Teori dan Aplikasi Statistika

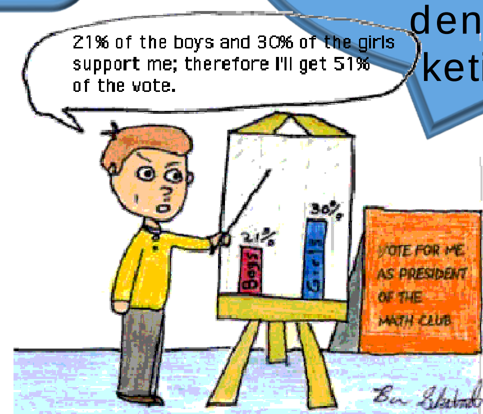
Dr.Ir. Supratomo

Sepanjang hidup, manusia menerka-nerka jawaban berdasarkan informasi yang tidak lengkap.

Banyak diantara manusia yang hidup nyaman dengan tingkat ketidakpastian tertentu .....

Keunikan statistika adalah kemampuan untuk menghitung ketidakpastian dengan tepat.

Dengan kemampuan itu, ahli statistika dapat membuat pernyataan yang tegas dan lengkap dengan jaminan ketidakpastian.



Larry Gonick and Woolicott Smith. 1993. The Cartoon Guide to Statistics.

Pokok Bahasan

- Dr.Ir. Supratomo
  - Pendahuluan
  - Statistika deskriptif
  - Statistika inferensia
- Prof.Dr.Ir. Kahar Mustari, MS
  - Perancangan percobaan
- Dr.Ir. Beta Putranto, M.Sc.
  - Regresi
  - Manova
  - Pengukuran berulang
- Prof.Dr.Ir. Didi Rukmana, M.Sc
  - Statistika non-parametric
  - Structural Equation Model (SEM)

Definition

- The word ' Statistics' and ' Statistical' are all derived from the Latin word *Status*, means a political state.
- **Statistics** is the art of learning from data.
- **Statistics** is concerned with
  - the collection of data,
  - their description, and
  - their analysis, which often leads to the drawing of conclusions.
- **Formal Definition of Statistics**
  - The science of *collecting, organizing, presenting, analyzing*, and *interpreting* data to assist in making more effective decisions.

Definition

- A **parameter** is a numerical description of a *population* characteristic.
- A **statistic** is a numerical description of a *sample* characteristic.

It is important to note that a sample statistic can differ from sample to sample whereas a population parameter is constant for a population.

Distinguishing Between a Parameter and a Statistic

Determine whether the numerical value describes a population parameter or a sample statistic. Explain your reasoning.

1. A recent survey of approximately 400,000 employers reported that the average starting salary for marketing majors is Rp 2,500,000,-.
2. The freshman class at a university has an average SAT math score of 514.
3. In a random check of 400 retail stores, the Balai POM found that 34 % of the stores were not storing fish at the proper temperature.

Solution

1. Because the average of Rp 2,500,000 is based on a subset of the population, it is a sample statistic.
2. Because the average SAT math score of 514 is based on the entire freshman class, it is a population parameter.
3. Because the percent, 34%, is based on a subset of the population, it is a sample statistic.

Peranan Statistika

1. **Deskripsi** : menggambarkan atau menjelaskan data
2. **Komparasi** : membandingkan data dari dua kelompok atau lebih
3. **Korelasi** : mencari kekuatan hubungan antar data
4. **Regresi** : memprediksikan pengaruh data yang satu terhadap yang lain
5. **Komunikasi** : merupakan alat penghubung antar pihak berupa laporan data atau analisa statistik

Data

TIPE DATA

- **Kualitatif**
  - **Data diskrit/kategorik** :
    - Nominal
    - Ordinal
    - Binomial

- **Kuantitatif**
  - **Data diskrit :** data dimana nilainya tidak dijumpai dalam bilangan pecahan.
  - **Data kontinu** : data yang nilainya pada selang tertentu dapat mengambil sembarang nilai.

Data

SKALA DATA

- ❖ **Nominal** : variabel pada tingkat yang paling sederhana, dimana angka, simbol atau lambang digunakan untuk mengklasifikasikan suatu objek, orang atau sifat.
- ❖ **Ordinal** : variabel yang dikelompokan menurut jenjang/ derajatnya.
- ❖ **Interval** : variabel yang mempunyai sifat-sifat nominal dan ordinal dan jarak antara dua angka dalam skala itu diketahui ukurannya.
- ❖ **Rasio** : variabel yg. mempunyai sifat² interval dan juga mempunyai nilai **<u>nol mutlak</u>** sebagai titik acuan..

Analisis Statistika

- Analis deskriptif (*Descriptive analysis*)
- Analis induktif (*Inferential analysis*)
- Analis perbedaan (*Differences analysis*)
- Analis hubungan (*Associative analysis*)
- Analis prediksi (*Predictive analysis*)

Metode Statistika

- **Statistika deskriptif** : Metode dan prosedur statistika yang digunakan untuk mengumpulkan, menyajikan dan menganalisa data dalam bentuk narasi, tabulasi atau diagram dan perhitungan numerik dari data sampel tanpa perlu adanya pembuktian terhadap populasinya.
- **Statistika induktif (inferensial)** : Metode dan prosedur statistika yang berhubungan dengan penarikan kesimpulan tentang keadaan populasi berdasarkan data sampel yang diambil, dengan menggunakan metode tertentu.

Deskriptif vs Inferensia

- Organize
- Summarize
- Simplify
- Presentation of data

Describing data

- Generalize from samples to populations
- Hypothesis testing

Make desicion

Statistika Deskriptif

Frequency Distributions

- Table
- Histogram
- Ogive
- Steam and Leaf

Graphical Representations

- Graphs
- Tables

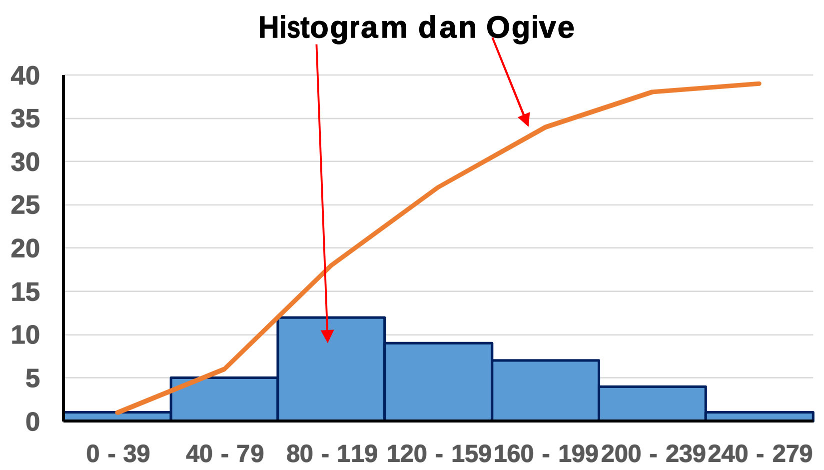
Summary Statistics

- Central tendency
- Variability

Frequency Distributions

Table Frequency distribution, Histogram and Ogive

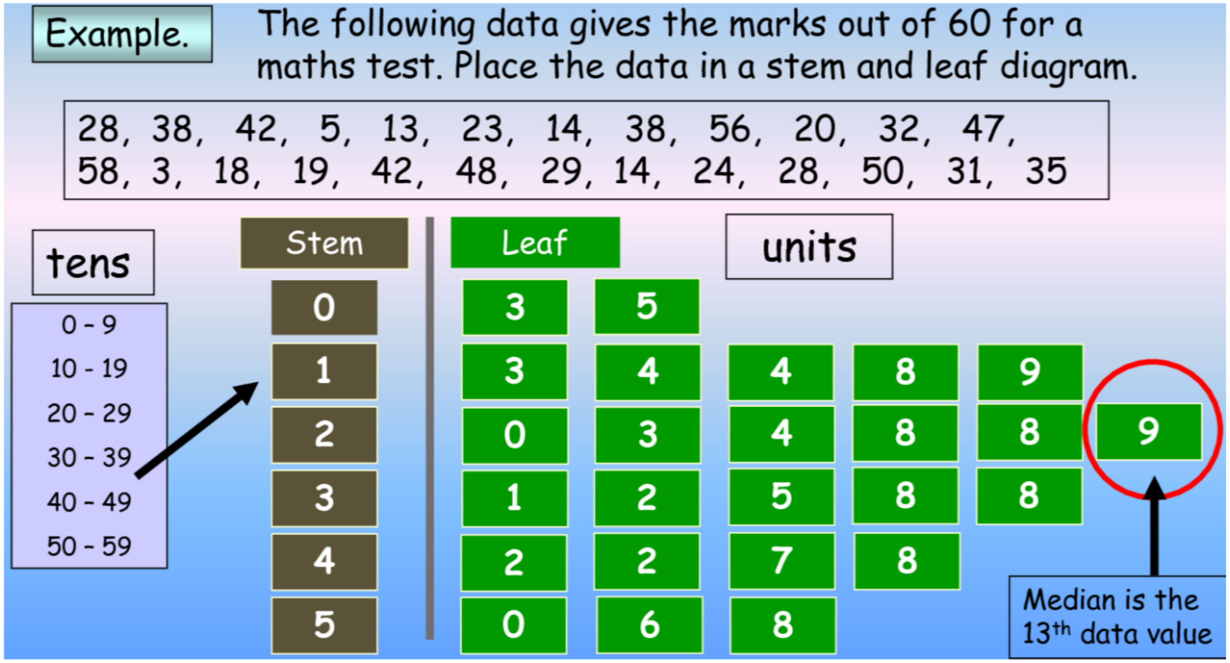
Class interval	Tally	Frequency	Cumulative frequency
0 - 39	|	1	1
40 - 79	~~||||~~	5	6
80 - 119	~~||||~~ ~~||||~~ ||	12	18
120 - 159	~~||||~~ ||||	9	27
160 - 199	~~||||~~ ||	7	34
200 - 239	||||	4	38
240 - 279	|	1	39
Total		39	



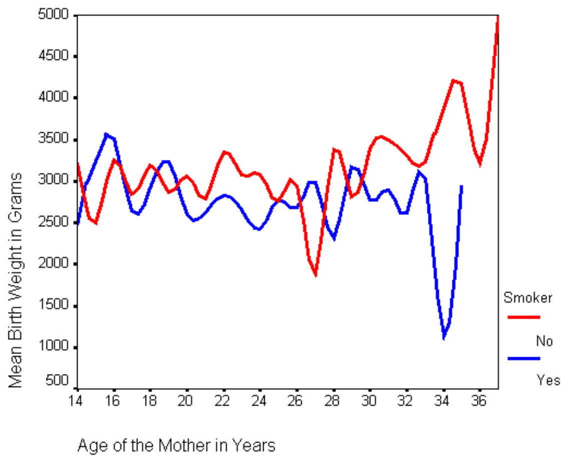
Frequency Distributions

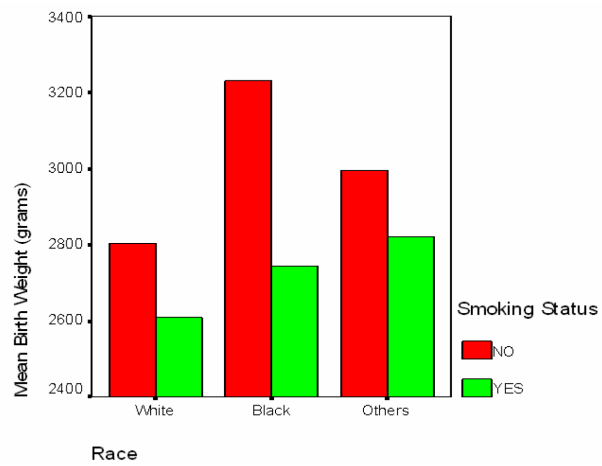
Steam and Leaf Plot

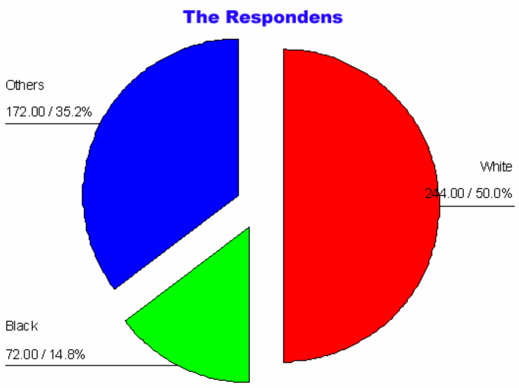
A steam and leaf plot is a frequency diagram in which the raw data is displayed together with its frequency. The data is then placed in suitable groups.

