Estimation

Estimation: An important problem of statistical inference is the estimation of population parameters from corresponding sample statistics.

Estimation is required to use the statistic obtained from the sample as estimate of the unknown parameter of the population from which sample is drawn.

Estimate: An estimate is a statement made to find an unknown population parameter.

Estimator: The procedure or rule to determine an unknown population parameter.

Ex Sample mean is an estimator of population mean.

Because, the sample mean is a method of determining population mean and it is a statistic of sample.

A parameter can have one or two or many estimators.

Types of estimation:

Basically, there are two types of estimate to determine the statistic of the population, namely.

(a) point estimation (b) Interval estimation.

(a) ~~Ex~~ point estimation:

A point estimation of a parameter is a ~~Sig~~ Single numerical value, which is computed from given sample and serves as an approximation of the unknown exact values of the parameter.

Ex The sample mean $\bar{x}$ is a point estimator of population mean.

→ point estimator: A point estimator is a statistic for estimating the population parameter θ and will be denoted by $\hat{\theta}$.

ex 'Sample mean $\bar{x}$ is an point estimator of μ.

Properties of Good Estimator:

i) Consistency

An estimator is not expected to estimate the population parameter without error. But it should give a value which is very close to the true value of unknown parameter. This is called consistency of estimator.

ii) Consistency

Thus, for smaller values of ϵ & η $\exists$ $n \geq n$ such that $P[|\hat{\theta}_n - \theta| < \epsilon] > 1 - \eta$

ie $\hat{\theta}_n \longrightarrow \theta$ as $n \longrightarrow \infty$.

ii) Unbiased estimator:

A statistic or a point estimator $\hat{\theta}$ is said to be unbiased estimator of the parameter θ if

$E(\hat{\theta}) = \theta$.

Otherwise it is biased.

Ex. Sample mean $\bar{x}$ is an unbiased estimator of the population mean.

Q1 Show that $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is an unbiased estimator of the parameter σ^2.

Sol Let $x_1, x_2 \ldots x_n$ be the elements of sample of size n and μ and σ^2 are the population mean and variance. So, $E(x_i)=\mu$ and $Var(x_i) = E[x_i-\mu]^2$ $= \sigma^2$; $i=1,2\ldots n$.

Let us write.

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}[(x_i-\mu)-(\bar{x}-\mu)]^2$$

$$= \sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu) + n(\bar{x}-\mu)^2$$

$$= \sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}-\mu)^2 \quad \text{---} (1).$$

Now consider $E(s^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$

$$\Rightarrow E(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} E(x_i-\mu)^2 - n\,E(\bar{x}-\mu)^2\right]$$

Using (1)

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma_{x_i}^2 - n\,\sigma_{\bar{x}}^2\right]$$

However $\sigma_{x_i}^2 = \sigma^2$, $i=1,2\ldots n$ and $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$.

$$E(s^2) = \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] = \sigma^2$$

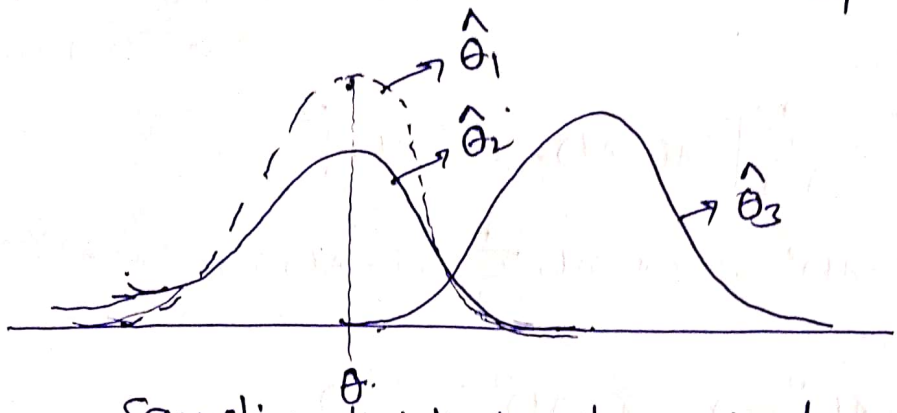
$\therefore E(s^2) = \sigma^2$, hence $\frac{1}{n-1}\sum(x_i-\bar{x})^2$ is an unbiased estimator for σ^2.

