruang sampel mengandung titik sampel yang tak hingga banyaknya sebanyak titik pada sepotong garis bilangan riil pada ruang sampel tersebut disebut *ruang sampel kontinu*. Fungsi $f(x)$ disebut fungsi kepadatan peluang peubah acak kontinu X, yang didefinisikan diatas himpunan semua bilangan riil R bila :

1. $f(x) \geq 0$ untuk semua $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a \leq x \leq) = \int_a^b f(x)dx$

Distribusi kumulatif $F(x)$ suatu peubah acak kontinu X dengan fungsi kepadatan peluang $f(x)$ diberikan oleh :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \text{ untuk } -\infty < x < \infty.$$

Pada distribusi peluang kontinu, ekspektasi dari X adalah

$$E[X] = \mu_x = \int_{-\infty}^{\infty} xf(x)dx$$

sedangkan variansi dari *X* adalah

$$Var\left(X\right)=\sigma^2=E\left[\left(X-\mu_x\right)^2\right]=\int_{-\infty}^{\infty}(X-\mu_x)^2 f(x)dx$$

Contoh data yang dikategorikan sebagai data kontinu misalnya berat badan, tinggi badan, volume air dan sebagainya. Beberapa distribusi kontinu yang terkenal adalah *Distribusi Normal, Distibusi Uniform, Distribusi Student (t), Distribusi Chi-Square, Distribusi Gamma, Distribusi Eksponensial,* dan *Distribusi Fisher.*

Berikut ini akan dibahas mengenai Distribusi Normal, Distribusi Uniform disertai sedikit informasi mengenai Distribusi Gamma dan Distribusi Eksponensial.

1. DISTRIBUSI NORMAL

Distribusi normal adalah distribusi peluang kontinu yang paling penting dalam statistika. Grafik dari distribusi normal disebut sebagai *kurva normal*. Kurva normal menggambarkan berbagai kumpulan data berdistribusi normal yang muncul di alam, industri, dan penelitian. Kurva ini merupakan kurva dari fungsi kepadatan peluang.

Suatu peubah acak berdistribusi normal ditulis $\boldsymbol{X \sim N(\mu, \sigma^2)}$.Distribusi ini bergantung pada dua parameter yaitu mean (μ) dan variansi populasi (σ^2). Fungsi kepadatan peluang (pdf) dari peubah acak X yang berdistribusi normal dengan mean μ dan variansi σ^2 adalah :

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$,-\infty< x<\infty,\ -\infty<\mu<\infty,\ \sigma^2\geq 0.$$

Sifat Kurva Normal :

1) Modus, titik pada sumbu datar yang memberikan nilai maksimum terdapat pada $x = \mu$.
2) Sumbu simetri kurva adalah garis $x = \mu$.
3) Kurva mempunyai titik belok pada $x = \mu \pm \sigma$, cekung ke bawah bila $\mu - \sigma < x < \mu + \sigma$ dan cekung ke atas untuk harga *x* lainnya.
4) Kedua ujung kurva mendekati asimtot sumbu datar bila harga *x* menuju $-\infty$ dan ∞.
5) Luas daerah yang terdapat diantara kurva normal dan sumbu datar adalah 1.

Langkah-langkah menghitung nilai peluang suatu peubah acak yang berdistribusi normal melalui SPSS:

1. Pilih menu Transform→compute
2. Pada Target variabel ketikkan nama variabel bebas
3. Pilih fungsi cdf.normal (q, mean, stddev) untuk dimasukkan dalam Numeric Expression.
4. Ganti quant dengan nilai titik yang akan dihitung peluangnya (x), mean dengan rataan (μ) dan stddev dengan standar deviasi (σ).
5. OK

2. DISTRIBUSI GAMMA DAN EKSPONENSIAL

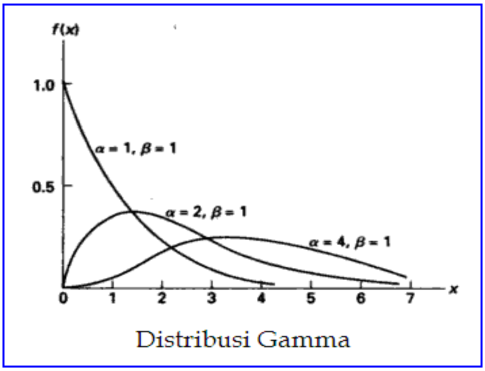
Misalkan X suatu peubah acak berdistribusi gamma dengan parameter $\alpha > 0$ dan $\beta > 0$. Maka fungsi kepadatan peluangnya sebagai berikut :

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}}, \qquad x > 0$$

Misalkan X suatu peubah acak berdistribusi gamma dengan parameter α=1 dan $\beta > 0$. Fungsi kepadatan peluangnya sebagai berikut :

$$f(x) = \frac{1}{\beta} e^{\frac{-x}{\beta}} \quad , x > 0$$

Fungsi kepadatan peluang di atas merupakan fungsi padat peluang dari distribusi eksponensial. Dari pernyataan di atas jelaslah bahwa distribusi Eksponensial merupakan hal khusus dari distribusi Gamma. Kedua distribusi di atas mempunyai terapan yang luas, misalnya keduanya memainkan peranan penting dalam teori antrian dan teori keandalan (reliabilitas). Jarak antara waktu tiba di fasilitas pelayanan dan lamanya waktu sampai rusaknya suku cadang dan alat listrik, sering menyangkut fungsi eksponensial.



Gambar III.1Fungsi peluang beberapa distribusi keluarga distribusi Gamma.

Fungsi kepadatan peluang dari distribusi eksponensial, juga dapat ditulis sebagai berikut:

$$f(x) = \lambda e^{-\lambda x} \text{ , } x > 0, \lambda > 0$$

dimana jika melihat eksponensial sebagai turunan distribusi Gamma, maka $\lambda = \frac{1}{\beta}$. Mean dan variansi dari distribusi eksponensial adalah:

$$\mu = \frac{1}{\lambda} \text{ dan } \sigma^2 = \frac{1}{\lambda^2}$$

Suatu variabel random X yang berdistribusi eksponensial dengan parameter λ dinotasikan dengan ***X* ~ exp (λ).**

Langkah-langkah menghitung nilai peluang suatu peubah acak yang berdistribusi Eksponensial melalui SPSS:

1. Pilih menu Transform→computevariable
2. Pada Target variabel, ketikkan nama variabel (bebas) sebagai tempat keluaran hasil perhitungan.
3. Pilih fungsi cdf.exp (quant, scale) untuk dimasukkan dalam Numeric Expression.
4. Ganti quant dengan nilai titik yang akan dihitung peluangnya (x) danscale dengan parameter λ.
5. OK

3. HAMPIRAN DISTRIBUSI NORMAL TERHADAP DISTRIBUSI BINOMIAL

Misalkan$X$ peubah acak berdistribusi binomial dengan rataan$\mu = np$dan variansi $\sigma^2 = np(1-p)$maka distribusi binomial dapat dihampiri dengan distribusi normal bila n menuju tak hingga. Apabila$n$ kecil tapi p cukup dekat pada nilai½ ,nilai hampiran masih cukup baik.

Sebagai contoh, apabila kita ingin menghitung peluang bahwa 30 orang dari 100 orang yang terkena penyakit kusta sembuh, dengan peluang seorang penderita sembuh adalah

0.2, maka kita dapat menggunakan hampiran distribusi normal.

Langkah-langkah melihat distribusi binomial dihampiri oleh distribusi normal.

1. Buka Microsoft excel
2. Pastikan menu statplus telah ada, jika belum ada *install* StatPlus dengan cara sebagai berikut:
   a. Klik ***Tools*** pada menu bar. Kemudian pilih ***Add-Ins***.
   b. Klik tombol ***Browse***.
   c. Cari folder Berk-Carey (atau folder lain yang telah digunakan untuk menyimpan file ***StatPlusV2*** kemudian klik folder tersebut.
   d. Buka folder ***StatPlusV2***.
   e. Klik ***StatPlusV2.xls*** kemudian klik ***OK***.
      StatPlus versi 2.0. sekarang telah terdapat dalam kotak dialog Add-Ins. Jika StatPlus versi 2.0 belum di tandai (◎), aktifkan dengan mengklik di kotak tulisan StatPlus versi 2.0.
   f. Klik ***OK***.
   g. Setelah mengklik tombol OK, kotak dialog add-ins akan tertutup dan menu baru, StatPlus, muncul di menu bar Excel.
3. Klik **Statplus** pada menu bar, kemudian pilih **create data→Random Namber**

   Maka di layar excel akan muncul ***dialog box***. Pada ***Type of distribution*** pilih binomial. Kita hanya melakukan penelitian tentang satu penyakit yaitu kusta maka ***Number Sampel to generate*** nya 1. Misalkan kita melakukan percobaannya sebanyak 150 kali maka ***size of each sample*** nya 150. masukan***p-value*** 0.2, dan terakhir karena setiap percobaan kita amati dari 100 orang maka ***number of trial*** nya 100.

4. Klik ***output*** pilih cell yang akan menjadi tempat keluarnya data.
5. Buat histogramnya dengan langkah sebagai berikut:
   a. Klik StatPlus > Single Variable Charts > Histogram.
   b. Klik Data Values > Use Range References > OK. (hilangkan tanda ceklist terlebih dahulu)
   c. Klik tombol Output > As a new chart sheet dan ketikkan Histogram pada kotak teks kemudian klik OK.
   d. Klik tombol OKuntuk membuat histogram.
6. Amati bentuk histogramnya

   Lakukan simulasi ini beberapa kali dengan jumlah penderita kusta yang diamati naik menjadi 200 dan 300. Selain itu coba untuk p = 0.4 tapi jumlah penderita kusta yang diamati hanya 10.

Apa kesimpulan dari simulasi tersebut?

Masalah lain:

Misalkan kita mengamati hasil ujian PG 10 soal, dengan peluang benar setiap soal 0.5, dari 5 orang anak. Maka Pada ***Type of distribution*** pilih binomial. Kita melakukan penelitian terhadap 10 soal maka ***Number Sampel to generate*** nya 10. Misalkan kita melakukan percobaannya sebanyak 1 kali maka ***size of each sample*** nya 1. masukan***p-value*** 0.5, dan terakhir karen percobaan kita amati dari 5orang anak maka ***number of trial*** nya 5. Hasil dari data bangkitan ini bukanlah sebuah sample acak *melainkan barisan sample acak* (setiap kolom menyatakan satu sample acak). Maka dalam kasus ini kita dapat bicara tentang ke-konvergen-an.

E. SOAL LATIHAN

1. Dalam pengujian sejenis ban truk yang melalui jalan yang kasar ditemukan bahwa 25% truk mengalami kegagalan karena ban pecah. Dari 15 truk yang diuji selanjutnya, carilah peluang :
   a. dari 3 sampai 7 truk mengalami kegagalan (ban pecah).
   b. minimal 5 truk mengalami kegagalan.
   c. kurang dari 8 truk berhasil dalam pengujian
2. Seorang pengawas lalu lintas melaporkan bahwa 70 % dari kendaraan yang melintasi suatu daerah (melalui pemeriksaan) adalah berasal dari Bandung.
   a. Jika ada 9 kendaraan mendatang yang melalui pemeriksaan tersebut. Berapakah peluang maksimal 4 kendaraan berasal dari luar Bandung ?
   b. Jika ada 15 kendaraan mendatang yang melalui pemeriksaan tersebut. Berapakah peluang 5 sampai 8 kendaraan berasal dari Bandung ?
3. Terdapat 40 siswa didalam suatu kelas, 28 orang diantaranya merupakan anak laki-laki.
   a. Jika dipilih 10 anak secara acak, berapakah peluang 4 diantaranya merupakan anak perempuan?
   b. Jika pada suatu waktu 4 anak laki-laki tidak dapat mengikuti kegiatan pembelajaran, dan dilakukan

pemilihan ketua kelas dengan 5 calon, berapa peluang 3 sampai 5 calon berjenis kelamin laki-laki?

