# MM104/MM106/BM110
# Statistics and Data Presentation

## Lecture 6-3:
## Confidence Intervals
## Sample Size

Chris Robertson

DEPARTMENT OF MATHEMATICS & STATISTICS

University of Strathclyde Science

# Sample size

# Setting

At Public Health Scotland I am planning to carry out a study to estimate the average length of time that a Covid 19 patient is in hospital for.

This is to be used for health service planning over the winter of 2020-21.

From previous data I know that patients are in hospital for varying lengths of time and that the standard deviation is about 7 days.

How large a sample of patients do I need to collect data on to ensure that I estimate the mean time for covid 19 patients to within plus or minus 1 day with 95% confidence

This is a practical problem in statistics – sample size calculation – and it is needed to inform policy

There is some previous data giving an estimate of the standard deviation to use

# Sample size when estimating a mean

- Remember: CI width depends on sample size.
- Given a precision and confidence levels, we can calculate how large sample size needs to be to reach that precision.
- We want sample mean to be within $B$ units of population mean,
  with confidence level $100(1-\alpha)\%$.

- In example of covid 19 patients B=1 day

# Sample size when estimating a mean

The Margin of error is the (CI width/2)

This is less than or equal to $B$:

$$\bar{x} \pm \boxed{z_{\alpha/2} * \left(\frac{\sigma}{\sqrt{n}}\right)} \leq B$$

The is the half width of the CI and is usually how you specify the precision – estimating the mean to within B units

$$z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \leq B$$

$$n \geq \left(z_{\alpha/2}\frac{\sigma}{B}\right)^2$$

# Covid 19 Example

$$z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le B$$

From the specification in the setting

$$n \ge \left(z_{\alpha/2}\frac{\sigma}{B}\right)^2$$

B = 1 day

95% CI so $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$

$\sigma = 1$

Put these quantities into the formula

$$n \ge \left(z_{\alpha/2} \times \frac{\sigma}{B}\right)^2 = \left(1.96 \times \frac{7}{1}\right)^2 = 188.2$$

So we need a sample of at least 189 patients.

# Sample size when estimating a proportion

- Exactly the same principles except that we need a guess as to the value of p
- We want sample proportion to be within $B$ of pop. proportion, with confidence level $100(1-\alpha)\%$.

Margin of error (CI width/2) less or equal to $B$:

$$\hat{p} \pm \boxed{z_{\alpha/2} * \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)} \le B \Rightarrow n \ge \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{\mathrm{B}}\right)^2$$

# Example - Proportion

- Last week we tried to assess how many people we would need to interview to get a reliable poll using the “Brexit” vote as a reference (48.1% electorate voted “Remain”). I want to achieve a confidence level of 90% that the test survey result is within 0.01 of the correct proportion.

What is the minimum sample size it needs to achieve this (assuming that nobody changed their minds)?

