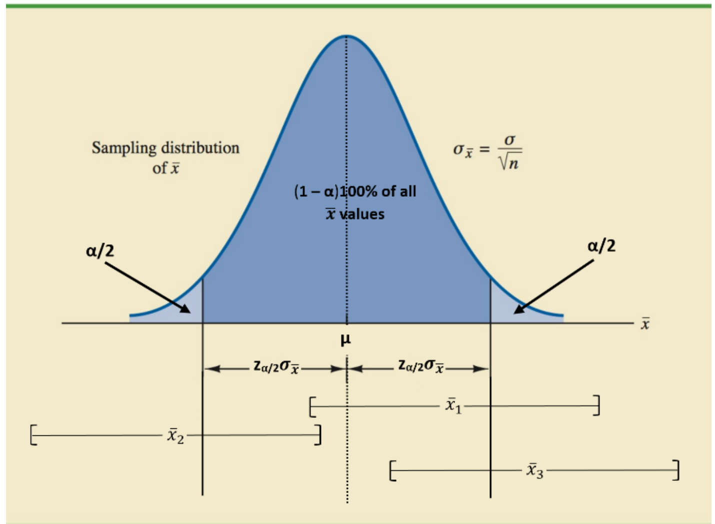


Chapter 8: Interval Estimation

Interval estimate = Point estimate ± Margin of error

8.1 Population mean: σ known

(1 − α)100% confidence interval:

$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$ (1)

where

(1 − α)100% = confidence level

(1 − α) = confidence coefficient

$z_{\alpha/2}\sigma_{\bar{x}}$ = margin of error

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$ (see Ch.7, assuming n/N ≤0.5)

$z_{\alpha/2}$ = z value such that $P(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = 1 - \alpha$

or $\{P(z \leq -z_{\alpha/2}) = \alpha/2$ and $P(z \geq +z_{\alpha/2}) = \alpha/2\}$

NOTE: When $P(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = 1 - \alpha$ and the sampling distribution of $\bar{x}$ is normally distributed, we also have $P(\mu - z_{\alpha/2}\sigma_{\bar{x}} \leq \bar{x} \leq \mu + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$: (1 − α)100% of the possible interval estimates, $\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$, will include the population mean μ.

Practical Advice: If the population follows a normal distribution, the above is an exact result. If not, it is a good approximate one, for, in most applications, sample sizes n ≥ 30. If the population is not normally distributed but is roughly symmetric, samples sizes as small as 15 can be expected to provide good approximate confidence intervals. With smaller sample sizes, (1) should only be used if the analyst believes, or is willing to assume, that the population distribution is at least approximately normal.

8.2 Population means: σ *un*known

Degrees of freedom:

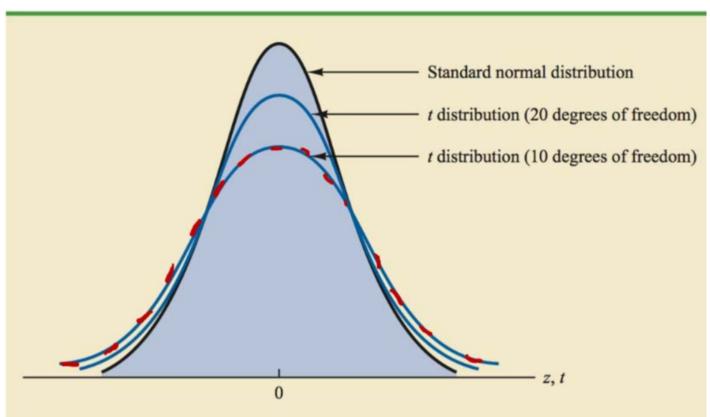
Refer to the number of independent pieces of information that go into the computation of a point estimator or population parameter.

Chi-square distribution:

If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then $Y = \Sigma_i Z_i^2$ is said to have the chi-squared (χ^2) distribution with n degrees of freedom (Ross, Example 3b, p. 242).

***t*-distribution**:

If Z and Y are independent, with Z having a standard normal distribution and Y having a chi-squared distribution with *n* degrees of freedom, then the random variable T defined by $T = \frac{Z}{\sqrt{Y/n}}$, is said to have a *t*-distribution with n degrees of freedom (Ross, Example 5c, p. 252).



(1 − α)100% confidence interval:

$\bar{x} \pm t_{\alpha/2}\, \sigma_{\bar{x}}$ (2)

where

$(1 - \alpha)100\%$ = confidence level

$(1 - \alpha)$ = confidence coefficient

$t_{\alpha/2}\, \sigma_{\bar{x}}$ = margin of error

$\sigma_{\bar{x}} = s/\sqrt{n}$, where s = sample standard deviation (see Ch.3)

$t_{\alpha/2}$ = t value such that $P(-t_{\alpha/2} \le t \le +t_{\alpha/2}) = 1 - \alpha$ or $\{P(t \le -t_{\alpha/2}) = \alpha/2$ and $P(t \ge +t_{\alpha/2}) = \alpha/2\}$ for the t-distribution with $n - 1$ degrees of freedom.