4. Suatu daerah di bagian timur Amerika Serikat, rata-rata ditimpa 6 angin topan setahun.
   Carilah peluang di suatu tahun tertentu
   a. tidak sampai 4 angin topan yang akan menimpa daerah tersebut;
   b. antara 6 sampai 8 angin topan akan menimpa daerah tersebut.
5. Suatu perusahaan kecil di Indonesia, telah memproduksi sebanyak 250 pasang sepatu dalam satu hari. Karena keterbatasan teknologi yang dipakai, ternyata rata-rata ditemukan 50 pasang sepatu yang cacat dalam satu produksi per hari. Jika pada suatu hari, 65 konsumen membeli sepatu di tempat tersebut (asumsi: setiap konsumen membeli 1 pasang sepatu).
   a. Berapa peluang 12 pasang sepatu cacat?
   b. Berapa kemungkinan 27 pasang sepatu yang tidak cacat?
6. Suatu percobaan menghitung banyaknya partikel–α yang luruh dalam selang 1 detik dari 1 gram radioaktif. Dari pengalaman sebelumnya diketahui bahwa rata-rata sebanyak 3,2 partikel– α yang akan luruh per detik. Berapakah taksiran peluangnya bahwa tidak lebih dari 2 partikel– α yang akan luruh ?
7. Diketahui bahwa 40% dari tikus yang disuntik dengan sejenis serum terlindung dari serangan sejenis penyakit. Bila 5 tikus disuntik, berapakah peluang bahwa:
   a. tidak ada yag terserang penyakit tersebut;

   b. kurang dari 2 yang terserang;
   c. lebih dari 3 yang terserang.

8. Peluang seseorang sembuh dari operasi jantung yang rumit adalah 0,9. Berapakah peluang tepat 5 dari 7 orang yang menjalani operasi ini akan sembuh?
   Hitung peluangnya dengan :
   a. SPSS
   b.EXCEL
   c. Tabel statistik
   d. Manual
9. Peluang seseorang meninggal karena suatu infeksi pernafasan adalah 0,002. Carilah peluang bila 2000 orang yang terserang infeksi tersebut, kurang dari 5 orang yang akan meninggal. Lakukan perhitungan peluang dengan distribusi Binomial dan Poisson! Bandingkan hasilnya!
10. Di suatu simpang jalan rata-rata terjadi 3 kecelakaan seminggu. Berapakah peluang pada suatu minggu tertentu
    a. tepat 5 kecelakaan akan terjadi?
    b. Kurang dari 3 kecelakaan akan terjadi?
    c. Paling sedikit 2 kecelakaan akan terjadi?
11. Misalkan nilai akhir mata kuliah statistika dasar mahasiswa departemen Tambang berdistribusi normal dengan rata-rata 59 dan standar deviasi 34. Jika standar nilai yang digunakan adalah sebagai berikut :
    A : Nilai Akhir > 85
    B : 70 < Nilai Akhir ≤ 85
    C : 55 < Nilai Akhir ≤ 70
    D : 40 < Nilai Akhir ≤ 55

E : Nilai Akhir ≤ 40

Pertanyaan :

a. Berapa peluang seorang mahasiswa mendapatkan nilai akhir B ?

b. Berapa peluang seorang mahasiswa lulus mata kuliah statistika dasar? (mahasiswa lulus jika nilainya A, B, atau C)

12. Diameter sebelah dalam suatu cincin torak berdistribusi normal dengan rataan 10 cm dan simpangan baku 0,03 cm.

a. Berapa proporsi cincin yang mempunyai diameter dalam melebihi 10,075 cm?

b. Berapa peluang suatu cincin torak berdiameter dalam antara 9.97 dan 10.03 cm?

c.Di bawah nilai diameter dalam berapakah terdapat 15% dari seluruh cincin torak?

13. Diketahui suatu distribusi normal baku, carilah nilai k sehingga

a.$P(Z < k) = 0.0427$

b.$P(Z > k) = 0.2946$

c.$P(-0.93 < Z < k) = 0.7235$

14.Pada tempat pemberhentian bus, jadwal kedatangan bus tercatat setiap 15 menit terhitung dari pukul 07.00. Bila kita amati sampai pukul 07.30 dan diasumsikan waktu kedatangan penumpang berdistribusi Uniform. Tentukan peluang :

a.Seseorang menuggu kurang dari 5 menit

b.Seseorang menunggu lebih dari 10 menit

15.Misalkan suatu sistem mengandung sejenis komponen yang daya tahannya dalam tahun dinyatakan oleh peubah acak T yang berdistribusi eksponensial dengan parameter waktu rata-rata sampai komponen tersebut rusak, $\frac{1}{\lambda} = 5$. Bila sebanyak 5 komponen tersebut dipasang dalam sistem yang berlainan, berapakah peluang paling sedikit 2 komponen masih akan berfungsi pada akhir tahun kedelapan ?

16.Misalkan waktu (dalam satuan jam) yang diperlukan untuk memperbaiki pompa panas berbentuk peubah acak X yang berdistribusi gamma dengan $\alpha = 2$ dan $\beta = 0.5$. Berapa peluangnya bahwa perbaikan berikutnya akan memerlukan waktu

a.paling lama 1 jam ?

b.paling sedikit 2 jam?

17.Misalkan lama pembicaraan telepon dapat dimodelkan oleh distribusi eksponensial, dengan rataan 10menit/ orang. Bila seseorang tiba-tiba mendahului anda di suatu telepon umum, carilah peluangnya bahwa anda harus menunggu:

a.lebih dari 10menit

b.antara 10 sampai 20 menit

III. INFERENSI STATISTIKA : PENAKSIRAN

A. TUJUAN

1. Mempelajari bagaimana melakukan penaksiran titik dan penaksiran selang untuk mean dan variansi.
2. Melakukan penaksiran selang untuk mean dan variansi terhadap beberapa contoh permasalahan.

B. PENDAHULUAN

Pengambilan suatu keputusan menjadi permasalahan yang harus dihadapi dengan tepat.Analisis terhadap masalah menjadi suatu peranan penting yang harus dipertimbangkan. Apabila langkah pertama yang diambil salah maka keputusan akan menjadi "salah", sebab langkahini merupakan tahap awal yang sangat menentukan langkah-langkah selanjutnya.

Jika tahap awal sudah benar maka tahap selanjutnya dalam statistik yakni*hipotesis statistik*. Suatu pernyataan/ anggapan, yang mungkin benar atau tidak, mengenai suatu masalah disebut sebagai hipotesis statistik. Kebenaran suatu hipotesis tidak diketahui dengan pasti kecuali seluruh populasi diamati, haltersebutmustahil terjadi karena kendala faktor biaya dan waktu. Kendala tersebut dapat diatasi dengan cara mengambil sampel acak dari suatu populasi untuk mewakili populasi tersebut.

Akibat dari permasalahan ini muncul *Statistika Inferensi*; yang mempelajari penaksiran dan pengujian hipotesis. Kedua bagian ini akan dibahas secara terpisah, teori penaksiran dan terapannya akan dibahas pada modul ini sedangkan teori pengujian hipotesis pada modul selanjutnya.

C. PENAKSIRAN

Sifat atau ciri penaksir yang baik adalah bias, variasi minimum, konsisten, dan cukup statistik.

1. Tak Bias

Penaksir tak bias bagi suatu parameter misalkan θ adalah jika $E(\theta^*) = \theta$, sedangkan penaksir bias bagi parameter θ jika $E(\theta^*) \neq \theta$. Namun, penaksir bias dapat diubah menjadi penaksir tak bias jika ruas kanan dikalikan atau ditambahkan dengan konstanta tertentu.

2. Variansi Minimum

Apabila terdapat dua buah penaksir tak bias, maka kedua penksir tersebut akan dibandingkan dalam hal variansi. Penaksir yang memiliki variansi terkecil dikatakan penaksir tak bias bervariansi minimum.

3. Konsisten

Jika θ_n^* adalah penaksir untuk θ yang didasarkan pada sampel acak berukuran *n*, maka θ_n^* dikatakan konsisten bagi parameter θ, apabila

$$\lim_{x \to \infty} p\left(\left|\theta_n^* - \theta\right| < \varepsilon\right) = 1$$

4.Sufficiency Statistics

Statistik $Y = Y(x_1, x_2, \ldots, x_n)$ dikatakan cukup bagi parameter, jika fungsi kepadatan peluang bersyarat : $P(x_1, x_2, \ldots, x_n) = x_n \mid Y(x_1, x_2, \ldots, x_n) = t$ tidak bergantung pada θ.

Penaksiran nilai suatu parameter (misalnya rataan, variansi, atau proporsi) dapat dilakukan dengan dua cara yaitu:

1.Penaksiran titik (*Point Estimation*)

Penaksiran titik digunakan apabila akan dicari nilai tunggal dari suatu parameter melalui pendekatan metode tertentu.

Misal : IP mahasiswa IAIN Kediri tahun pertama pada semester ganjil kira-kira 3.5, seekor tikus melahirkan anak kira-kira 3 ekor, sebatang rokok merek tertentu mempunyai kadar nikotin kira-kira 0,325 mg, rataan $\bar{x}$ merupakan salah satu nilai dari penaksir rataan populasi $\bar{X}$, dan sebagainya.

2. Penaksiran selang (*Interval Estimation*)

Penaksiran selang digunakan untuk mencari nilai sesungguhnya dari suatu parameter, dimana semua nilai yang mungkin dari parameter tersebut berada pada kisaran selang tertentu.

Misal : nilai kalkulus mahasiswa IAIN Kediri tahun pertama antara 53 sampai 55, bola lampu 5 watt berumur rata-rata antara 780 jam sampai 785 jam, tinggi badan mahasiswa IAIN Kediri tahun pertama antara 145 cm sampai 182 cm, satu kaleng minuman bersoda berisi rata-rata antara 2,25 desiliter sampai 2,50 desiliter, dan sebagainya.

Walaupun hasil penaksiran selang tidak memberikan suatu nilai tunggal dari nilai parameter sesungguhnya, tetapi mengenai keberadaan nilai tersebut dapat diperkirakan.

Penaksiran selang kepercayaan untuk rataan

Terdapat dua kasus untuk mencari nilai kepercayaan $(1-\alpha)$ untuk rataan satu populasi, yakni :

1. **Kasus variansi populasi diketahui**

Selang kepercayaan kepercayaan $(1-\alpha)$ untuk rataan populasi adalah

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

dengan $\bar{x}$ menyatakan rataan sampel berukuran n dan $z_{1-\frac{\alpha}{2}}$ menyatakan nilai tabel normal bakuyang diperoleh dengan memanfaatkan Teorema Limit Pusat, yaitu

$$P\left(Z \le z_{1-\frac{\alpha}{2}} \right) = 1 - \frac{\alpha}{2}.$$

2. **Kasus variansi populasi tidak diketahui**

Untuk kasus variansi populasi tidak diketahui maka variansi populasi tersebut ditaksir oleh S^2, sehingga selang kepercayaan $(1-\alpha)$ untuk rataan populasi adalah

$$\bar{x} - t_{\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}}$$

dengan $\bar{x}$ menyatakan rataan sampel berukuran n dan $t_{\frac{\alpha}{2},(n-1)}$ menyatakan nilai tabel t (*student-t*)dengan derajat kebebasan $(n-1)$ yang menggunakan distribusi sampel dari peubah acak T, yaitu $P\left(T > t_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ (persamaan hanya diganti z oleh t dan σ oleh S).

Selanjutnya, ada empat kondisi yang harus dicermati untuk mencari rumus selang kepercayaan untuk dua populasi.

1. **Kasus variansi dari populasi 1 dan populasi 2 diketahui**

   Selang kepercayaan $(1-\alpha)$ untuk selisih rataan kedua populasi adalah

$$(\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

   dengan n_1 dan n_2 masing-masing menyatakan banyaknya observasi populasi 1 dan populasi 2.

2. **Kasus variansi dari populasi 1 dan populasi 2 tidak diketahui tetapi dianggap sama** $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$

   σ^2 ditaksir dengan S_p^2 yang adalah:

$$\hat{\sigma}^2 = S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

   Maka, selang kepercayaan $(1-\alpha)$ untuk selisih rataan kedua populasi adalah

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2},(n_1+n_2-2)}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2},(n_1+n_2-2)}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

3. **Kasus variansi dari populasi 1 dan populasi 2 tidak diketahui dan dianggap tidak sama** $(\sigma_1^2 \neq \sigma_2^2)$.

   Dengan kata lain σ_1^2 dan σ_2^2 ditaksir dengan S_1^2 dan S_2^2.

   Selang kepercayaan $(1-\alpha)$ untuk selisih rataan kedua populasi adalah

$$\left(\bar{x}_1-\bar{x}_2\right)-t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\left(\mu_1-\mu_2\right)<\left(\bar{x}_1-\bar{x}_2\right)+t_{\frac{\alpha}{2},v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

dengan v menyatakan derajat kebebasan yang diperoleh dari rumus:

$$v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{1}{(n_1-1)}\left(\frac{S_1^2}{n_1}\right)^2+\frac{1}{(n_2-1)}\left(\frac{S_2^2}{n_2}\right)^2}$$

Jika pada soal tidak dituliskan variansi kedua populasi, maka variansi kedua populasi dianggap tidak diketahui dan <u>tidak sama</u> besarnya.

