Basic Statistics

Probability and Confidence Intervals

Learning, Teaching
and Student Engagement

Probability and Confidence Intervals

JAMES COOK UNIVERSITY AUSTRALIA

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

- A Point estimate is a single value that best describes the population of interest
- Sample mean is the most common point estimate
- An Interval estimate provides a range of values that best describes the population

Image accessed: http://metropole.com.au/property-investors-need-know-population-growth/

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

Image accessed: http://www.carclipart.com/free_car_clipart/point.html

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

Image accessed: http://blog.thewellnesstrain.com/add-interval-training-to-maximize-time-and-fitness/

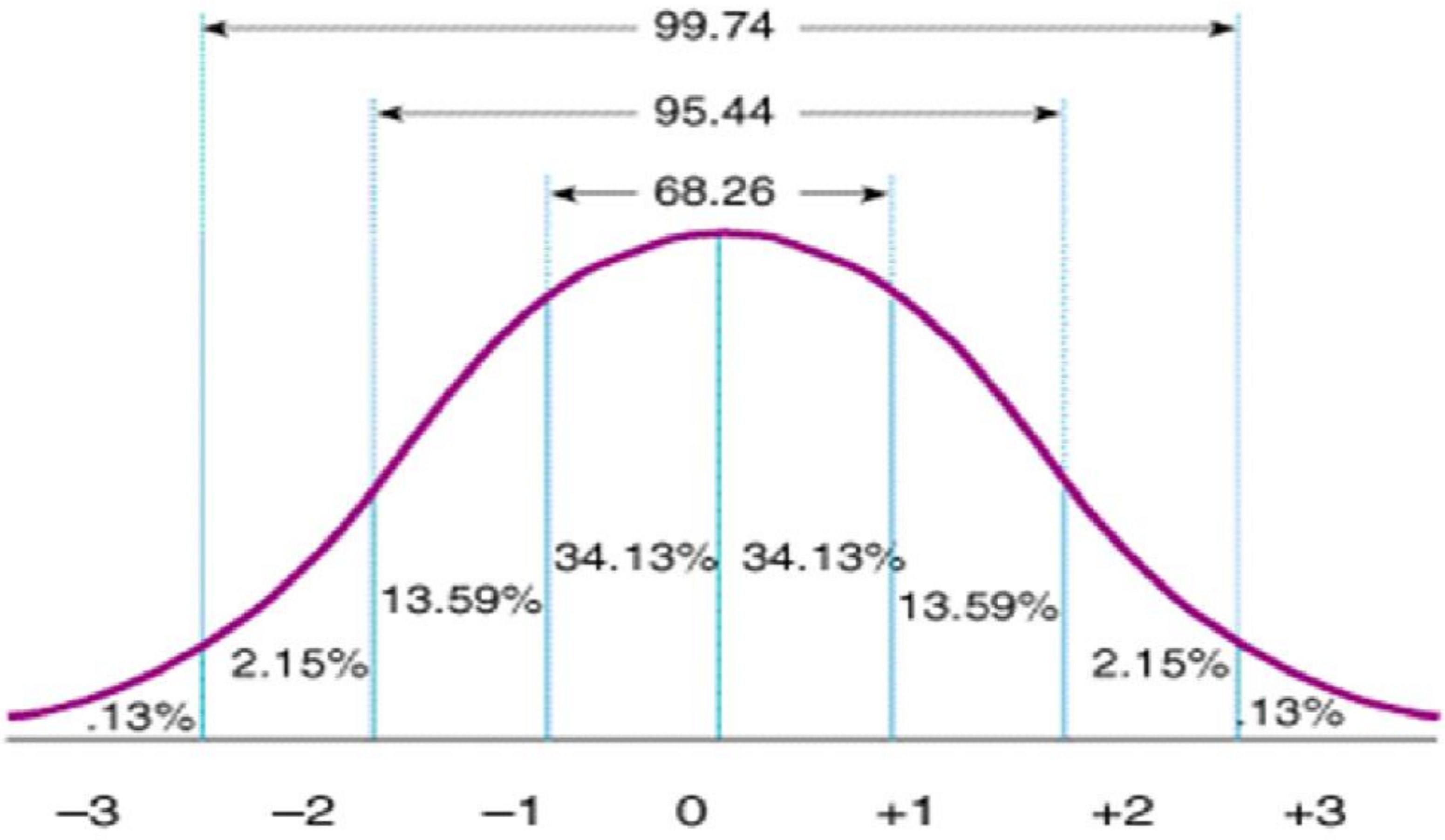
Confidence Levels

- 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

Image accessed: http://paul7brown.blogspot.com.au/2013/05/check-your-confidence-levels.html

*Remember - Standard Normal Distribution



- Normal distribution with μ = 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

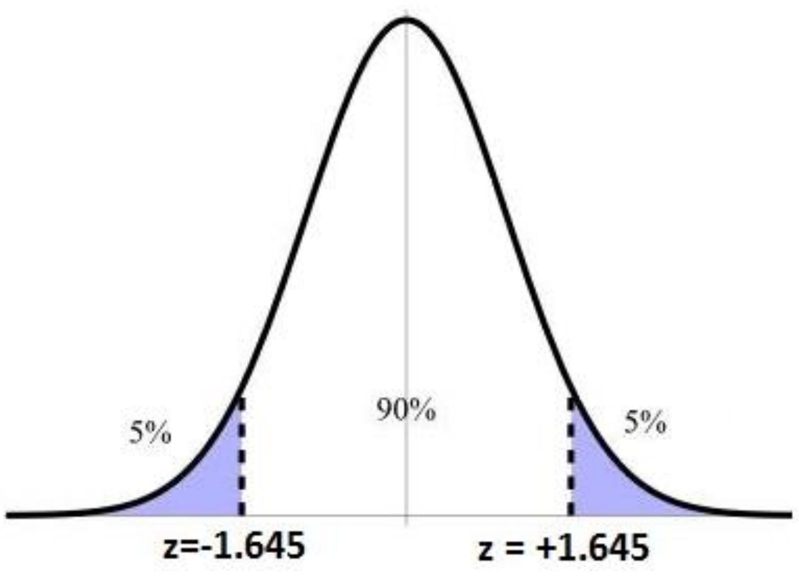


Image accessed: http://www.mathandstatistics.com/learn-stats/finding-z-critical-values

Confidence Intervals

- 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:

$$\overline{x} \pm z\sigma_{\bar{X}}$$

Confidence Intervals

$$\overline{x} \pm z\sigma_{\bar{X}}$$

Where:

$\overline{x}$ = 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

Image accessed: http://www.mathandstatistics.com/learn-stats/finding-z-critical-values