NOTE: Although the mathematical development of the *t* distribution is based on the assumption of a normal distribution for the population we are sampling from, research shows that the t distribution can be successfully applied in many situations where the population deviates significantly from normal.
NOTE: The reason the number of degrees of freedom associated with the *t* value in (2) is $n - 1$ concerns the use of *s* as an estimate of the population standard deviation σ: for details see p. 357.

Practical Advice: If the population follows a normal distribution, (2) is exact. If not, it is a good approximate one, for, in most applications, sample sizes n ≥ 30. If the population distribution is highly skewed, most statisticians would recommend increasing the sample size to 50 or more. If the population is not normally distributed but is roughly symmetric, samples sizes as small as 15 can be expected to provide good approximate confidence intervals. With smaller sample sizes, (2) should only be used if the analyst believes, or is willing to assume, that the population distribution is at least approximately normal.

Substitute planning value for σ, if it is unknown. See p. 364 for details.

8.3 Determining the sample size to provide a desired margin of error E, at a chosen confidence level

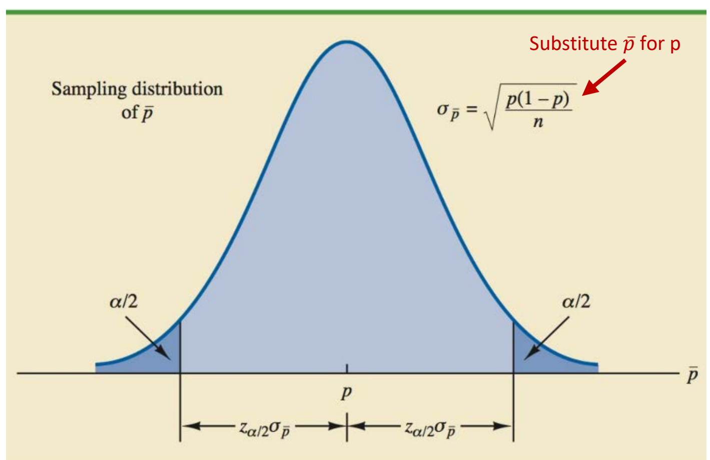
$E = z_{\alpha/2}\, \sigma_{\bar{x}} = z_{\alpha/2}(\sigma/\sqrt{n}) \Leftrightarrow n = [z_{\alpha/2}(\sigma/E)]^2 = (z_{\alpha/2})^2\sigma^2/E^2$

Substitute p* for p, where p* is its planning value. See pp. 368-69 for details.

$E = z_{\alpha/2}\, \sigma_{\bar{p}} = z_{\alpha/2}(\sqrt{\bar{p}(1-\bar{p})}/\sqrt{n}) \Leftrightarrow n = [z_{\alpha/2}(\sqrt{\bar{p}(1-\bar{p})}/E)]^2 = (z_{\alpha/2})^2\bar{p}(1-\bar{p})/E^2$

Note: If *n* is not an integer, round up to the next integer value to get the recommended sample size.

8.4 Population proportion (*p* unknown)



(1 − α)100% confidence interval:

$\bar{p} \pm z_{\alpha/2}\sigma_{\bar{p}}$ (3)

where

$(1 - \alpha)100\%$ = confidence level

$(1 - \alpha)$ = confidence coefficient

$z_{\alpha/2}\sigma_{\bar{p}}$ = margin of error

$\sigma_{\bar{p}} = \sqrt{\bar{p}(1-\bar{p})/n}$ (see Ch.7, assuming $n/N \le 0.5$; because *p* is *un*known, we substitute $\bar{p}$ for p)

$z_{\alpha/2}$ = z value such that $P(-z_{\alpha/2} \le z \le +z_{\alpha/2}) = 1 - \alpha$ or $\{P(z \le -z_{\alpha/2}) = \alpha/2$ and $P(z \ge +z_{\alpha/2}) = \alpha/2\}$

NOTE: When $P(-z_{\alpha/2} \le z \le +z_{\alpha/2}) = 1 - \alpha$ and the sampling distribution of $\bar{p}$ is normally distributed, we also have $P(p - z_{\alpha/2}\sigma_{\bar{p}} \le \bar{p} \le p + z_{\alpha/2}\sigma_{\bar{p}}) = 1 - \alpha$: $(1 - \alpha)100\%$ of the possible interval estimates, $\bar{p} \pm z_{\alpha/2}\sigma_{\bar{p}}$, will include the population proportion *p*.