MM104/MM106/BM110
Statistics and Data Presentation

Lecture 6-2: Confidence Intervals for the population proportion

Chris Robertson

DEPARTMENT OF MATHEMATICS & STATISTICS

University of Strathclyde Science

Confidence interval for a proportion

Original (binomial) variable, X, with mean np and standard deviation $\sqrt{np(1-p)}$

X is the number of positive responses out of a series of n independent trials each of which can have two responses – positive or negative response

Sample proportion ($\hat{p}$) distribution approximately normal with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$

Large samples $np > 5$, $n(1-p) > 5$

Confidence interval for a proportion

Every sample, of size , n, generates one value for point estimator $\hat{p}$.

All samples share the same S.E., $\sqrt{\frac{p(1-p)}{n}}$

The true population proportion (p) is fixed and this is unknown.

Use $\hat{p}$ in the formula for the standard error

Estimated $\mathrm{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Confidence interval for a proportion

Point estimator

Standard error estimator

(Sample proportion) $\pm z_{\alpha/2} *$(Standard error sample proportion)

$$\hat{p} \pm z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

UPPER CONFIDENCE LIMIT

$$\hat{p} + z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

LOWER CONFIDENCE LIMIT

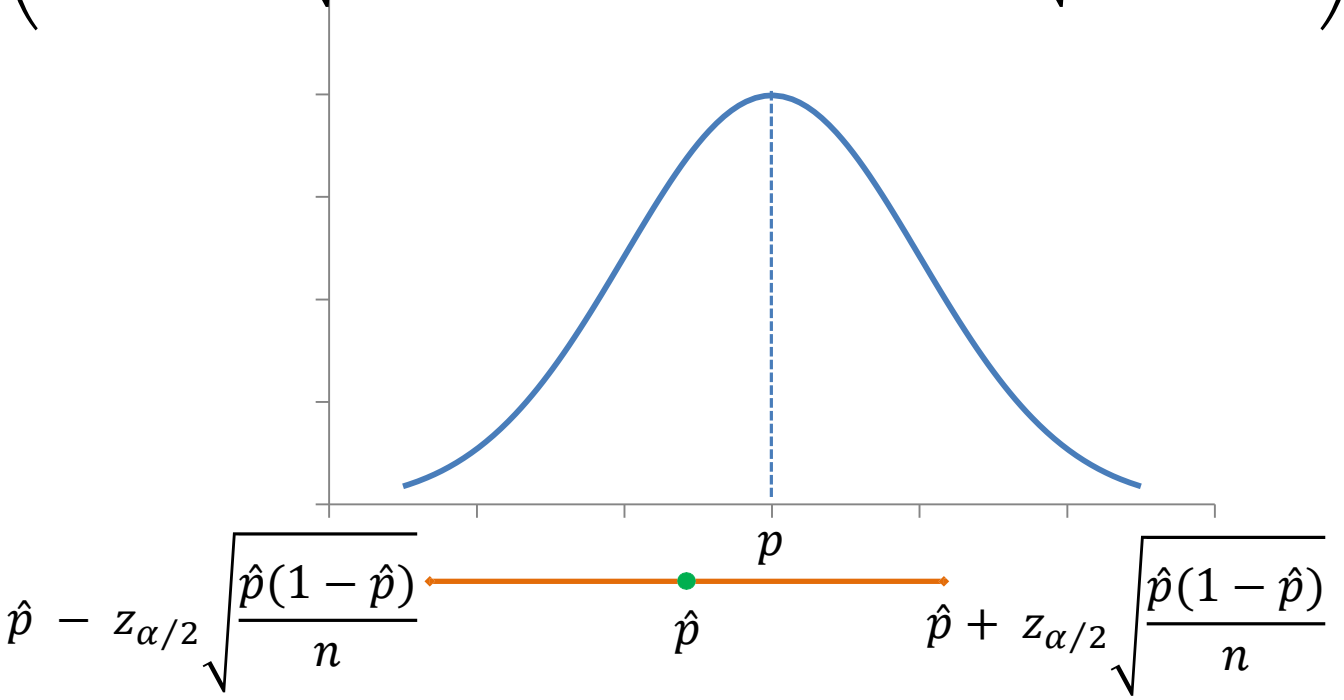
$$\hat{p} - z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Confidence interval for a proportion

$$P(LCL < p < UCL) = 1 - \alpha$$

Note: LCL and UCL vary, p is fixed

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$



$\hat{p} = x/n$

x is number of positive responses

Example

— In a survey of randomly selected students, a sample of 274 was obtained. Each student was asked if they like Game of Thrones©. The results below record 1 where the response was `yes', and 0 where the response was `no'. Find a 90% CI for the proportion who like Game of Thrones©:

0	1	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	1	0	0	0	1	1	0	0	0
0	0	1	0	0	1	0	1	0	0
0	0	0	0	1	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0
0	1	1	0	1	1	1	0	0	1
0	1	1	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	1	0	1
0	1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0
0	1	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	0
0	1	0	1	1	1	1	0	0	0
0	0	0	0	0	0	1	1	0	0
1	1	0	0	0	0	0	0	0	0
1	1	0	0	0	0	1	0	1	0
0	0	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	1
0	1	0	1	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	1	1	0	0	0	1	0
1	1	0	0						

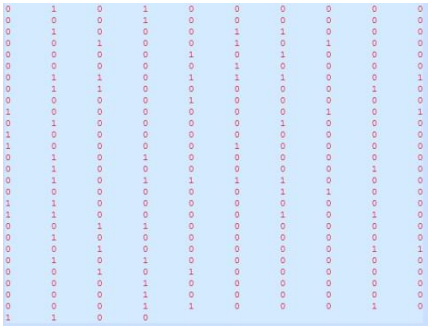
Example

This bit of the question is telling you that $n = 274$

In a survey of randomly selected students, a sample of 274 was obtained. Each student was asked if they like Game of Thrones©. The results below record 1 where the response was "yes", and 0 where the response was "no". Find a 90% CI for the proportion who like Game of Thrones©:

This bit of the question is telling you that

$1 - \alpha = 0.90$



Success: picking a student who likes Game of Thrones©

Variable: proportion of successes after n students

$\hat{p}$ for this sample??

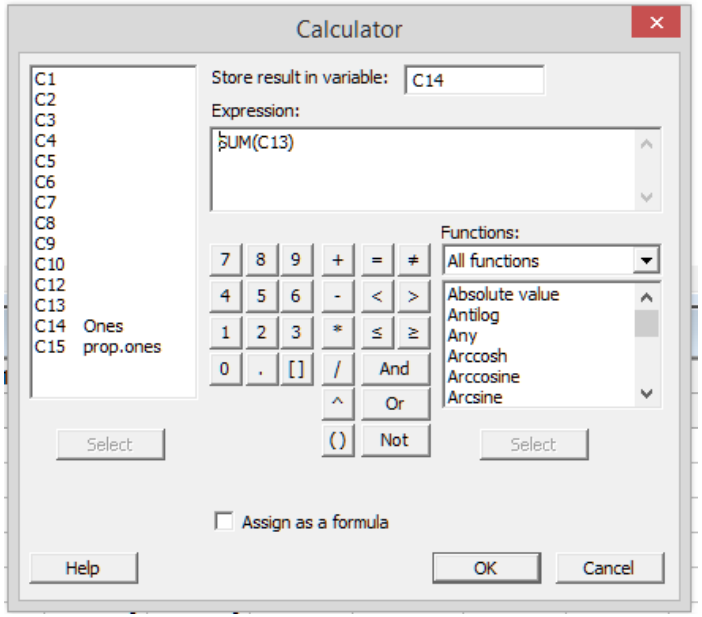
This is going to be the number of 1's in the sample divided by the total number in the sample (274)

Example

0	1	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	1	0	0	0	1	1	0	0	0
0	0	1	0	0	1	0	1	0	0
0	0	0	0	1	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0
0	1	1	0	1	1	1	0	0	1
0	1	1	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	1	0	1
0	1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0
0	1	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	0
0	1	0	1	1	1	1	0	0	0
0	0	0	0	0	0	1	1	0	0
1	1	0	0	0	0	0	0	0	0
1	1	0	0	0	0	1	0	1	0
0	0	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	1
0	1	0	1	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	1	1	0	0	0	1	0
1	1	0	0						

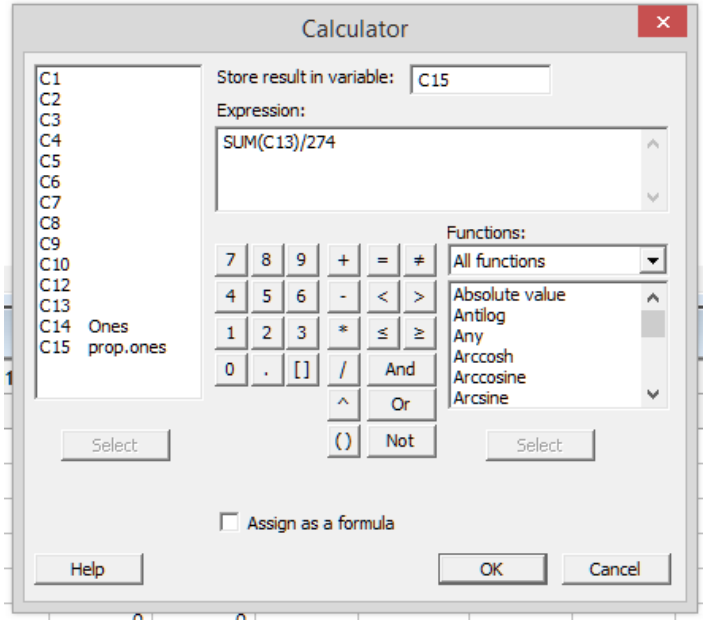
Use Minitab
Read Data into C1-C10 and stack into column c13
Use Minitab CALCULATOR
Store results in C14 and C15