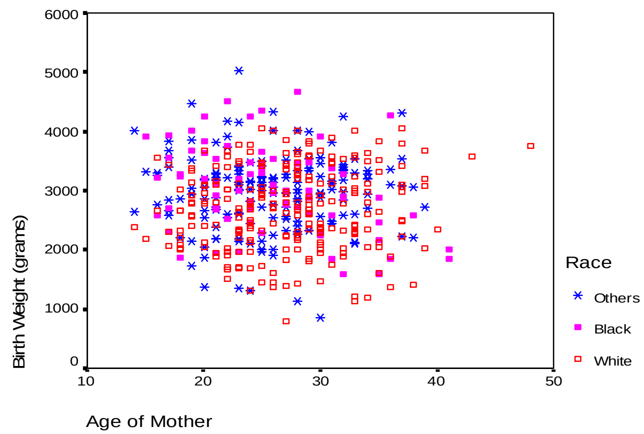


Graphical Representations





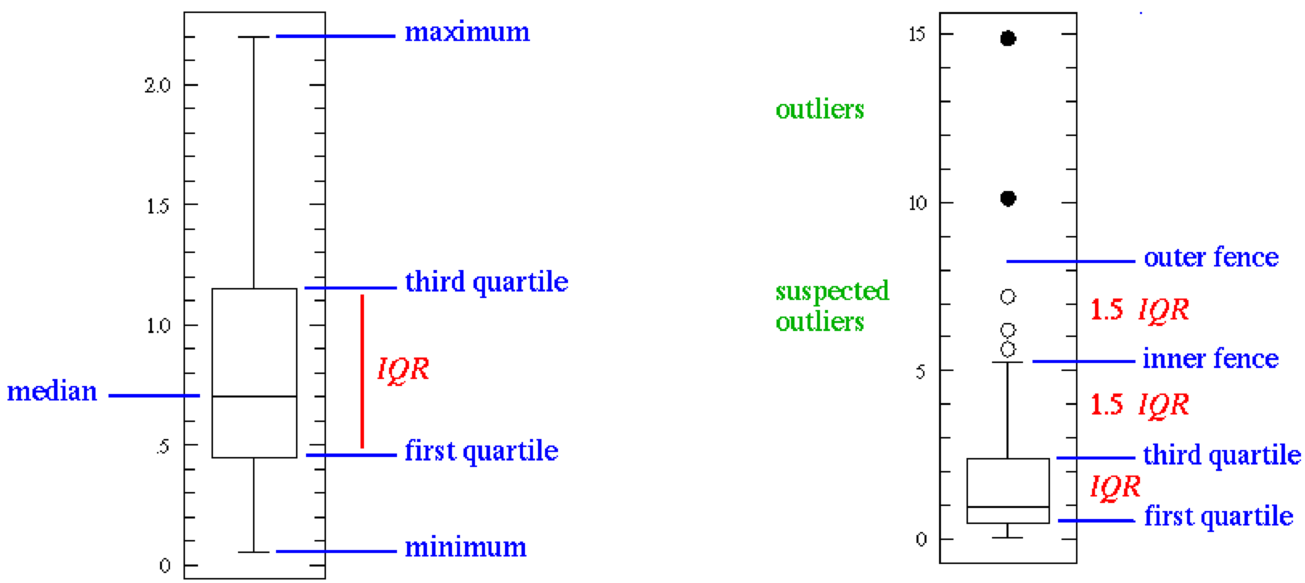




Graphical Representations

Box-whisker plots

A Box-and-Whisker Plot shows the distribution of a set of data along a number line, dividing the data into four parts using the median and quartiles.



IQR : Inter Quartile Range

Summary Statistics

Central Tendency

- Mean
- Median
- Modus

Variabilty

- Range
- Interquartile range
- Variance
- Standard deviation
- Coefficient of variation

Summary Statistics → Central tendency

Mean

- The mean of a data set is the sum of the data entries divided by the number of entries.
- Means can be badly affected by outliers. They can make the mean a bad measure of central tendency.

Median

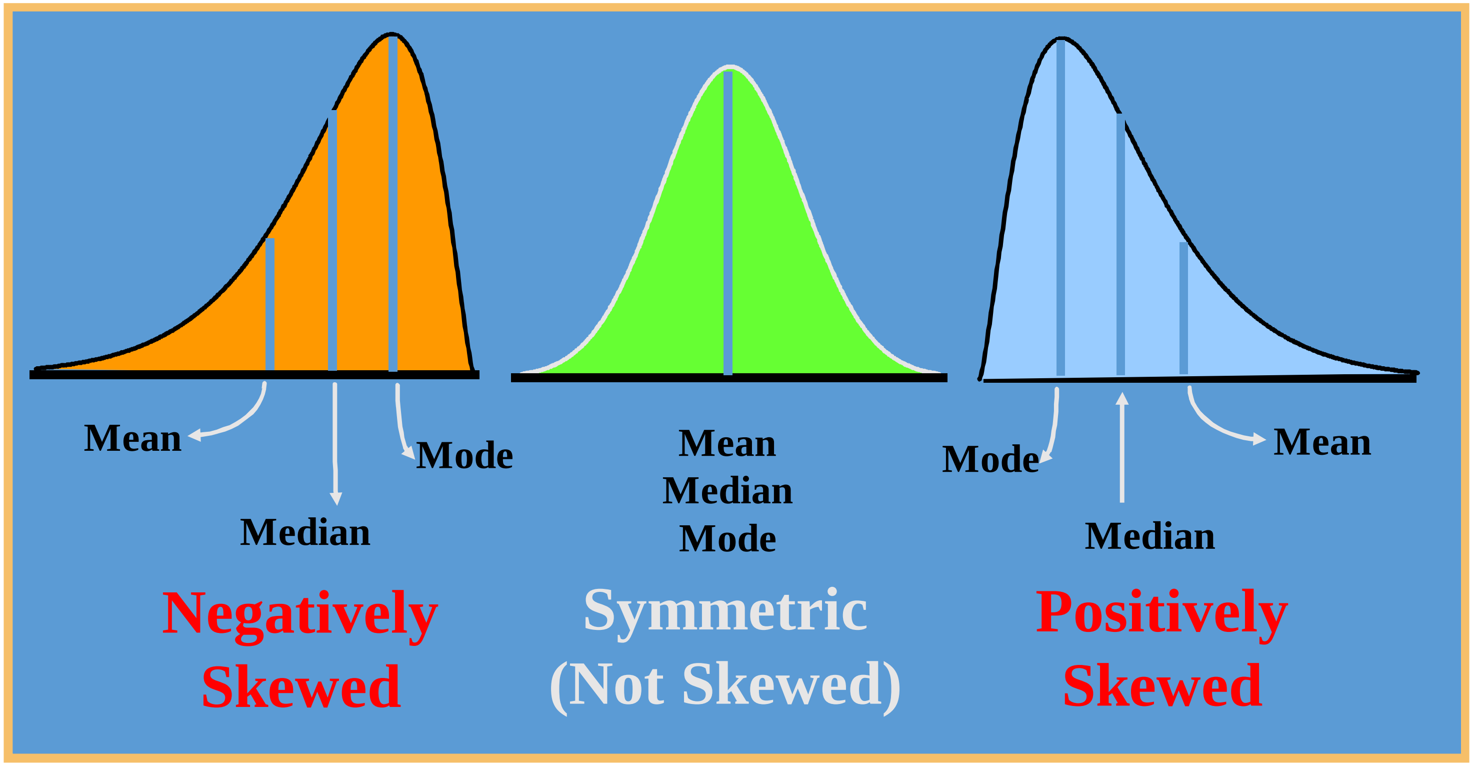
- The median of a data set is the value that lies in the middle of the data when the data set is ordered.
- The median is unaffected by outliers, making it a better measure of central tendency, better describing the "typical person" than the mean when data are skewed.

Mode

- The mode of a data set is the data entry that occurs with the greatest frequency.

Summary Statistics → Central tendency

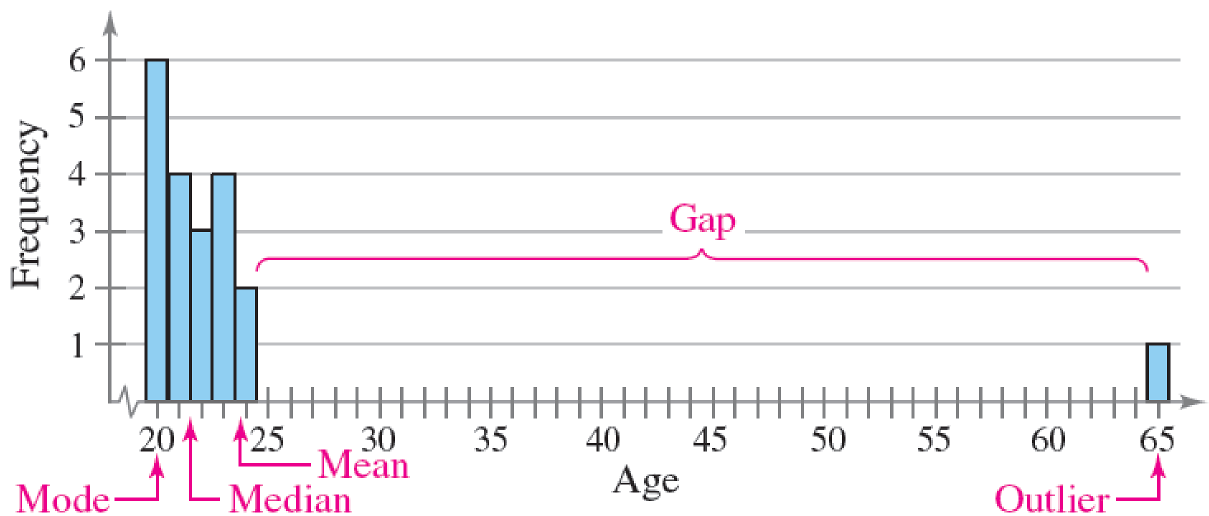
Mean, Median and Mode



Outliner

- An **outlier** is a data entry that is far removed from the other entries in the data set.
- While some outliers are valid data, other outliers may occur due to data-recording errors.
- A data set can have one or more outliers, causing **gaps** in a distribution.
- Conclusions that are drawn from a data set that contains outliers may be flawed

Ages in a class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	



Summary Statistics → Variablity

Range	• The spread, or the distance, between the lowest and highest values of a variable. • Highly sensitive to outliers, insensitive to shape.
Interquartile Range	• A quartile is the value that marks one of the divisions that breaks a series of values into four equal parts.
Standard Deviation	• A measure of the dispersion of a set of data from its mean.
Variance	• A measure of the spread of the recorded values on a variable. • Square of standard deviation.
Coefficient of Variance	• Describes the standard deviation as a percent of the mean. • For comparing the degree of variation from one data series to another, even if the means are drastically different from each other.

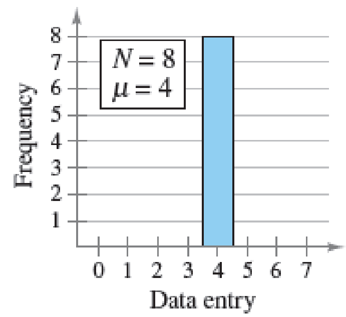
Summary Statistics → Variablity

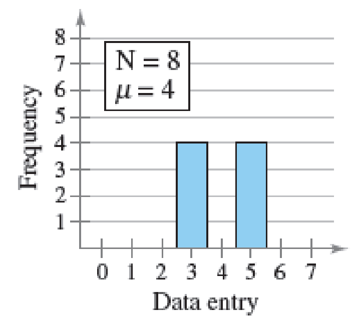
Standard deviation

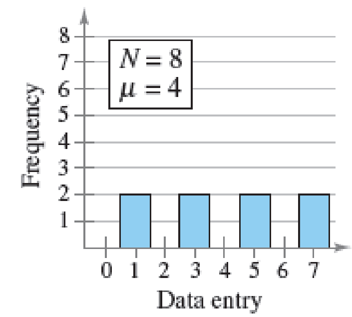
Population $\sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{N}}$ **Sample** $s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}$

INTERPRETING STANDARD DEVIATION

Estimating Standard Deviation : Without calculating, estimate the population standard deviation of each data set.





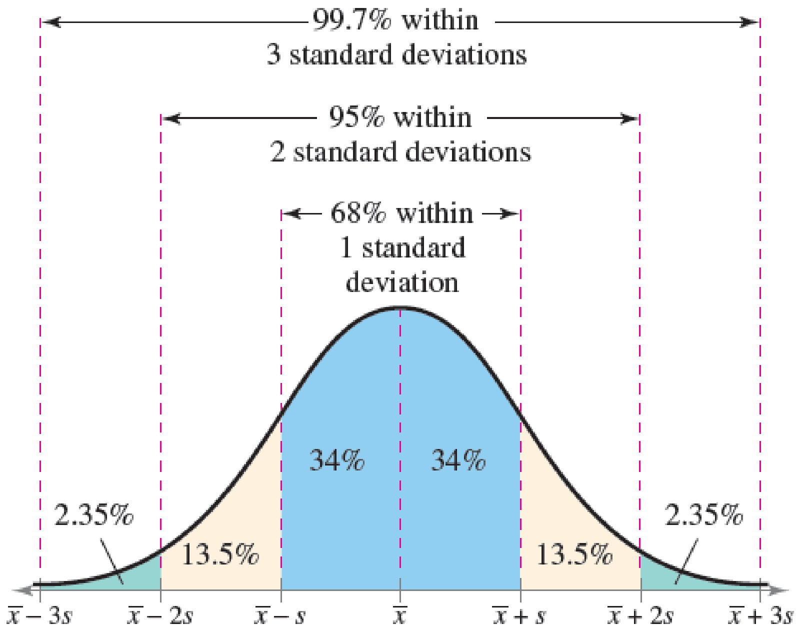


1. Each of the eight entries is 4. The deviation of each entry is 0, so $\sigma = 0$.
2. Each of the eight entries has a deviation of ±1. So, the population standard deviation should be 1. By calculating, you can see that $\sigma = 1$.
3. Each of the eight entries has a deviation of ±1 or ±3. So, the population standard deviation should be about 2. By calculating, you can see that σ is greater than 2, with $\sigma \approx 2.2$.

Summary Statistics → Variablity

Standard deviation

Bell-Shaped Distribution



EMPIRICAL RULE (OR 6 8 – 9 5 – 9 9 . 7 RULE)

For data sets with distributions that are approximately symmetric and bell-shaped, the standard deviation has these characteristics.

1. About 68% of the data lie within one standard deviation of the mean.
2. About 95% of the data lie within two standard deviations of the mean.
3. About 99.7% of the data lie within three standard deviations of the mean.

Summary Statistics → Variablity

Chebychev's Theorem

The portion of any data set lying within *k* standard deviations ($k > 1$) of the mean is at least $1 - \frac{1}{k^2}$

- $k = 2$: In any data set, at least $1 - \frac{1}{2^2} = \frac{3}{4}$ or 75%, of the data lie within 2 standard deviations of the mean.
- $k = 3$: In any data set, at least $1 - \frac{1}{3^2} = \frac{8}{9}$ or 88.9%, of the data lie within 3 standard deviations of the mean.