Untuk mencari selang kepercayaan rataan dua populasi, hitung secara manual menggunakan persamaan-persamaan pada Ms. Excel atau

4. Data berpasangan

Seringkali peneliti ingin melihat efek sebelum dan setelah perlakuan, misal:

a. efek pemberian obat penurun berat badan sebelum dan setelah minum,
b. pengujian metode pembelajaran terhadap tingkat penguasaan materi ajar pada mahasiswa dinilai dari banyak PR yang dikerjakan dan nilai ujian,
c. lama waktu untuk membuka pegangan pintu yang diputar ke kanan dan diputar ke kiri (dalam detik),
d. penjualan beberapa stand mi instan dengan melakukan *promotion mix* dan penjualan biasa, dst.,

sehingga populasi yang diukur tetap sama hanya berbeda perlakuan. Data seperti itu disebut *data berpasangan*. Ciri dari pengamatan data berpasangan adalah:

- setiap anggota sampel populasi memiliki setiap perlakuan yang diamati
- setiap data dari suatu perlakuaan berpasangan satu-satu dengan data dari perlakuaan yang lain (setiap data berat badan sebelum minum obat penurun badan memiliki pasangan data berat badan sesudah minum obat penurun)

Selang kepercayaannya $(1-\alpha)$ merupakan selisih dari data sebelum perlakuan dan data setelah perlakuan (misal D = $X_{setelah} - X_{sebelum}$ atau D = $X_{sebelum} - X_{setelah}$) menyerupai selang kepercayaan untuk kasus satu populasi dengan variansi tidak diketahui.

Selang kepercayaan $(1-\alpha)$ untuk data berpasangan adalah

$$\bar{d} - t_{\frac{\alpha}{2},(n-1)} \frac{S_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\frac{\alpha}{2},(n-1)} \frac{S_d}{\sqrt{n}}$$

dengan $\bar{d}$ dan S_d adalah rataan dan simpangan baku dari selisih n pasangan dan $t_{\frac{\alpha}{2},(n-1)}$ menyatakan nilai tabel *t* (*student-t*)dengan derajat kebebasan $(n-1)$.

Langkah-langkah dalam SPSS

Untuk satu populasi

1. Pilih menu Analyze → compare means → one sample T test.

2. Masukkan variabel yang akan dicari selang kepercayaannya pada test variable(s).
3. Klik option, kemudian pada confidence interval, tuliskan nilai kepercayaan yang akan dicari selangnya, lalu klik continu.
4. Klik *OK*.

Untuk dua populasi (data tidak berpasangan)

1. Masukkan data untuk sampel acak 1 dan sampel acak 2 dalam satu kolom, kemudian pada kolom berikutnya tuliskan 1 untuk semua data sampel 1 dan 2 untuk semua data sampel 2 dan beri nama indeks.
2. Pilih menu Analyze→compare means→independent sample T-test.
3. Masukkan variabel yang akan dicari selang kepercayaannya pada test variable(s).
4. Masukkan variabel index pada grouping variable, klik define groups, masukkan 1 pada group 1 dan masukkan 2 pada group 2, klik continue.
5. Klik option, kemudian pada confidence interval, tulis nilai kepercayaan yang akan dicari selangnya, lalu continue. Dilanjutkan denganmemilih*OK*.

Untuk dua populasi (data berpasangan)

1. Pilih menu Analyze→compare means→paired-sample T test.
2. Masukkan variabel yang akan dicari selang kepercayaannya secara berpasangan pada paired variable.

3. Klik option, kemudian pada confidence interval, tulis nilai kepercayaan yang akan dicari selangnya, lalu continue.
4. klik *OK.*

Contoh Soal

Enam belas botol yang mirip masing-masing berisi cairan asam sulfat sebanyak
9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6, 11.20, 10.30, 11.60, 9.40, 9.20, 9.60, 10.60, 9.00 dan 9.20 liter.
Carilah selang kepercayaan 95% untuk rataan isi botol semacam itu, bila distribusinya dianggap hampir normal.
dengan menggunakan langkah tersebut di atas diperoleh hasil sebagai berikut :

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
ISIBOTOL	55.758	15	.000	10.006	9.624	10.389

selang kepercayaan untuk rataan isi botol adalah $9.624 < \mu < 10.389$ atau sering ditulis
$\mu = (9.6237 , 10.389)$.

Penaksiran selang kepercayaan untuk variansi

1. Variansi satu populasi

Selang kepercayaan $(1-\alpha)$ untuk variansi satu populasi adalah

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},(n-1)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}}$$

dengan S^2 adalah variansi sampel, n, adalah ukuran sampel, $\chi^2_{\frac{\alpha}{2},(n-1)}$ dan $\chi^2_{1-\frac{\alpha}{2},(n-1)}$ adalah nilai dari tabel **distribusi khi-kuadrat** dengan derajat kebebasan $(n-1)$.Cara penghitungan selang kepercayaan variansi dapat dilakukan secara manual menggunakan persamaan di atas pada Ms. Excel atau SPSS.

2. Nisbah variansi dua populasi

Selang kepercayaan$(1-\alpha)$untuk nisbah variansi dua populasi adalah

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{\frac{\alpha}{2},(v_1,v_2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2},(v_2,v_1)}$$

Dengan S_1^2 adalah variansi sampel dari populasi pertama, S_2^2 adalah variansi sampel dari populasi kedua, $f_{\frac{\alpha}{2},(v_1,v_2)}$ dan $f_{\frac{\alpha}{2},(v_2,v_1)}$adalah nilai dari tabel **distribusi Fisher** dengan derajat kebebasan $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$. Cara penghitungan selang kepercayaan untuk nisbah dua variansi dapat dilakukan secara manual menggunakan persamaan di atas pada Ms. Excel atau SPSS.

Langkah-langkah dalam SAS

1. Ketik data dalam excel dansimpan.
2. Kemudianketik syntax berikutpada editor.

```
procmeansdata=work.Kadarnmeanstderrclm;

varkadar;
```

Output SAS

testingkadar 12:15 Tuesday, July 9, 2013 21

The UNIVARIATE Procedure

Variable: Kadar (Kadar)

Extreme Observations

----Lowest---- ----Highest---

Value Obs Value Obs

9.0 15 10.3 9

9.2 16 10.4 3

9.2 12 10.6 14

9.4 11 11.2 8

9.6 13 11.6 10

95% confidence interval for Kadar

The MEANS Procedure

Analysis Variable : Kadar Kadar				
N	**Mean**	**Std Error**	**Lower 95% CL for Mean**	**Upper 95% CL for Mean**
16	10.0062500	0.1794595	9.6237410	10.3887590

D. SOAL LATIHAN

1. Pengukuran berikut memberikan waktu mengering, dalam jam, sejenis cat lateks merek tertentu.

3,4	2,5	4,8	2,9	3,6
2,8	3,3	5,6	3,7	2,8
4,4	4,0	5,2	3,0	4,8

Jika dimisalkan pengukuran menyatakan sampel acak yang diambil dari populasi normal, hitunglah selang kepercayaan 95% rata – rata hasil pengukuran (waktu mengering) sejenis cat lateks tertentu!

2. Suatu riset dilakukan untuk membuktikan bahwasuhu pisau mempengaruhi kekuatan luka pada pemotongan dalam pembedahan. Delapan anjing dipilih dan pembedahan dilakukan dengan pisau panas dan dingin pada setiap anjing. Berikut hasil pengukururan kekuatannya pada setiap anjing:

Anjing	1	2	3	4	5	6	7	8
Panas	5.120	10.000	10.000	10.000	10.000	7.900	510	1.020
Dingin	8.200	8.600	9.200	6.200	10.000	5.200	885	460

Uji apakah rata-rata perbedaan antara kekuatan pemotongan panas dan dingin lebih dari 700, gunakan taraf keberartian 5%!

3. Sembilan sampel zat yang mengandung besi digunakan untuk menentukan adanya perbedaan kandungan besi (dalam ppm) antara analisis dengan X-ray dan Kimia. Dari penelitian ini diperoleh data sebagai berikut :

AnalisisBesi	**Sampel (ppm)**								
	1	2	3	4	5	6	7	8	9
X – ray	2.2	1.9	2.5	2.3	2.4	1.8	2	2.1	2.3
Kimia	2	2.2	2.7	2	2.4	2.1	2.3	1.9	2.1

a. Berapa selang kepercayaan 95 % untuk rataan pada masing-masing analisis ?

b. Berapa selang kepercayaan 95 % untuk selisih rataan kedua analisis tersebut ?

4. Dalam penelitian untuk menentukan apakah suatu serum baru dapat memperlambat leukemia digunakan 9 tikus yang telah terkena penyakit tersebut pada tahap lanjut. Lima tikus diberikan serum sedangkan sisanya tidak. Dari penelitian ini diperoleh data sebagai berikut :

Perlakuan	**Umur** **1 2 3 4 5**
Diberi Serum	2,1 5,3 1,4 4,6 0,9
Tidak Diberi Serum	1,9 0,5 2,8 3,1

a. tentukan mean lamanya umur tikus dari masing-masing perlakuan!

b. tentukan selang kepercayaan 95% untuk selisih rataan umur tikus dari kedua perlakuan!

c. tentukan selang kepercayaan 95% untuk nisbah variansi umur tikus dari kedua perlakuan!

5. Seorang guru ingin menguji penguasaan siswa terhadap suatu bahan yang diajarkannya dengan menguji siswa pada awal dan akhir pertemuan. Pengujiandilakukan terhadap 15 orang siswadan diperoleh nilai tes sebagai berikut :

Siswa	tesawal	tesakhir
1	66,7	53,3
2	66,7	73,3
3	26,7	60
4	33,3	80
5	26,7	80

6	66,7	93
7	53,3	73,3
8	40	80
9	66,7	86
10	60	73,3
11	66,7	66,7
12	66,7	66,7
13	60	66,7
14	26,7	86
15	33,3	86

Tentukan selang kepercayaan 95% untuk rataan selisih nilai kedua tes tersebut!

6. Suatu mesin oli mobil diatur sedemikian rupa sehingga volume oli yang dikeluarkannya berdistribusi hampir normal. Suatu sampel acak diambil dan hasilnya adalah sebagai berikut: (dalam desiliter)

2,1 2,2 2,4 2,2 2,0 2,1 2.3 2,0 2,2

Tulis selang kepercayaan 95% masing-masing untuk rataan dan variansi volume oli!

7. Diketahui data produksi volumen minyak bumi beberapa sumur di Jatibarang selama bulan Desember 1990 (satuan: m^3)adalah sebagai berikut:

150 203 259 269 137 135 140 144 161 136 86

Diasumsikan bahwa volume produksi minyak bumi berdistribusi hampir normal.

a. Tentukan selang kepercayaan 98% untuk rataan data produksi minyak bumi!

b. Tentukan selang kepercayaan 95% untuk variansi data produksi minyak bumi!

Apakah anda sependapat jika seorang ahli mengatakan bahwa rata-rata volume produksi minyak bumi di reservoir Jatibarang kurang dari 165 m^3. (gunakan tingkat keberartian , $\alpha = 5\%$)?

8. Tigabelas merek rokok yang dipilih secara acak dan kadar nikotin yang diukur dalam mg adalah sebagai berikut:

7,3	8,6	10,4	9,5	16,5	12,2	11,5
8,1	11,5	9,3	9,2	8,5	10,5	

Tentukan selang kepercayaan 95% untuk rataan dan variansi nikotin dalam rokok!

9. Suatu percobaan untuk menentukan apakah atmosfer yang bercampur CO mempengaruhi kemampuan bernafas dilakukan terhadap 9 orang. Petugas dari Jurusan Pendidikan Jasmani *Virginia Polytechnic Int.* dan *State Univ.*memasukkan setiap peserta berturut-turut ke dua ruang pernafasan yang salah satunya mengandung kadar COtinggi dan mengukur pernafasan masing-masing peserta di setiap ruang dengan urutan acak. Data berikut adalah banyaknya pernafasan per menit:

Tanpa CO	32	25	27	37	30	51	43	42	32
Dengan CO	32	53	27	36	47	28	48	34	32

a. Menurut Anda apakah dua kelompok tersebut berasal dari dua kelompok yang berpasangan atau bukan?

b. Tentukan selang kepercayaan 98% untuk rataan dan variansi dari <u>beda</u> banyaknya pernafasan dengan CO dan tanpa CO!

