

Interval Estimation

Alternative way to point estimation is interval estimation
A interval estimate of θ is an interval of the form

$$\hat{\theta}_1 < \theta < \hat{\theta}_2$$

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \underbrace{1 - \alpha}_{\rightarrow \text{some specified prob.}}$$

↳ confidence interval for θ

$1 - \alpha \Rightarrow$ degree of confidence
End points of interval, $\hat{\theta}_1$ & $\hat{\theta}_2$ } lower & upper confidence limits

Estimation of means

Suppose that mean of a r.v. is to be used to estimate mean of a normal population with known variance σ^2
So, the sampling distribution of $\bar{x}$ for a random sample of size n from a normal population with mean μ & variance σ^2 has a normal distribution with

$$\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Thus, it can be written

$$P(|z| < \frac{z_{\alpha}}{2}) = 1 - \alpha$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$z_{\alpha/2} \rightarrow$ such that integral of standard normal density from $\frac{z_{\alpha}}{2}$ to ∞ equals $\alpha/2$

Then,

$$P(|\bar{x} - \mu| < \frac{z_{\alpha}}{2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

to construct a confidence interval form for estimating mean of a normal population with known variance σ^2
we can write

$$P(\bar{x} - \frac{z_{\alpha}}{2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{z_{\alpha}}{2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Thus, if $\bar{X}$ is value of sample mean of a random sample having n observations from a normal population & population variance is known, then $100(1-\alpha)\%$ confidence for population mean is given by

$$\bar{X} - \frac{z_{\alpha}}{2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{\alpha}}{2} \cdot \frac{\sigma}{\sqrt{n}}$$

Where $P(z < \frac{z_{\alpha}}{2}) = \frac{\alpha}{2}$ & random variable z has a standard normal distribution

Ex: If a random sample of size $n=20$ from a normal population with variance $\sigma^2=225$ has a sample mean $\bar{X}=64.3$. Construct a 95% confidence interval for population mean μ

$$1-\alpha = 0.95 \quad \alpha = 0.05 \quad \alpha/2 = 0.025$$

$$z_{0.025} = 1.96$$

$$P\left[64.3 - 1.96\left(\frac{15}{\sqrt{20}}\right) < \mu < 64.3 + 1.96\left(\frac{15}{\sqrt{20}}\right)\right] = 0.95$$

$$P(57.7 < \mu < 70.9) = 0.95$$

By help of central limit theorem, we can construct confidence intervals for non-normal populations if n is sufficiently large ($n \geq 30$)

In this case, we use sample standard deviation as an approximation as a population standard deviation

$$P\left(\bar{X} - \frac{z_{\alpha}}{2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{\alpha}}{2} \cdot \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

Suppose the population standard deviation σ is unknown, we want to estimate mean of a random sample from a normal population ($n < 30$)

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

$$P\left(-\frac{t_{\alpha}}{2}, n-1 < T < \frac{t_{\alpha}}{2}, n-1\right) = 1-\alpha$$

By substituting T , then

$$P\left(\bar{X} - \frac{t_{\alpha}}{2} \frac{S}{\sqrt{n}} < \mu < \bar{X} + \frac{t_{\alpha}}{2} \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

is a $100(1-\alpha)\%$ confidence interval for mean of population

Ex: A random sample of six cars from a particular model year had mean 19.4833 and $s^2 = 0.96$ fuel consumption. Find a 90% confidence interval for population mean fuel consumption for cars of this model year, assuming population distribution is normal

$$\alpha = 0.10 \rightarrow t_{0.05, 5} = 2.015$$

$$n = 6$$

$$n-1 = 5$$

$$s^2 = 0.96 \quad s = \sqrt{0.96} = 0.98$$

$$P\left(19.4833 - 2.015 \left(\frac{0.98}{\sqrt{6}}\right) < \mu < 19.4833 + 2.015 \left(\frac{0.98}{\sqrt{6}}\right)\right)$$

$$= 0.9$$

$$P(18.67 < \mu < 20.29) = 0.90$$

Estimation of Differences Between Means for Independent Random Samples from Normal Populations

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

has a standard normal distribution

$$P\left(-\frac{z_{\alpha}}{2} < Z < \frac{z_{\alpha}}{2}\right) = 1-\alpha$$

$$P \left[\bar{X}_1 - \bar{X}_2 - \frac{z_{\alpha}}{2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + \frac{z_{\alpha}}{2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] = 1 - \alpha$$

is $100(1-\alpha)\%$ confidence interval for differences bw. two population means with known variances. By central limit theorem this confidence interval formula can be used independent random samples from non-normal populations with known variances when n_1 & n_2 are large ($n_1 \geq 30$, $n_2 \geq 30$)

In order to construct a confidence interval for differences between two means when $n_1 \geq 30$, $n_2 \geq 30$ but σ_1^2 & σ_2^2 are unknown, then we substitute sample S_1^2 & S_2^2 for σ_1^2 & σ_2^2

$$P \left[\bar{X}_1 - \bar{X}_2 - \frac{z_{\alpha}}{2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + \frac{z_{\alpha}}{2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right] = 1 - \alpha$$

When σ_1^2 & σ_2^2 are unknown and samples are small ($n_1 < 30$, $n_2 < 30$)

Pooled estimator for variance is used

$$s_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

$$\sigma_1^2 = \sigma_2^2 \Rightarrow \text{Assume}$$

If $\bar{x}_1$ & $\bar{x}_2$, S_1 & S_2 are values of s.t. devs of independent random samples of size n_1 & n_2 from normal populations with equal variances then

$$P(\bar{X}_1 - \bar{X}_2) - \frac{t_{\alpha}}{2} \cdot (n_1 + n_2 - 2) \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 <$$

$$(\bar{X}_1 - \bar{X}_2) + \frac{t_{\alpha}}{2} \cdot (n_1 + n_2 - 2) \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1 - \alpha$$

is $100(1-\alpha)\%$ confidence interval for differences between two population means

Ex: In a study of research & development activities on profit of banks, a random sample of six banks having a high level of R & D expenditures showed mean annual percentage increase in profit of 9.972 & s.d. of 7.470. An independent random sample of nine banks with a low level of R & D expenditures show a mean annual % increase in profit of 2.098 & s.d. of 10.834. Assuming two population distributions are normal with same variance, find 0.90 confidence interval for difference between their means.

$$\begin{aligned} n_1 &= 6 & \bar{x}_1 &= 9.972 & S_1 &= 7.470 \\ n_2 &= 9 & \bar{x}_2 &= 2.098 & S_2 &= 10.834 \\ \alpha &= 0.10 & t_{\alpha/2, n_1+n_2-2} &= t_{0.05, 13} = 1.771 \end{aligned}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{(6 - 1)(7.470)^2 + (9 - 1)(10.834)^2}{9 + 6 - 2} = 93.693$$

$$S_p = \sqrt{93.693} = 9.680$$

$$P \left[(9.972 - 2.098) - (1.771)(9.680) \sqrt{\frac{1}{6} + \frac{1}{9}} < (\mu_1 - \mu_2) < \right.$$

$$\left. (9.972 - 2.098) + (1.771)(9.680) \sqrt{\frac{1}{6} + \frac{1}{9}} \right] = 0.90$$

$$\left[-1.161 < \mu_1 - \mu_2 < 16.909 \right] = 0.90$$