Summary Statistics → Variablity

Coefficient of Variation

Population $$CV = \frac{\sigma}{\mu} \times 100\%$$

Sample $$CV = \frac{s}{\bar{x}} \times 100\%$$

Heights	Weights
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

Comparing Variation in Different Data Sets

The mean height is μ = 72.8 " with a standard deviation of σ = 3.3 ".
The coefficient of variation for the heights is :

$$CV = \frac{\sigma}{\mu} \times 100\% = \frac{3.3}{72.8} x100\% = 4.5\%$$

The mean weight is μ = 187.8 lbs with a standard deviation of σ = 17.7 lbs.
The coefficient of variation for the weights is :

$$CV = \frac{\sigma}{\mu} \times 100\% = \frac{17.7}{187.8} x100\% = 9.4\%$$

Skewness

Pearson's Index of Skewness. The English statistician Karl Pearson (1857–1936) introduced a formula for the skewness of a distribution.

$$\alpha_3 = \frac{\dfrac{\sum (x_i - \overline{x})^3}{N}}{\left(\sqrt{\dfrac{(x_i - \overline{x})^2}{N}}\right)^3}$$

- Most distributions have an index of skewness between -3 and 3.
- When $\alpha_3 > 0$, the data are skewed right.
- When $\alpha_3 < 0$, the data are skewed left.
- When $\alpha_3 = 0$, the data are symmetric.

Kurtosis

Momen koefisien kurtosis (α_4) menunjukkan derajat keruncingan (*peakedness*) suatu kurva distribusi data.

$$\alpha_4 = \frac{n\sum(x_i - \bar{x})^4}{\left(\sum(x_i - \bar{x})^2\right)^2}$$

Jika $\alpha_4 = 3$, puncak kurva normal (mesokurtic)

Jika $\alpha_4 < 3$, puncak kurva landai (platykurtic)

Jika $\alpha_4 > 3$, puncak kurva runcing (leptokurtic)

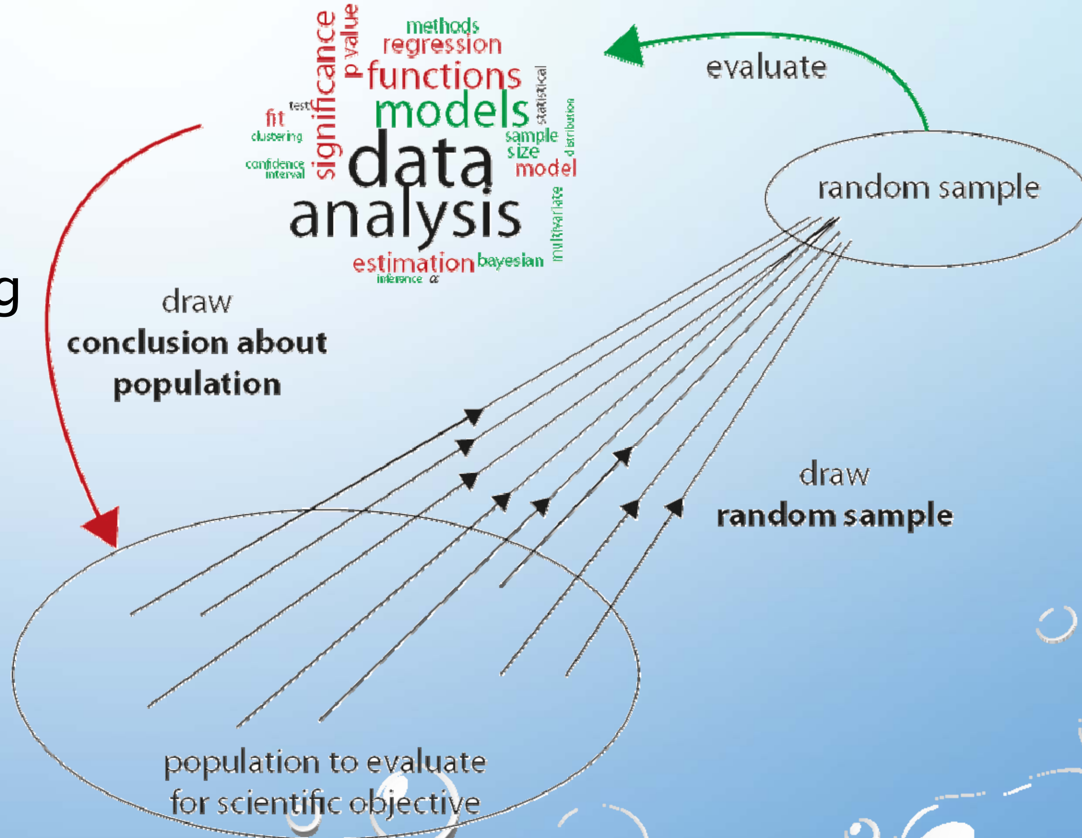
supratomodosen@yahoo.com

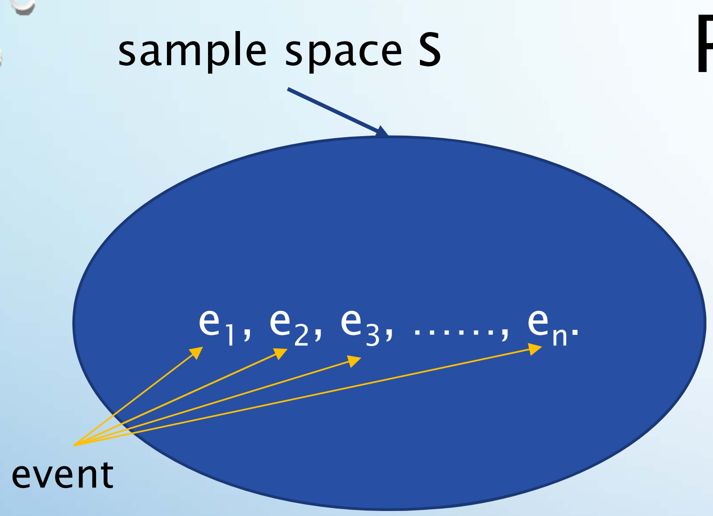
STATISTIKA INFERENSIA

Dr. Supratomo

STATISTICAL INFERENCE (INDUCTIVE)

The theory, methods, and practice of forming judgments about the parameters of a population, usually on the basis of random sampling.





PROBABILITY

$p(e_i) \rightarrow$ the ***probability*** of e_i such that:

1. $0 \leq p(e_i) \leq 1$
2. $p(S) = 1$
3. If e_1, e_2, e_3,…are *pairwise mutually exclusive* events in S then :

$$P(e_1 \cup e_2 \cup e_3 \ldots) = \sum_{i=1}^{\infty} P(e_i) = 1$$

PROBABILITY DISTRIBUTION

Random variable : a quantity that may take any of a range of values, either continuous or discrete, which cannot be predicted with certainty but only described probabilistically

Probability distribution : a distribution of all possible values of a random variable together with an indication of their probabilities.

TWO TYPES:

- **Discrete random variables distributions** are represented by a finite (or countable) number of values

$$P(X=X) = P(X)$$

- **Continuous random variables distributions** are be represented by a real-valued interval

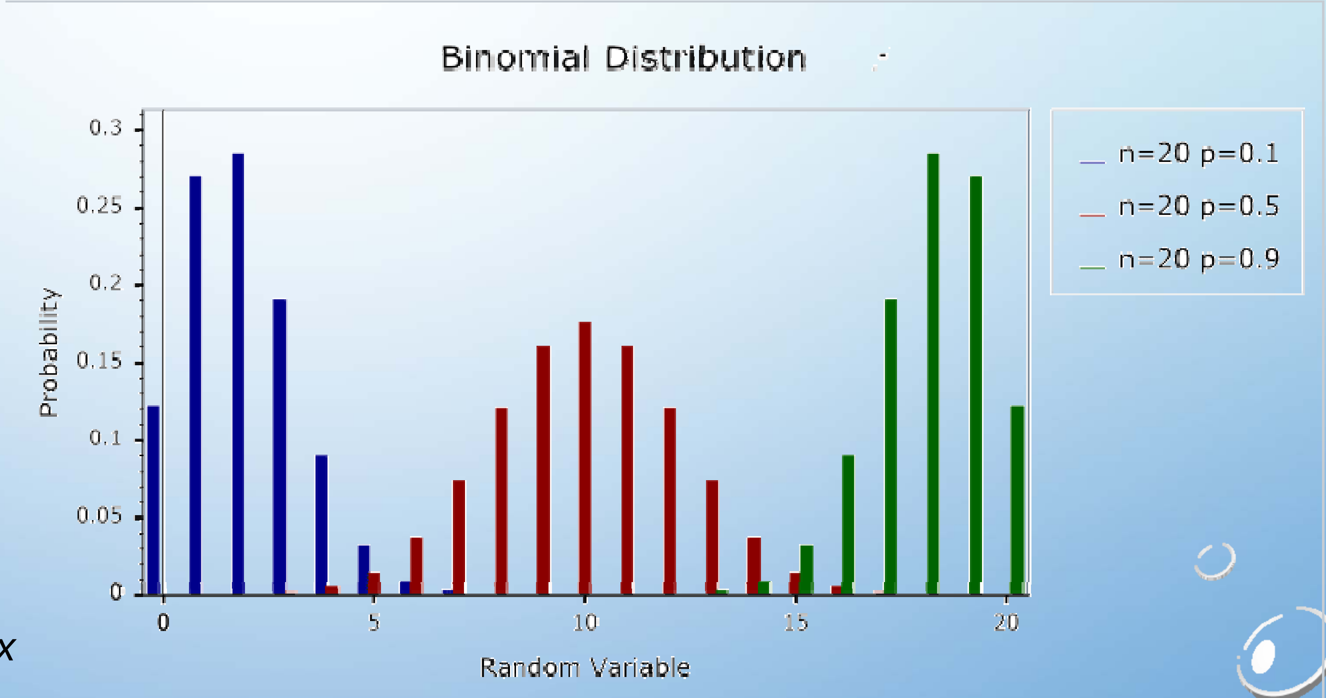
$$P(X_1<X<X_2) = F(X_2) - F(X_1)$$

DISCRETE DISTRIBUTION – *EXAMPLE*

BINOMIAL:

- Experiment consists of *n* identical trials,
- Each trial has only 2 outcomes: success (S) or failure (F),
- P(S) = *p* for a single trial; P(F) = 1−*p* = *q*,
- Trials are *independent*,
- The number of successes in *n* trials →

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$



CONTINOUS DISTRIBUTION – *EXAMPLE*

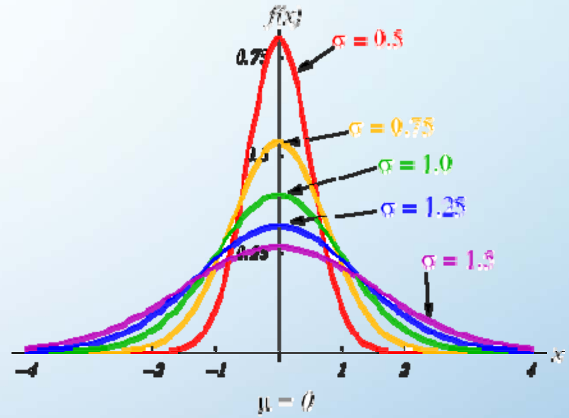
NORMAL (GAUSSIAN):

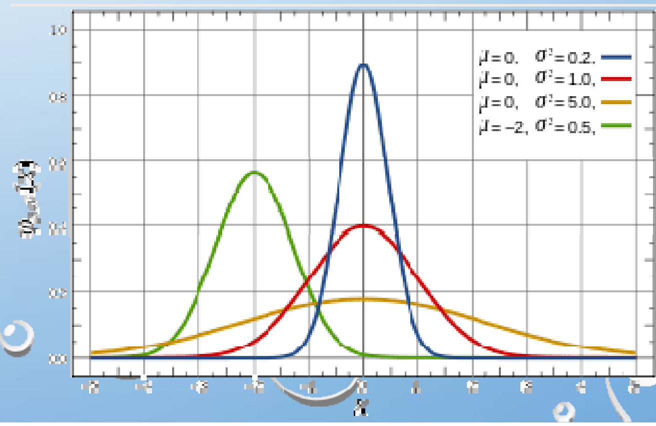
- The normal distribution is defined by its probability density function, which is given as

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty$$

For parameters μ and σ, where $\sigma > 0$.

$X \sim N(\mu, \sigma^2)$, $E(X) = \mu$ and $V(X) = \sigma^2$

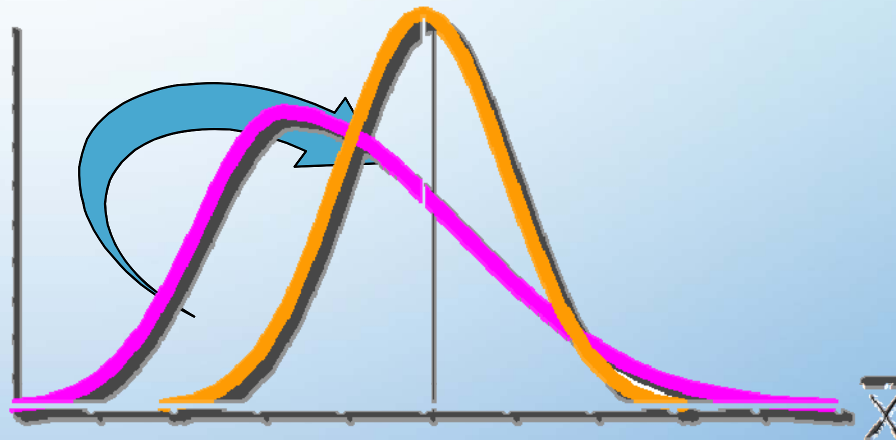




CENTRAL LIMIT THEORY

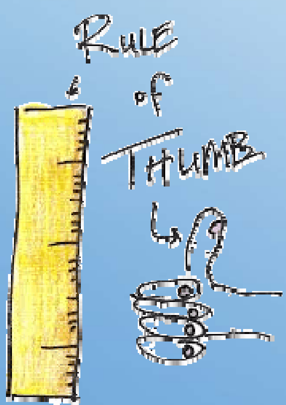
The sample mean will be approximately normally distributed for large sample sizes, regardless of the distribution from which we are sampling.

As sample size gets large enough, sampling distribution becomes almost normal, regardless of shape of population.



