Estimators & Estimates

Point and Confidence Estimates

Hypothesis Testing

- Introduction to Estimates

Introduction

Introduction

A specific value is called an **Estimate**.

There are 2 types of Estimates

- Point Estimates
- Confidence Interval Estimates

Point Estimates is a single number
Confidence Interval Estimate is an interval

The Point Estimate and Confidence Intervals are very closely related.

Point Estimate is located exactly in the middle of ‘Confidence Interval’

However confidence intervals provide much more information and are preferred in making inferences.

Introduction

Examples:
A sample mean ($\bar{x}$) is the point estimate of population mean (μ).

A sample variance (s^2) is the point estimate of population variance (σ^2)

Estimator (how to estimate)	Parameter (What to Estimate)	Estimate (Concrete Result)
$\bar{x}$ (Formula)	μ	52.55
S^2 (Formula)	σ^2	1724.93

Estimates have the following 2 properties

- Bias
- Efficiency (Efficiency denotes variance. So far, no problematic estimators in Statistics)

Confidence Interval

A confidence interval is a more accurate representation of reality.

Confidence Intervals have some level of uncertainty, these are measured in ‘Levels of Confidence’

Confidence Level = 1 – α
Where,
0 <= α <=1
α is called ‘Level of Significance’

Formula :
Confidence Interval
[PointEstimate ± (Reliability Factor * Standard Error)]

(Reliability Factor * Standard Error) also called ‘Margin of Error’

Confidence Interval

A confidence interval can be calculated for, when:

- Population Variance is known (Z statistics)
- Population Variance is Unknown (t statistics)

Confidence Interval : Z statistics

Assumptions:
Population is normally distributed. Even it is not, you should use large sample and normalize it using CLT.

Confident Interval Formula:

$$\bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{(n)}}$$

Where,
$\bar{x}$ = sample mean
$Z_{\alpha/2}$ = static for standard normal distribution ($Z \sim N(0,1)$)
σ/root(n) = Standard Error
α = significance level
i.e. for confidence level 95%, α = 5% or 0.05

Confidence Interval : t statistics

It is used to get inference through small samples when population variance is unknown.

Named after an English statistician William Gosset.
Later Ronal Fischer introduced t-statistics.

Test statistic

$$t = (\bar{x} - \mu_0) / (s / \sqrt{n})$$

where:

- $\bar{x}$ = the sample mean
- μ_0 = the hypothesized population mean
- s = the sample standard deviation
- n = the sample size

Confidence Interval : t statistics

Degrees of Freedom: For a sample size ‘n’, degree of freedom is (n-1)

Formula:

Confidence Interval:

$$\overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

Where,

$\bar{x}$ = sample mean

$t_{(n-1),\ \alpha/2}$ = static for standard normal distribution ($Z \sim N(0,1)$)

s/root(n) = Standard Error

α = significance level

i.e. for confidence level 95%, α = 5% or 0.05

Confidence Interval : 2 Samples

t statistics:

$$Lower\ bound: (\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_{12}^2}{n_2}}$$

$$Upper\ bound: (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_{12}^2}{n_2}}$$

Z statistics: Same as above, only difference is we use 'z' statistics here.