Sum(c13) gives you the number of 1's in the sample – stored in c14



Ones	Proportion
64	0.233577

$$\hat{p} = \frac{64}{274} = 0.233577$$



$z_{\alpha/2}$ when $\alpha = 0.1$

Statistical Tables | Probability | Inverse

Normal | t

Tail

- ○ Lower
- ◉ Upper
- ○ Both

p

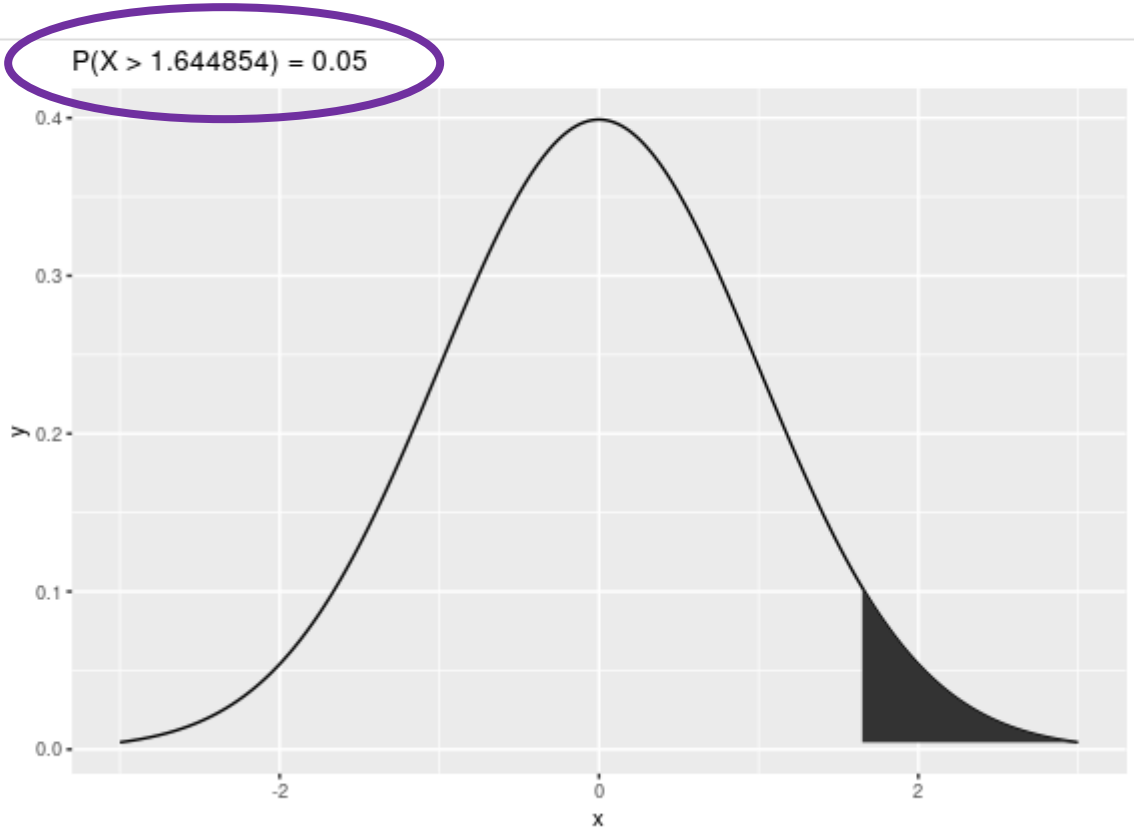
0.05

Mean

0

sd

1



Example

$$\sqrt{\frac{0.2336 \times (1 - 0.2336)}{274}} = 0.02556$$

$$\hat{p} = \frac{64}{274} = 0.233577$$

$$\hat{p} \pm z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$(1 - \alpha) = 0.90;$

$z_{\alpha/2} = 1.6448536$

UPPER C.L

$$\hat{p} + z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 0.276$$

LOWER C.L.

$$\hat{p} - z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 0.192$$

The proportion of students who like Game of Thrones is 0.234 with 90% confidence interval of (0.192, 0.276)

The interval (0.192, 0.276) contains the true, unknown, proportion with 90% confidence.

Key Points

- Confidence limits for a proportion are given by

$$\hat{p} \pm z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

- The population proportion is unknown so use the sample proportion in the calculation of the estimate the standard error
- The width of the confidence interval will decrease as the sample size increases.
- As n gets bigger the SE decreases and the width will decrease.
- Hence the precision of the estimate will increase