Note $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is a biased estimator.

iii) Most efficient estimator:

Consider all possible unbiased estimators of parameter θ, The one with the smallest variance is called the most efficient estimator.

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of the same population parameter θ. Then $\hat{\theta}_1$ is more efficient of θ than $\hat{\theta}_2$ if $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$.



Sampling distribution of estimators.

In the above figure, we illustrate the sampling distribution of three different estimators $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ all estimating θ. It is clear that only $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased since their distributions are centered θ ie $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$. But, variance of estimator $\hat{\theta}_1$ is smaller than variance of $\hat{\theta}_2$ ie $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$. Therefore $\hat{\theta}_1$ is more efficient among the three considered estimators.

iv) Sufficient estimator

An estimator is said to be sufficient for a parameter, if contains all the information in the sample regarding the parameter.

Thus a good estimator is one which is close to true value of the parameter as possible and should have i) consistency ii) unbiased iii) Efficient and iv) sufficient in all directions.

Interval estimation

Even the most efficient unbiased estimator cannot estimate the population parameter exactly. It is true that accuracy increases with large samples. But there is still no reason point estimate from a given sample to be exactly equal to the population parameter, it is supposed to estimate.

Therefore, in many situations it is preferable to determine an interval within which we would expect exact value of parameter.

An estimate of a population parameter given by two magnitudes within which a parameter can lie is called interval estimator of the parameter.

Let θ denote a population parameter. An interval estimator of θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$. where $\hat{\theta}_L$ & $\hat{\theta}_U$ depend on the value of θ for a particular sample and also on the sampling distribution of $\hat{\theta}$.

Let $\hat{\theta}$ be the random variable of estimator. Then samples will generally yield different values of $\hat{\theta}$.

From the sampling distribution of $\hat{\theta}$, we shall able to find $\hat{\theta}_L$ & $\hat{\theta}_U$ such that

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = \text{percentage of confidence}.$$

If $1-\alpha$ is a confidence about interval of $\hat{\theta}$ then we can find $\hat{\theta}_L$ and $\hat{\theta}_U$ such that

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1-\alpha.$$

where $0 < \alpha < 1$ (i.e critical value).

Here the interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ is called a $(1-\alpha)100\%$ confidence interval.

The fraction $1-\alpha$ is ~~of~~ confidence coefficient or degree of confidence. Also the end points, θ_L and θ_U are called lower and upper confidence limits.

Thus, when $\alpha = 0.05$, we have 95% confidence interval and $\alpha = 0.01$ we obtain a confidence interval with 99% confidence.

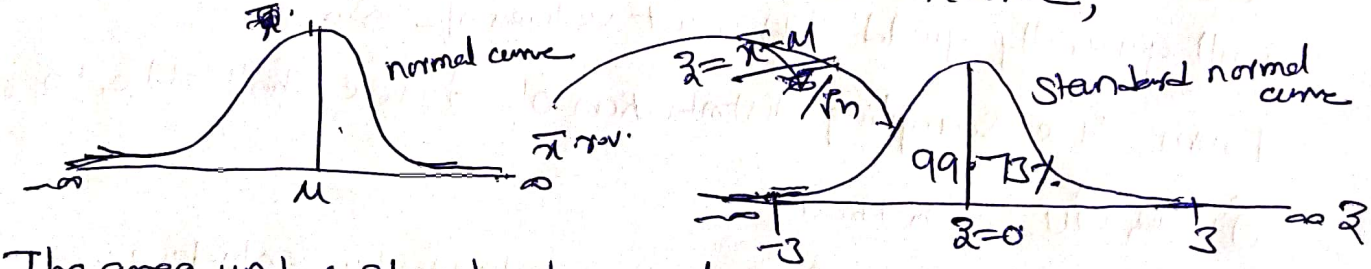
Confidence limits for population mean (μ)

Let $\bar{x}$ be the sample mean of sampling distribution of means of random sample of size n ($n \geq 30$) drawn from a population having mean μ and S.D σ is approximated by normal distribution with mean $\mu_{\bar{x}} = \mu$ and S.D $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

ie By central limit theorem $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}})$ for large samples ~~with a~~

Hence, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ has standard normal distribution with mean 0 and S.D 1. and it is used for r.v $\bar{x}$

Form the standard normal distribution curve,



The area under standard normal curve between $z=3$ and $z=3$ is 99.73%

ie $P(-3 \leq z < 3) = 0.9973$ where $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Then the inequality $-3 < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < 3$ — ① shows that 99.73% of cases are favourable to μ and 0.27% cases are not favourable.

