METODOLOGI PENELITIAN

Ir. Nurly Gofar, MSCE, PhD
Program Studi Teknik Sipil
Program Pascasarjana
Universitas Bina Darma
Sem Genap 2018/2019

Sesi 8

Interpretion & Results Analysis

Interpret / Analyze Results

- What did your experiment show?
- Qualitative data analysis.
- Quantitative data analysis.
  - Descriptive and inferential statistics, clustering, …
- What might weaken your confidence in the results (critical spirit)?
- Discussion regarding
  - Literature
  - Research objectives
  - Research questions.

Consider multiple perspectives !

Data Analysis

- Data analysis is about manipulating and presenting results
- Data need to be organized, summarized, and analyzed in order to draw/infer conclusion
- Commonly used approaches or tools
  - Statistics
  - Models
  - Standards

Types of Data Analysis

- Estimation of parameter mean values
- Estimation of parameter variability
- Comparison of parameter mean values
- Comparison of parameter variability
- Modeling the dependence of dependent variable on several quantitative and qualitative independent variables
- Accelerated experimentation

Data checking

Before doing data analysis and interpretation :

- Watch for invalid data. This could be done continuously throughout data gathering process
- Bad data can bias results and interpretation
- Repeat data gathering or experimentation if there exist suspicious data.

Source of error in experiment

- Mistake in reading and calculation (Random error)
- Human error and capability and quality of measuring instrument
- Observation error
- External effect e.g. wind, temperature, humidity
- Statistical fluctuation
- Error due to unrepresentative sample.
- When taking a sample for measurement to conclusion for the whole population, the selected sample must represent the population.

Use log book to minimize error

- Record anything about the test
- State measurement, observations and comments directly into the note books.
- Record the date and time of the experiment
- Write down all mathematical calculation on the note book, do not just doing calculation in your mind. Eg. To find the difference in weight of two objects, write down the weight of each the object, do not just write the calculated difference.
- When taking a set of readings, fill in them in a table form.

Analysis and Interpretation

- Summary and Tabulation (Session 11)
- Statistical
- Numerical - software
- Graphical (Session 11)
- Combination

Statistical analysis

Statistics is the science of conducting studies to collect, organise, summarize and draw conclusions from data. The objective is to maximise the understanding of such information.

Types of statistics

- Descriptive statistics
- Inferential statistics

Descriptive Statistics

Used to describe situation which involved :

- data collection
- organization
- summation

Examples of descriptive statistics used in everyday life

- A typical one-a-day vitamin pill boosted certain immune responses in older people by 64%.
- Rain covering 30 to 40% of Palembang in the late afternoon
- Of 1000 households polled in Palembang, 40% households owned at least one car, 9% had two or more.

Inferential Statistics

To make inferences from **sample** to **populations** based on probability theory. The following steps are required to generalize from samples to populations

- performing hypothesis testing
- determining relationships among variables, and
- making predictions

Note: A **population** consists of all subjects (human or otherwise) that the information are being studied.
A **sample** is a subgroup of the population.

Method of sampling

- Random sampling : Each item has equal chance to be selected.
- Non-random sapling :
  - Multi stage
  - Quata – by type
  - Cluster – combination of multi stage and quata
- Systematic sampling – take every n^{th} of item.

Regression Lines

- Correlation – individual judgement
- Regression – independent of individual judgement

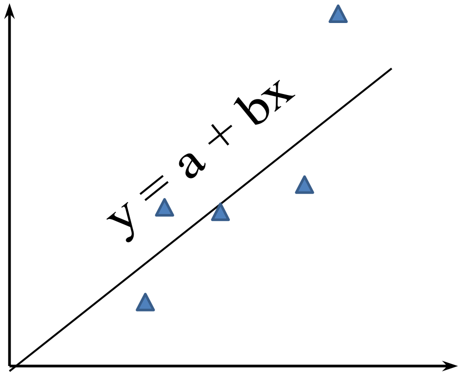
Method of least square is based on:

- Vertical deviation
- Horizontal deviation

Resulted in 2 different regression lines i.e.:

(A) Minimising the total of squared vertical deviation

(B) Minimising the total of squared horizontal deviation



Regression Lines

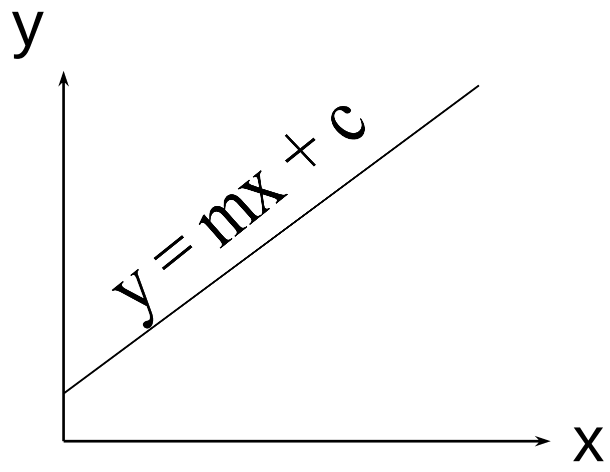
Regression of y on x:

$$C = \frac{\sum y - m\sum x}{n}$$

$$m = \frac{n\sum(xy) - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$y = mx + c$$

Where n is number of data



Regression Lines, Example

Calculate and plot the regression line of y on x for the set of data shown below.

x	y
2	60
5	100
4	70
6	90
3	80

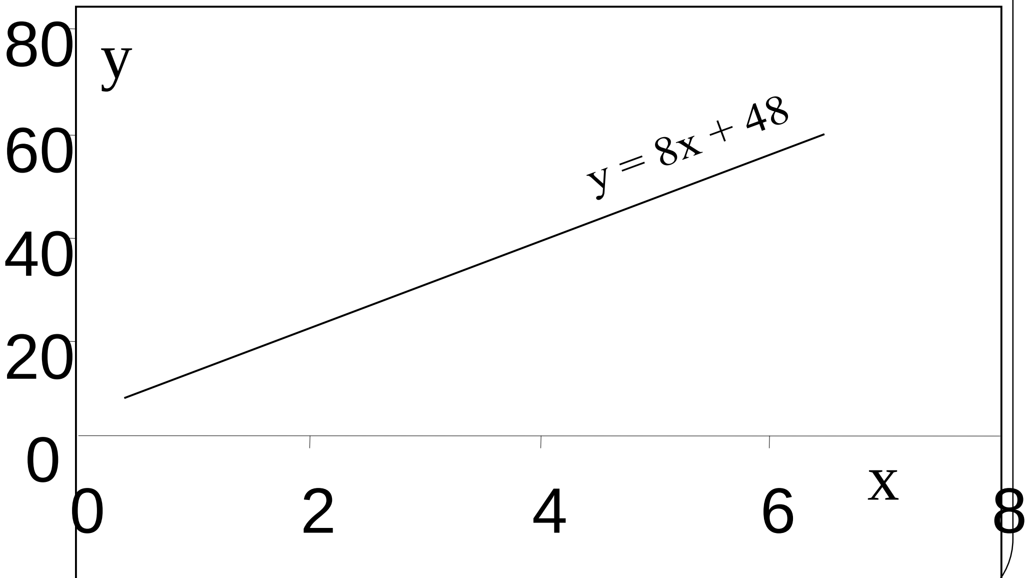
	x	y	xy	x^2
	2	60	120	4
	5	100	500	25
	4	70	280	16
	6	90	540	36
	3	80	240	9
Σ	20	400	1680	90

Regression Lines, Example

$$m = \frac{n\sum(xy) - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} = \frac{5(1680) - (20)(400)}{5(90) - (20)^2} = 8$$

$$C = \frac{\sum y - m\sum x}{n} = \frac{400 - 8(20)}{5} = 48$$

$y = mx + c = 8x + 48$



Measures of variation (range)

- Sometime the mean is not good enough to describe a data set as in the following example.
- Eg.: A testing lab wishes to test two experimental brands of outdoor paint (A and B) to see how long each would last before fading. Different chemical agents are added in each group and only six cans are involved. These two groups constitute two small populations.