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

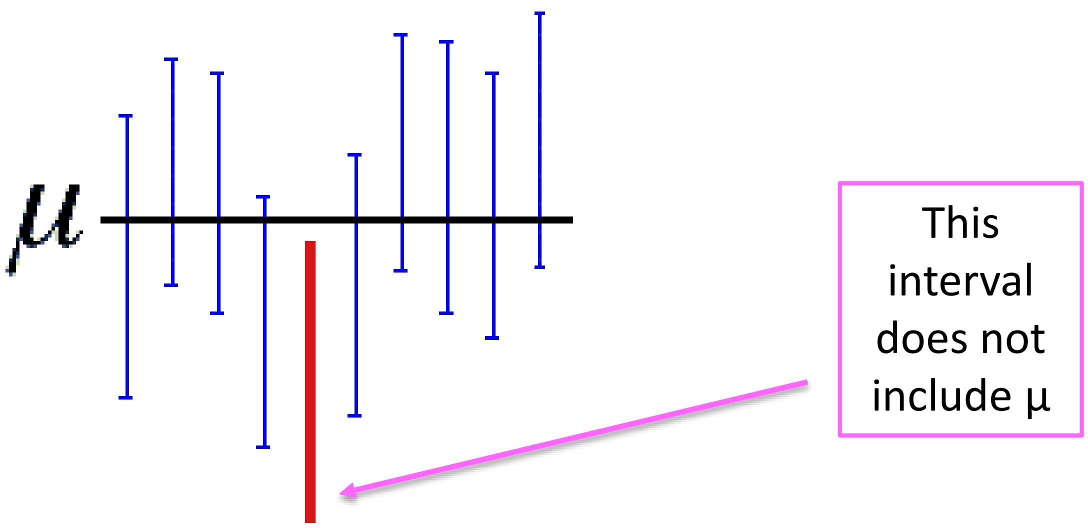


Image accessed: http://blog.minitab.com/blog/adventures-in-statistics/understanding-hypothesis-tests%3A-confidence-intervals-and-confidence-levels

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 (α) and is represented by the purple shaded area

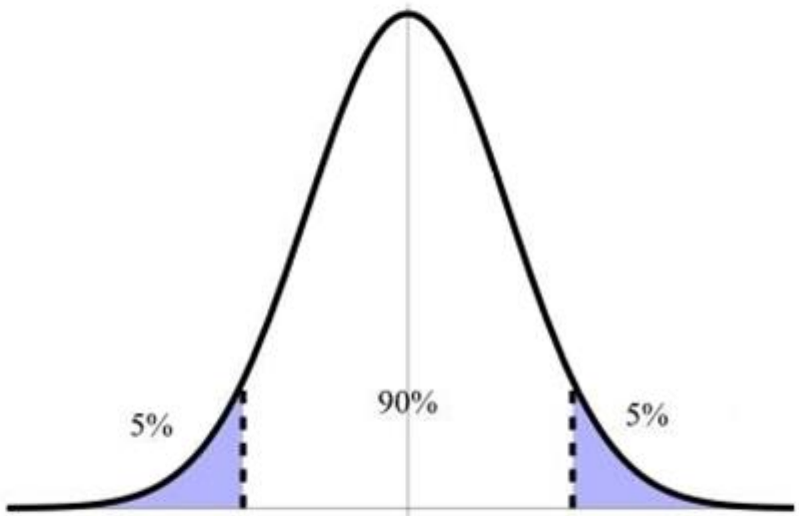


Image accessed: http://www.mathandstatistics.com/learn-stats/finding-z-critical-values

Level of Significance

- Level of significance (α) 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 – α) confidence interval has a significance level equal to α

Image accessed: http://www.vwmin.org/significance-level-definition-statistics-and-probability.html