Central Tendency : $\mu_{\bar{x}} = \mu$

Variation : $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

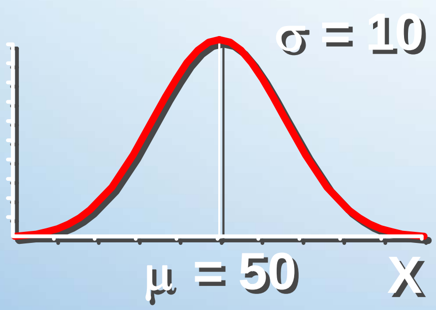


Rule of thumb: normal approximation for $\overline{X}$ will be good if $n \geq 30$. If $n < 30$, the approximation is only good if the population from which you are sampling is not too different from normal.

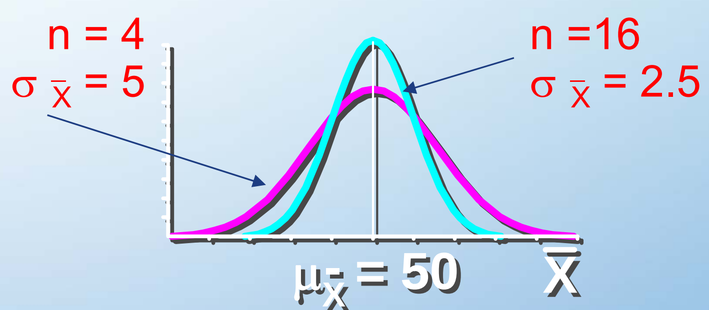
CENTRAL LIMIT THEORY

When the population is normal then the sampling distribution is also normal

Population distribution



Sampling distributions



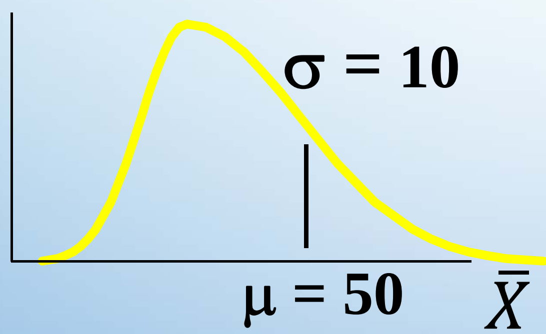
Central Tendency : $\mu_{\bar{x}} = \mu$

Variation : $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

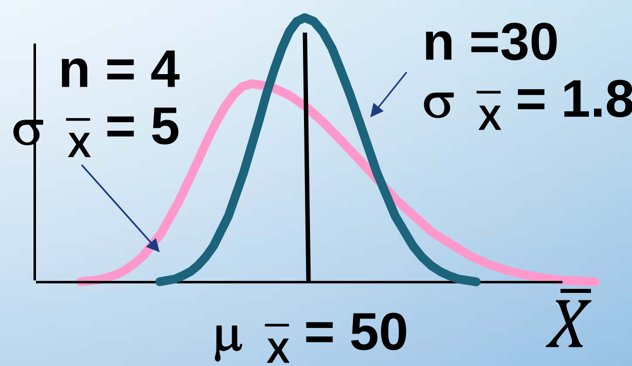
CENTRAL LIMIT THEORY

When the Population is not Normal

Population Distribution



Sampling Distributions



Central Tendency : $\mu_{\bar{x}} = \mu$

Variation : $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$\mu_{\bar{X}} = 50$ $\bar{X}$

HYPOTHESIS TEST

A tentative explanation for an observation, phenomenon, or scientific problem that can be tested by further investigation (*The American Heritage Dictionary of the English Language)*.

- About population parameters
- Null is status quo or no effect

Hypothesis Test

Terms	Steps
Null hypothesis	State H_o
Alternative hypothesis	State H_A
Type I error	Choose α
Type II error	Choose N
Statistical power	Choose test
One or two tailed test	Set-up critical value
Critical region	Collect data
Critical value	Compute test statistic
Level of significance	Make statistical decision
p-value	Express decision

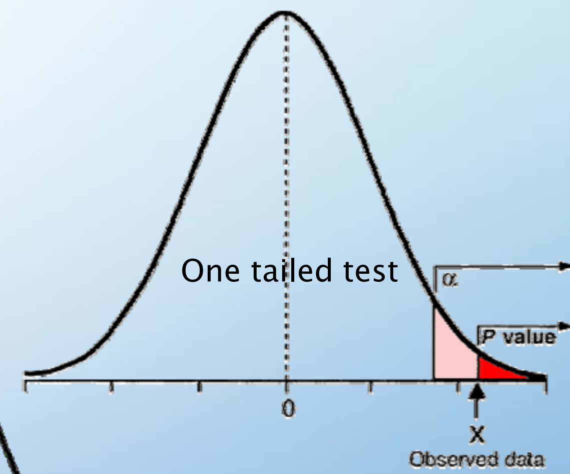
STATISTICAL DECISION

		Actual Situation	
		H_o TRUE	H_o FALSE
Statistical Decision	Do Not reject H_o	TRUE	Type II error β
	Reject H_o	Type I error α	TRUE

- A Type I error occurs when we reject a true null hypothesis.
- A Type II error occurs when we don't reject a false null hypothesis.

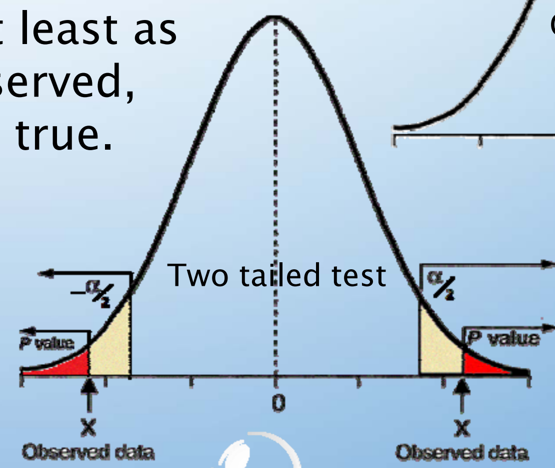
STATISTICAL SIGNIFICANCE

- α *(level of significance)* → the probability of incorrectly rejecting a given null hypothesis (H_0) in favor of an alternative hypothesis (H_A).
- *p*–value (*<u>observed</u> level of significance*) → the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis (H_0) is true.



If the *p*–value is as small or smaller than the level of significance α, → the data are statistically significant at level α.

The smaller the *p–value*, the more statistical evidence exists to support the alternative hypothesis.



Statistically Significant

does not mean

Scientically Significant

The test statistics will tell you if the collected data confirm or deny your expectation (hypothesis)

Statistical Test Choise

Data

Propose of Test	Nominal	Ordinal	Scale (interval/ratio)
Single Sample Test	Chi-square test of goodness of fit	Kolmogorov-Smirnov test	One-sample t test
Compare Two Independent Samples Test	Chi-square test of independence	Mann-Whitney U test	Independent two-sample t test
Compare Two Related Samples Test	McNecmar test	Wilcoxon pairs or Signed-ranks test	Paired-samples t test
Compare > 2 Independent Samples Test		Kruskall-Willis test	One-way or Factorial ANOVA
Compare > 2 Related Samples Test		Friedman test	Repeated measures ANOVA
Relationships		Spearman's coefficient of correlation	Pearson's coefficient of correlation

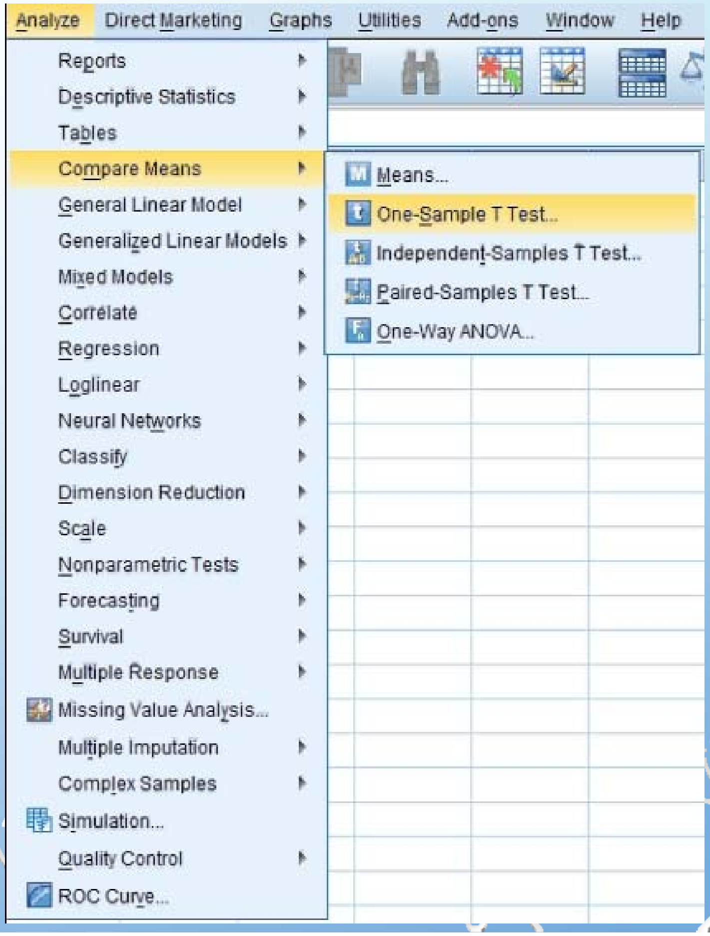
@Dr.Ir. Supratomo

Statistical Non-parametric

COMPARING THE MEANS

One-Sample T Test

1. The one-sample *t*-test is used for comparing sample results with a known value.
2. The purpose is to determine whether there is sufficient evidence to conclude that the mean of the population from which the sample is taken is different from the specified value.
3. The assumption is that the population of observation is normally distributed



One-Sample T Test : Example

SPSS example :

brakes.sav : This is a hypothetical data file that concerns quality control at a factory that produces disc brakes for high-performance automobiles. The data file contains diameter measurements of 16 discs from each of 8 production machines. The target diameter for the brakes is 322 millimeters.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Disc Brake Diameter (mm)	128	322.002009	.0108224	.0009566

One-Sample Test

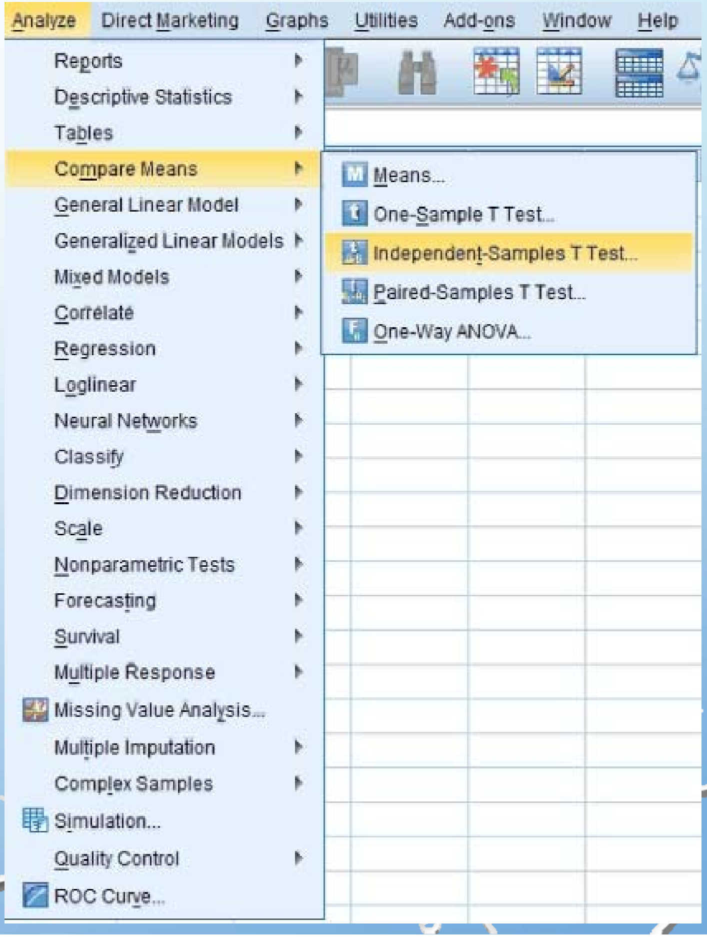
	Test Value = 322					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Disc Brake Diameter (mm)	2.101	127	.038	.0020094	.000116	.003902

H_o rejected

COMPARING THE MEANS

Two-Independent Samples T Test

- To determine whether the unknown means of two populations are different from each other based on independent samples from each population.
- Sample sizes may or may not be equal.
- The assumtions :
  - the two samples are independent (i.e., unrelated to each other).
  - each of the two populations of observations is normally distributed.
  - the two populations of observations are equally variable ($\sigma_1^2 = \sigma_2^2$).



Two-Independent Samples T Test : Example

SPSS example :

creditpromo.sav: This is a hypothetical data file that concerns a department store's efforts to evaluate the effectiveness of a recent credit card promotion. To this end, 500 cardholders were randomly selected. Half received an ad promoting a reduced interest rate on purchases made over the next three months. Half received a standard seasonal ad.