Test No	1	2	3	4	5	6	Mean
Brand A	10	60	50	30	40	20	35
Brand B	35	45	30	35	40	25	35

- The experiment showed that Brand A and B gave similar mean value = 35. Thus one might conclude that both brand of paint last equally well.
- But a different conclusion might be withdrawn when the data set are examined based on variation. Variation for Brand A: 60-10 = 50 month; for Brand B: 45-25 = 20 month

Measures of variation

Measures indicating the degree of spread/variation

- Variance average of the squares of the distance each value id from the mean,

  Symbol used : σ^2, s^2

- Standard Deviation Square root of the variance

  Symbol used : σ, s

$$\sigma = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

where

x = individual value

$\overline{x}$ = population mean

n = population size

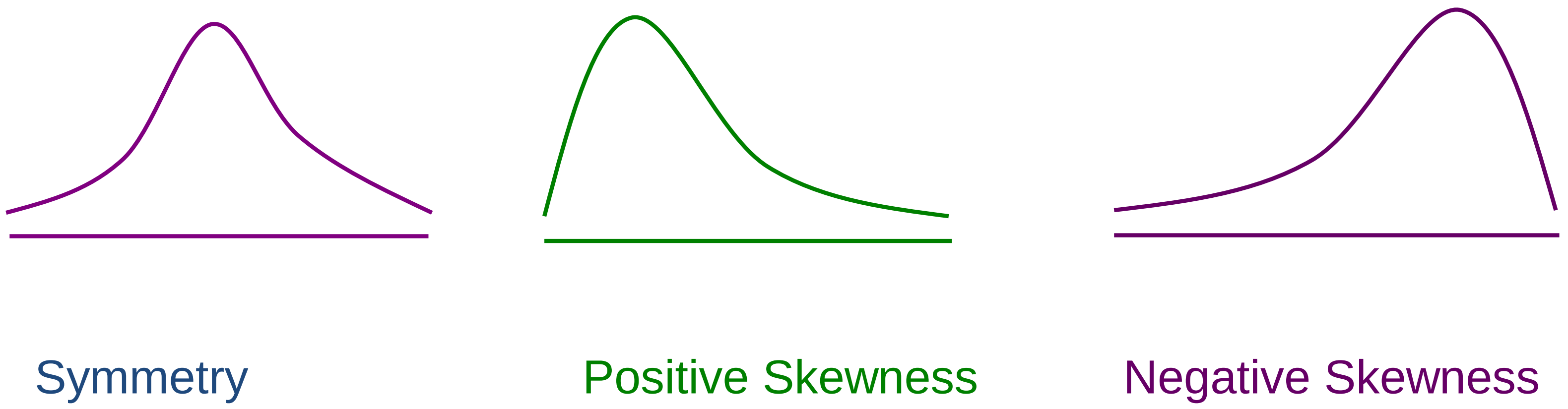
Measures of variation

Coefficient of variation,

$$\text{C.O.V.} = \frac{\sigma}{\bar{x}} \times 100\%$$

Symmetry and Skewness

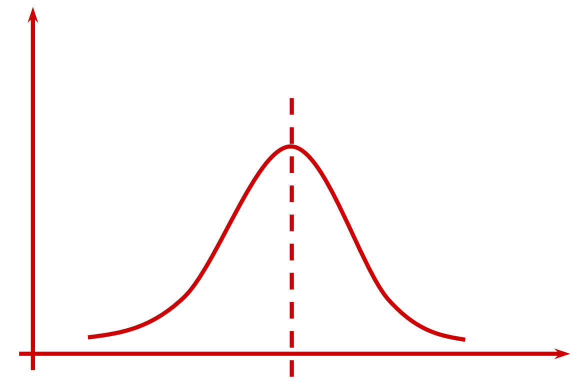
Coefficient of skewness, C.O.S = 3(mean – median) / σ



Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\bar{x})^2 / 2\sigma}$$

$$\int_{-\infty}^{\infty} f(x) = 1.0$$

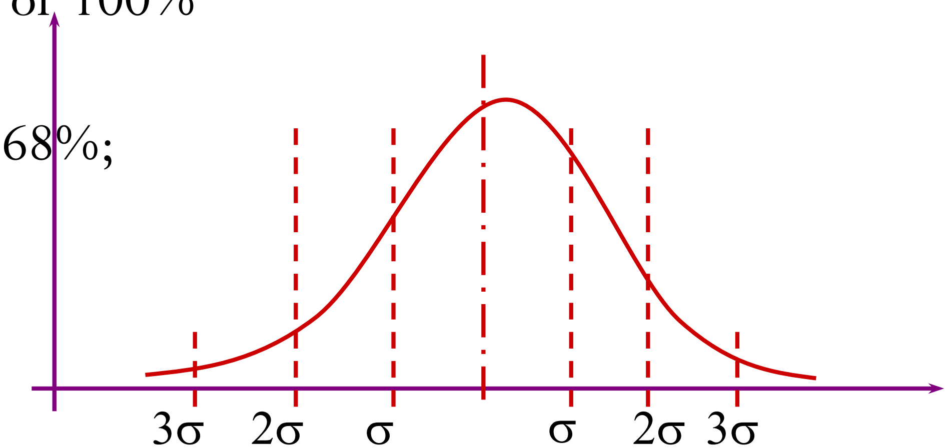


- Bell shaped; Data values are evenly distributed about the mean – symmetrical'
- The mean, median, and mode are equal and located at the center of the distribution
- Unimodal – only one mode
- Symmetrical about the mean – the shape is the same for both side
- The curve is continuous, asymptotic to x- axis
- Total area under the curve = 1.0 or 100%
- The area under the curve :

  within 1 std. deviation = 0.68 or 68%;

  within 2 std deviation = 95%

  within 3 std deviation = 99.7



Error

Error = X_i - X_{true}

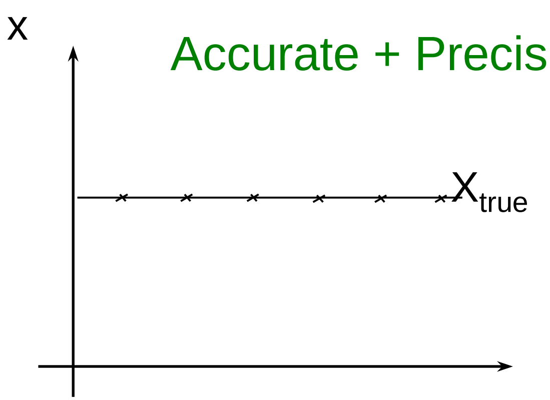
where X_i = measurement value

X_{true} = true value of the measurement

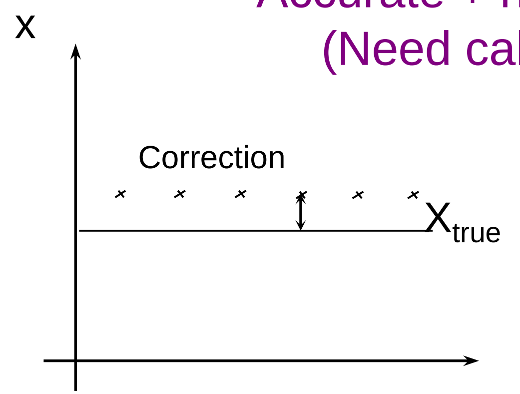
- *Accuracy* – Difference between measured and true value typically specified by a maximum value.
- *Precision* – Difference between measured and true value during repeated measurement.