Catatan: Distribusi banyaknya pernafasan permenit diasumsikan hampir normal.

IV. INFERENSI STATISTIKA UJI HIPOTESIS

A. TUJUAN

1. Mempelajari pengujian hipotesis terhadap distribusi normal.
2. Melakukan uji hipotesis terhadap beberapa contoh permasalahan.

B. PENDAHULUAN

Pengujian hipotesis digunakan sebagai alat pengambilan keputusan dari suatu pernyataan yang telah dirumuskan sebelumnya (hipotesis) berdasarkan suatu galat tertentu (biasanya disebut tingkat keberartian (α)).

Hipotesis statistik ialah suatu pernyataan/anggapan, yang mungkin benar atau tidak, mengenai suatu masalah (dalam hal ini menyangkut populasi).Pernyataan yang telah dirumuskan (hipotesisstatistik) selanjutnya akan dikelom-pokkan menjadi *hipotesisnol* (H_0) dan *hipotesis tandingan* (H_1).

Secara matematika, yang dimaksud *hipotesisnol* adalah pernyataan-pernyataan yang biasanya mengandung makna kesamaan (tidak lebih dari ($\leq$), tidak kurang dari ($\geq$) atau sama dengan (=)). Sedangkan *hipotesis tandingan* adalah semua pernyataan yang merupakan kebalikan dari hipotesis nol.

Secara praktisnya,

- H_0 adalah pernyataan yang selalu ditulis sederhana (arti kalimat tunggal), tidak mengandung lebih kecil (<), tidak lebih besar (>) atau tidak sama dengan ($\neq$), tetapi sama dengan (=).
- H_0 adalah lawan dari yang ingin kita buktikan.
- H_0 berdasarkan penelitian sebelumnya.
- H_0**tidak boleh** dibangun berdasarkan data sampel. Jika H_0 dibuat berdasarkan data sampel berarti mengintip data. Berbeda halnya dengan selang kepercayaan yang memanfaatkan data sampel untuk menaksir populasi.
- Karena bekerja di taraf kesalahan tipe 1 (α), maka H_0 selalu sama dengan (=).

Penolakan terhadap suatu hipotesis (H_0 atau H_1) tidak dapat disimpulkan bahwa hipotesis yang telah dirumuskan tersebut salah (tidak benar), tetapi hal itu menunjukkan bahwa data atau sampel yang terambil tidak (atau belum) mendukung hipotesis yang telah dirumuskan.

Langkah-langkah pengujian hipotesis adalah sebagai berikut :

1. Tentukan hipotesis nol dan tandingannya (H_0 dan H_1).
2. Tetapkan nilai α, kemudian carilah daerah dan titik kritis.

Daerah kritis adalah daerah penolakan H_0 dan titik kritis adalah titik batas dimana suatu hipotesis H_0 akan diterima atau ditolak.

3. Hitung nilai statistik uji.
4. Uji nilai statistik uji dengan titik kritis. Perhatikan jatuh di daerah mana.
5. Kesimpulan (H_0 ditolak atau tidak ditolak).
6. Tuliskan kalimat non matematika.

Dengan demikian, tampak sekali bahwa pengambilan keputusan sangat dipengaruhi oleh data atau sampel yang terambil dan nilai α . Pengambilan sampel (teknik sampling) yang baik dan sesuai dengan kaidah statistik akan memberikan informasi data yang baik. Akibatnya kesimpulan yang diperoleh dari data yang baik akan lebih objektif. Catatan : Jika hipotesis nol tidak jatuh pada daerah kritis tidak berarti H_0tidak ditolak.

C. UJI HIPOTESIS: RATAAN DAN VARIANSI

Uji Hipotesis Rataan Populasi

Bentuk hipotesis nol dan tandingannya untuk kasus **rataan satu populasi** adalah

1. $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0 \longrightarrow dwi\ arah$
2. $H_0 : \mu \geq \mu_0 \quad vs \quad H_1 : \mu < \mu_0 \longrightarrow eka\ arah$
3. $H_0 : \mu \leq \mu_0 \quad vs \quad H_1 : \mu > \mu_0 \longrightarrow eka\ arah$

Bentuk hipotesis nol dan tandingannya untuk kasus **rataan dua populasi** adalah

1. $H_0 : \mu_1 - \mu_2 = \mu_0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq \mu_0 \longrightarrow dwi\ arah$
2. $H_0 : \mu_1 - \mu_2 \geq \mu_0 \quad vs \quad H_1 : \mu_1 - \mu_2 < \mu_0 \longrightarrow eka\ arah$
3. $H_0 : \mu_1 - \mu_2 \leq \mu_0 \quad vs \quad H_1 : \mu_1 - \mu_2 > \mu_0 \longrightarrow eka\ arah$

Dengan μ_0 menyatakan suatu konstanta mengenai rataan yang diketahui.

Bentuk hipotesis dwi arah mengatakan ada dua daerah kritis dengan luas $\frac{\alpha}{2}$ sedangkan bentuk hipotesis eka arah mengatakan bahwa hanya ada satu daerah kritis dengan luas α.

Selanjutnya bentuk statistik uji untuk **rataan satu populasi** yang digunakan adalah

1. Kasus variansi populasi σ^2 diketahui.

   Statistik uji adalah

$$z_H = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

2. **Kasus variansi populasi σ^2 tidak diketahui** atau dengan kata lain variansi sampel S^2 diketahui atau harus dicari.

   Statistika uji adalah

$$t_H = \frac{\bar{x} - \mu_0}{S / \sqrt{n}}$$

Berikut adalah persamaan statistik uji untuk **rataan dua populasi**

1. Kasus variansi populasi 1 dan populasi 2 diketahui

   Statistik uji adalah

$$z_H = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

2. Kasus variansi populasi 1 dan populasi 2 tidak diketahui tapi dianggap sama

   ($\sigma_1^2 = \sigma_2^2$)

   Statistik uji adalah :

$$t_H = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

3. Kasus variansi populasi 1 dan populasi 2 tidak diketahui tapi dianggap berbeda ($\sigma_1^2 \neq \sigma_2^2$)

   Statistik uji adalah :

$$t_H = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

   dengan derajat kebebasan *v* seperti kasus yang sama pada selang kepercayaan.

4. Kasus data berpasangan.

   Statistik uji menyerupai statistik untuk kasus satu populasi dengan variansi tidak diketahui.

$$t = \frac{\bar{d} - d_0}{s_d\sqrt{n}}$$

   Daerah kritis dibuat dengan menggunakan distribusi *t* dengan derajat kebebasan $n-1$.

P-value

Metode lain yang dapat digunakan untuk menguji hipotesis adalah *p-value*. Secara definisi, p-value adalah taraf signifikansi terkecil sehingga statistik uji pengamatan bernilai signifikan, maka, nilai p-value dari sebuah statistik uji dengan distribusi X adalah$P(X > X_{test})$. Nilai p-value dari perhitungan sebuah statistik uji menentukan nilai taraf signifikansi (α) terkecil agar nilai statistik uji tersebut signifikan. Contoh,misal $H_0: \mu = 10$ dan $H_1: \mu \neq 10$, statistik uji yang diperoleh adalah $z = 2.73$ dan p-value yang diperoleh dari perhitungan adalah 0.0064, maka dapat dikatakan bahwa $z = 2.73$ terjadi hanya 64 kali dalam 10.000 kemungkinan peristiwa. Jika p-value dibandingkan dengan taraf signifikansi yang diberikan, misal $\alpha = 0.05$, karena $0.0064 < 0.05$, maka jelas peristiwa $z = 2.73$ berpeluang muncul tidak signifikan, sehingga hipotesis nol ditolak.

Karena perhitungan dari nilai p-value tidak sederhana, p-value biasanya diperoleh di perhitungan pada perangkat lunak, sehingga sangat dianjurkan untuk menggunakan uji statistik p-value jika diberikan.

Langkah-langkah dalam SPSS

- Untuk kasus satu populasi
  1. Lakukan langkah yang sama seperti menentukan selang kepercayaan untuk rataan satu populasi.
  2. Ketikkan nilai μ_0 pada *test value.*
- Untuk kasus dua populasi

  Lakukan langkah yang sama seperti saat menentukan selang kepercayaan untuk dua populasi.
- Untuk data berpasangan

  Lakukan langkah yang sama seperti saat menentukan selang kepercayaan untuk data berpasangan.

Contoh Soal

Sebuah mesin menghasilkan potongan logam berbentuk silinder. Selanjutnya ukuran diameter dari potongan logam tersebut (dalam mm) adalah :

1,01	0,97	1,03	1,04	0,99	0,98	0,99	1,01	1,03

Uji pernyataan yang mengatakan bahwa rata-rata potongan logam tersebut lebih kecil dari 1,09 mm. Gunakan $\alpha = 0{,}05$.

Pertama ditentukan terlebih dahulu hipotesis nol dan tandingannya.

$$H_0 : \mu \geq 1,09$$
$$H_1 : \mu < 1,09$$

Kemudian dengan tingkat signifikansi 0,05 diperoleh titik kritis $-1,86$ dan daerah kritis yakni $t < -1,86$.

Dengan uji t untuk satu populasi diperoleh:

One-Sample Test

	Test Value = 1.09					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
DIAMETER	-10.318	8	.00000671	-8.4444E-02	-.1033	-6.5572E-02

Maka didapatkan t_{hitung} adalah -10,318. Sedangkan nilai t_{tabel} dapat diketahui dengan pilih menu *transform* →*compute* kemudian pada bagian *function* pilih **IDF.T(p,df)** dengan p adalah selang kepercayaan $1-2\alpha$dan df adalah derajat kebebasan$(n-1)$.

Oleh karena nilai statistika uji $t = -10,3184 < -1,86$ maka H_0 ditolak. Jika ditinjau p-value (Sig.(2-tailed)) karena $p_{value} = 671 \times 10^{-8} \ll 0.05 = \alpha$, peristiwa $H_0 : \mu \geq 1,09$ tidak signifikan, sehingga H_0 ditolak. Kesimpulan yang diambil adalah rata-rata potongan logam tersebut lebih kecil dari 1,09 mm.

Uji Hipotesis Variansi Populasi

Bentuk hipotesis nol dan tandingannya untuk kasus **variansi satu populasi** adalah

1. $H_0 : \sigma^2 = \sigma_0^2 \;\; vs \;\; H_1 : \sigma^2 \neq \sigma_0^2 \;\; dwi \;\; arah$

2. $H_0 : \sigma^2 \geq \sigma_0^2$ *vs* $H_1 : \sigma^2 < \sigma_0^2$ *eka arah*
3. $H_0 : \sigma^2 \leq \sigma_0^2$ *vs* $H_1 : \sigma^2 > \sigma_0^2$ *eka arah*

Dengan σ_0 menyatakan suatu konstanta mengenai variansi yang diketahui. Bentuk hipotesis dwi arah mengatakan ada dua daerah kritis dengan luas $\frac{\alpha}{2}$ sedangkan bentuk hipotesis eka arah mengatakan bahwa hanya ada satu daerah kritis dengan luas α.

Statistik uji yang digunakan untuk menguji ketiga hipotesis di atas adalah

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Dibawah H_0 , statistik uji tersebut berdistribusi *chi-kuadrat* dengan derajat kebebasan $(n-1)$.

Untuk hipotesis:

$H_0 : \sigma^2 = \sigma_0^2$ *vs* $H_1 : \sigma^2 \neq \sigma_0^2$, tolak H_0 jika $\chi^2 < \chi^2_{1-\frac{\alpha}{2},(n-1)}$ atau $\chi^2 > \chi^2_{\frac{\alpha}{2},(n-1)}$.

$H_0 : \sigma^2 \geq \sigma_0^2$ *vs* $H_1 : \sigma^2 < \sigma_0^2$,tolak H_0 jika $\chi^2 < \chi^2_{1-\alpha,(n-1)}$.

$H_0 : \sigma^2 \leq \sigma_0^2$ *vs* $H_1 : \sigma^2 > \sigma_0^2$, tolak $\chi^2 > \chi^2_{\alpha,(n-1)}$.

