

BASIC STATISTIC FOR STUDENTS

fransiscus fendy novento

PENDAHULUAN

lihat kasus berikut:

1. terkumpul nilai ulangan matematika suatu kelas sbb:
7,8,9,4,5,8,7,9,10,5
2. dikumpulkan lagi nilai ulangan matematika diatas sbb:

no	nama	nilai
01	fendy	7
02	hatma	8
03	yudhie	9
04	hari	4
05	yoelianto	5
06	samuel	8
07	airlangga	7
08	yoyok	9
09	agus	10
10	adi	5

3. kemudian dilakukan analisis sbb:

descriptive statistic	
Mean	7.2
Standard Error	0.628932075
Median	7.5
Mode	7
Standard Deviation	1.988857852
Sample Variance	3.955555556
Kurtosis	-1.061351381
Skewness	-0.343204151
Range	6
Minimum	4
Maximum	10
Sum	72
Count	10
Largest(1)	10
Smallest(1)	4
Confidence Level(95.0%)	1.422744284

Dari ketiga kasus diatas dapat kita tarik suatu kesimpulan sbb:
 kasus 1 disebut sebagai **data**: bahan mentah yang tidak memiliki arti
 kasus 2 disebut sebagai **informasi**: data diolah menjadi bermakna
 kasus 3 disebut sebagai **knowledge**: informasi diolah menjadi pengetahuan yang selanjutnya dapat digunakan untuk membuat penafsiran atau inferensi dan pengambilan keputusan.

ilmu yang mempelajari proses perjalanan dari data menjadi knowledge dinamakan ilmu *statistik*. sedang ilmu yang digunakan untuk membuat inferensi atau pengambilan keputusan dinamakan *probabilitas*. oleh karena kedua ilmu tersebut yaitu statistik dan probabilitas tidak dapat dipisahkan karena merupakan satu kesatuan utuh. ada pembahasan tentang statistik, tentu saja ada pembahasan tentang probabilitas, seperti saudara sekandung.

DEFINISI AWAL

Statistik adalah ilmu yang berhubungan dengan pengumpulan data, pengolahan dan penyajian data, dan cara pengambilan keputusan dengan probabilitas.

elemen statistik terdiri atas:

- populasi
adalah kumpulan semua elemen yang akan diteliti
- sensus
pengambilan seluruh elemen populasi
- sampling
pengambilan sebagian kecil dari seluruh elemenn populasi

Tipe data

- Data kuantitatif
Data interval: 50 - 60 kg
Data rasio: 27 roti
- Data kualitatif
Data nominal: jenis kelamin(pria atau wanita)
Data ordinal: kepuasan pelanggan(puas,tidak puas, sangat tidak puas)

penggolongan statistik

- statistik deskriptif
berusaha menjelaskan atau menggambarkan berbagai karakteristik data, seperti berapa rata-ratanya, seberapa jauh data-data bervariasi, dsb
- statistik induktif (inferensi)
berusaha membuat berbagai inferensi terhadap sekumpulan data yang berasal dari suatu sampel. Tindakan inferensi tersebut seperti melakukan perkiraan, peramalan, pengambilan keputusan, dsb ada dua metode yang bisa dilakukan:
 1. statistik parametrik
untuk populasi yang parameternya telah memenuhi persyaratan tertentu (misal sebaran data mengikuti distribusi normal)
 2. statistik non-parametrik
untuk populasi yang parameternya tidak memenuhi persyaratan tertentu.

STATISTIK DAN KOMPUTER

Pada dasarnya ilmu statistik memiliki 1001 rumus yang kalau disuruh menghafal ataupun menghitung untuk jumlah data kecil tidak akan mengalami masalah, tetapi apabila data sangatlah besar hampir tidak mungkin apabila kita menghitungnya secara manual karena akan menghabiskan waktu dan lagi tingkat kesalahan manusia juga tinggi. ("mbrodol sirah" kalau orang Jawa bilang).

Seiring dengan penggunaan komputer untuk perhitungan matematis maka muncullah program-program statistik mulai dari yang sederhana sampai yang kompleks. aplikasi yang sering digunakan untuk analisis statistik adalah *MS Excel* dan *SPSS*.

Sebelum kita melangkah lebih jauh lagi, kita perlu memahami dahulu suatu konsep paling penting dalam statistik, yaitu: **distribusi normal**.

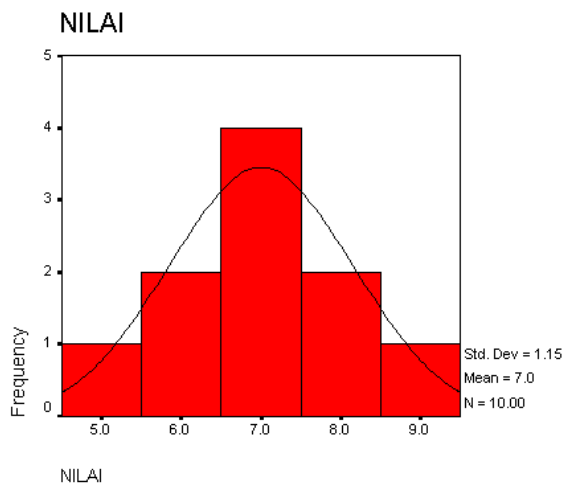
DISTRIBUSI NORMAL

apabila kita memiliki sekelompok data kemudian kita gambarkan dalam bentuk grafik, maka bentuknya bisa bermacam-macam. apabila grafik tersebut berbentuk seperti lonceng maka kita sebut bahwa sekelompok data tersebut mempunyai distribusi normal (distribusi gauss).

untuk selanjutnya sekelompok data memiliki distribusi normal apabila:

- datanya terukur
- jumlah data yang nilainya ekstrem (sangat besar / kecil) tidak terlalu banyak
- nilai rata-rata=nilai median=nilai modus dan grafik berbentuk seperti lonceng

sebagai contoh, kita bisa melihat grafik nilai ulangan matematika sbb:



tampak bahwa nilai rata-rata = nilai modus = nilai median = 7 dan grafik berbentuk lonceng sehingga memenuhi **distribusi normal**

catatan:

Dalam kehidupan sehari-hari sangat sulit atau hampir tidak pernah dijumpai kejadian-kejadian yang benar-benar mempunyai distribusi normal, tetapi ada banyak terjadi kejadian-kejadian yang dapat dianggap mendekati distribusi normal, misalnya berat badan murid, hasil ujian, kekayaan penduduk di suatu tempat, dll. Dalam statistik jika jumlah data melebihi 30 sudah dianggap mempunyai distribusi normal.

