Confidence Intervals

- A confidence interval for the mean is an estimate interval constructed with respect to the sample mean by which the likelihood that the interval includes the value of the population mean can be specified.
- The *level of confidence* associated with a confidence interval indicates the long-run percentage of such intervals which would include the parameter being estimated.
  - Confidence intervals for the mean typically are constructed with the unbiased estimator $\bar{x}$ at the midpoint of the interval.
  - The $\pm Z\sigma_x$ or $\pm Zs_x$ frequently is called the ***margin of error*** for the confidence interval.
- We indicated that use of the normal distribution in estimating a population mean is warranted for any large sample ($n > 30$), **and** for a small sample ($n \leq 30$) only if the population is normally distributed and σ is known.
- Now we consider the situation in which the sample is small and the population is normally distributed, but σ is not known.
- The distribution is a family of distributions, with a somewhat different distribution associated with the degrees of freedom (df). For a confidence interval for the population mean based on a sample of size n, $df = n - 1$.
- A confidence interval gives us some idea of the range of values which an unknown population parameter (such as the mean or variance) is likely to take based on a given set of sample data.
- Sometimes we are interested in the proportion of responses that fall into one of two categories.
  - For example, a firm may wish to know what proportion of their customers pay by credit card as opposed to those who pay by cash;
  - The manager of a TV station may wish to know what percentage of households in a certain town have more than one TV set;
  - A doctor may be interested in the proportion of patients who benefited from a new drug as opposed to those who didn't, etc.

  A confidence interval for a proportion would specify a range of values within which the true population proportion may lie, for such examples.
- The procedure for obtaining such an interval is based on the proportion, p of a sample from the overall population.

0.0.1 Confidence Intervals for Sample Proportion

Unlike confidence intervals for sample means, there is only one type of confidence interval when dealing with sample proportions.

Optional

- It is often easier to work in terms of percentages, rather than proportions. If you are working in terms of percentages, remember to use the appropriate ***complement value*** in the standard error formula (i.e. $100 - \hat{p}$)

- The standard error, in the case of sample proportions, is

$$\text{S.E.}(\hat{p}) = \sqrt{\frac{\hat{p} \times (100 - \hat{p}}{n}}$$

- Complement Values:

  - When working in terms of proportions, for the the value $\hat{p} = 0.35$, the complement value is $1 - \hat{p} = 0.65$.
  - When working in terms of percentages, for the value $\hat{p} = 35\%$, the complement value is $100 - \hat{p} = 65\%$.

0.1 Confidence Intervals for Sample Proportion

Confidence Intervals may also be computed for **Sample Proportions**.
The sample proportion is used to estimate the value of a population proportion. The sample proportion is denoted $\hat{p}$. The population proportion is denoted π.

Confidence Intervals for Sample Proportion

- The Structure of a confidence interval for sample proportion is

$$\hat{p} \pm z_{(\alpha/2)} \times \text{S.E.}(\hat{p})$$

- The standard error, in the case of sample proportions, is

$$\text{S.E.}(\hat{p}) = \sqrt{\frac{hatp \times (1 - \hat{p})}{n}}$$

- (When computing this interval with statistical software, it is common to enhance the solution using a **continuity correction** . This is not part of our syllabus.)

Point Estimates for proportions Sample Percentage

$$\hat{p} = \frac{x}{n} \times 100\%$$

- $\hat{p}$ - sample proportion.
- x - number of "successes".
- n - the sample size.

- The general structure of confidence intervals is as follows

$$\text{Point Estimate } \pm [\text{ Quantile } \times \text{ Standard Error }]$$

- The structure of a confidence interval for sample proportion is

$$\hat{p} \pm \left[z_{(\alpha/2)} \times \text{S.E.}(\hat{p})\right]$$

Point Estimate:

- The point estimate is the sample proportion, denoted $\hat{p}$.
- The sample proportion is calculated as the number of 'successes' (x) divided by the total number of cases, in other words, the sample size n.

Quantile:

- In the cases of large samples ($n > 30$) , the standard normal ('Z') distribution is used.

Standard Error for Proportions:

The standard error for proportions is computed using this formula.

$$S.E.(\hat{p}) \;=\; \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}}$$

When expressing the proportion as a percentage, we adjust the standard error accordingly.

$$S.E.(\hat{p}) \;=\; \sqrt{\frac{\hat{p} \times (100-\hat{p})}{n}}$$

0.1.1 Sample Proportion : Example

Point Estimate The sample proportion is computed as follows

$$\hat{p} = \frac{x}{n} = \frac{84}{120} = 0.70$$

Quantile We are asked for a 95% confidence interval. We have a large sample ($n = 120$). The quantile is therefore 1.96.

$$z_{\alpha/2} = 1.96$$

Standard Error The standard error, with sample size n=120 is computed as follows

$$\text{S.E.}(\hat{p}) = \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}} = \sqrt{\frac{0.70 \times 0.30}{120}}$$

Unlike confidence intervals for sample means, there is only one type of confidence interval when dealing with sample proportions.