When σ is Unknown

- So far our examples have assumed we know σ - the population standard deviation

- If σ is unknown we can substitute *s* (sample standard deviation) for σ

- $n \geq 30$

- We use $\hat{\sigma}_{\bar{X}}$ to show we have approximated the standard

error of the mean by using *s* instead of σ

Using Excel

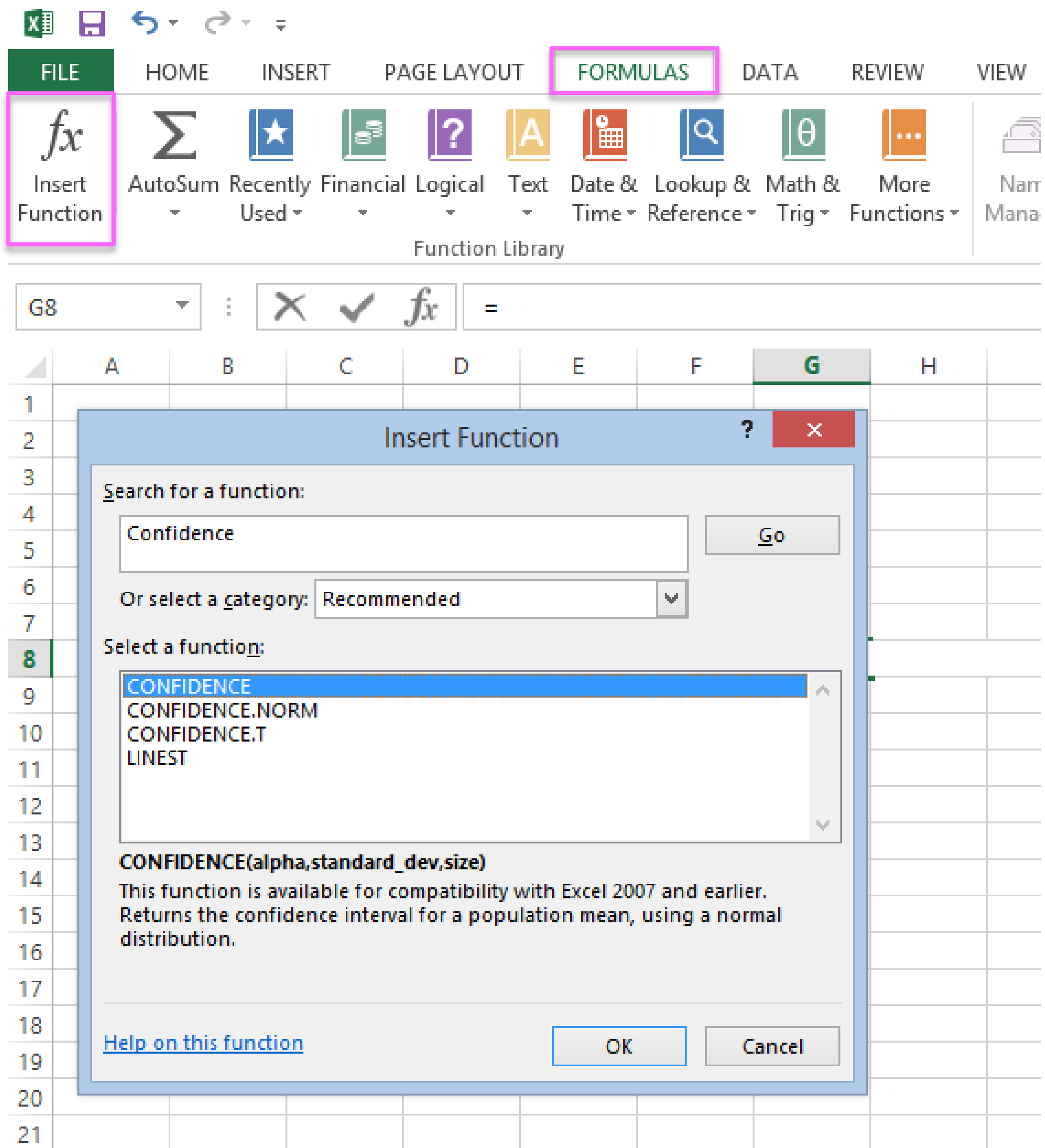
- You can calculate confidence intervals in Excel

- CONFIDENCE(alpha, standard_dev, size)

Where:

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

Using Excel


FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
REVIEW
VIEW
Insert Function
AutoSum
Recently Used
Financial
Logical
Text
Date & Time
Lookup & Reference
Math & Trig
More Functions
Function Library
G8
=
Insert Function
Search for a function:
Confidence
Go
Or select a category:
Recommended
Select a function:
CONFIDENCE
CONFIDENCE.NORM
CONFIDENCE.T
LINEST
CONFIDENCE(alpha,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
OK
Cancel

Confidence Intervals for the Mean with Small Samples

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

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

- When σ is unknown (more often the case!) we make adjustments

When σ is Unknown – Small Samples

JAMES COOK UNIVERSITY AUSTRALIA

- Substitute *s*, sample standard deviation, for σ
- 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