[soal latihan distribusi normal](#)

Untuk selanjutnya pokok bahasan di bawah ini akan menyesuaikan dengan menu program SPSS

STATISTIK DESKRIPTIF

statistik deskriptif erat hubungannya dengan pengumpulan data, peringkasan data, dan penyajian data. data disajikan dalam bentuk tabel, grafik, ataupun ringkasan statistik.

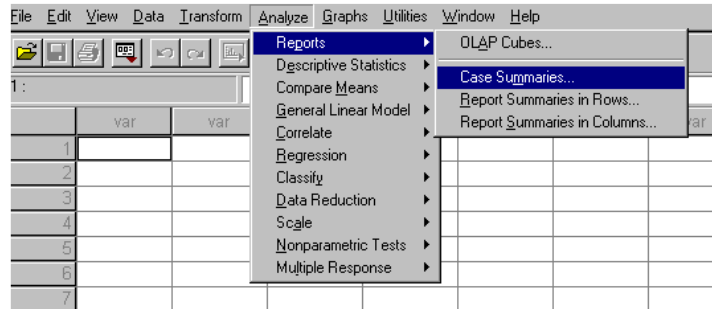
ringkasan statistik memiliki ukuran yang sering dipakai sbb:

1. ukuran pemusatan (*central tendency*): seperti mean, median, modus
2. ukuran penyebaran (*dispersi*): seperti standard deviasi, varians

selain kedua ukuran diatas SPSS juga mengenal *skewness dan kurtosis* untuk mengetahui kemiringan data atau memenuhi syarat distribusi normal atau tidak.

Pada statistik deskriptif beberapa menu yang biasanya digunakan dalam SPSS adalah:

- **case summaries**



contoh:

disajikan data tinggi badan 25 responden yang diambil secara acak sbb:

tinggi	gender	tinggi	gender	tinggi	gender
170.2	pria	170.4	wanita	170.4	pria
172.5	pria	161.3	wanita	168.9	pria
180.3	pria	172.5	pria	164.8	wanita
172.5	pria	170.4	wanita	167.2	wanita
159.6	wanita	168.9	wanita	167.2	wanita
168.5	wanita	168.9	wanita		
168.5	pria	177.5	wanita		
172.5	pria	174.5	pria		
174.5	pria	168.6	wanita		
150.6	wanita	164.8	wanita		

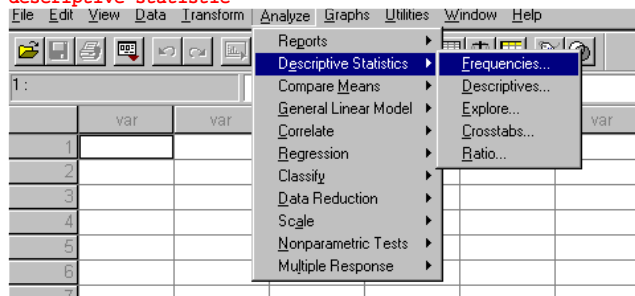
dengan case summarise akan memberikan output sbb:

Case Summaries				
			Case Number	TINGGI
GENDER	pria	1	1	170.20
		2	2	172.50
		3	3	180.30
		4	4	172.50
		5	7	168.50
		6	8	172.50
		7	9	174.50
		8	13	172.50
		9	18	174.50
		10	21	170.40
		11	22	168.90
	Total	N	11	
		Mean	172.4818	
		Std. Deviation	3.27866	
wanita	1	5	159.60	
	2	6	168.50	
	3	10	150.60	
	4	11	170.40	
	5	12	161.30	
	6	14	170.40	
	7	15	168.90	
	8	16	168.90	
	9	17	177.50	
	10	19	168.60	
	11	20	164.80	
	12	23	164.80	
	13	24	167.20	
	14	25	167.20	
Total	N	14		
	Mean	166.3357		
	Std. Deviation	6.23385		
Total	N	25		
	Mean	169.0400		
	Std. Deviation	5.93500		

analisis:dari ouput diatas tampak bahwa:

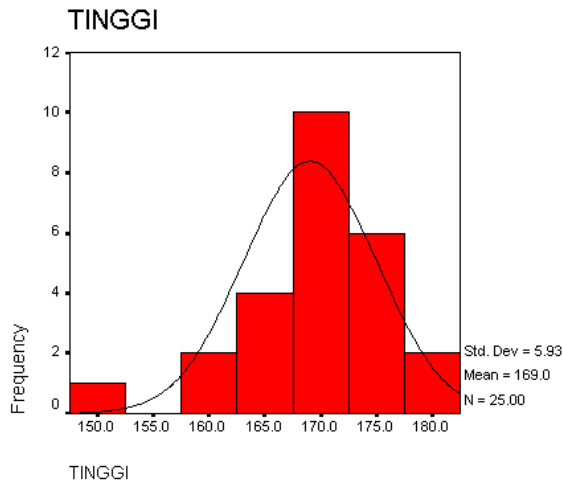
- ada 11 pria yang diproses (N=11) yang memiliki mean 172.4818 dan std deviasi 3.27866
- ada 14 wanita yang diproses (N=14) yang memiliki mean 166.3357 dan std deviasi 6.23385
- total ada 25 responden yang diproses (N=25) dengan mean 169.0400 dan std deviasi 5.93500