0.1.2 Confidence Interval for a proportion

n = 400
240

$$\hat{P} = \frac{240}{400} = 0.60$$

$$S.E.\ (\hat{P}) = \sqrt{\frac{\hat{P}\ \times\ (1 - \hat{P})}{n}} = \sqrt{\frac{0.60\ \times\ 0.40}{400}}$$

$$S.E.\ (\hat{P}) = \sqrt{\frac{60\%\ \times\ 40\%}{400}}$$

0.1.3 Standard Error

- The standard error measures the dispersion of the sampling distribution.
- For each type of point estimate, there is a corresponding standard error.
- A full list of standard error formulae will be attached in your examination paper.
- The standard error for a proportion (for confidence intervals only) is

$$S.E(p) = \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$$

- When expressing the proportion as a percentage, we adjust the standard error accordingly.

$$S.E.(p) \ = \ \sqrt{\frac{\hat{p} \times (100 - \hat{p})}{n}}$$

Standard Error for Proportions

- When computing the standard error for computing the confidence intervals, one would use the point estimate $\hat{p}$ in their calculation.
- When computing the standard error for computing the hypothesis test, one would use the value for the proportion expected under the null hypothesis π_0 in that calculation.
- For Hypothesis tests The standard error for proportions is computed using this formula.

$$S.E.(p) \ = \ \sqrt{\frac{\pi_o \times (1 - \pi_o)}{n}}$$

In an election campaign, a campaign manager requests that a sample of votes be polled to determine public support for a candidate. In a sample of 150 votes 72 expressed plans to support the candidate.

What is the point estimate of the proportion of the voters who will support the candidate in the election?

contruct and interpret a 95% confidence interval for the proportion of votes in the population that support the candidate.

given the confidennce interval, is the campaign manager justified in feeling confident that the candidate has at least 50

$$S.E.(\hat{P}) = \sqrt{\frac{\hat{p}(1-\hat{p}}{n}}$$

0.2 CI for Proportion: Example (1)

- $\hat{p} = 0.62$
- Sample Size $n = 250$
- Confidence level $1 - \alpha$ is 95%

0.3 CI for Proportion: Example (2)

- First, lets determine the quantile.
- The sample size is large, so we will use the Z distribution.
- (Alternatively we can uses the $t-$ distribution with ∞ degrees of freedom.

Although the sample mean is useful as an unbiased estimator of the population mean, there is no way of expressing the degree of accuracy of a point estimator. In fact, mathematically speaking, the probability that the sample mean is exactly correct as an estimator of the population mean is $P = 0$.

0.4 Computing the Standard Error

$$S.E.(\hat{p}) \;=\; \sqrt{\frac{\hat{(p)} \times (100 - \hat{p})}{n}}$$

$$\hat{p} = 144/200 \times 100\% = 0.72 \times 100\%. = 72$$

$100\% - \hat{p} = 100\% - 72\% = 28\%$
Computing the Standard Error

$$S.E.(\hat{p}) \;=\; \sqrt{\frac{72 \times 28}{200}}$$

Standard Error for Proportions

The standard error for proportions is computed using this formula.

$$S.E.(\hat{p}) \;=\; \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$$

When expressing the proportion as a percentage, we adjust the standard error accordingly.

$$S.E.(\hat{p}) \;=\; \sqrt{\frac{\hat{p} \times (100 - \hat{p})}{n}}$$

0.5 Confidence Intervals: Example

From a sample of 500 analytics professional, 372 reported that `R` was their primary programming language. Let p be the proportion pf all analytics professionals who regard `R` as their primary programming language. Estimate p and provide a 95% confidence interval for p.

0.5.1 Example

In an election campaign, a campaign manager requests that a sample of votes be polled to determine public support for a candidate. In a sample of 150 votes 72 expressed plans to support the candidate.

- What is the point estimate of the proportion of the voters who will support the candidate in the election?
- Contruct and interpret a 95% confidence interval for the proportion of votes in the population that support the candidate.
- Given the confidennce interval, is the campaign manager justified in feeling confident that the candidate has at least 50% support

$$S.E.(\hat{P}) = \sqrt{\frac{\hat{p}(1 - \hat{p}}{n}}$$