Types of error

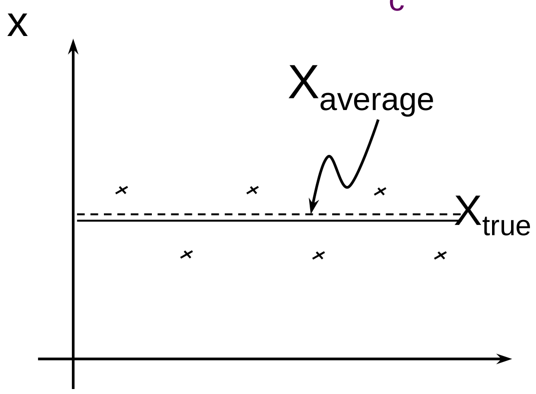
Accurate + Precision



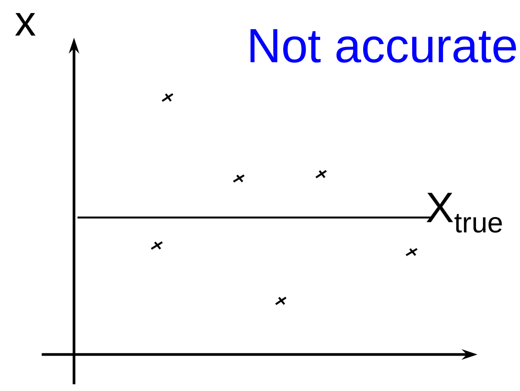
Accurate + no Precision
(Need calibration)



