Basic Statistics

Probability and Confidence Intervals
Probability and Confidence Intervals

Learning Intentions

Today we will understand:

- Interpreting the meaning of a confidence interval
- Calculating the confidence interval for the mean with large and small samples
Probability and Confidence Intervals

- An important role of statistics is to use information gathered from a sample to make statements about the population from which it was chosen.

- Using samples as an estimate of the population.

- How good of an estimate is that sample providing us with?

Image accessed: https://www.youtube.com/watch?v=rckB8T8WthM
Estimators of a Population

- **Point estimate** is a single value that best describes the population of interest.

- Sample mean is the most common point estimate.

- **Interval estimate** provides a range of values that best describes the population.

Point Estimate

- Single value that best describes the population of interest
- Sample mean is most common point estimate
- Easy to calculate and easy to understand
- Gives no indication of how accurate the estimation really is

Interval Estimate

- To deal with uncertainty, we can use an interval estimate
- Provides a range of values that best describe the population
- To develop an interval estimate we need to learn about confidence levels

Confidence Levels

- A confidence level is the probability that the interval estimate will include the population parameter (such as the mean)

- A parameter is a numerical description of a characteristic of the population

*Remember - Standard Normal Distribution

- Normal distribution with $\mu = 0$ and $SD = 1$
Confidence Levels

- Sample means will follow the normal probability distribution for large sample sizes \( (n \geq 30) \)

- To construct an interval estimate with a 90% confidence level

- Confidence level corresponds to a z-score from the standard normal table equal to 1.645

A confidence interval is a range of values used to estimate a population parameter and is associated with a specific confidence level.

Construct confidence interval around a sample mean using these equations:

\[
\bar{x} \pm z \sigma_x
\]
Confidence Intervals

\[ \bar{x} \pm z \sigma_{\bar{x}} \]

Where:

\( \bar{x} \) = the sample mean

\( z \) = the z-score, which is the number of standard deviations based on the confidence level

\( \sigma_{\bar{x}} \) = the standard error of the mean
Confidence Intervals

- A **confidence interval** is a range of values used to estimate a population parameter and is associated with a specific confidence level.

- Associated with specific confidence level.

- Needs to be described in the context of several samples.

Confidence Intervals

- Select 10 samples and construct 90% confidence intervals around each of the sample means.

- Theoretically, 9 of the 10 intervals will contain the true population mean, which remains unknown.

Confidence Intervals

- Careful not to misinterpret the definition of a confidence interval

- NOT Correct – “there is a 90 % probability that the true population mean is within the interval”

- CORRECT – “there is a 90 % probability that any given confidence interval from a random sample will contain the true population mean”
Level of Significance

- As there is a 90% probability that any given confidence interval will contain the true population mean, there is a 10% chance that it won’t.

- This 10% is known as the level of significance ($\alpha$) and is represented by the purple shaded area.

Level of Significance

- Level of significance ($\alpha$) is the probability of making a type 1 error (next week)

- The probability for the confidence interval is a complement to the significance level

- A $(1 - \alpha)$ confidence interval has a significance level equal to $\alpha$
When $\sigma$ is Unknown

- So far our examples have assumed we know $\sigma$ - the population standard deviation.

- If $\sigma$ is unknown we can substitute $s$ (sample standard deviation) for $\sigma$.

- $n \geq 30$

- We use $\frac{\hat{\sigma}}{\bar{X}}$ to show we have approximated the standard error of the mean by using $s$ instead of $\sigma$. 
Using Excel

- You can calculate confidence intervals in Excel

- `CONFIDENCE(alpha, standard_dev, size)`

Where:

Alpha = the significance level
Standard_dev = standard deviation of the population
Size = sample size
Using Excel

![Excel screenshot showing the 'Insert Function' dialog box]

- Click the 'fx' button in the 'FORMULAS' tab to open the 'Insert Function' dialog box.
- Search for a function: 'Confidence' and select 'Recommended' category.
- Select 'CONFIDENCE' function.
- CONFIDENCE(α, standard_dev, size)
- This function is available for compatibility with Excel 2007 and earlier.
- Returns the confidence interval for a population mean, using a normal distribution.

[Help on this function]
Confidence Intervals for the Mean with Small Samples

- So far we have discussed confidence intervals for the mean where \( n \geq 30 \)

- When \( \sigma \) is known, we are assuming the population is normally distributed and so we can follow the procedure for large sample sizes

- When \( \sigma \) is unknown (more often the case!) we make adjustments
When $\sigma$ is Unknown – Small Samples

- Substitute $s$, sample standard deviation, for $\sigma$

- Because of the small sample size, this substitution forces us to use the t-distribution probability distribution

- Continuous probability distribution

- Bell-shaped and symmetrical around the mean

- Shape of curve depends on degrees of freedom (d.f) which equals $n - 1$
T-distribution

- Flatter than normal distribution

- As degrees of freedom increase, the shape of t-distribution becomes similar to normal distribution

- With more than 30 d.f. (sample size of 30 or more) the two distributions are practically identical