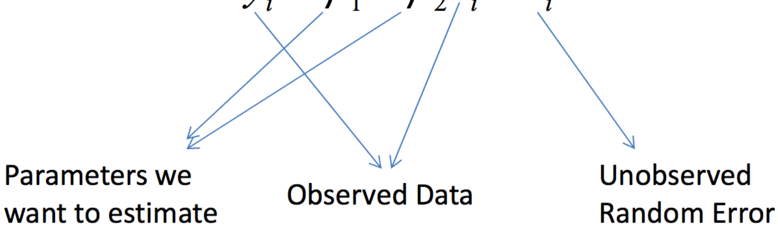
Quantitative Methods 2

Lecture 1 - 03/03/2015

Simple regression - relationship between two or more variables e.g - expenditure and income of households. The best line that fits the linear relationship of the data is utilised to show the regression of the two variables.

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

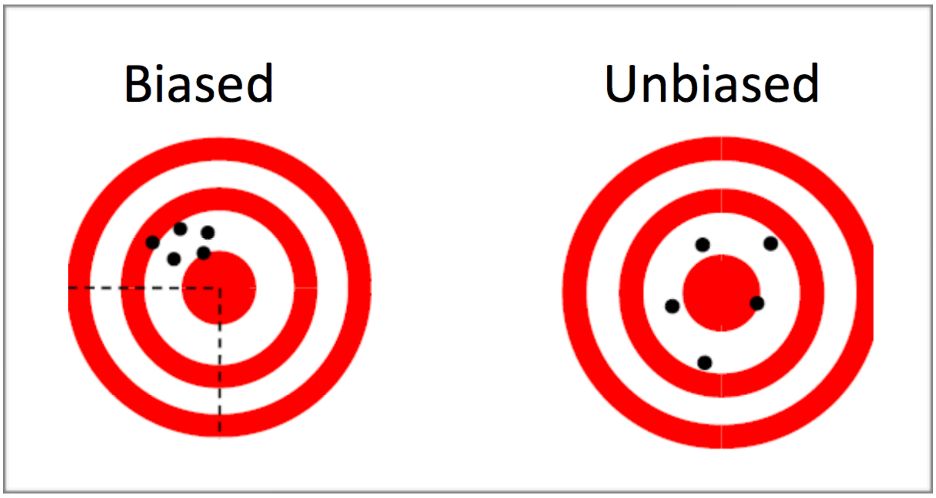


Cross section - data collected at a particular period of time.
Panel - data collected over several time periods.
Time series- data that is collected at intervals of time that depicts the trend of the variables over time. Thus, with the trend of the time series we can forecast future values of them.

Lecture 2 - 06/03/2015

Statistical inference
-A sample is drawn from a population and thus features can be observed from this sample that may reflect upon the features of the population.

Unbiased estimator - is one which produces estimates centred around the true value of the population. When the estimator is biased, it may not be centred towards the target where if it is unbiased the estimators may be equally distributed towards the true value of the population.



Minimum variance - when the variance is low, the sample values are all quite close and does not vary away greatly from the true value.

Random variables - a probability density function for a discrete variable can be plotted that displays the probability of a random variable, e.g tomorrow's weather.

Estimator - thus, an estimator is a random variable because it can take different values over different samples.
The entrap limit theorem states that with an increase in sample size, the distribution of the sample mean becomes more and more like a normal distribution with the mean equal to population mean and variance closer to the population variance.

Hypothesis testing

-E.g - you have a sample from QM2 students marks in 2015 with a sample mean = 60, Can the mean of the population be 70? - The answer depends on the difference between sample mean and hypothesised value, the variance of the population, sample size and confidence level.

Suppose

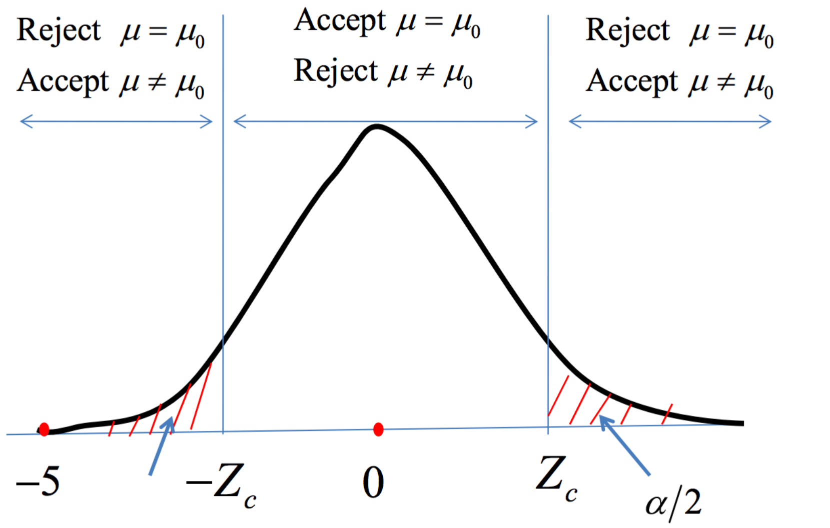
$$\bar{x} = 60$$
$$\sigma = 12$$
$$n = 36$$

Test

$$\begin{cases} H_0 : \mu = 70 \\ H_1 : \mu \neq 70 \end{cases}$$

Compute

$$\sigma_{\bar{x}} = \sigma / \sqrt{n} = 12/6 = 2 \Rightarrow Z_c = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{60 - 70}{2} = -5$$



Thus, for this hypothesis, we reject it because it is very unlikely that the population mean is 70. Looking at above, if the sample mean is within each of the shaded areas, then it is very unlikely the hypothesis is true and thus we reject it.