• **descriptive statistic**



menggunakan descriptive statistics / frequencies

Statistics		
TINGGI		
N	Valid	25
	Missing	0
Mean		169.0400
Std. Error of Mean		1.18700
Median		168.9000
Mode		172.50
Std. Deviation		5.93500
Variance		35.22417
Skewness		-1.128
Std. Error of Skewness		.464
Kurtosis		3.115
Std. Error of Kurtosis		.902
Range		29.70
Minimum		150.60
Maximum		180.30
Percentiles	10	160.6200
	25	167.2000
	50	168.9000
	75	172.5000
	90	175.7000



analisis: dari output diatas tampak bahwa:

- mean adalah 169.0400 dan standard deviasinya 5.93500
- $\text{ratio skewness} = \text{skewness} / \text{std error of skewness} = -1.128 / 0.464 = -2.43$
- $\text{ratio kurtosis} = \text{kurtosis} / \text{std error of kurtosis} = 3.115 / 0.902 = 3.45$
- apabila ratio skewnes ataupun kurtosis berada diantara -2 dan 2 maka sample berdistribusi normal
- ratio skewness maupun kurtosis sample diatas berada di atas 2 sehingga sample tidak berdistribusi normal
- untuk lebih menguatkan ratio skewness maupun kurtosis ini kita bisa melihat grafiknya yang notabene agak menceng ke kanan, sehingga tidak memenuhi kaidah distribusi normal.

menggunakan descriptive statistics / descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
TINGGI	25	150.60	180.30	169.0400	5.93500
Valid N (listwise)	25				

	tinggi	ztinggi	ve
1	170.20	.19545	
2	172.50	.58298	
3	180.30	1.89722	
4	172.50	.58298	
5	159.60	-1.59057	
6	168.50	-.09099	
7	168.50	-.09099	
8	172.50	.58298	
9	174.50	.91997	
10	150.60	-3.10699	
11	170.40	.22915	
12	161.30	-1.30413	
13	172.50	.58298	
14	170.40	.22915	
15	168.90	-.02359	
16	168.90	-.02359	
17	177.50	1.42544	
18	174.50	.91997	
19	168.60	-.07414	
20	164.80	-.71441	
21	170.40	.22915	
22	168.90	-.02359	
23	164.80	-.71441	
24	167.20	-.31003	
25	167.20	-.31003	

analisis: dari output diatas tampak bahwa

- o jumlah dataa N=25, nilai minimum 150.6 dan maximum 180.3, mean 169.0400, dan std deviasi 5.93500
- o terlihat tampilan kolom baru di sebelah kolom tinggi yaitu ztinggi
SPSS menggunakan tingkat kepercayaan 95 % sehingga jika z berada diantara - 1.96 dan 1.96 maka data memiliki distribusi normal. pada tabel diatas terlihat ada satu data yang termasuk outlier (di luar 1,96) yaitu tinggi 150.6 (z = - 3.10699)

menggunakan descriptive statistics / explore

Descriptives					
GENDER				Statistic	Std. Error
TINGGI	pria	Mean		172.4818	.98855
		95% Confidence Interval for Mean	Lower Bound	170.2792	
			Upper Bound	174.6845	
		5% Trimmed Mean		172.2687	
		Median		172.5000	
		Variance		10.750	
		Std. Deviation		3.27866	
		Minimum		168.50	
		Maximum		180.30	
		Range		11.80	
		Interquartile Range		4.3000	
		Skewness		1.263	
		Kurtosis		2.546	
					1.279
		wanita		Mean	
95% Confidence Interval for Mean	Lower Bound			162.7364	
	Upper Bound			169.9350	
5% Trimmed Mean				166.5897	
Median				167.8500	
Variance				38.861	
Std. Deviation				6.23385	
Minimum				150.60	
Maximum				177.50	
Range				26.90	
Interquartile Range				5.3500	
Skewness				-1.020	
Kurtosis				2.662	
					.597
					1.154

analisis: tampak bahwa pada pria memiliki mean 172.4818, standar deviasi 3.27866, ratio skewness = 1,263/0.661 = 1,91 , ratio kurtosis = 2.546/1,279= 1.99, karena kedua hasil ini tidak melebihi angka 2 maka tinggi badan laki-laki mendekati distribusi normal. analisis yg sama bisa dilakukan pada wanita

M-Estimators

GENDER		Huber's M-Estimator ^a	Tukey's Biweight ^b	Hampel's M-Estimator ^c	Andrews' Wave ^d
TINGGI	pria	171.9653	171.7213	171.8320	171.7011
	wanita	167.3600	167.5792	167.4276	167.5672

- a. The weighting constant is 1.339.
 b. The weighting constant is 4.685.
 c. The weighting constants are 1.700, 3.400, and 8.500.
 d. The weighting constant is $1.340 \cdot \pi$.

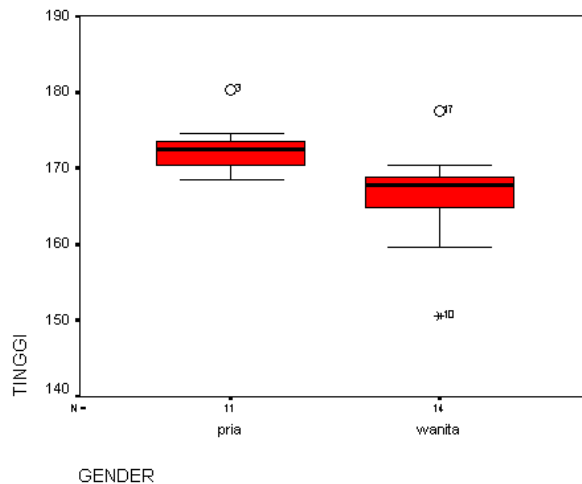
analisis: SPSS memberi bobot terhadap data, sehingga dari output tersebut didapatkan mean pria menurut Huber adalah 171.9653, dsb. M-estimator ini digunakan sebagai alternatif pengukuran pusat, apabila mean normal tidak menggambarkan ukuran pusat daripada data tersebut karena mengandung beberapa data yang menyimpang