Group Statistics

	Type of mail insert received	N	Mean	Std. Deviation	Std. Error Mean
$ spent during promotional period	Standard	250	1566.3890	346.67305	21.92553
	New Promotion	250	1637.5000	356.70317	22.55989

Independent Samples Test

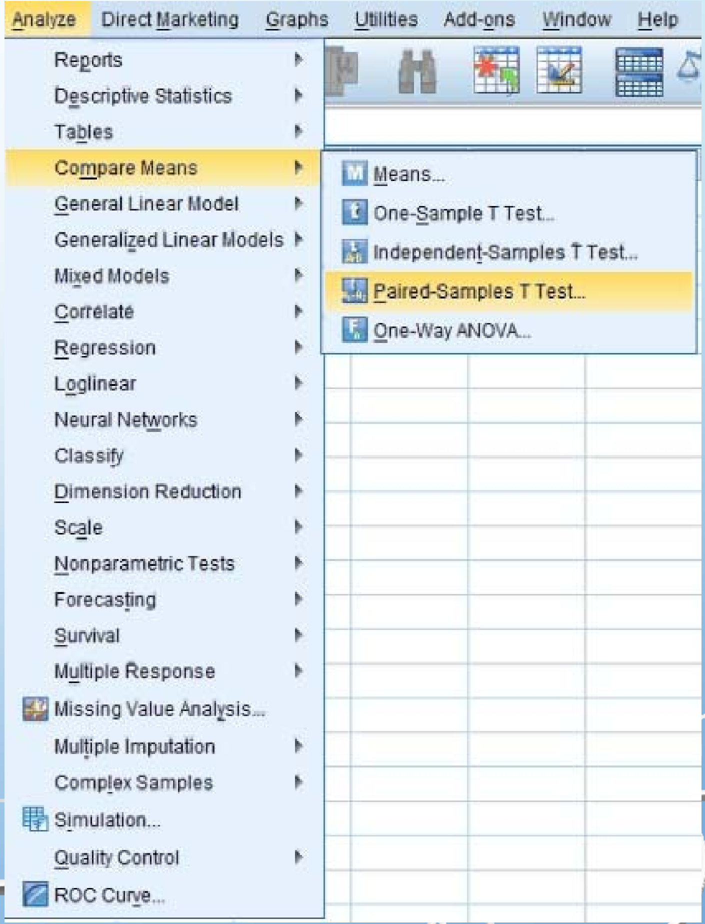
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
$ spent during promotional period	Equal variances assumed	1.190	.276	-2.260	498	.024	-71.11095	31.45914	-132.920	-9.3020
	Equal variances not assumed			-2.260	497.595	.024	-71.11095	31.45914	-132.920	-9.3018

H_o rejected

COMPARING THE MEANS

Paired Samples T Test

- The paired *t*–test is appropriate for data in which the two samples are paired in some way.
  - Pairs consist of before and after measurements on a single group of subjects.
  - Two measurements on the same subject or entity are paired.
- The assumption is that the population of observations is normally distributed.
- Sample sizes must be equal.



Paired Samples T Test : Example

SPSS example :

dietstudy.sav: This hypothetical data file contains the results of a study of the "Stillman diet" (Rickman, Mitchell, Dingman, and Dalen, 1974). Each case corresponds to a separate subject and records his or her pre- and post-diet weights in pounds and triglyceride levels in mg/100 ml.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Triglyceride	138.44	16	29.040	7.260
	Final triglyceride	124.38	16	29.412	7.353
Pair 2	Weight	198.38	16	33.472	8.368
	Final weight	190.31	16	33.508	8.377

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Triglyceride & Final triglyceride	16	-.286	.283
Pair 2	Weight & Final weight	16	.996	.000

Paired Samples Test

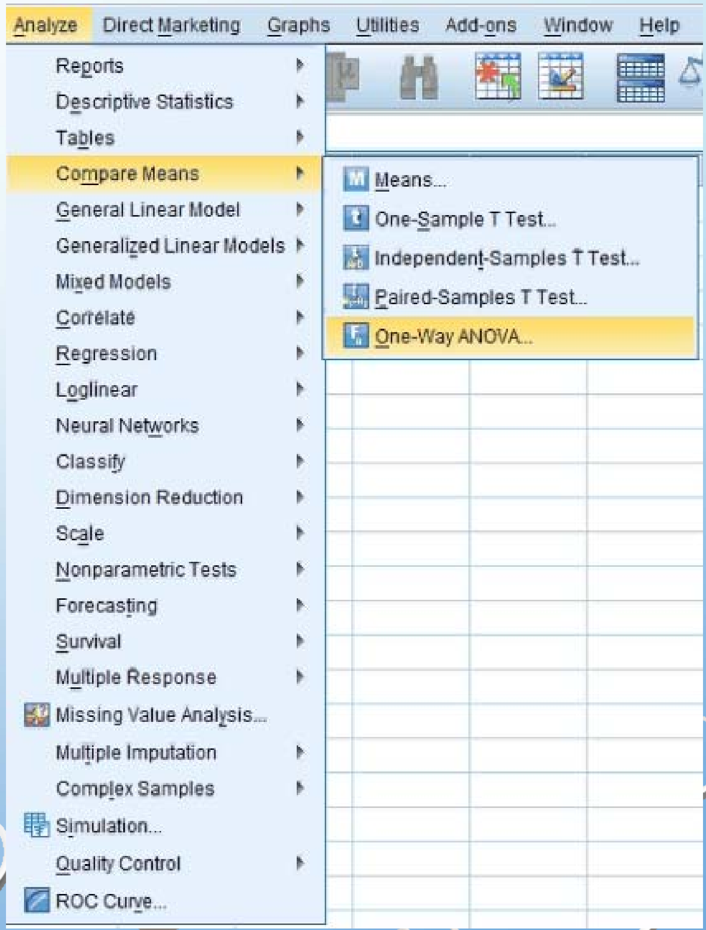
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	Triglyceride - Final triglyceride	14.063	46.875	11.719	-10.915	39.040	1.200	15	.249
Pair 2	Weight - Final weight	8.063	2.886	.722	6.525	9.600	11.175	15	.000

H_o accepted

COMPARING THE MEANS

One-Way ANOVA

- **One-Way Analysis of Variance (ANOVA)** is an inferential method that is used to test the equality of three or more population means.
- **Requirements of a One-Way ANOVA Test**
  1. There are *k* simple random samples from *k* populations.
  2. The *k* samples are independent of each other → the subjects in one group cannot be related in any way to subjects in a second group.
  3. The populations are normally distributed.
  4. The populations have the same variance (homogeneity of variance) → each treatment group has population variance σ^2.



POST HOC COMPARISONS

Post hoc range tests can determine which means differ, once determined that differences exist among the means.

- Tests that Assume Homogeneity of Variance :
  - The most frequently used tests :
  - **Fisher's Least Squared Difference** (LSD) **test** is a nice option if it is exactly 3 treatments.
  - **Bonferroni test** based on Student's t statistic.
  - **Tukey's Honestly Significant Difference** (HSD) **test** uses the Studentized range (q distribution) statistic, and was designed for a situation with equal sample sizes per group, but can be adapted to unequal sample sizes as well.
    - For a large number of pairs of means, Tukey's honestly significant difference test is more powerful than the Bonferroni test.
    - For a small number of pairs, Bonferroni is more powerful.
- Tests that Do Not Assume Homogeneity of Variance :
  - **Dunnett's C test** is the best known that the correct for the problem of not having homogeneity of variance.

One-Way ANOVA Test : Example

- SPSS example :
  dvdplayer.sav : This is a hypothetical data file that concerns the development of a new DVD player. Using a prototype, the marketing team has collected focus group data. Each case corresponds to a separate surveyed user and records some demographic information about them and their responses to questions about the prototype.

ANOVA

Total DVD assessment

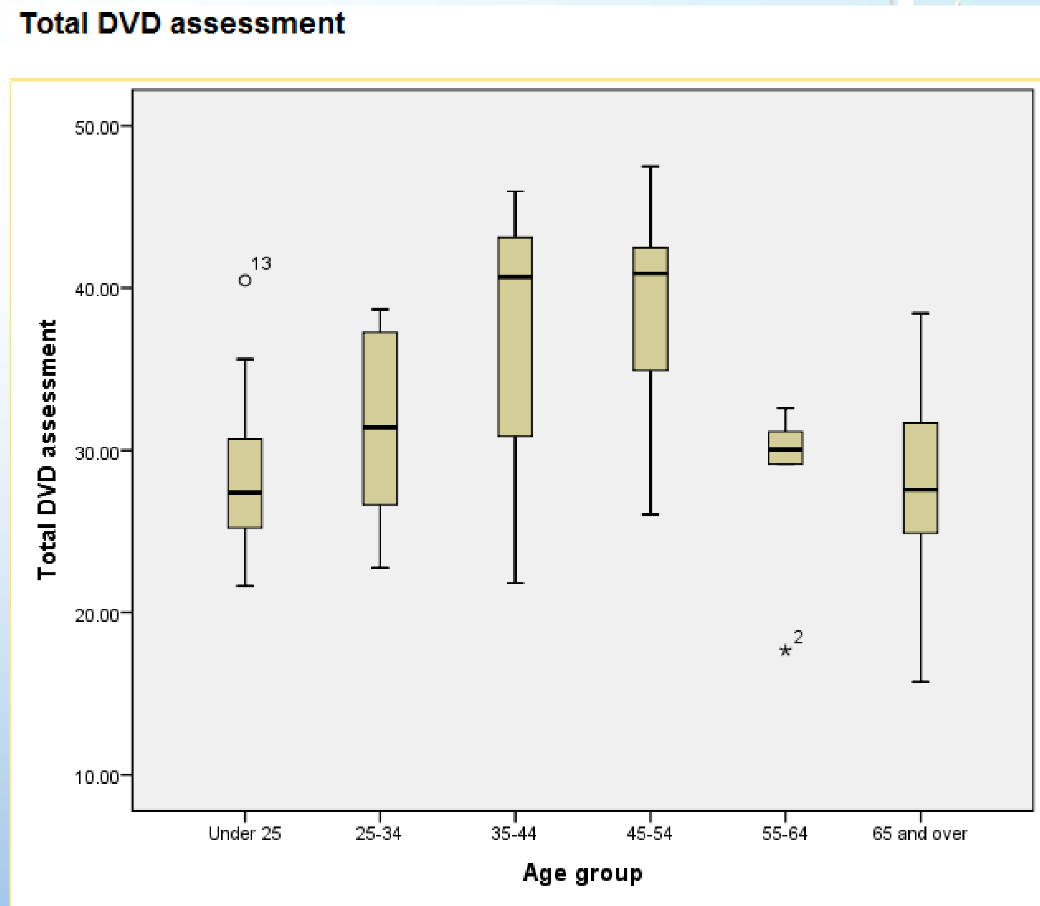
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1294.481	5	258.896	6.993	.000
Within Groups	2295.532	62	37.025		
Total	3590.013	67			

Test of Homogeneity of Variances

Total DVD assessment

Levene Statistic	df1	df2	Sig.
1.292	5	62	.279

accept the null hypothesis that the group variances are equal.



POST HOC COMPARISONS

EXAMPLE : TUKEY HSD

Homogeneous Subsets

Total DVD assessment

Tukey HSD[a,b]

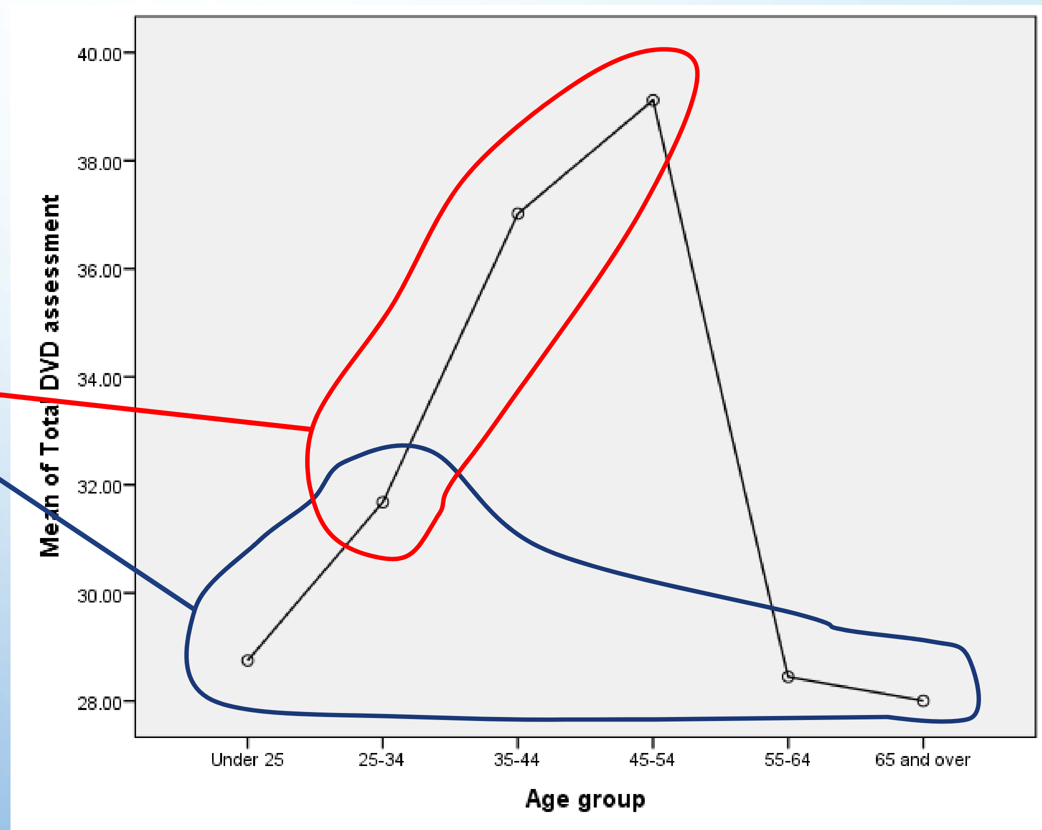
Age group	N	Subset for alpha = 0.05	
		1	2
65 and over	17	28.0028	
55-64	6	28.4473	
Under 25	13	28.7492	
25-34	10	31.6780	31.6780
35-44	12		37.0181
45-54	10		39.1184
Sig.		.746	.077

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.243.