Bentuk hipotesis nol dan tandingannya untuk kasus **variansi dua populasi** adalah

1. $H_0 : \sigma_1^2 = \sigma_2^2$ *vs* $H_1 : \sigma_1^2 \neq \sigma_2^2$ *dwi arah*
2. $H_0 : \sigma_1^2 \geq \sigma_2^2$ *vs* $H_1 : \sigma_1^2 < \sigma_2^2$ *eka arah*
3. $H_0 : \sigma_1^2 \leq \sigma_2^2$ *vs* $H_1 : \sigma_1^2 > \sigma_2^2$ *eka arah*

Dengan σ_1^2 adalah variansi dari populasi 1 dan σ_2^2 adalah variansi dari populasi 2.

Statistisk uji yang digunakan untuk menguji ketiga hipotesis di atas adalah

$$F = \frac{S_1^2}{S_2^2}$$

Dibawah H_0, statistik uji tersebut berdistribusi F dengan derajat kebebasan $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$.

Untuk hipotesis $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, tolak H_0 jika $F < f_{1-\frac{\alpha}{2},(v_1,v_2)}$ atau $F > f_{\frac{\alpha}{2},(v_1,v_2)}$ Untuk hipotesis $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs $H_1 : \sigma_1^2 < \sigma_2^2$, tolak H_0 jika $F < f_{1-\alpha,(v_1,v_2)}$. Untuk hipotesis $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 > \sigma_2^2$, tolak $F > f_{\alpha,(v_1,v_2)}$.

Cara pengujian hipotesis tentang variansi populasi dengan menggunakan Ms. Excel secara manual menggunakan persamaan statistik uji.

Langkah-langkah dalam SAS

1. Simpan data observasi.
2. Kemudianketiksyntax :

```
Procunivariatedata=work.Kadarods select test
for location;

Varkadar;
```

Output SAS

Testing kadar 12:15 Tuesday, July 9, 2013 20
The UNIVARIATE Procedure
Variable: Kadar (Kadar)
Moments

N 16 Sum Weights 16
Mean 10.00625 Sum Observations 160.1
Std Deviation 0.71783819 Variance 0.51529167
Skewness 0.74524318 Kurtosis 0.3079324
Uncorrected SS 1609.73 Corrected SS 7.729375
Coeff Variation 7.17389821 Std Error Mean 0.17945955
Basic Statistical Measures
Location Variability
Mean 10.00625 Std Deviation 0.71784
Median 9.90000 Variance 0.51529
Mode 9.20000 Range 2.60000
Interquartile Range 0.85000

Note: The mode displayed is the smallest of 4 modes with a count of 2.

Tests for Location: Mu0=1
Test -Statistic- -----p Value------
Student's t t 50.1854 Pr> |t| <.0001
Sign M 8 Pr>= |M| <.0001
Signed Rank S 68 Pr>= |S| <.0001
Quantiles (Definition 5)
Quantile Estimate
100% Max 11.60
99% 11.60
95% 11.60
90% 11.20
75% Q3 10.35
50% Median 9.90
25% Q1 9.50
10% 9.20
5% 9.00
1% 9.00
0% Min 9.00

testing kadar

The UNIVARIATE Procedure
Variable: Kadar (Kadar)

Moments			
N	16	**Sum Weights**	16
Mean	10.00625	**Sum Observations**	160.1
Std Deviation	0.71783819	**Variance**	0.51529167
Skewness	0.74524318	**Kurtosis**	0.3079324
Uncorrected SS	1609.73	**Corrected SS**	7.729375
Coeff Variation	7.17389821	**Std Error Mean**	0.17945955

Basic Statistical Measures			
Location		**Variability**	
Mean	10.00625	**Std Deviation**	0.71784
Median	9.90000	**Variance**	0.51529
Mode	9.20000	**Range**	2.60000
		Interquartile Range	0.85000

Note: The mode displayed is the smallest of 4 modes with a count of 2.

Tests for Location: Mu0=1				
Test	**Statistic**		**p Value**	
Student's t	**T**	50.1854	**Pr> |t|**	<.0001
Sign	**M**	8	**Pr>= |M|**	<.0001
Signed Rank	**S**	68	**Pr>= |S|**	<.0001

Quantiles (Definition 5)	
Quantile	**Estimate**
100% Max	11.60
99%	11.60
95%	11.60
90%	11.20
75% Q3	10.35
50% Median	9.90
25% Q1	9.50
10%	9.20
5%	9.00
1%	9.00

Quantiles (Definition 5)	
Quantile	**Estimate**
0% Min	9.00

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
9.0	15	10.3	9
9.2	16	10.4	3
9.2	12	10.6	14
9.4	11	11.2	8
9.6	13	11.6	10

D. UJI HIPOTESIS PENTING LAINNYA

Uji Kenormalan

Dalam menghadapi data, seringkali kita ingin mengetahui apakah data tersebut berdistribusi normal atau tidak. Contoh klasik adalah kita ingin mengetahui apakah data nilai ujian mata kuliah Biostatik berdistribusi normal atau tidak. Maka, perlu dilakukan uji kenormalan. Uji kenormalan yang terkenal adalah uji Kolomogorov-Smirnov dan uji Shapiro-Wilk. Hipotesis nol dan tandingan untuk uji kenormalan adalah:

$$H_0\text{: variabel berdistribusi normal}$$
$$H_1\text{: variabel tidak berdistribusi normal}$$

Langkah uji kenormalan di SPSS adalah sebagai berikut

1. Masukkan variabel pada satu kolom
2. Klik Analyze→Descriptive Statstics→Explore
3. Masukkan variabel yang ingin diuji pada kotak Dependent List
4. Klik tombol Statistics. Pada jendela Explore: Statistics, ceklis Descriptives dan masukkan selang kepercayaan $1-\alpha$. Klik Continue
5. Klik tombol Statistics. Pada jendela Explore: Statistics, ceklis Normality plots with tests dan masukkan selang kepercayaan $1-\alpha$. Ceklis Histogram jika ingin dilihat grafik dari data, lalu klik Continue
6. Klik OK

Uji Kebebasan

Kita seringkali ingin mengecek kebebasannya dari dua variabel (contoh: tingkat pendidikan vs jenis kelamin karyawan perusahaan A). Hal ini dapat diperiksa dengan melakukan uji kebebasan. Jika dua atau lebih variabel terbukti saling bebas, maka tidak ada keterkaitan antar nilai-nilai dari satu kelompok data dengan data yang lain. Hal ini penting untuk diperiksa sebelum melakukan uji-uji statistik lain seperti uji data berdistribusi tertentu. Uji kebebasan bisa dilakukan salah satunya menggunakan uji chi-kuadrat.

Hipotesis nol dan tandingan dari uji kebebasan dari dua variabel, A dan B, adalah:

$$H_0\text{: variabel A dan B saling bebas}$$
$$H_1\text{: variabel A dan B tidak saling bebas}$$

Langkah uji kebebasan chi-kuadrat di SPSS adalah sebagai berikut

7. Masukkan dua data variabel pada kolom yang berbeda
8. Klik Analyze→Descriptive Statstics→Crosstabs
9. Masukkan masing-masing variabel pada kotakRow(s) dan Column(s), tergantung dari bagaimana data ingin ditinjau
10. Klik tombol Statistics. Pada jendela Crosstabs: Statistics, ceklis Chi-square dan klik Continue
11. Klik tombol Cells. Pada jendela Crosstabs: Cell Display, ceklis Observed, Row, Column, dan Total. Klik Continue
12. Klik Continue

Pengamatan statistik uji untuk uji kenormalan dan kebebasan sama dengan uji hipotesis.

E. SOAL LATIHAN

1. Dari Penelitian *"Comparison of Sorbic Acid in Country Ham Before and After Storage"* yg dilakukan di Virginia Polytchnic Institute and State university di tahun 1983, diperoleh data berikut yang menyangkut perbandingan sisa asam sorbat yang dinyatakandalam per sejuta. Berikut adalah data perbandingan sisa asam sorbatpada daging ham yang setelah dicelupkan dalam larutan

sorbat dan daging ham yang disimpan 60 hari setelah proses pencelupan asam sorbat:

Sisa Asam Sorbat dalam ham	Potongan daging ham							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Sebelum disimpan	224	270	400	444	590	660	1400	680
Setelah disimpan	116	96	239	329	437	597	689	576

Bila dianggap kedua populasinya berdistribusi normal, apakah terdapat kenyataan yang cukup untuk menyatakan bahwa lamanya penyimpanan mempengaruhi konsentrasi sisa asam sorbat jika digunakan taraf keberartian 0.05?

2. Data berikut memberikan waktu putar film yang dihasilkan oleh dua perusahaan film gambar hidup :

Perusahaan	**Waktu (menit)**						
A	102	86	98	109	92		
B	81	165	97	134	92	87	114

Ujilah hipotesis bahwa rata – rata waktu putar film hasil perusahaan B lebih 10 menit dari rata – rata waktu putar film perusahaan A serta lawan tandingan selisih kedua perusahaan tersebut melebihi 10 menit. Gunakan taraf keberartian 0.05 dan anggaplah kedua distribusi tersebut hampir normal.

3. Produksi sumur minyak di suatu daerah tercatat sebagai berikut :

1992	5,6	5,4	4,3	7,4	5	6,67	6,3	4,8	7,62	4,56	6,43	5,5
1995	6,16	5,86	4,9	7,68	5,54	7,2	6,7	5,23	8,26	5,04	6,79	5,12

Ujilah asumsi penelitian di atas yang mengatakan bahwa rataan hasil minyak selama 4 tahun bertambah sebesar 0,5 barrel (gunakan tingkat signikansi 0,01).

4. Sebuah mesin otomatis distel untuk mengisi minuman ke dalam botol sebanyak 0.65 liter. Setiap periode tertentu fungsi mesin itu dipantau dengan cara mengukur volume 10 buah botol yang diambil secara acak. Pada suatu tertentu, ternyata diperoleh data berikut sebagai berikut:

   0.63 0.61 0.62 0.60 0.61 0.66 0.67 0.64 0.68 0.64

   berdasarkan data ini, tuliskan rumusan uji hipotesisnya dan ujilah apakah volume minuman dalam botol di bawah 0.65, jika diketahui bahwa volume minuman dalam botol berdistribusi normal dengan deviasi standar sebesar 0.025 liter dan tingkat signifikansi α=5 %.
5. Akan diuji apakah rata-rata isi kaleng sejenis minyak pelumas adalah **10 liter** bila isi sampel acak 10 kaleng adalah sebagai berikut (dalam liter):

   10.2 9.7 10.1 10.3 10.1 9.8 9.9 10.4 10.3 9.8

   dengan tingkat signifikansi $\boldsymbol{\alpha}$**=0.01**

V. ANALISIS VARIANSI (ANOVA)

A. TUJUAN

1. Membandingkan beberapa populasi (angkatan) yang saling bebas (*independent*).
2. Melakukan pengujian terhadap rataan beberapa populasi dengan studi kasus dari beberapa contoh permasalahan.

B. PENDAHULUAN

Pada penaksiran dan pengujian hipotesis sebelumnya Anda telah melakukan uji hipotesis untuk dua buah sampel dengan menggunakan uji *t* (*t*-test). Namun, pembahasan tiap hal terbatas hanya sampai dua parameter. Banyakpenelitian yang memerlukan pengujian lebih dari dua kelompok populasi. Dalam kasus membandingkan <u>lebih dari dua kelompok populasi</u>, salahsatucara yang umum digunakan untuk menguji rataan populasi tersebut adalah **Analisis Variansi (ANOVA**).

C. ANALISIS VARIANSI (ANOVA)

Fungsi dari ANOVA adalah untuk menguji apakah terdapat perbedaan yang signifikan antara rata-rata beberapa kelompok (lebih dari dua), melalui ukuran-ukuran penyebaran (variansi) dari masing-masing kelompok populasi tersebut. Variansi hasil penggabungan semua angkatan data terdiri atas rata-rata variansi setiap angkatan dan variansi dari semua rata-rata angkatan.

Analisis variansi, seperti semua uji konfirmasi lainnya, didasarkan atas beberapa anggapan mengenai sifat data sebagai berikut :

a. Populasi-populasi yang akan diuji berdistribusi normal.
b. Variansi dari populasi-populasi tersebut adalah “sama”.
c. Sampel tidak berhubungan satu dengan yang lainnya (saling bebas).

Sampel acak ukuran *n* diambil masing-masing dari *k* populasi. Ke-*k* populasi yang berbeda ini diklasifikasikan menurut perlakuan atau grup yang berbeda. Populasi-populasi tersebut dianggap saling bebas dan berdistribusi normal dengan rataan μ_1, μ_2, …, μ_k dan variansi σ^2 yang “sama”. Hipotesis yang digunakan dalam metode ini adalah sebagai berikut :

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$.

H_1 : Paling sedikit dua diantara rataan tersebut tidak sama.