Extreme Values

GENDER			Case Number		Value
TINGGI	pria	Highest	1	3	180.30
			2	9	174.50
			3	18	174.50
			4	13	172.50
			5	8	a
	wanita	Lowest	1	7	168.50
			2	22	168.90
			3	1	170.20
			4	21	170.40
			5	4	b
	pria	Highest	1	17	177.50
			2	11	170.40
			3	14	170.40
			4	15	168.90
			5	16	168.90
	wanita	Lowest	1	10	150.60
			2	5	159.60
			3	12	161.30
			4	20	164.80
			5	23	164.80

- a. Only a partial list of cases with the value 173 are shown in the table of upper extremes.
 b. Only a partial list of cases with the value 173 are shown in the table of lower extremes.

analisis: disajikan lima data terbesar dan terkecil dari masing-masing gender



analisis: tampak bahwa garis tengah hitam (median) agak ke atas pada pria maupun wanita, hal ini menunjukkan distribusi adalah normal miring ke kanan. ada tiga data outlier (menyimpang) yaitu data nomor 3 pada pria (180.3) , pada wanita data nomor 17 (177.5) dan 10 (150.6) . Tanda o=outlier dan *=far outside outlier.

Stem-and-Leaf Plots

TINGGI Stem-and-Leaf Plot for
GENDER= pria

```

Frequency      Stem & Leaf

      2.00      16 . 88
      8.00      17 . 00222244
      1.00 Extremes      (>=180)

Stem width:      10.00
Each leaf:      1 case(s)

```

TINGGI Stem-and-Leaf Plot for
GENDER= wanita

```

Frequency      Stem & Leaf

      1.00 Extremes      (<=151)
      1.00      15 . 9
      3.00      16 . 144
      6.00      16 . 778888
      2.00      17 . 00
      1.00 Extremes      (>=178)

Stem width:      10.00
Each leaf:      1 case(s)

```

analisis: pada pria terlihat ada dua pria (frequency=2) yang memiliki tinggi 160 an (steam 16) dan leaf 88 hal ini dapat dibaca sebagai (steam+leaf), yaitu 168 dan 168. ada 8 pria yang memiliki tinggi badan (170,170,172,172,172,172,174,174),dan ada satu pria yang extreme (tinggi >= 180).bandingkan dengan boxplot sebelumnya. analisis yang sama bisa dilakukan pada wanita dimana kita mendapatkan dua data extreme

menguji normalitas data dan varians menggunakan descriptive statistics / explore

Tests of Normality						
GENDER	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TINGGI pria	.225	11	.125	.882	11	.110
wanita	.198	14	.142	.900	14	.113

a. Lilliefors Significance Correction

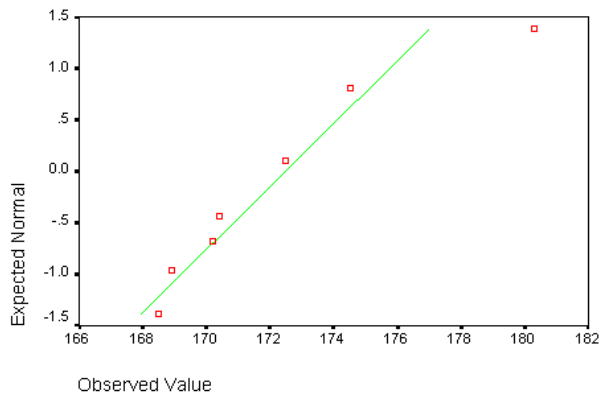
analisis: nilai sig ini disebut sebagai nilai probabilitas, apabila nilainya > 0.05 maka distribusi adalah normal. menggunakan uji kolmogorov-smirnov kita melihat nilai probabilitas pria 0.125 dan wanita 0.142 (> 0.05) sehingga dikatakan distribusi kedua sample adalah normal. menggunakan uji shapiro-wilk kita melihat nilai probabilitas pria 0.110 dan wanita 0.113 (> 0.05) sehingga dikatakan bahwa distribusi kedua sample adalah normal

Test of Homogeneity of Variance					
		Levene Statistic	df1	df2	Sig.
TINGGI	Based on Mean	2.334	1	23	.140
	Based on Median	1.528	1	23	.229
	Based on Median and with adjusted df	1.528	1	17.520	.233
	Based on trimmed mean	2.014	1	23	.169

analisis: nilai sig ini disebut sebagai nilai probabilitas, apabila nilainya > 0.05 maka data berasal dari populasi yang mempunyai varians sama. menggunakan uji levene kita melihat nilai probabilitas 0.140 (> 0.05) sehingga dapat dikatakan bahwa data berasal dari varians yang sama.