From (1), we have

$$-3 < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < 3$$

$$\Rightarrow -3\frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < 3\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \bar{x}-3\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+3\frac{\sigma}{\sqrt{n}}$$

$$\rightarrow \bar{x}-3\,S.E < \mu < \bar{x}+3.S.E$$

Since S.E of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$.

$\therefore [\bar{x}-3.S.E,\ \bar{x}+3.S.E]$ is called 99.73% confidence interval ~~of~~ for the population mean μ.

Similarly $[\bar{x}-2\,S.E, \bar{x}+2\,S.E]$ is 95.45% confidence interval for μ.

$[\bar{x}-S.E, \bar{x}+S.E]$ is 68.26% confidence interval.

In general, confidence limits for population mean μ are $\bar{x} \pm z_*$ (S.E of $\bar{x}$). where z_* value of z.

(a) Confidence limits for population mean with.

i) 99% confidence are $\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$.

ii) 95% Confidence are $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

iii) 90% confidence are $\bar{x} \pm 1.64\frac{\sigma}{\sqrt{n}}$.

Thus (1−α)100% confidence interval of population mean is

$\bar{x} \pm z_{\alpha/2}^{*}\frac{\sigma}{\sqrt{n}}$ where $P(-z_{\alpha/2}^{*} \le z \le z_{\alpha/2}^{*}) = 1-\alpha$.

Where $z_{\alpha/2}$ is a value of z leaving an area of $\alpha/2$ to right ~~and left~~ tail of standard normal distribution curve.

Confidence limits for population proportion:

By central limit theorem, the sampling distribution of the sample proportions p will be approximated by normal distribution with a statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$ for a large samples.

Here Z is a S.D variable with mean 0 and S.D 1.

Then the confidence limits for population proportion with

i) 99% of confidence are $p \pm 2.58$ (S.E of p)

ii) 95% of confidence are $p \pm 1.96$ (S.E of p)

iii) 90% of confidence are $p \pm 1.64$ (S.E of p)

Here S.E of p is $\sqrt{\frac{PQ}{n}}$.

In general, the $(1-\alpha)100\%$ confidence interval for population proportion is $p \pm Z_{\alpha/2}$ (S.E of p).

Here $Z_{\alpha/2}$ is a value of Z leaving an area of $\alpha/2$ to right tail of standard normal curve.

Question

① A random sample of 200 measurements from a large population gave mean value of 50 and S.D of 9. Determine the 95% of confidence interval for the mean of the population.

Sol Given that a sample of 200 is drawn from normal population having mean 50 and S.D 9.

ie. $n = 200$, $\bar{x} = 50$, $\sigma = 9$.

The 95% confidence interval for population means.

$\bar{x} \pm 1.96$ (S.E of $\bar{x}$) ie $\left(50 \pm 1.96 \frac{9}{\sqrt{200}}\right)$

② * Maximum Error of estimate E with $(1-\alpha)100\%$ confidence is

$$E_{max} = |\bar{x} - E(\bar{x})|$$

(i) If Maximum Error in sampling distribution of mean is

$$E_{max} = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$ for a normal population.

ie one can assert a confidence interval for population mean with $(1-\alpha)100\%$ confidence that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Then $$n = \left(\frac{z_{\alpha/2}\,\sigma}{E_{max}}\right)^2.$$

(ii) Maximum error of sample proportion is

$$E_{max} = z_{\alpha/2}\sqrt{\frac{PQ}{n}}, \quad Q = 1-P.$$

Questions

① To estimate the average amount of time visitors take to move from one building to another in an office complex, the mean of a random sample of size n is used. Given $\sigma = 1.40$ minutes, determine how large should be the sample size if it ascertained with 99% confidence that the maximum error is at most 0.25.

② A random sample of size 100 has a standard diviation of 5. What can you say about the maximum error with 95% confidence.

③ Construct a 99% confidence interval for the true mean weight loss of If 40 persons on diet control after one month had mean weight loss 5 Kgs with S.D of 1.2 kgs.