Misalkan y_{ij} menyatakan pengamatan ke-j dalam perlakuan ke-i dan susunan datanya seperti tabel berikut

	PERLAKUAN	
	1 2 …. i …. k	
	y_{11} y_{21} …. y_{i1} …. y_{k1} y_{12} y_{22} …. y_{i2} …. y_{k2} ⋮ ⋮ ⋮ ⋮	

	$y_{1,n1}\ y_{2,n2}\ \ldots\ y_{i,ni}\ \ldots\ y_{k,nk}$	
Jumlah	$\sum_{j=1}^{n_1} y_{ij}\ \sum_{j=1}^{n_2} y_{ij}\ \ldots\ \sum_{j=1}^{n_i} y_{ij}\ \ldots \sum_{j=1}^{n_k} y_{ij}$	$\sum_{i=1}^{k}\sum_{j=1}^{n_1} y_{ij}$
Rataan	$\bar{y}_{1\bullet}\ \bar{y}_{2\bullet} \ldots\ \bar{y}_{i\bullet}\ \ldots\ \bar{y}_{k\bullet}$ $\frac{1}{n_1}\sum_{j=1}^{n_1} y_{ij}\ \ \frac{1}{n_2}\sum_{j=1}^{n_2} y_{ij}\ \ldots \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}\ \ldots$ $\frac{1}{n_k}\sum_{j=1}^{n_k} y_{ij}$	$\bar{y}_{\bullet\bullet}$ $\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_1} y_{ij}$

Keterangan Tabel

- $\bar{y}_{i\bullet}$ = rataan semua pengamatan dalam sampel dari perlakuan ke-i.
- $N = (n_1 + n_2 + \ldots + n_i + \ldots + n_k)$
- $\bar{y}_{\bullet\bullet}$ menyatakan rataan semua pengamatan.
- Jumlah ulangan perlakuan tidak harus sama.

Besaran-besaran dalam tabel ANOVA :

1. $a = \dfrac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_1} y_{ij}\right)^2}{N}$ (kuadrat jumlah semua data dibagi banyaknya data)

   $b = \sum_{i=1}^{k}\sum_{j=1}^{n_1} y_{ij}^2$ (jumlah semua kuadrat data)

   $c = \sum_{i=1}^{k} \dfrac{\left(\sum_{j=1}^{n_1} y_{ij}\right)^2}{n_i}$

2. JK$_T$ = $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_{\bullet\bullet}\right)^2 = b-a$ (jumlah kuadrat total)

JK$_P$ = $\sum_{i=1}^{k} n_i\left(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}\right)^2 = c-a$ (jumlah kuadrat perlakuan)

JK$_G$ = $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_{i\bullet}\right)^2$ = JK$_T$ - JK$_P$ = $b-c$ (jumlah kuadrat galat)

TabelV-1 Tabel Analisis Variansi (ANOVA)

Sumber Variasi	**Jumlah Kuadrat**	**Derajat Kebebasan**	**Rataan Kuadrat**	**F (hitung)**
Perlakuan **Galat**	JK$_P$ JK$_G$ = JK$_T$ - JK$_P$	$k-1$ $N-k$	RK$_P$ = $\dfrac{JK_P}{k-1}$ RK$_G$ = $\dfrac{JK_G}{N-k}$	F$_{(hitung)}$ = $\dfrac{RK_P}{RK_G}$
Total	JK$_T$	$N-1$		

Keterangan :

- RK$_P$ = $\dfrac{JK_P}{k-1}$ merupakan rataan kuadrat untuk perlakuan yaitu jumlah kuadrat perlakuan dibagi derajat kebebasan perlakuan
- RK$_G$ = $\dfrac{JK_G}{N-k}$ merupakan rataan kuadrat untuk galat yaitu jumlah kuadrat galat dibagi derajat kebebasan galat.
- Nilai F$_{(hitung)}$ merupakan nilai yang akan dibandingkan dengan nilai F pada tabel dalam pengambil keputusan.

Contoh Soal

Kita ingin mengetahui apakah hormon Auksin, Sitokinin, dan Giberelin memberikan perbedaan yang berarti secara statistik, dilihat dari pengaruhnya terhadap pertambahan panjang kecambah *Phaseolus vulgaris*. Dengan kata lain kita ingin mengetahui apakah terdapat perbedaan yang nyata antara populasi yang diberi hormone Auksin, Sitokinin, dan Giberelin.

Misalkan kita memiliki 15 biji *Phaseolus vulgaris*. Kemudian kita bagi dalam kelompok secara acak. Masing-masing kelompok diberi perlakuan dengan penambahan hormone Auksin, Sitokinin, dan Giberelin masing-masing 4 ppm. Setelah 14 hari pengamatan, didapatkan data panjang tanaman (dalam cm) sebagai berikut :

Auksin	32	37	34	33	30
Sitokinin	36	38	37	30	34
Giberelin	35	30	36	29	31

Data diatas dianalisis dengan mengggunakan SPSS.

Langkah-langkah untuk melakukan uji Anova dengan SPSS sebagai berikut:

1. Masukkan seluruh data perlakuan pada satu kolom. Dan kemudian buat pada kolom berikutnya kolom indeks.
2. Pilih menu Analyze →Compare Means→ One-way ANOVA.
3. Pada kotak Dependent List masukkan data yang akan diuji, variable pada *dependent list* seharusnya sebuah data kuantitatif.

4. Pada kotak Factor, variabel pada kotak *factor* seharusnya sebuah data kualitatif.
5. Pada submenu Option pilih Descriptive dan Homogenity-of-variance. Homogenity-of-variance dipakai untuk menguji berlaku tidaknya salah satu asumsi untuk ANOVA yaitu apakah ketiga sampel mempunyai variansi sama.
6. Untuk mendapatkan $F_{(tabel)}$ pada SPSS gunakan IDF.F (p,df1,df2) dengan *p* adalah selang kepercayaan $(1-\alpha)$, df1 adalahderajatkebebasanuntukperlakuan, df2 adalah derajat kebebasan untuk galat.

Hasil dari output padaSPSS :

Oneway

Descriptives

PANJANG

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Auksin	5	33.2000	2.5884	1.1576	29.9860	36.4140	30.00	37.00
Sito kinin	5	35.0000	3.1623	1.4142	31.0735	38.9265	30.00	38.00
Gibe relin	5	32.2000	3.1145	1.3928	28.3329	36.0671	29.00	36.00
Total	15	33.4667	2.9968	.7738	31.8071	35.1263	29.00	38.00

Test of Homogeneity of Variances
PANJANG

Levene Statistic	df1	df2	Sig.
.404	2	12	.676

ANOVA
PANJANG

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.133	2	10.067	1.144	.351
Within Groups	105.600	12	8.800		34
Total	125.733	14			

Berdasarkan tabel ANOVA didapatkan $F_{(hitung)}$ = 1,144 sedangkan $F_{(tabel)}$ = 3.98, sehingga $F_{(hitung)} < F_{(tabel)}$, *$H_0$ tidak ditolak* artinya rata–rata panjang tanaman yang diberikan perlakuan berbeda (pemberian hormon auksin, sitokinin, giberelin) adalah sama.

D. SOAL LATIHAN

1. Untuk menentukan kestabilan vitamin C dalam sari air jeruk pekat beku disimpan dalam lemari es selama waktu sampai satu minggu, telah dilakukan penelitian *'Vitamin C Retention in Reconstituted Frozen Orange Juice'* oleh Jurusan Gizi dan Makanan di Virginia Polytechnic Institute dan State University di tahun 1975. Sari air jeruk pekat beku diuji dalam 3 jangka waktu berbeda. Jangka waktu menyatakan jumlah hari sejak air jeruk diperas

sampai diuji. Hasilnya tercatat sebagai berikut (dalam mg asam askarbonat per liter) :

Waktu (hari)		
0	3	7
52.6	49.4	42.7
54.2	42.8	48.8
49.8	49.2	40.4
46.5	53.2	47.6
56.0	48.8	49.2
48.0	44.0	42.0
49.6	42.4	43.2
48.4	48.0	48.5
52.5	48.2	43.3
52.0	49.6	45.2
51.8		47.6
53.6		

Gunakan taraf keberartian 5 % untuk menguji hipotesis bahwa tidak ada perbedaan kadar asam askarbonat untuk jangka waktu yang berlainan.

2. Pupuk magnesium amonium fosfat, $MgNH_4PO_4$, merupakan pemasok makanan efektif yang diperlukan bagi pertumbuhan tanaman. Suatu penelitian yang dilakukan oleh George Mason mengenai *'Magnesium Ammonium Phosphate on Height of Chrysanthemums'* dilakukan untuk menentukan taraf pemupukan optimum

yang mungkin, berdasarkan rangsangan respon pertumbuhan vertikal bunga krisan. 40 bibit bunga krisan dibagi dalam 4 kelompok masing-masing 10 tanaman. Tiap kelompok ditanam di dalam pot yang serupa yang mengandung media pertumbuhan yang seragam. Pada tiap kelompok ditambahkan konsentrasi $MgNH_4PO_4$ yang menaik, diukur dalam satuan gram per goni. Ke empat kelompok bunga ditanam pada keadaan seragam dalam rumah kaca selama 4 mingggu. Perlakuan dan perubahan tinggi masing-masing diperlihatkan pada tabel berikut (dalam cm) :

Perlakuan			
50 gr/goni	100 gr/goni	200 gr/goni	400 gr/goni
13.2	16.0	17.8	21.0
12.4	12.6	14.4	14.8
12.8	14.8	20.0	19.1
17.2	13.0	15.8	15.8
13.0	14.0	17.0	18.0
14.0	23.6	27.0	26.0
14.2	14.0	19.6	21.1
21.6	17.0	18.0	22.0
15.0	22.2	20.2	25.0
20.0	24.4	23.2	18.2

a. Lakukan analisis variansi untuk data diatas !
b. Apakah dapat disimpulkan pada taraf keberartian 5 % bahwa konsentrasi $MgNH_4PO_4$ yang berlainan

mempengaruhi tinggi rata-rata yang dicapai bunga krisan ?

3. Irritable Bowel Syndrome (IBS) is a nonspesific intestinal disorder characterized by abdominal pain and irregular bowel habits. Each person in a random sample of 24 patients having periodic attacks of IBS was randomly assigned to one of three treatment groups, A, B, and C. The number of hours of relief while on therapy is recorded for each patient in Table.

Treatment	A	4,2	2,3	6,6	6,1	10,2	11,7	7	3,6
	B	4,1	10,7	14,3	10,4	15,3	11,5	19,8	12,6
	C	38,7	26,3	5,4	10,3	16,9	43,1	48,6	29,5

Test for differences among the population variances. Use $\alpha = 0{,}05$.

4. Tiga kelas mata kuliah Statistika Dasar diberikan oleh tiga pengajar yang berbeda. Nilai akhir yang diberikan oleh pengajar dari mata kuliah tersebut tercatat sebagai berikut :

Pengajar		
A	B	C
73	88	68
89	78	79
82	48	56
43	91	91
80	51	71
73	85	71
66	74	87
60	77	41
45	31	59

Pengajar		
A	B	C
93	78	68
36	62	53
77	76	79
	96	15
	80	
	56	

Adakah perbedaan yang berarti dalam nilai rata-rata yang diberikan oleh ketiga pengajar? Gunakan taraf keberartian 0.05.

5. Data pada tabel di bawah ini menyatakan lamanya jam kesembuhan yang diakibatkan oleh 5 merek obat demam yang berlainan, yang diberikan kepada 25 penderita demam dengan panas 38^0C atau lebih. Lakukan analisis variansi dan uji hipotesis pada taraf keberartian 0.05 bahwa rataan lamanya jam kesembuhan yang diakibatkan oleh kelima merek tablet sama saja.

Tablet				
Panadol	Paramex	Antangin	Inza	Decolgen
5	9	3	2	7
4	7	5	3	6
8	8	2	4	9
6	6	3	1	4
3	9	7	4	7

VI. ANALISIS DERET WAKTU

A. TUJUAN

1. Memahami mekanisme stokastik yang berkembang menjadi observasi barisan.
2. Menentukan karakteristik data seperti kestasioneran, tren, dan musiman melalui scatter plot data terhadap waktu.
3. Menentukan model deret waktu stasioner.