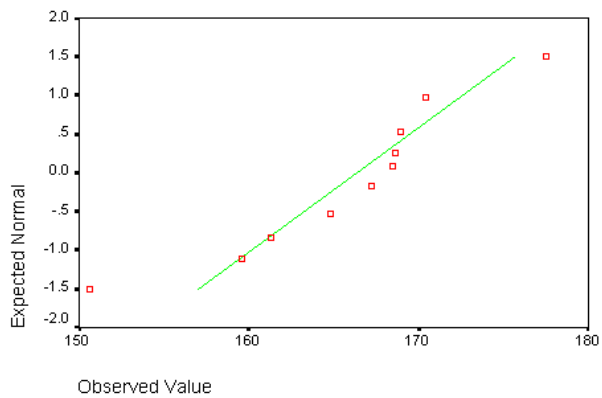
Normal Q-Q Plot of TINGGI

For GENDER= pria



Normal Q-Q Plot of TINGGI

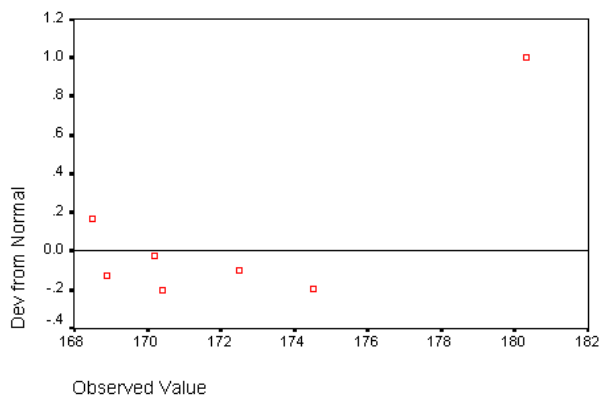
For GENDER= wanita



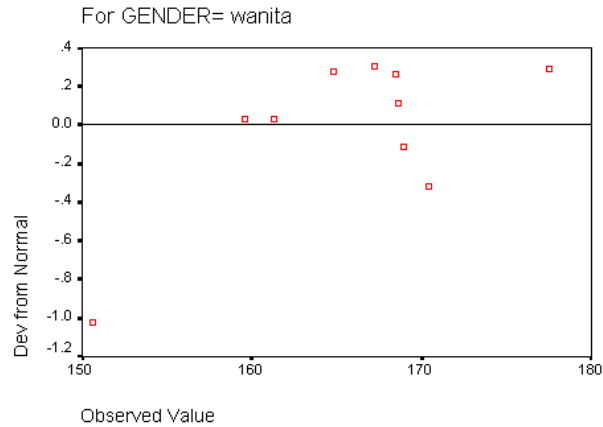
analisis: terlihat garis lurus dari kiri ke kanan, garis tersebut berasal dari nilai z. jika distribusi adalah normal maka data tersebut akan tersebar di sekeliling garis. terlihat bahwa data tersebar di sekeliling garis (kecuali satu pria dan dua wanita yang outlier) sehingga distribusi adalah normal

Detrended Normal Q-Q Plot of TINGGI

For GENDER= pria



Detrended Normal Q-Q Plot of TINGGI



analisis: sama seperti sebelumnya, distribusi adalah normal apabila data tersebar merata di sekitar garis. terlihat bahwa hanya ada satu data pada pria yang di luar garis dan dua data pada wanita yang di luar garis. sehingga distribusi adalah normal.

menggunakan descriptive statistics / crosstab

diketahui data sebagai berikut:

no	kerja	didik	gender
1	karyawan	akademi	pria
2	petani	sarjana	pria
3	wiraswasta	sma	wanita
4	petani	sma	wanita
5	wiraswasta	akademi	wanita
6	karyawan	sarjana	pria
7	wiraswasta	sma	wanita
8	wiraswasta	sma	pria
9	petani	akademi	wanita
10	petani	akademi	wanita
11	karyawan	sarjana	pria
12	karyawan	sarjana	pria
13	petani	sma	wanita
14	wiraswasta	sarjana	pria
15	wiraswasta	akademi	wanita
16	karyawan	sarjana	pria
17	petani	sma	wanita
18	karyawan	akademi	pria
19	karyawan	sma	wanita
20	petani	akademi	pria
21	wiraswasta	sarjana	wanita
22	petani	sarjana	wanita
23	petani	sarjana	pria
24	karyawan	sma	pria
25	karyawan	sma	pria

HUBUNGAN KERJA DENGAN GENDER

GENDER * KERJA Crosstabulation

Count		KERJA			Total
		karyawan	wiraswasta	petani	
GENDER	pria	8	2	3	13
	wanita	1	5	6	12
Total		9	7	9	25

analisis: tampak table silang antara kerja dengan gender. sebagai contoh baris pertama tampak bahwa ada 8 orang pria yang bekerja sebagai karyawan, 2 orang wiraswasta, dan 3 orang sebagai petani.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.702 ^a	2	.021
Likelihood Ratio	8.505	2	.014
Linear-by-Linear Association	5.342	1	.021
N of Valid Cases	25		

a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is 3.36.

analisis: nilai asymp sig disebut juga sebagai nilai probabilitas. apabila nilai probabilitas ini kurang dari 0.05 maka ada hubungan antar variabel tersebut. pada output terlihat nilai probabilitasnya 0.021 (< 0.05) sehingga ada hubungan antara kerja dengan gender.

HUBUNGAN KERJA DENGAN TINGKAT PENDIDIKAN

DIDIK * KERJA Crosstabulation

Count

		KERJA			Total
		karyawan	wiraswasta	petani	
DIDIK sma		3	3	3	9
akademi		2	2	3	7
sarjana		4	2	3	9
Total		9	7	9	25

analisis: tampak tabel silang antara kerja dengan tingkat pendidikan, sebagai contoh baris pertama tampak bahwa ada 4 orang karyawan, 3 orang wiraswasta, dan 3 petani yang berpendidikan sma

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.630 ^a	4	.960
Likelihood Ratio	.625	4	.960
Linear-by-Linear Association	.074	1	.785
N of Valid Cases	25		

a. 9 cells (100.0%) have expected count less than 5. The minimum expected count is 1.96.

analisis: pada output terlihat nilai probabilitasnya 0.960 (> 0.05) sehingga tidak ada hubungan antara kerja dengan tingkat pendidikan

CROSSTAB UNTUK UJI KORELASI DUA VARIABEL BERSKALA NOMINAL

dari analisis sebelumnya kita telah mengetahui bahwa ada hubungan antara kerja dengan gender, tetapi seberapa kuat hubungan tersebut, kita akan melihat output berikut:

Symmetric Measures

		Value	Asymp. Std. Error ^b	Approx. ^a	Approx. Sig.
Nominal by Nominal	Phi	.555			.021
	Cramer's V	.555			.021
	Contingency Coefficient	.485			.021
Interval by Interval	Pearson's R	.472	.167	2.566	.017 ^c
Ordinal by Ordinal	Spearman Correlation	.472	.173	2.566	.017 ^c
N of Valid Cases		25			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

analisis: karena variabel kerja dan gender bertipe nominal maka kita perhatikan korelasi nominal by nominal. ketiga besaran memiliki nilai probabilitas 0.021 (< 0.05) yang berarti ada hubungan antara kerja dan gender. sementara valuenya 0.555 dan 0.485 (semakin dekat ke 1 maka hubungan semakin kuat).