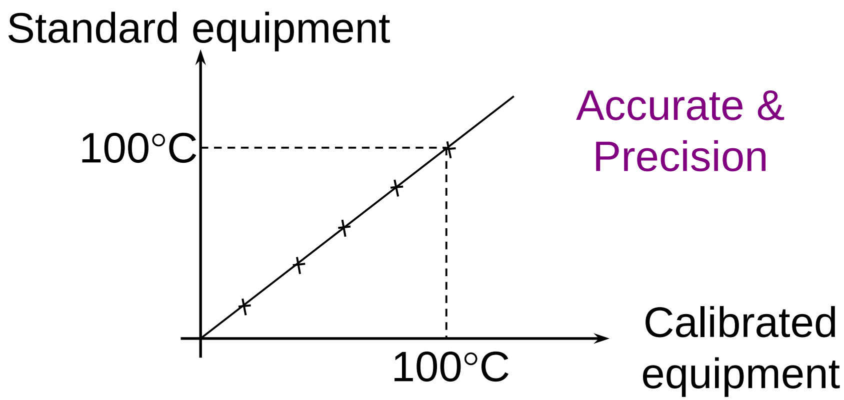
Accurate x_c ± Uncertainty

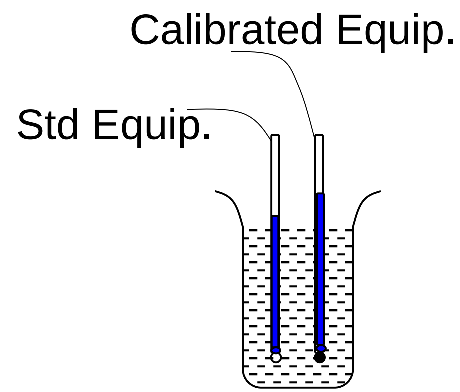


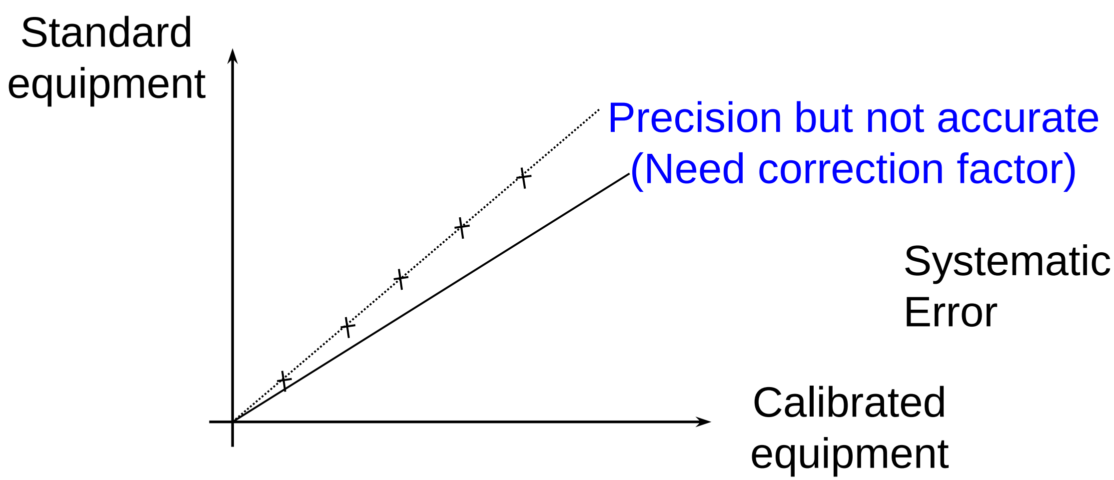
Not accurate + No precision



Types of error

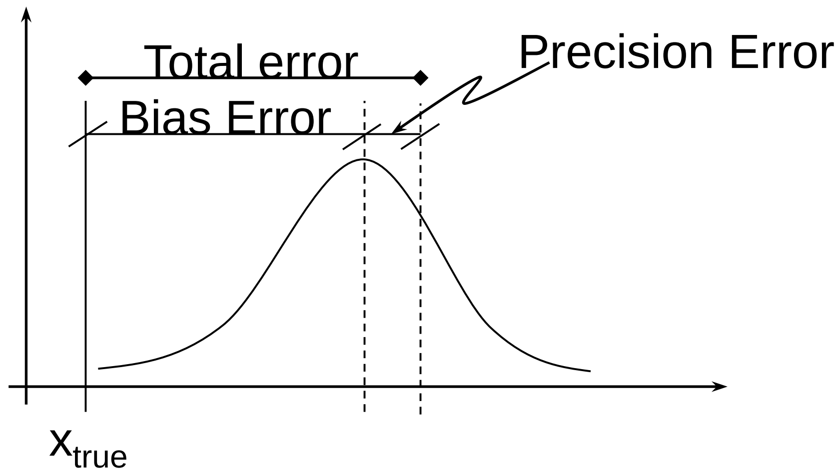




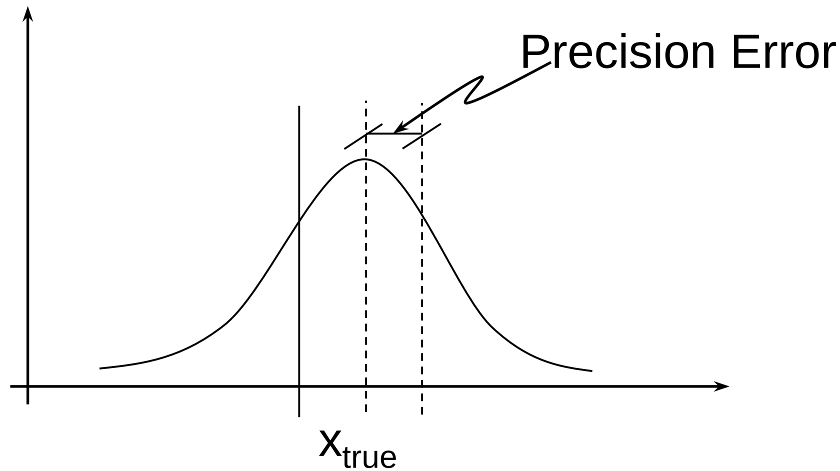


Systematic
Error

Components of error



General



No Systematic Error

Confidence Intervals

- Confidence level of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter.
- The parameter is specified as falling between two values. Fpr example the average age of all students

  $26.9 < \mu < 27.7$ OR 27.3 ± 0.4
- The probability of being correct can be assigned. e.g 95% confidence interval means that it is 95% sure/chance that the population mean is contained within the range.
- This mean that a person is 95% confident that the interval 26.9 to 27.7 includes the population mean, μ

Hypothesis

A hypothesis is a conjecture about a population parameter.