B. PENDAHULUAN

Dalam kehidupan sehari-hari banyak fenomena yang terjadi memiliki pola dan keteraturan tertentu. Fenomena yang terjadi tersebut dapat dimisalkan sebagai barisan peubah acak yang memiliki indeks parameter tertentu berupa waktu, lokasi, atau waktu dan lokasi sekaligus. Sebagai contoh, fenomena indeks curah hujan tahunan di suatu propinsi Jawa Barat, penutupan harga saham IBM harian, produksi hasil panen perkebunan teh tahunan, kandungan coal bed metan (CBM) di suatu blok pertambangan setiap 5 tahun sekali. Contoh di atas merupakan contoh data deret waktu yang dapat dianalisis menggunakan analisis deret waktu (*time series analysis*).

C. ANALISIS DERET WAKTU

Analisis deret waktu bertujuan untuk membuat model yang dapat berguna untuk memprediksi nilai pada waktu yang akan datang berdasarkan observasi-observasi yang telah ada. Ada beberapa langkah untuk membangun model deret waktu sebagai berikut :

1. Spesifikasi atau identifikasi model

   Pada tahap ini, kita memiliki model dengan prinsip *parsimony* (model sederhana) dengan jumlah parameter yang sedikit. Prosesnya dengan membuat plot observasi terhadap waktu kemudian diamati apakah grafiknya sudah stasioner, memiliki tren naik atau turun, dan mengandung unsur musiman.
2. Pencocokan (*fitting*) model

   Pada tahap ini, kita menemukan estimasi atau taksiran terbaik dari parameter yang tidak diketahui. Dalam hal ini, kita dapat menggunakan SPSS untuk menentukan model yang sesuai dengan data.
3. Diagnosa model

   Disini kita akan menganalisa kualitas model, apakah model cukup layak dengan melihat plot error dan plot normal.

Pengolahan Data Observasi

Misalkan observasi Z_t merupakan peubah acak maka barisan peubah acak $\{Z_t, t=1,2,...,n\}$ adalah proses stokastik. Ada beberapa statistik yang dapat dihitung yaitu:

- Fungsi mean : $\mu_t = E[Z_t], t = 1, 2, ..., n$

- Fungsi autokovariansi :

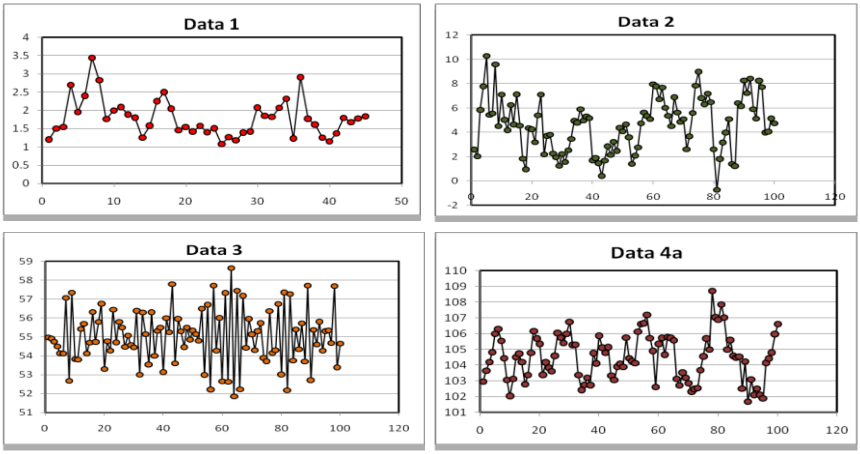
$$\gamma_{t,s} = \text{Cov}(Z_t, Z_s) = \text{E}[(Z_t - \mu_t)(Z_s - \mu_s)] = \text{E}[Z_t Z_s] - \mu_t \mu_s$$

- Fungsi auto korelasi (ACF) :

$$\rho_{t,s} = \text{Corr}(Z_t, Z_s) = \frac{\text{Cov}(Z_t, Z_s)}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$$

Model Deret Waktu

Model deret waktu dibagi menjadi 2 yaitu model stasioner dan model tidak stasioner. Dalam modul ini hanya akan dibahas model deret waktu yang stasioner. Kestasioneran data dapat langsung dilihat dari plot observasi terhadap waktu, jika grafik bersifat acak dan tidak terdapat unsur trend dan musiman, maka kita dapat membuat model stasioner untuk data tersebut. Secara statistika, data yang stasioner memiliki sifat mean dan variansi yang konstan. Berikut ini berbagai grafik data yang dapat diidentifikasikan kestasionerannya :



GambarVII. 1 Beberapa jenis plot data *time series*.

Pada grafik Data 2 terdapat trend turun dan naik, Data 3 data tidak stasioner karena variansinya tidak konstan, sedangkan pada Data 4a, terdapat unsur musiman.

Terdapat dua model deret waktu stasioner yaitu model AR (*Auto Regressive*), model MA (*Moving Average*), dan gabungan keduanya yaitu model ARMA. Berikut ini masing-masing penjelasan tentang model-model tersebut:

1. Model MA : $Z_t = a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - ... - \psi_q a_{t-q}$
2. Model AR : $Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + ... + \varphi_p Z_{t-p} + a_t$
3. Model ARMA :
   $Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + ... + \varphi_p Z_{t-p} + a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - ... - \psi_q a_{t-q}$

Cara membedakan model yang digunakan tersebut, kita dapat membuat grafik ACF dan PACF (*Partial* ACF).

TabelVII-1Pola teoritis grafik ACF dan PACF.

Model	ACF	PACF
AR(p)	Eksponensial turun atau seperti gelombang sinus	Terpotong setelah laq-p
MA(q)	Terpotong setelahlaq-q	Eksponensial turun atau seperti gelombang sinus
ARMA(p,q)	Eksponensial turun	Eksponensial turun

Pemilihan model terbaikberdasarkan AIC dan BIC

Akakike info criterion (AIC)

Suatu ukuran informasi yang dikembangkan oleh Hirotsugu Akaike pada tahun 1971 mengenai ukuran terbaik dalam kelayakan pengukuran estimasi model. Akaike info criteration biasanya disingkat AIC. AIC bukan penyajian hipotesis, melainkan adalah tes antara model-model atau sebagai alat ukur untuk memilih model yang tepat. Semakin kecil nilai AIC maka model yang digunakan semakin baik. Adapun rumus AIC adalah sebagai berikut:

$$AIC = 2k - 2\ln(L)$$

Dimana k adalah jumlah parameter dalam model statistik, dan L adalah nilai maksimum dari fungsi likelihood dari estimasi model.

Schwarz criterion

Biasanya disebut sebagai Bayesian Information Criterion (BIC) yang dikembangkan oleh Gideon E. Schwartz dan diadopsi oleh Bayesian. Kriteria ini hamper sama dengan Akakike info criterion (AIC). Namun pengukurannya lebih baik dari Akakike info criterion (AIC). Semakin kecil nilai Schwartz criterion maka model yang digunakan semakin baik. Adapun rumus BIC adalah sebagai berikut:

$$-2\ln p(x|k) \approx BIC = -2.\ln L + k\ln(n)$$

Dimana

x = Objek penelitian.

n = Banyak data dalam objek x, banyaknya jumlah yang diobservasi, atauukuransampel.

k = jumlah dari parameter bebas yang diestimasi.

$p(x|k)$= likelihood dari observasi data yang menunjukkan jumlah dari parameter.

L = jumlah maksimum dari fungsi likelihood dari estimasi model.

Langkah-langkah pengerjaan dalam SPSS sebagai berikut :

1. Pilih menu Analyze →Forecasting → Create models
   Masukkan data yang akan diolah pada dependent variables, dan data waktu pada independent variable.
   Pilih Model Type : All models.
   Pada toolbar Statistics, checklist : goodness of fit untuk kebaikan model serta parameter estimates, checklist Normalized BIC untuk mengetahui model yang paling baik. Klik OK.
2. Membuat grafik ACF dan PACF untuk menentukan model deret waktu yang sesuai :
   Pilih menu Analyze →Forecasting → Autocorrelations
   Masukkan data yang akan dioleh pada variables, dan checklist pada display autocorrelationsdan partial autocorrelations.
3. Jika ingin memunculkan hasil berdasarkan model yang telah dipilih gunakan:
   Lakukan langkah 1, tambahkan :
   - Variables → Method : pilih ARIMA → criteria → ARIMA Orders→ Masukkan bilangan pada model yang sesuai →Continue.
   - pada Statistics : klik display forecasts.
   - pada Plots : klik observed values, forecasts, dan fit value.
   - klik OK.
4. Jika grafik data observasi terhadap waktu menunjukkan adanya trend dan musiman, maka yang perlu diubah-ubah adalah method pada variablesdengan mengganti

bilangan pada ARIMA Orders dan memasukkan unsur seasonal jika ada.

Contoh

Berikut ini Data penjualan suatu perusahaan elektronik yang dinyatakan dalam juta rupiah selama 100 hari.

54,97	53,80	54,77	56,38	56,00	55,14	57,33	54,29	57,36	55,37
54,91	55,42	54,27	53,00	55,25	54,80	52,62	55,31	52,17	54,60
54,75	55,69	56,44	56,29	57,80	56,49	58,64	55,73	57,27	55,81
54,50	54,12	54,71	55,15	53,59	52,98	51,85	53,88	53,75	54,27
54,13	54,69	55,80	53,53	55,96	56,71	57,44	53,71	55,39	55,31
54,12	56,32	55,50	56,31	55,29	52,21	52,22	56,37	54,35	55,35
57,07	54,73	54,47	54,00	54,46	57,72	57,18	54,15	55,72	54,67
52,67	55,80	55,07	55,32	55,49	54,27	54,42	54,29	53,70	57,69
57,35	56,76	54,57	55,50	54,86	56,01	55,96	56,75	57,72	53,38
53,82	53,29	54,44	53,13	55,33	52,65	55,13	53,01	52,71	54,65

Tentukan model deret waktu yang sesuai dengan data di atas.

Hasilnya adalah :

1. Langkah pertama menghasilkan :

Time Series Modeler

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	.554	.	.554	.554	.554	.554	.554	.554	.554	.554	.554
R-squared	.554	.	.554	.554	.554	.554	.554	.554	.554	.554	.554
RMSE	97.846	.	97.846	97.846	97.846	97.846	97.846	97.846	97.846	97.846	97.846
MAPE	1.404	.	1.404	1.404	1.404	1.404	1.404	1.404	1.404	1.404	1.404
MaxAPE	5.095	.	5.095	5.095	5.095	5.095	5.095	5.095	5.095	5.095	5.095
MAE	77.269	.	77.269	77.269	77.269	77.269	77.269	77.269	77.269	77.269	77.269
MaxAE	294.491	.	294.491	294.491	294.491	294.49	294.49	294.49	294.49	294.49	294.49
Normalized BIC	9.259	.	9.259	9.259	9.259	9.259	9.259	9.259	9.259	9.259	9.259

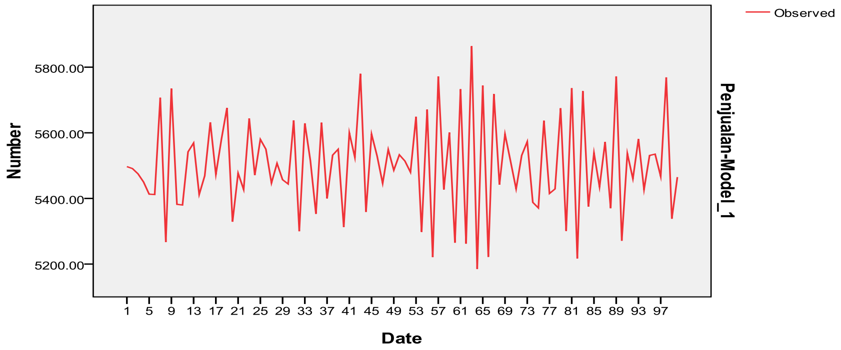
ARIMA Model Parameters

					Estimate	SE	T
Penjualan-Model_1	Penjualan	No Transformation	Constant		5502.27	5.656	9
			AR	Lag 1	-.737	.067	-

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Penjualan-Model_1	0	.554	15.528	17	.558	0

Scatterplot dari data penjualan di atas sebagai berikut:



GambarVII. 2 Scatter plot dari data penjualan. Mean data hampir konstan sedangkan variabilitasnya relatif besar.