Directional Measures

			Value	Asymp. Std. Error ^b	Approx. ^a	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	.393	.163	2.003	.045
		GENDER Dependent	.500	.236	1.572	.116
		KERJA Dependent	.313	.137	2.041	.041
	Goodman and Kruskal tau	GENDER Dependent	.308	.165		.025 ^c
		KERJA Dependent	.160	.095		.021 ^c
	Uncertainty Coefficient	Symmetric	.191	.114	1.673	.014 ^d
		GENDER Dependent	.246	.147	1.673	.014 ^d
		KERJA Dependent	.156	.093	1.673	.014 ^d

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on chi-square approximation

d. Likelihood ratio chi-square probability.

analisis:

ada tiga korelasi yang dapat dilihat:

1. korelasi lambda
2. korelasi goodman and kruskal tau
3. korelasi uncertainty coefficient

dari ketiga korelasi tersebut kita melihat bahwa angka signifikansinya < 0.05 yang berarti hubungannya signifikan, tetapi kalau dilihat dari angka korelasinya sangat jauh dari 1, sehingga variabel gender tidak bisa memprediksi secara kuat pekerjaan, demikian pula sebaliknya.

CHI SQUARE UNTUK MULTI TABEL

kalau sebelumnya kita menggunakan crosstab untuk dua variabel, sekarang kita akan menggunakan tiga variabel. yaitu kita ingin mencari tahu hubungan antara kerja dengan tingkat pendidikan dengan variabel gender sebagai control.

Chi-Square Tests

GENDER		Value	df	Asymp. Sig. (2-sided)
pria	Pearson Chi-Square	2.063 ^a	4	.724
	Likelihood Ratio	3.035	4	.552
	Linear-by-Linear Association	.357	1	.550
	N of Valid Cases	13		
wanita	Pearson Chi-Square	1.200 ^b	4	.878
	Likelihood Ratio	1.588	4	.811
	Linear-by-Linear Association	.149	1	.699
	N of Valid Cases	12		

a. 9 cells (100.0%) have expected count less than 5. The minimum expected count is .46.

b. 9 cells (100.0%) have expected count less than 5. The minimum expected count is .17.

analisis: dari output terlihat nilai probabilitasnya adalah 0.724 untuk pria dan 0.878 untuk wanita, keduanya > 0.05 yang berarti gender tidak menentukan hubungan antara kerja dengan tingkat pendidikan

CROSSTAB UNTUK UJI KORELASI DUA VARIABLE BERSKALA ORDINAL

menggunakan data sbb:

sikap_km=sikap thd kemasan, sikap_rs=sikap thd rasa					
no	kerja	didik	gender	sikap_km	sikap_rs
1	karyawan	akademi	pria	tidaksuka	suka
2	petani	sarjana	pria	cukupsuka	tidaksuka
3	wiraswaste	sma	wanita	tidaksuka	tidaksuka
4	petani	sma	wanita	tidaksuka	suka
5	wiraswaste	akademi	wanita	tidaksuka	tidaksuka
6	karyawan	sarjana	pria	suka	suka
7	wiraswaste	sma	wanita	cukupsuka	cukupsuka
8	wiraswaste	sma	pria	tidaksuka	tidaksuka
9	petani	akademi	wanita	cukupsuka	suka
10	petani	akademi	wanita	tidaksuka	cukupsuka
11	karyawan	sarjana	pria	suka	cukupsuka
12	karyawan	sarjana	pria	suka	tidaksuka
13	petani	sma	wanita	suka	cukupsuka
14	wiraswaste	sarjana	pria	tidaksuka	suka
15	wiraswaste	akademi	wanita	cukupsuka	cukupsuka
16	karyawan	sarjana	pria	suka	suka
17	petani	sma	wanita	suka	suka
18	karyawan	akademi	pria	tidaksuka	cukupsuka
19	karyawan	sma	wanita	cukupsuka	cukupsuka
20	petani	akademi	pria	suka	cukupsuka
21	wiraswaste	sarjana	wanita	tidaksuka	cukupsuka
22	petani	sarjana	wanita	cukupsuka	tidaksuka
23	petani	sarjana	pria	suka	suka
24	karyawan	sma	pria	suka	suka
25	karyawan	sma	pria	suka	suka

didapatkan output sbb:

SIKAP_KM * SIKAP_RS Crosstabulation

Count		SIKAP_RS			Total
		suka	cukup suka	tidak suka	
SIKAP_KM	suka	5	3	2	10
	cukup suka	1	3	2	6
	tidak suka	3	3	3	9
Total		9	9	7	25

analisis: merupakan tabel silang dari sikap terhadap kemasan dan rasa, sebagai contoh pada baris pertama terdapat 5 orang yang suka terhadap kemasan dan rasa.

Directional Measures

		Value	Asymp. Std. Error ^a	Approx. Φ	Approx. Sig.
Ordinal by Ordinal	Somers' d	.156	.184	.848	.396
	SIKAP_KM Dependent	.155	.182	.848	.396
	SIKAP_RS Dependent	.157	.185	.848	.396

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

analisis: tampak bahwa angka signifikasinya > 0.05 dan angka korelasinya jauh dibawah 1, sehingga tidak ada hubungan antara sikap terhadap rasa dan sikap terhadap kemasan.

Symmetric Measures		Value	Asymp. Std. Error ^a	Approx. χ^2	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.156	.184	.848	.396
	Kendall's tau-c	.154	.181	.848	.396
	Gamma	.232	.269	.848	.396
	Spearman Correlation	.172	.202	.837	.411 ^c
Interval by Interval	Pearson's R	.168	.201	.820	.421 ^c
N of Valid Cases		25			

a. Not assuming the null hypothesis.