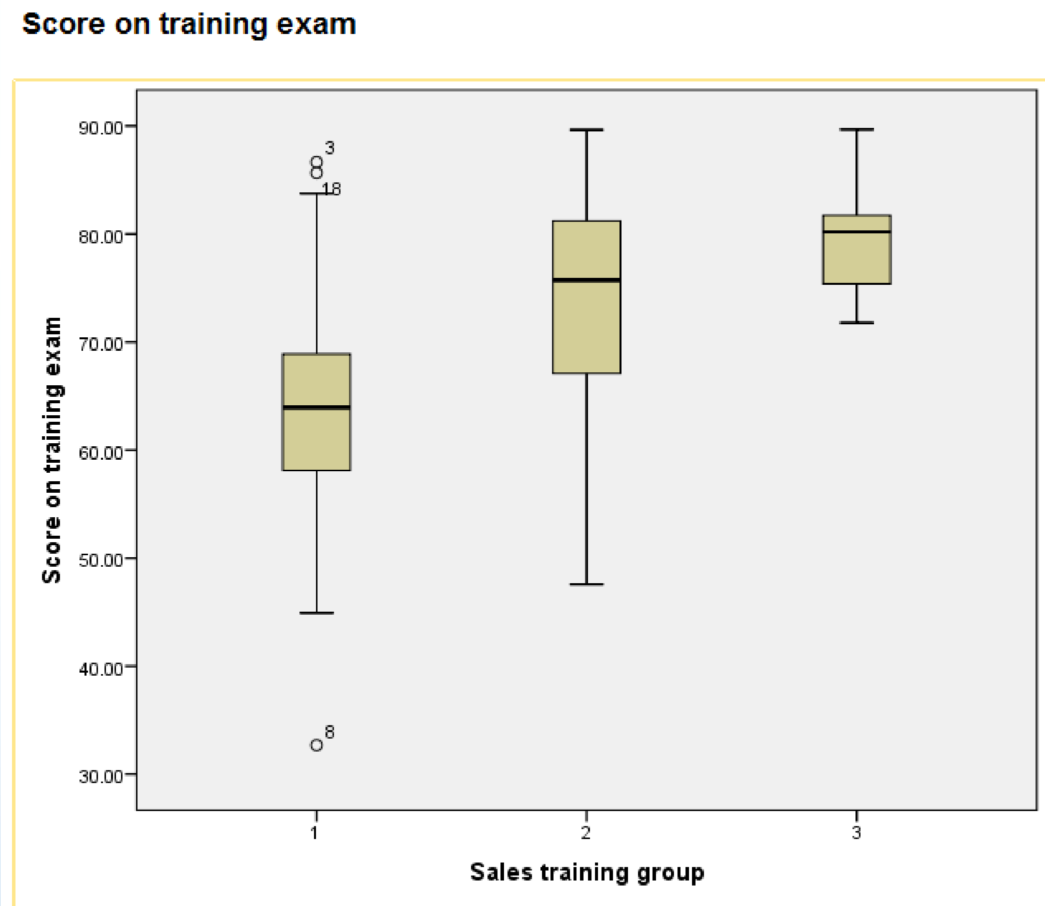
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Means Plots



One-Way ANOVA Test : Example

SPSS example : *salesperformance.sav:* This is a hypothetical data file that concerns the evaluation of two new sales training courses. Sixty employees, divided into three groups, all receive standard training. In addition, group 2 gets technical training; group 3, a hands-on tutorial. Each employee was tested at the end of the training course and their score recorded. Each case in the data file represents a separate trainee and records the group to which they were assigned and the score they received on the exam.



Test of Homogeneity of Variances

Score on training exam

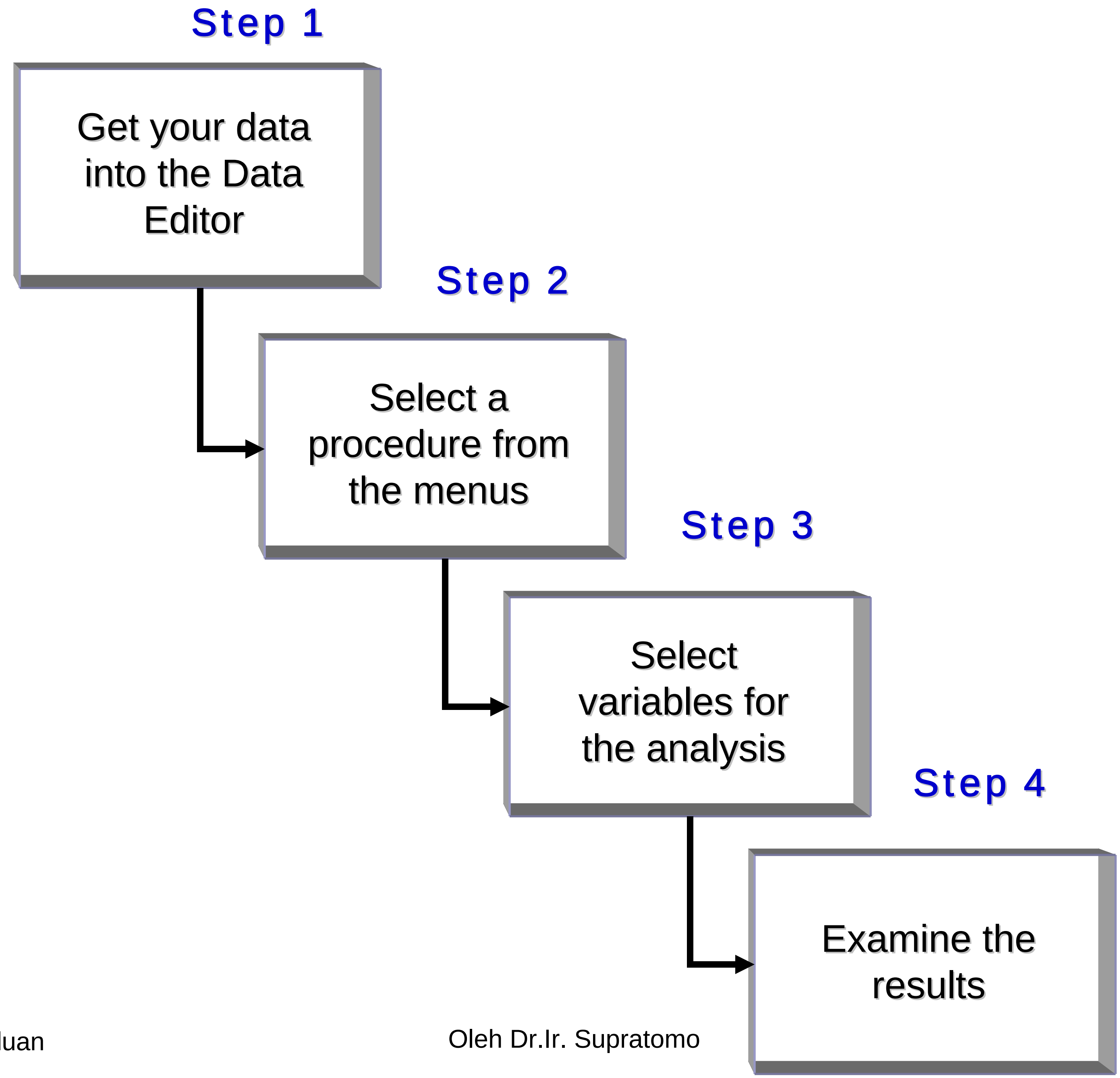
Levene Statistic	df1	df2	Sig.
4.637	2	57	.014

rejects H_o that the group variances are equal.

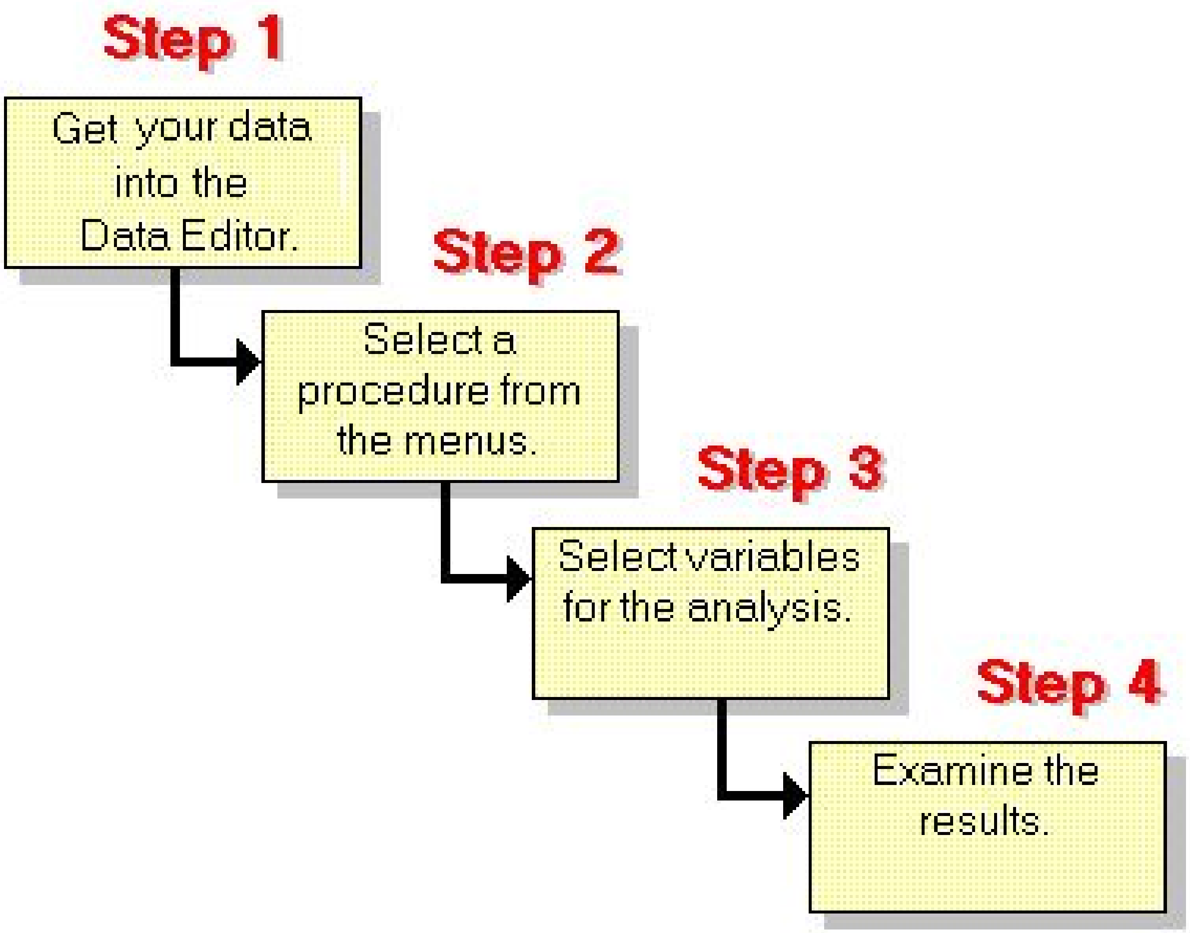
→ choose to transform the data or perform a nonparametric test that does not require this assumption.

Terima Kasih

Tahapan Analisis Statistik dengan SPSS

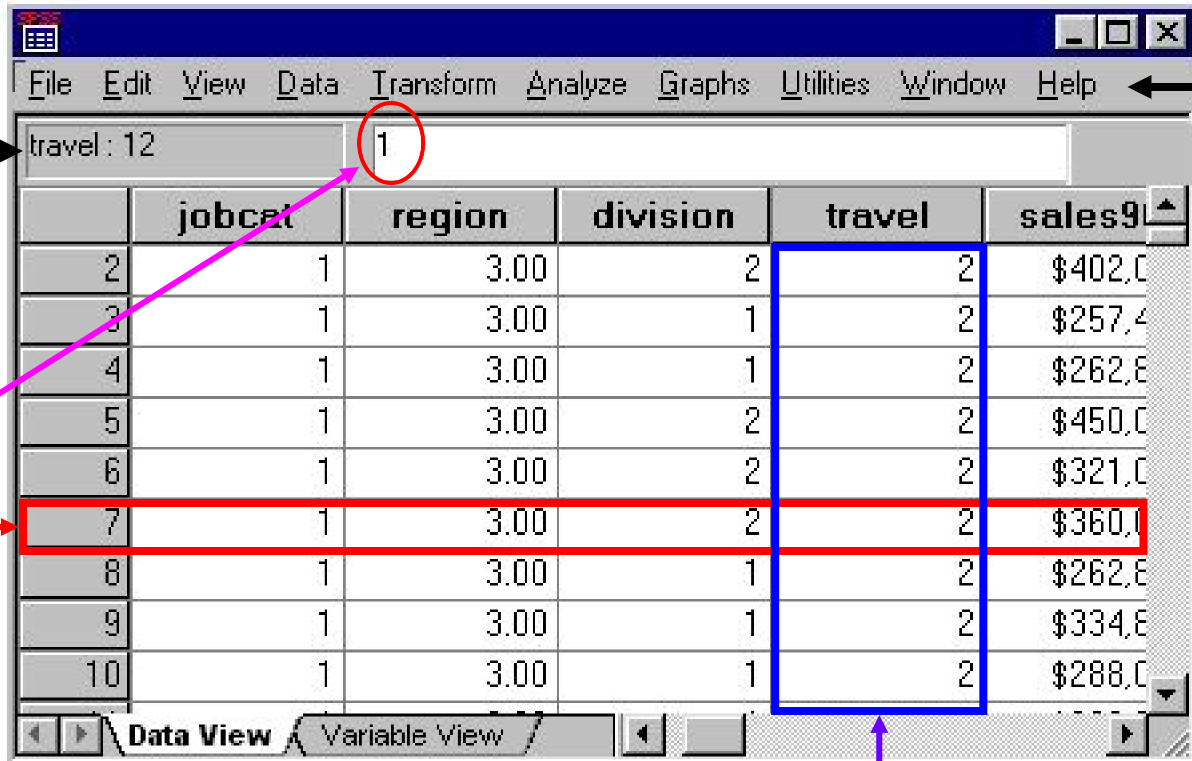


TAHAPAN-TAHAPAN untuk ANALISIS DATA dengan SPSS *(Statistical Package for Social Sciences)*