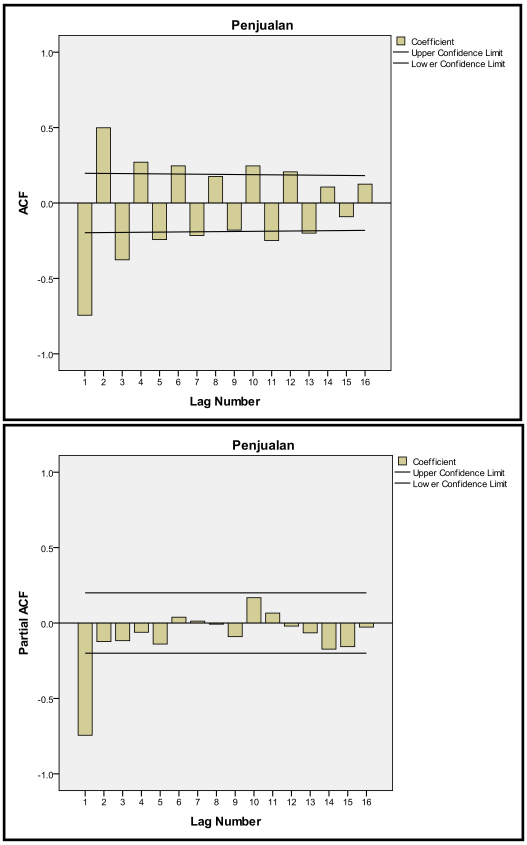
Berdasarkan output di atas, diperoleh beberapa kesimpulan sebagai berikut:

- Model yang dipilihkan oleh software dan sesuai dengan data penjualan di atasadalah model AR(1) dengan parameter model yang tercantum pada tabel ARIMA Model Parameters. Model deret waktu sebagai berikut:

$$Z_t = 5502.27 - 0.737 Z_{t-1} + a_t$$

- Kecocokan model dengan data penjualan sebenarnya diperoleh dari nilai R-squared sebesar 0,554 atau 55,4% artinya kecocokan model dan data cukup cocok.

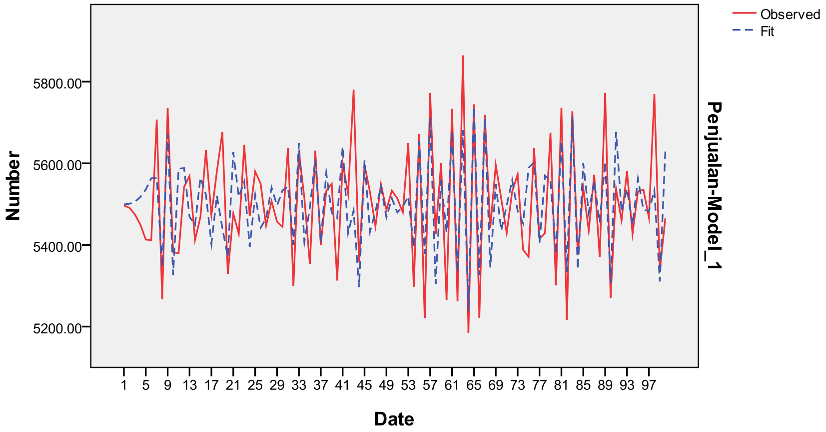
2. Langkah kedua menghasilkan grafik ACF dan PACF sebagaiberikut:



GambarVII. 3 Plot ACF dan PACF. Seiring bertambahnya lag waktu yang diamati, nilai ACF menurun dengan pola sinus teredam. Sedangkan nilai PACF patah (*cut-off*) setelah lag pertama.

Dari grafik ACF dan PACF di atas, model yang sesuai untuk data keuntungan di atas adalah AR(1). Hal ini sesuai dengan hasil yang diperoleh pada Langkah 1.

3. Grafikperbandingan data observasidenganhasil model AR(1) sebagaiberikut:



GambarVII. 4 Plot data observasi (garis merah) dengan hasil pemodelan AR(1) (garis biru, putus-putus).

Dari grafik di atas, tampakbahwa data hasil model AR(1) sebagian kecil sesuai dengan data observasi karena memang hanya 55,4% kecocokan model dan observasi.

D. SOAL LATIHAN

1. Berikut ini Data pendapatan seorang pegawai Pabrik Konveksi XXX selama 45 hari.

1,20	2,40	2,09	2,25	1,42	1,27	1,85	2,91	1,37
1,50	3,44	1,89	2,50	1,57	1,18	1,82	1,77	1,79
1,54	2,83	1,80	2,05	1,40	1,39	2,07	1,61	1,68
2,70	1,76	1,25	1,46	1,51	1,42	2,32	1,25	1,78
1,95	2,00	1,58	1,54	1,08	2,08	1,23	1,15	1,84

Tentukan model deret waktu yang sesuai untuk data di atas.

2. Data berikut ini adalah data diameter suatu bahan karet yang memuai jika terkena panas yang diamati selama 75 hari.

25,86	23,07	22,66	21,95	24,92
22,21	23,45	24,85	25,78	24,31
25,62	23,84	24,04	24,27	23,08
24,98	24,72	25,91	24,75	23,83
22,96	22,98	23,57	23,02	25,51
23,73	24,61	25,20	26,11	22,85
23,89	24,66	24,86	24,91	23,50
22,44	23,42	24,29	23,87	22,79
23,60	24,11	22,58	23,43	23,72
24,38	23,90	24,20	22,11	24,09
24,26	25,14	24,89	24,81	25,23
24,36	23,53	25,34	23,60	21,52
21,93	26,70	21,22	24,03	23,36
24,77	22,43	24,52	25,13	23,35
25,36	24,91	24,62	23,65	25,08

Tentukan model deret waktu yang sesuai dengan data diameter bahan karet di atas.

3. Data inflasi untuk bahan makanan mulai Januari 2009 sampai dengan Agustus 2012 sebagai berikut:

0.76	-0.13	1.49	-0.35
0.95	1.73	2.81	0.59
-0.26	0.86	2.21	1.62
-1.33	-0.91	-0.33	1.85
-0.25	0.33	-1.94	-0.73
-0.18	0.49	-1.90	-0.33
1.14	3.20	-0.28	0.12
1.29	4.69	1.27	-0.15
2.43	0.47	1.84	1.57
0.28	0.44	1.07	1.68

-0.82	-0.85	-0.09	1.48

Tentukanjenis data di atas apakah memiliki tren atau pola musiman serta tentukan model yang sesuai untuk data inflasi tersebut.

4. Data berikut ini adalah data kecelakaan yang terjadi di Indonesia mulai dari tahun 1992 sampai dengan 2010. Tentukan model deret waktu untuk data tersebut.

19,920	17,101	12,267	49,553
17,323	14,858	13,399	59,164
17,469	12,675	17,732	62,960
16,510	12,649	91,623	66,488
15,291	12,791	87,020	

5. Data nilaiekspor Indonesia mulaidariJanuari 2009 sampaidengan April 2012 dinyatakandalam US $ diperolehdaripusat data statistikabpps

NilaiEkspor Indonesia (dalam US $)			
7,280,109,646	10,775,361,672	12,181,628,292	17,418,472,565
7,134,319,273	13,348,131,454	14,399,644,857	18,647,825,151
8,614,725,871	11,595,867,120	15,633,275,868	17,543,408,243
8,453,957,057	11,166,450,436	16,829,888,773	16,957,743,283
9,208,774,059	12,774,365,884	14,606,249,454	17,235,463,273
9,381,479,071	12,035,247,591	14,415,278,398	17,077,694,229
9,684,145,879	12,619,125,277	16,365,953,469	15,570,069,320
10,543,777,892	12,330,114,499	16,554,240,767	15,695,443,242
9,842,571,682	12,486,972,905	18,287,435,825	17,251,519,437
12,242,672,525	13,726,521,968	18,386,855,403	16,173,190,978

Tentukan model deretwaktuuntuknilaiekspor Indonesia tersebut.

DAFTAR PUSTAKA

Cryer, Jonathan D., *Kung-Sik Chan, Time Series Analysis With Applications in R*, 2nd Ed.,Springer,2008.

Montgomery, Douglas C., *Statistical Quality Control*, John Wiley & Sons, 2009.

Walpole, Ronald E., dan Myers, Raymond H., *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*, Edisi 4, Bandung: Penerbit ITB, 1995.

Walpole, Ronald E., et.al, *Statistic for Scientist and Engineering*, 8th Ed.,2007.

BIODATA DIRI

Saya Mahfudhotin, M.Si., seorang dosen PNS IAIN Kediri. Saya S1 Matematika di UNAIR tahun 2009 dan Saya lanjut program studi magister di ITB Beasiswa Calon Dosen BPPDN Dikti 2013. Saya aktif dalam organisasi Ikatan Sarjana Nahdlatul Ulama' Kab.Tuban, di sana saya merasa sudah mulai mendapatkan efek pengembangan diri sebagai scientist, selain pendidik, yang bercita-cita menjadi seorang yang sukses tanpa saya lupa akan kondisi masyarakat yang tidak mampu untuk memperolah hidup yang lebih layak. Saya juga tergabung dalam suatu komunitas Dunia Santri Community yang bergerak dalam bidang literasi, anggotanya terdiri dari santri atau alumni santri dari berbagai latar belakang keilmuan. Saya alumni Ponpes Sunan Bejagung Tuban dengan pengasuh Kiai Matin Jawahir dan PP Al Fithroh kedinding Surabaya. Didalamnya saya seperti menemukan keluarga, selain saya bisa membangun jaringan dengan orang-orang yang berkompeten dibidangnya. Saya bersyukur ini menjadi bagian dari wiridan saya di era digital. Di samping itu, saya juga menulis di http://www.asmatuqa.blogspot.com/.

A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Mahfudhotin, M.Si.
2	Jenis Kelamin	Perempuan
3	Tempat dan Tanggal Lahir	Kediri, 3 Oktober 1990
4	E-mail	mahfudhotin@iainkediri.ac.id
5	Alamat Kantor	Jln. Sunan Ampel No.7 Ngronggo Kota Kediri
6	Matakuliah Yang Diampu	1. Metode Penelitian
		2. Statistika
		3. Matematika Ekonomi
		4. Kewirausahaan

B. Riwayat Pendidikan

Uraian	S-1	S-2	S-3
Nama Perguruan Tinggi	Universitas Airlangga	Institut Teknologi Bandung	-
Bidang Ilmu	Matematika	Matematika	-
Tahun Masuk-Lulus	2009-2013	2013-2015	-
Judul Skripsi/ Tesis/ Disertasi	Estimasi Model Regresi Panel Komponen Error Satu Arah Dengan Metode Generalized Least Square	Kebergantungan Ekor Copula Untuk Prediksi Nilai Batas Dalam	-

C. Pengalaman Penelitian Dalam 5 Tahun Terakhir

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber*	Jml (Juta Rp)
1	2020	Quantile-based estimative VaR forecast and dependence measure: A simulation approach	"Hibah Berbasis Kompetensi" from the Ministry of Research, Technology and Higher Education of the Republic of Indonesia	100
2	2016	Regresi Generalized Poisson Untuk Memodelkan Jumlah Penderita Gizi Buruk Pada Balita di Surabaya	DIKTI	7,5

D. Publikasi Artikel Ilmiah Dalam Jurnal dalam 5 Tahun Terakhir

No.	Judul Artikel Ilmiah	Nama Jurnal	Volume/ Nomor/Tahun
1	Quantile-based estimative VaR	Journal of Applied	Volume 2020 |Article ID 8276019 |

	forecast and dependence measure: A simulation approach	Mathematics : Hindawi 2020	https://doi.org/10.1155/2020/8276019
2	Regresi Generalized Poisson Untuk Memodelkan Jumlah Penderita Gizi Buruk Pada Balita di Surabaya	Jambura Journal of Probability and Statistics	Volume 1 Nomor 1, Mei 2020
3	Analisa Pertumbuhan Tenaga Kerja dan Jaringan Kantor Terhadap Perkembangan Aset Perbankan Syariah	El-Qist: Journal of Islamic Economics and Business (JIEB)	Volume 9 Nomor 1, 2019

E. Pemakalah Seminar Ilmiah (*Oral Presentation*) dalam 5 Tahun Terakhir

No.	Nama Temu ilmiah / Seminar	Judul Artikel Ilmiah	Waktu dan Tempat
1	International Conference in Mathematics :	Value-at-Risk Prediction : A Copula-based	2016. Institut Teknologi Sepuluh

	Pure, Applied, and Computation (ICoMPAC)	Tail Dependence Approach	Nopember (ITS), Surabaya

F. Perolehan HKI dalam 10 Tahun Terakhir

No.	Judul/Tema HKI	Tahun	Jenis	Nomor P/ID
1	Sample Schedulling (Si Dull) Program di Balai Riset dan Standardisasi Industri Surabaya	2021	Program Komputer	EC00202131073, 1 Juli 2021