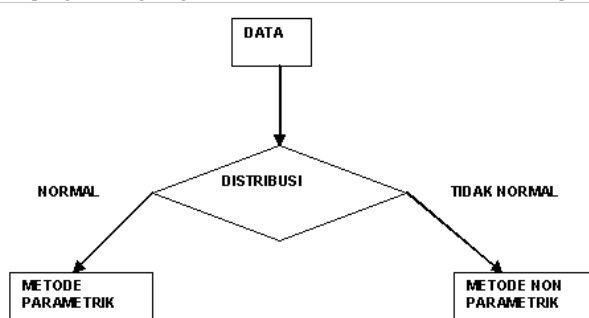
b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

analisis: karena variabel sikap kemasan dan sikap rasa bertipe ordinal maka kita perhatikan korelasi ordinal by ordinal. dari empat besaran tampak bahwa nilai signifikansinya > 0.05 dan angka korelasinya jauh di bawah 1, sehingga tidak ada hubungan antara sikap terhadap kemasan dan sikap terhadap rasa.

STATISTIK INFERENSI

statistik inferensi erat hubungannya dengan pengambilan keputusan. ilmu yang dipakai untuk mengambil keputusan ini menggunakan saudara kandung dari statistik, yaitu probabilitas. dua metode yang sering digunakan dalam statistik inferensi dapat digambarkan sbb:



berikut adalah beberapa menu yang biasa dipakai di SPSS, tetapi sebelumnya kita perlu mengenal konsep **hipotesis**

HIPOTESIS

apabila saya mengatakan bahwa tahun depan penjualan produk A akan meningkat, maka saya akan ditanya atas dasar apa, beruntunglah ada ilmu statistik dan probabilitas yang membuat pernyataan saya bisa diterima secara ilmiah.

hipotesa adalah pernyataan tentang sesuatu yang perlu dibuktikan atau diuji kebenarannya. Setelah diadakan pengujian, baru dapat disimpulkan apakah suatu hipotesa diterima atau ditolak. Pengujian melalui analisa statistik menggunakan teori probabilitas.

Dalam statistik, hipotesa yang akan diuji dilambangkan dengan H_0 dan harus dibandingkan dengan hipotesa alternatif H_1 .

Level of significance

Jika setelah pengujian data sample, kita menerima atau menolak suatu hipotesa, belum tentu kesimpulan kita tersebut benar 100 % mengingat bahwa selalu terjadi kekeliruan sampling atau kesalahan uji hipotesa.

Ada dua jenis kesalahan uji hipotesa yang mungkin terjadi:

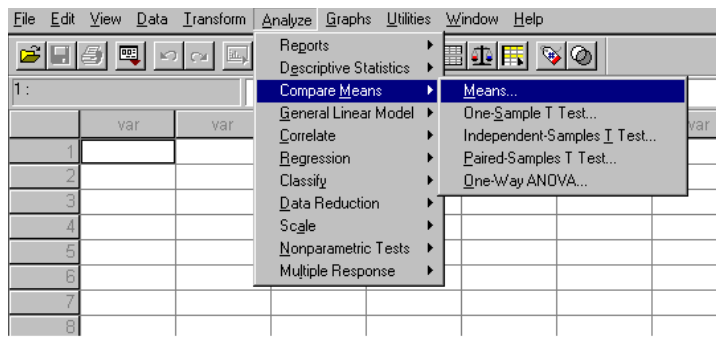
- Kesalahan alpha : kesalahan karena menolak hipotesa H_0 padahal hipotesa benar
- Kesalahan beta : kesalahan karena menerima hipotesa H_0 padahal hipotesa salah

Kesalahan alpha disebut **level of significance** dari kesalahan. Jika harga $\alpha = 5\%$ berarti bahwa tingkat kepercayaan kita = 95 %. Dalam statistik biasanya kita mengambil **tingkat kepercayaan** 95 % yang berarti kita percaya bahwa kesimpulan yang kita ambil 95 % benar, jadi kemungkinan kita membuat kesalahan menolak hipotesa yang seharusnya diterima besarnya hanya 5 %.

Dalam uji hipotesa ada beberapa faktor yang perlu kita perhatikan:

- menggunakan uji z jika distribusi normal (sample ≥ 30) dan uji t jika distribusi tidak normal (sample < 30)
- tingkat signifikan (α) yang digunakan
- jenis pengujian yang dilakukan (dua sisi, satu sisi: kanan atau kiri)

METODE PARAMETRIK



menggunakan compare means / means

diketahui data sbb:

tinggi	berat	gender	tinggal
174.5	65.8	pria	desa
178.6	62.7	pria	kampung
170.8	66.4	pria	kota
168.2	68.9	pria	kampung
159.7	67.8	pria	kampung
167.8	67.8	pria	kota
165.5	65.8	pria	kota
154.7	48.7	wanita	desa
152.7	45.7	wanita	desa
155.8	46.2	wanita	desa
154.8	43.8	wanita	kampung
157.8	58.1	wanita	kota
156.7	54.7	wanita	kota
154.7	49.7	wanita	kota

dengan means kita mendapatkan output sbb:

Report

GENDER	TINGGAL		TINGGI	BERAT
pria	desa	Mean	174.5000	65.8000
		N	1	1
		Std. Deviation	.	.
	kampung	Mean	168.8333	66.4667
		N	3	3
		Std. Deviation	9.46590	3.30807
	kota	Mean	168.0333	66.6667
		N	3	3
		Std. Deviation	2.65769	1.02632
wanita	desa	Mean	154.4000	46.8667
		N	3	3
		Std. Deviation	1.57162	1.60728
	kampung	Mean	154.8000	43.8000
		N	1	1
		Std. Deviation	.	.
	kota	Mean	156.4000	54.1667
		N	3	3
		Std. Deviation	1.57162	4.22532
Total	desa	Mean	159.4250	51.6000
		N	4	4
		Std. Deviation	10.13159	9.55720
	kampung	Mean	165.3250	60.8000
		N	4	4
		Std. Deviation	10.43883	11.65075
	kota	Mean	162.2167	60.4167
		N	6	6
		Std. Deviation	6.66436	7.37819
Total	Total	Mean	162.3071	58.0071
		N	14	14
		Std. Deviation	8.44270	9.54209

analisis: kita dapatkan tabel silang dimana data dipecah menjadi pria yang bertempat tinggal di desa, kampung, dan kota serta wanita yang bertempat tinggal di desa, kampung, dan kota.
menggunakan compare means / one sample T Test

berikut data berat badan sebelum dan sesudah minum obat:

sebelum	sesudah
76.85	76.22
77.95	77.89
78.65	79.02
79.25	80.21
82.65	82.65
88.15	82.53
92.54	92.56
96.25	92.33
84.56	85.12
88.25	84.56

menggunakan one sample t test didapat output sbb:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
SEBELUM	10	84.5100	6.63931	2.09953