# Example - Proportion

This is an estimate $\hat{p}$ from a previous study

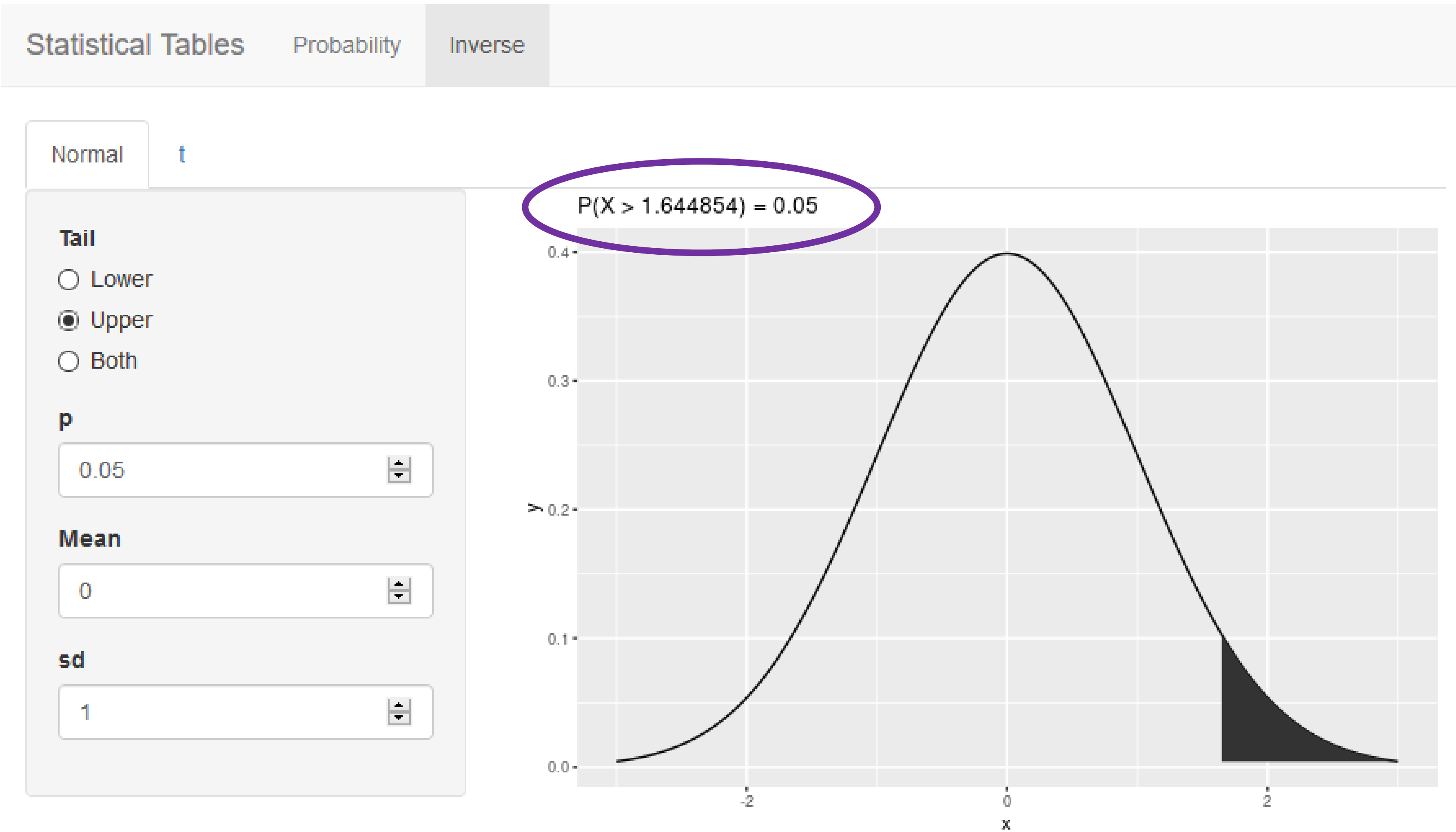
$(1-\alpha) = 0.90$

- Last week we tried to assess how many people we would need to interview to get a reliable poll using the "Brexit" vote as a reference (48.1% electorate voted "Remain"). I want to achieve a confidence level of 90% that the test survey result is within 0.01 of the correct proportion.

Margin of error $B$

What is the minimum sample size it needs to achieve this (assuming that nobody changed their minds)?

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



# Example - proportion

What is the minimum sample size it needs to achieve this (assuming that nobody changed their minds)?

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

$\hat{p} = 0.481$ $\quad B = 0.01$

$$n \geq \hat{p}\,(1-\hat{p})\left(\frac{z_{\alpha/2}}{B}\right)^2$$

$$n \geq 0.481 * (1-0.481) * \left(\frac{1.6448536}{0.01}\right)^2 = 6754.1$$

$$\boxed{N_{min} = 6755}$$

Be careful about proportions and percentages – these are often interchanged but do all the calculations on proportions

# Key Points

- Sample Sizes can be calculated by specifying a precision (CI width) for the estimate and a confidence – usually 90% or 95%
- This CI width is based upon knowledge of the setting
- For a sample mean the formula is

$$n \geq \left( z_{\alpha/2} \frac{\sigma}{B} \right)^2$$

- For a proportion

$$n \geq \hat{p}\,(1-\hat{p}) \left( \frac{z_{\alpha/2}}{\mathrm{B}} \right)^2$$

- In both cases use previous information for the unknown values , standard deviation and proportion to estimate the standard error