Steps of the hypothesis

1 - Set up the null (H0) and alternative (H1) hypothesis for the question at hand
2- determine the appropriate test statistic and its sampling distribution
3- specify the value of alpha, also called the significance level
4- Define the decision rule - for what values of the statistics will we reject the hypothesis
5- calculate the value of the test statistic
6- make the decision to answer the hypothesis

Example:

1- Null & Alternative hypothesis

$$\begin{cases} H_0 : \mu = 70 \\ H_1 : \mu \neq 70 \end{cases}$$

2- Test stat and its Distn $Z = \dfrac{\overline{x} - \mu_0}{\sigma_{\overline{x}}} \sim N(0,1)$

3- Significance Level $\alpha = 0.05$

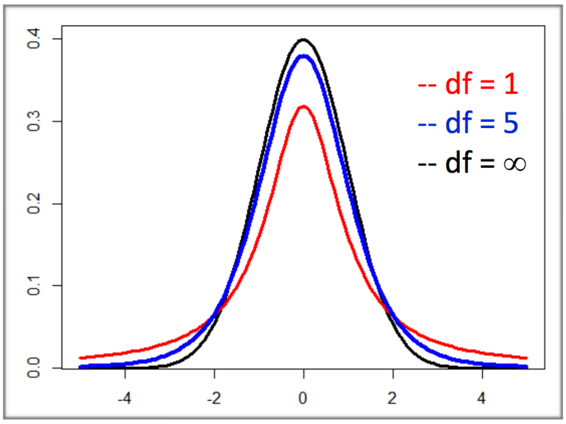
4- Null is rejected if $Z \leq -1.96 \quad or \quad Z \geq 1.96$

5- Test statistics $Z = \dfrac{\overline{x} - \mu_0}{\sigma_{\overline{x}}} = \dfrac{60 - 70}{2} = -5$

6- Since $Z \leq -1.6$ null hypothesis is rejected

	H_0 is **true**	H_0 **is false**
Reject H_0	Type I Error P(Type I Error)=α Size of the test	Correct Decision P(Reject H_0 | H_0 false)=1-β Power of the test
Do not reject H_0	Correct Decision P(Correct Decision)=1-α	Type II Error P(Type II Error)=β

-When we use the sample variance s^2 when the population variance is unknown, this makes the distribution is different, in this case the T-distribution is used, where the T distribution has increased degrees of freedom.



The T distribution looks like a standard distribution and when the degrees of freedom are increased, the closer to a normal distribution it becomes. Example:

Example

Are mean salaries for Commerce Graduates below $50,000?

We have a sample information

$$\bar{x} = \$48,918$$
$$\text{n} = 50$$
$$\text{s} = 6271$$

<u>Solution</u>

1- Null & Alternative hypothesis

$$\begin{cases} H_0 : \mu = 50000 \\ H_1 : \mu < 50000 \end{cases}$$

2- Test stat and its Distn $Z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} \sim t(n-1)$

3- Significance Level $\alpha = 0.05$

4- Decision Rule: Null is rejected if $t_{0.05,49} \leq -1.676$

5- Test statistics

$$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{48918 - 50000}{6271 / \sqrt{50}} = -1.22$$

6- Since $Z \geq -1.6$ Null hypothesis is accepted

There is not sufficient evidence to conclude that mean salaries of commerce graduates are below $50,000

Lecture 3 - 10/03/2015

Testing the means and variance of two independent populations.
E.g - is average household income higher in Sunshine or South Yarra?
-Who is a more popular leader: Tony Abbot or Bill Shorten?

What Test to use?
1- Type of data - Quantitative data? Ranked/ordinal data? or categorical?
2- Normality assumption - is normal distribution a good approx or not, if it is we can use normal or T stat.
3- Type of sample - independent or dependent samples

Classical means test - SSK - 12.1, 12.2, 14.1

Examples
-For the South Yarra vs Sunshine resident income independent samples, we can construct a test of the difference of the mean incomes
-The information we need, is the mean, sample sizes, and sample means and variances of both populations, the variance of the population may be unknown.

-It can be shown that the best estimator is the difference in the sample averages

- It is **unbiased** $E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$

- It is **consistent** $Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

-It can be shown that it has the minimum variance among all potential estimators.

6 Steps of the test

1) Null and alternative