This conjecture may or may not be true.

- An educated guess based on theory and background information
- Hypothesis Testing is a process of using sample data and statistical procedures to decide whether to reject or not reject a hypothesis (statement) about a population parameter value.
- Examples of Hypothesis testing:
  - Adding a chemical X will improve water quality
  - Rubbish from construction activity will reduce the infiltration capacity of ground
  - The average life expectancy in the next decade for man will be more than 100 years

Hypothesis

Two types of statistical hypothesis

- The Null Hypothesis: symbolised by Ho, states that there is no difference between a parameter and a specific value OR that there is no difference between two parameters. NULL means NO CHANGE. Statement of equality
- The Alternative Hypothesis: symbolised by Ha, states a specific difference between parameter and a specific value OR states that there is a difference between two parameters. TEST or Research Hyphothesis.

Hypothesis - example

- **Situation A**: A researcher is interested in finding out whether a new medicine will have any undesirable side effects on the pulse rate of the patient. Will the pulse rate increase, decrease or remain unchanged. Since the researcher knows the pulse rate of the population under study is 82 beats per minute, the hypothesis will be
- Ho : $\mu = 82$ (remain unchanged)
- H1 : $\mu \neq 82$ (will be different)
- This is a **two-tailed test** since the possible effect could be to raise or lower the pulse

Hypothesis - example

- Situation B: A chemist invents an additive to increase the life of an automobile battery. The mean life time of ordinary battery is 36 months. The hypothesis will be:

  Ho : $\mu \leq 36$

  Ha : $\mu > 36$
- The chemist is interested only in increasing the lifespan of the battery. His alternative hypothesis is that the mean is larger than 36. Therefore the test is called **right-tailed**, interested in the increase only.

Hypoythesis - example

- Situation C: A contractor wishes to lower heating bill by using a special type of insulation in house. If the average monthly bill is IDR 500.000,- his hypothesis will be

  Ho : $\mu \geq$ IDR 500.000,-

  H1 : $\mu <$ IDR 500.000,-

- This is a **left-tailed test** since the contractor is only interested in reducing the bill

Hypothesis testing

- General Procedure for testing the hypothesis. Can be done statistically.
- Step 1: State the hypothesis
- Step 2: find critical value for a selected level of significant (α) e.g. 0.1, 0.05, 0.01. Consider case for one-tailed or two-tailed
- Step 3: Compute the test value using z-test or t-test
- Step 4: Make the decision to reject or not to reject the hypothesis. If test value < critical value accept Ho. test value > critical value reject Ho. Refer

Hypothesis – significant difference

- **What is Significant Difference??**
- A significant difference occurs if the difference between the hypothesized (null) value and the sample statistic value is too large to be attributed to chance. A significant difference strongly suggests that the null hypothesis is not true.
- **Significant difference at $p<0.05$** means, 95% of the time the sample mean is larger than the hypothesised value.

Methods for Test of significance

z-test

For detecting difference between two means for large sample (two samples)

Assumptions required

The sample must be independent, that is no relationship between the subject in the sample

The sample must be normally distributed

F-test

For the comparison of two variances or standard deviations. E.g variation in cholesterol level in man and women

Assumptions

The population from which the samples were obtained must be normally distributed

Samples must be independent of each other

t-test

To test the difference between two means for small independent sample ($n<30$)

Assumptions

Sample must be independent

The populations are normally distributed