WINDOWS : DATA EDITOR

Kursor berada di baris ke-12 variabel travel **dengan nilai 1**

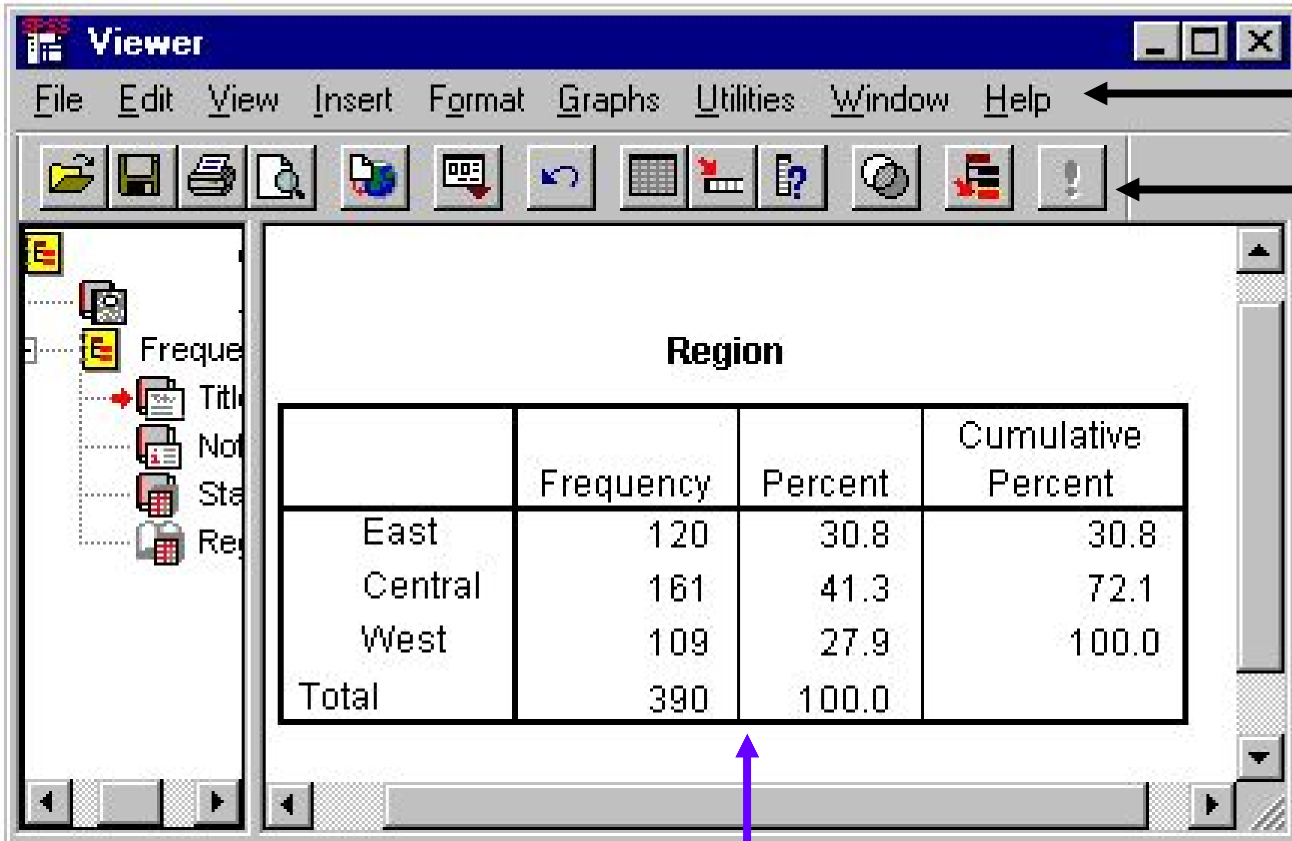


Menu

BARIS : menunjukan nilai untuk satu PENGAMATAN

KOLOM : menunjukkan nilai-nilai untuk satu VARIABEL

WINDOWS : OUTPUT VIEWER



Menu

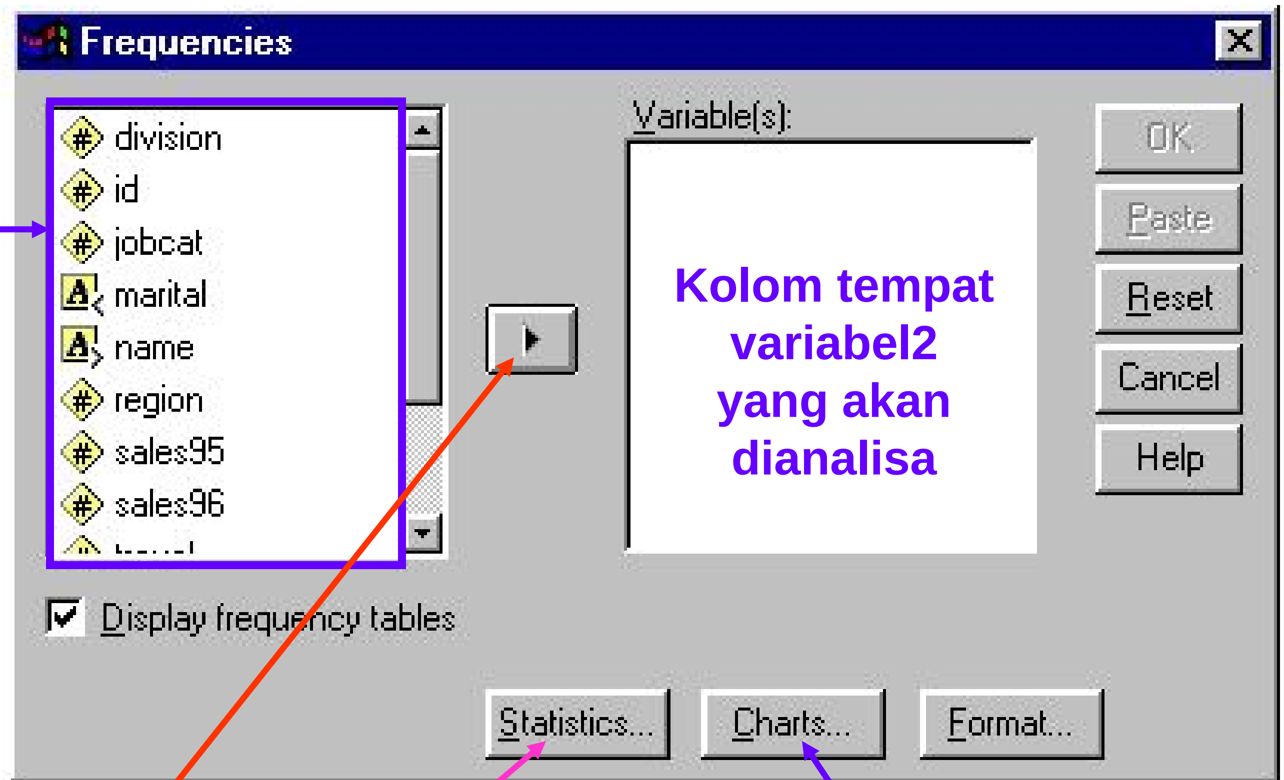
Toolbar

Hasil perhitungan

KOTAK DIALOG *(DIALOG BOX)*

Untuk memilih variabel yang akan dihitung nilai statistiknya atau yang akan dibuat grafiknya

Daftar nama-nama variabel yang ada



Variabel dengan data teks (*string*)

Variabel dengan data numerik

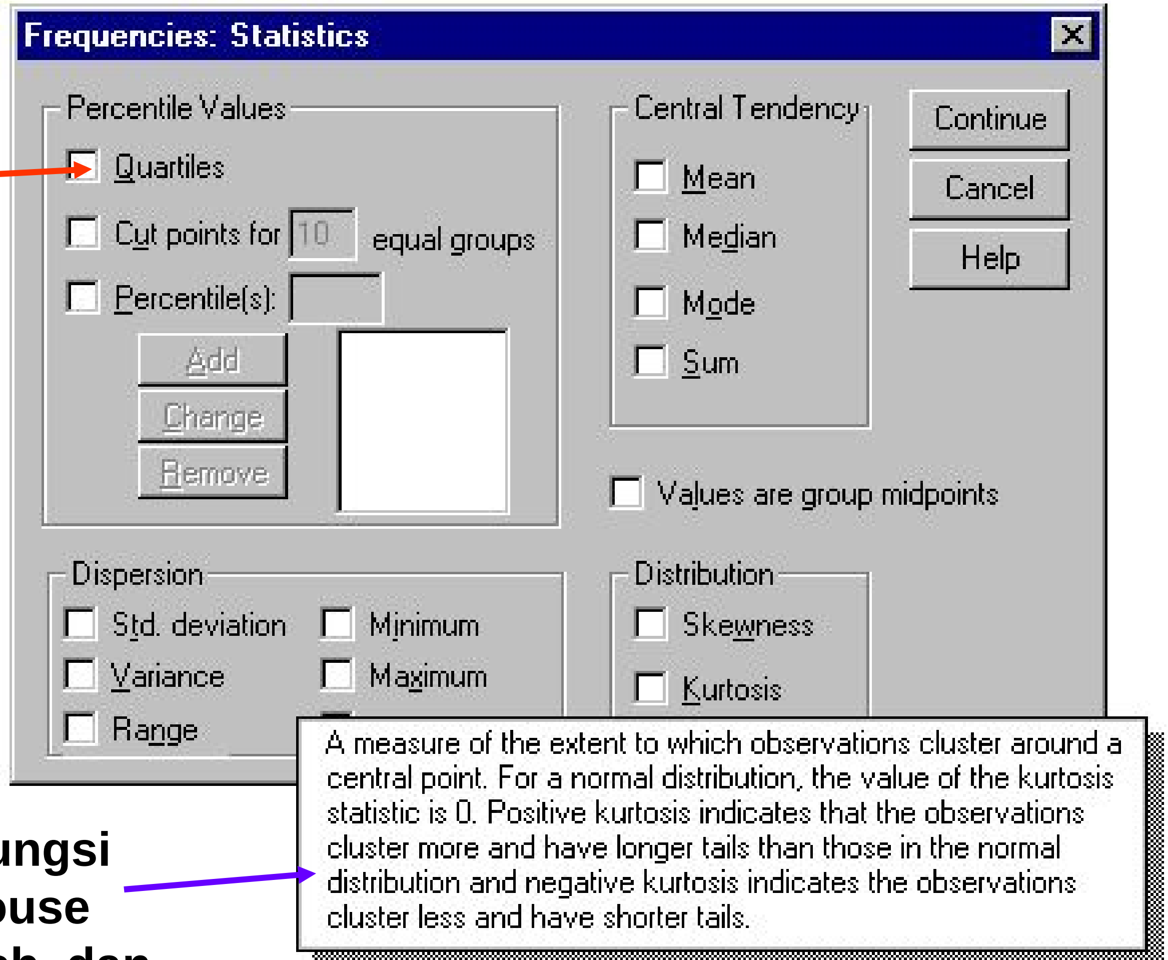
Tombol pemindah dari daftar variabel ke kolom analisa

Tombol fungsi statistik

Tombol fungsi grafik

KOTAK DIALOG *(DIALOG BOX)*

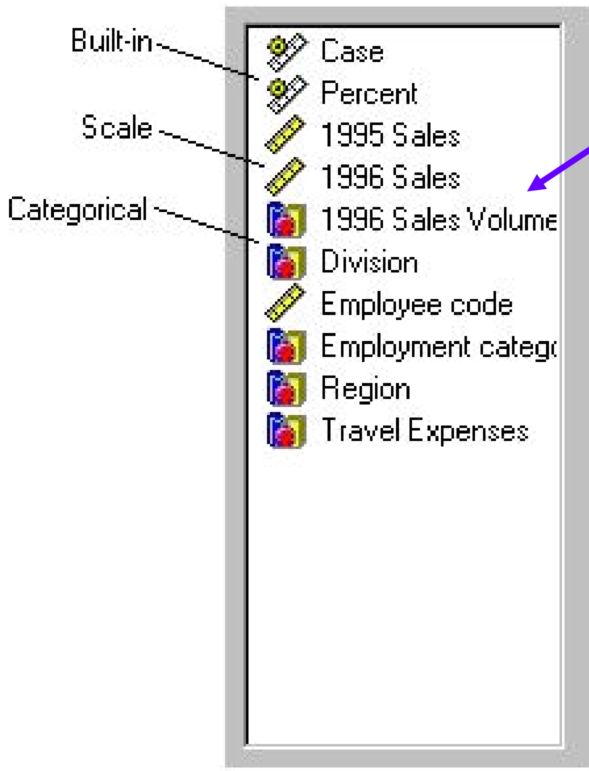
Fungsi - fungsi statistik



Keterangan tentang fungsi statistik jika ujung mouse diarahkan ke fungsi tsb. dan tombol kanannya ditekan

KOTAK DIALOG *(DIALOG BOX)*

Untuk memilih variabel-variabel yang akan dibuat grafiknya



FILE DATA

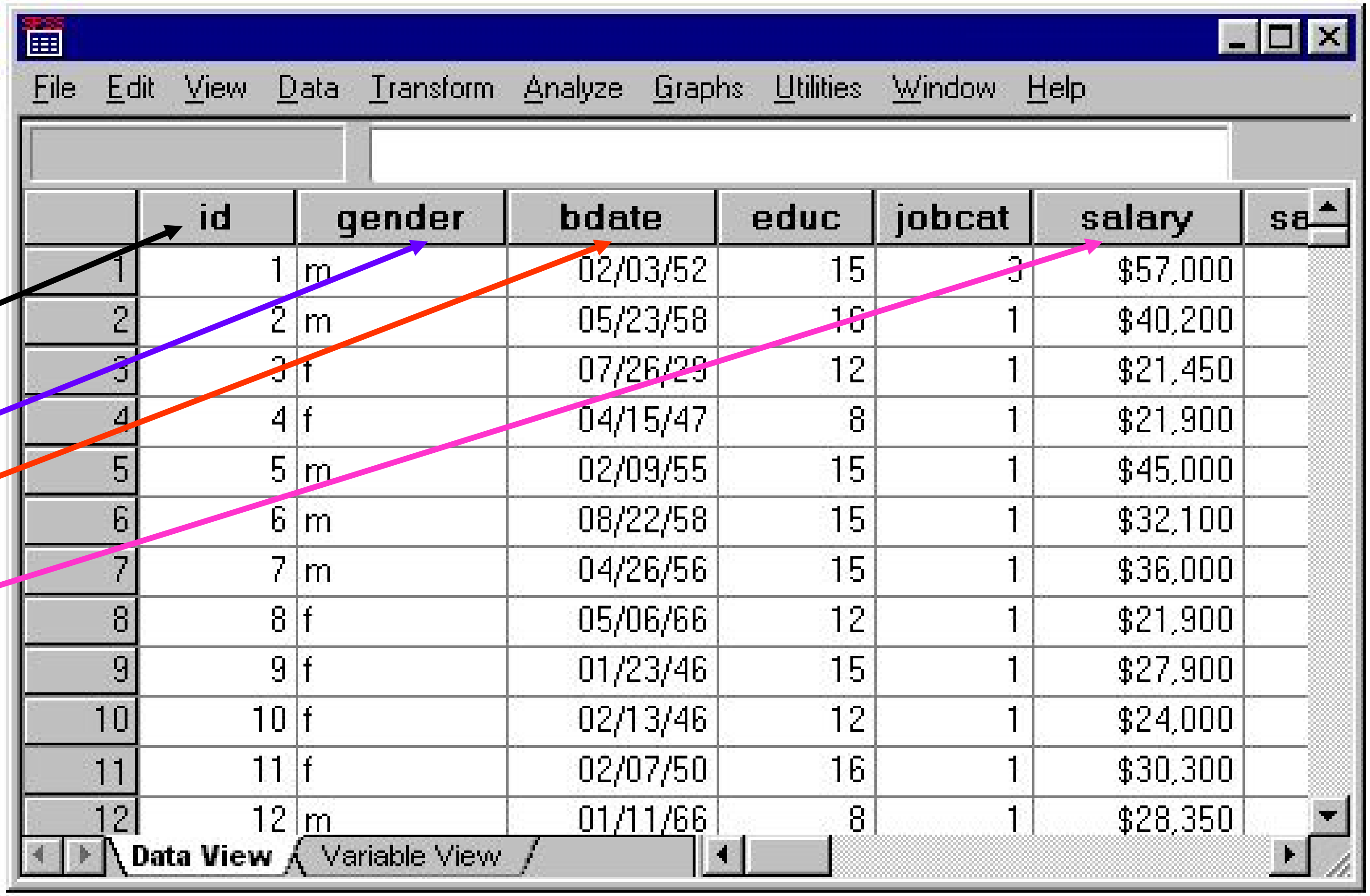
Jenis data :