One-Sample Test

	Test Value = 90					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SEBELUM	-2.615	9	.028	-5.4900	-10.2395	-.7405

analisis: tampak bahwa nilai probabilitas 0.028 (< 0.05) yang berarti berat kelompok memang berbeda dengan berat rata-rata populasi sebelum minum obat
menggunakan compare means / independent samples T Test

tinggi	berat	gender
174.5	65.8	pria
178.6	62.7	pria
170.8	66.4	pria
168.2	68.9	pria
159.7	67.8	pria
167.8	67.8	pria
165.5	65.8	pria
154.7	48.7	wanita
152.7	45.7	wanita
155.8	46.2	wanita
154.8	43.8	wanita
157.8	58.1	wanita
156.7	54.7	wanita
154.7	49.7	wanita

menggunakan independent sample t test didapat output sbb:

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
TINGGI	pria	7	169.3000	6.13514	2.31887
	wanita	7	155.3143	1.64259	.62084
BERAT	pria	7	66.4571	2.02308	.76465
	wanita	7	49.5571	5.15553	1.94861

Independent Samples Test

		TINGGI		BERAT	
		Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	5.475		4.345	
	Sig.	.037		.059	
t-test for Equality of Means	t	5.826	5.826	8.074	8.074
	df	12	6.856	12	7.805
	Sig. (2-tailed)	.000	.001	.000	.000
	Mean Difference	13.9857	13.9857	16.9000	16.9000
	Std. Error Difference	2.40054	2.40054	2.09327	2.09327
	95% Confidence Interval of the Difference	Lower Upper	8.75539 19.21604	12.33916 21.46084	12.05185 21.74815

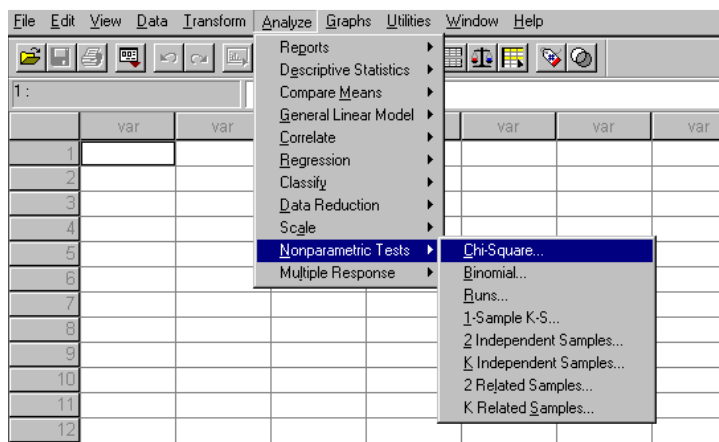
analisis:

- diuji F test dahulu (Lavene test) apakah varians sama atau tidak, apabila nilai probabilitas < 0.05 maka varians adalah berbeda. pada kasus tinggi terlihat angka 0.037 (< 0.05) yang berarti kedua varians tinggi pria maupun wanita adalah berbeda, sedang pada kasus berat 0.059 (> 0.05) yang berarti kedua varians berat pria maupun wanita adalah sama.
- jika varians memang berbeda maka dilakukan uji t untuk membandingkan means dengan asumsi varians tidak sama. pada kasus tinggi terlihat probabilitas 0.01 (< 0.05) yang berarti kedua mean tinggi pria mapun wanita adalah berbeda.

menggunakan compare means / Paired Samples T Test

menggunakan compare means / One Way ANOVA

METODE NON-PARAMETRIK



menggunakan nonparametrik test / chi square

menggunakan nonparametrik test / binomial

menggunakan nonparametrik test / runs

menggunakan nonparametrik test / 1 sample K-S

menggunakan nonparametrik test / 2 independent samples

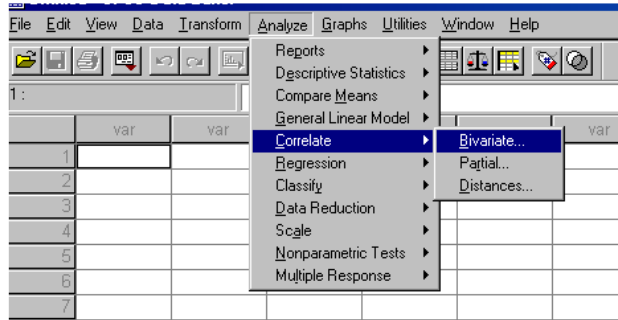
menggunakan nonparametrik test / K independent samples

menggunakan nonparametrik test / 2 Related samples

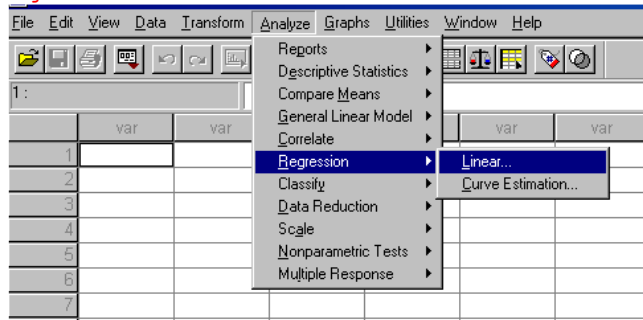
menggunakan nonparametrik test / K Related samples

selain menu diatas SPSS juga memiliki menu tersendiri yaitu:

- korelasi



- regresi



REFERENSI

- Santoso, Singgih, SPSS Mengolah Data Secara Profesional, Elex Media Komputindo, Jakarta, 2000
- Kuswadi, Mutiara, Erna, Statistik berbasis komputer untuk orang-orang nonstatistik, Elex Media Komputindo, Jakarta, 2004