Numerik

Teks (string)

Tanggal

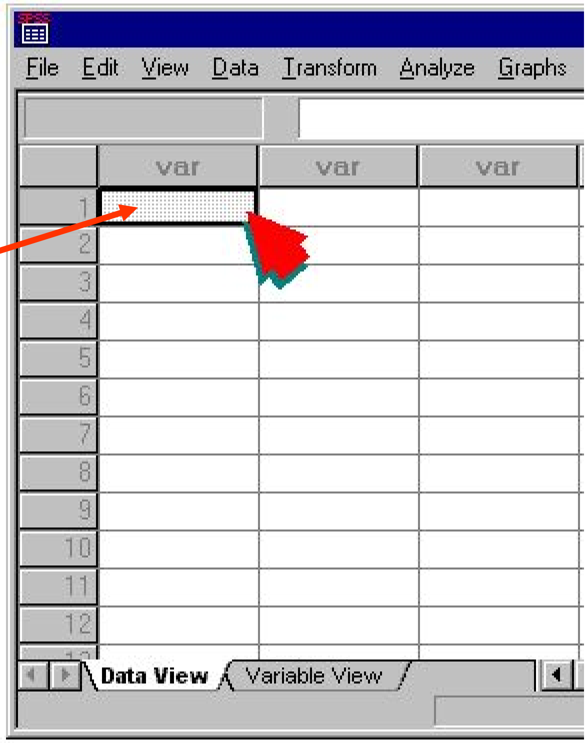
Currency



	id	gender	bdate	educ	jobcat	salary	sa
1	1	m	02/03/52	15	3	$57,000	
2	2	m	05/23/58	16	1	$40,200	
3	3	f	07/26/29	12	1	$21,450	
4	4	f	04/15/47	8	1	$21,900	
5	5	m	02/09/55	15	1	$45,000	
6	6	m	08/22/58	15	1	$32,100	
7	7	m	04/26/56	15	1	$36,000	
8	8	f	05/06/66	12	1	$21,900	
9	9	f	01/23/46	15	1	$27,900	
10	10	f	02/13/46	12	1	$24,000	
11	11	f	02/07/50	16	1	$30,300	
12	12	m	01/11/66	8	1	$28,350	

Memasukkan data NUMERIK

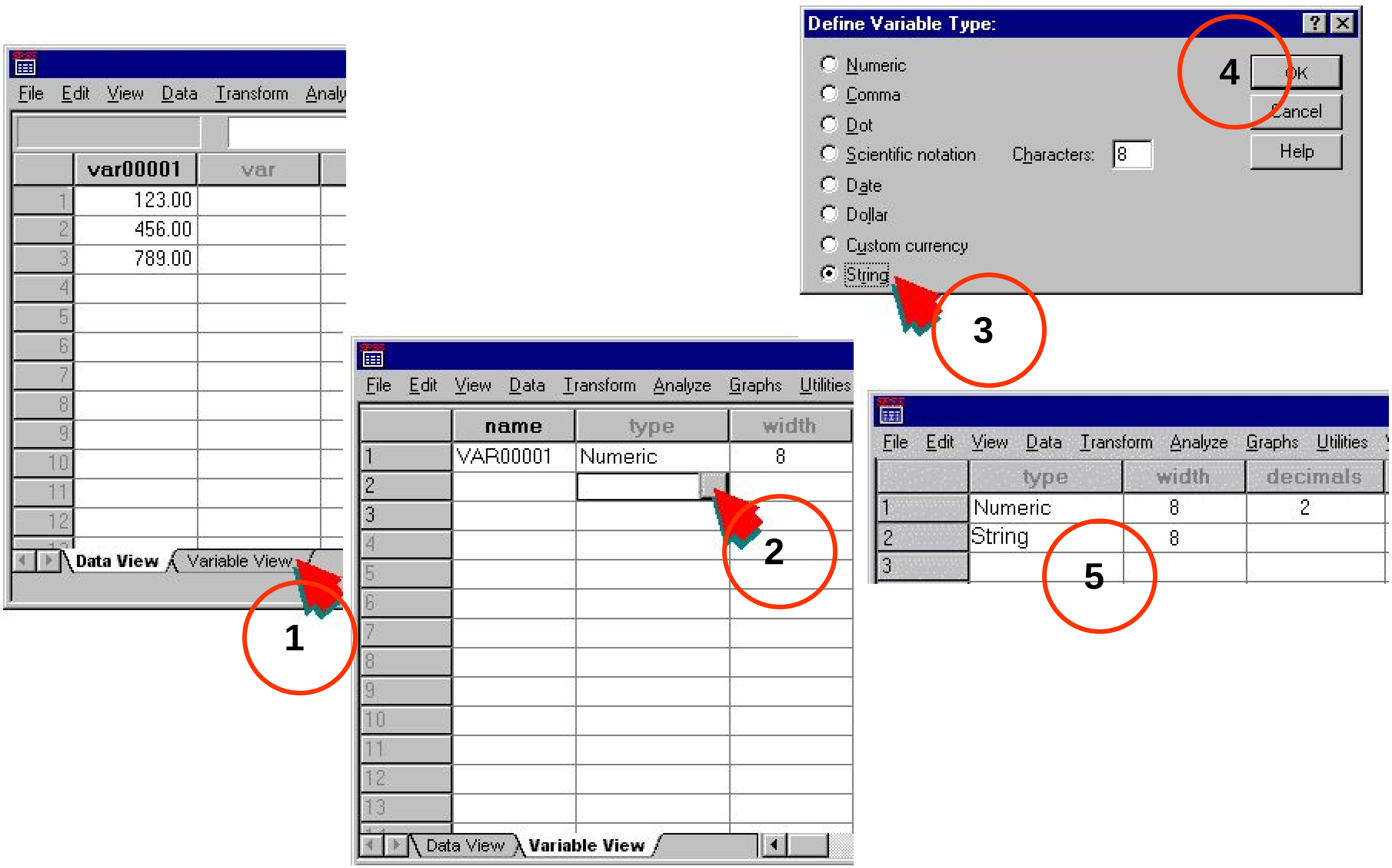
Letakkan kursor di sel dimana data akan dimasukkan



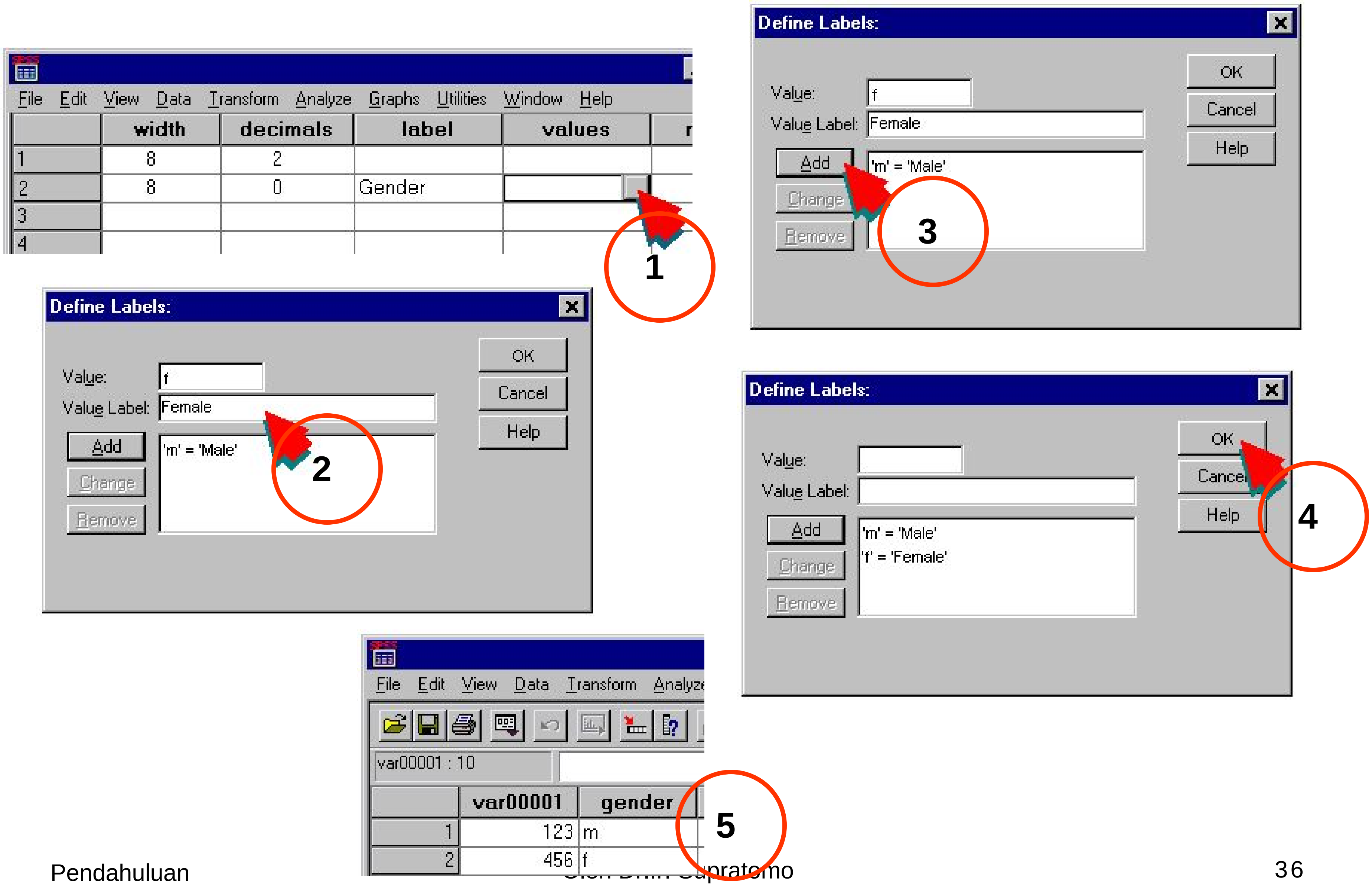
Nilai data yang ada di sel

Ketik nilai data di sel tsb. kemudian tekan ENTER

Memasukkan data AlphaNUMERIK (teks/string)



Mendefinisikan data AlphaNUMERIK (teks/string)



No	GENDER	BIDANG	GAJI	USIA	KERJA	No	GENDER	BIDANG	GAJI	USIA	KERJA
1	Wanita	Administrasi	715.00	23	2	21	Pria	Produksi	1,259.00	30	6
2	Wanita	Administrasi	745.00	21	2	22	Wanita	Marketing	745.00	20	2
3	Pria	Administrasi	850.50	26	3	23	Pria	Marketing	785.00	21	2
4	Wanita	Administrasi	1,087.00	29	5	24	Wanita	Marketing	985.00	22	5
5	Pria	Administrasi	1,285.00	30	7	25	Pria	Marketing	1,485.00	22	10
6	Wanita	Administrasi	835.00	30	4	26	Pria	Marketing	1,275.00	22	6
7	Wanita	Administrasi	995.00	30	7	27	Wanita	Marketing	985.00	25	2
8	Pria	Administrasi	912.50	31	6	28	Wanita	Marketing	1,067.50	25	5
9	Pria	Personalia	925.00	20	4	29	Pria	Marketing	745.00	26	2
10	Pria	Personalia	980.00	24	6	30	Wanita	Marketing	1,085.00	26	5
11	Pria	Personalia	851.50	29	5	31	Wanita	Marketing	1,082.50	26	5
12	Pria	Personalia	820.00	31	6	32	Pria	Marketing	954.00	26	4
13	Pria	Personalia	1,106.50	30	7	33	Wanita	Marketing	887.50	26	4
14	Pria	Personalia	1,465.00	33	5	34	Pria	Marketing	1,022.50	28	5
15	Pria	Produksi	745.00	20	1	35	Wanita	Marketing	1,150.00	28	6
16	Pria	Produksi	915.00	22	4	36	Wanita	Marketing	705.00	24	1
17	Wanita	Produksi	755.50	24	3	37	Wanita	Marketing	785.00	25	2
18	Pria	Produksi	1,065.00	26	6	38	Pria	Marketing	905.00	25	5
19	Pria	Produksi	915.00	27	3	39	Pria	Marketing	974.50	25	4
20	Pria	Produksi	985.00	30	4	40	Wanita	Marketing	1,245.00	29	6