ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Royal Statistical Society – Statistics for Journalists module

Note: excludes instructions on “How to go through this online course”, which are specific to the online implementation.

Storyboard of Statistics for Journalists module, published at
http://www.statslife.org.uk/resources/for-journalists
Students should use the interactive course if possible, for the optimum learning experience. This is provided for the benefit of those who require a text-only version of the course.

Section 1: Welcome and introduction

Screen 1: Welcome and introduction

Learning outcomes:

- Users receive a general welcome to the course.
- Outline the Society’s commitment to encouraging excellence in the use of statistics and data in journalism.
- Give credit to the Department of Business Innovation and Skills for funding.

Screen type:

Text, graphics and audio.

Screen content:

Graphics:

- Royal Statistical Society Logo with strapline of “Data Evidence Decisions” (Hotspot 1).
- Department for Business Innovation and Skills Logo (Hotspot 2).

Audio:

None.

Text:

Welcome to the Statistics for Journalists e-learning module. This module has been developed to help you in reporting on stories with statistics and numbers.

Click on the logos to find out more about the module, and the organisations involved.

Text for Hotspot 1:

This module forms part of the Royal Statistical Society's Science Journalism Programme, a programme launched in 2010, as part of the wider RSS GetStats Campaign.

Text for Hotspot 2:

The programme was funded by the Department for Business Innovation and Skills from 2010-2014 and aims to improve the scientific and statistical understanding of journalism professionals, and journalism students, through workshops delivered by science and statistics professionals across the UK.

As part of the programme, the RSS have developed e-learning modules to support journalists and journalism students in effective science, data and statistics reporting.

This module, Statistics for Journalists, is one of two e-learning modules on the RSS Science Journalism Programme.

Resources:

None.

Screen 2: Module structure

Learning outcomes:

- Users receive an overview of the course.

Screen type:

Text.

Screen content:

Graphics:

None.

Audio:

This Module covers six key areas. You will learn about

- the importance of putting numbers into context,
- the difference between a percentage and a percentage point,
- the importance reliability of surveys and polls, and how to exercise judgement when interpreting results.

You will also find out about

- the difference between absolute risk and relative risk,
- the distinction between correlation and causation, and regression to the mean,
- and the importance of communicating all of these in the best way when presenting news stories.

Text:

- Communicating numbers.
- Communicating percentages.
- Surveys and polls.
- Risk – absolute risk and relative risk.
- Correlation and causation.
- Regression to the mean.

Resources

None.

Section 2: Communicating numbers

Screen 3: Communicating numbers

Learning outcomes:

You will be able to understand:

- the importance of putting numbers into context
- that a number is more than just a string of digits
- the importance of breaking down big numbers and finding out what they are trying to communicate

Screen type:

Text and images.

Screen content:

Graphics:

Five stock photographs of people in various settings in home and at work appear, overlaid.

Audio:

Journalists just love big numbers! But, it is important for you to break them down and put them in to context, so that these numbers can be more meaningful to your readers.

Text:

Journalists love big numbers. Headlines are sprinkled with millions, billions and even trillions, but it sometimes seems that once a number has an ‘illion’ on the end it’s just used as shorthand for ‘big’. Break them down to make them more meaningful for your audience

Resources

None.

Screen 4: Communicating numbers - quiz

Learning outcomes:

As above.

Screen type:

Text with image hotspots. Clicking on each image brings up a text response.

Screen content:

Graphics:

Photographs of bundles of cash, labelled with cash values.

Audio:

Have a go at this exercise to find out more about communicating numbers.

Text:

We will now explore how you communicate numbers in the news in a bit more detail. Look at the hotspots numbered from £1 million to £10 trillion. Can you select the sum that you think is closest to the annual NHS Budget?

- £1 Million
- £10 Million
- £100 Million
- £1 Billion
- £10 Billion
- £100 Billion
- £1 Trillion
- £10 Trillion

Feedback for each statement tested:

- £1million: An annual NHS budget of £1 million would equate to 1.6p per person per year, or 0.03p per person per week. That's not quite right, have another guess.
- £10 million: An annual NHS budget of £10 million equates to 16p per person per year, or 0.3p per person per week. That's not quite right, have another guess.
- £100 million: An annual NHS budget of £100 million equates to £1.60 per person per year, or 3p per person per week. That's not quite right, have another guess.
- £1 billion: An annual NHS budget of £1 billion equates to £16 per person per year, or 30p per person per week. That's not quite right, have another guess.

- £10 billion: An NHS budget of £10 billion equates to £160 per person per year, or £3 per person per week.
- £100 billion: That is the correct answer, well done! An annual NHS budget of £100 billion equates to £1,600 per person per year, or £30 per person per week. The NHS Budget for the 2012/2013 period was approximately £108.9 billion. Big numbers are hard to relate to. If we break them down, putting them on a scale we can relate to, then their story becomes clearer.
- £1 trillion: An NHS budget of £1 trillion equates to £16,000 per person per year, or £300 per person per week. That's too high, have another guess.
- £10 trillion: An NHS budget of £10 trillion equates to £160,000 per person per year, or £3,000 per person per week. That's too high, have another guess.

Resources:

None.

Section 3: Communicating percentages

Screen 5: Talking about percentages

Learning outcomes:

Users will be able to:

- explain and describe the difference between a percentage and a percentage point

Screen type:

Graphics, text, and two quizzes, each with two possible answers.

Screen content:

Graphics:

A diagram showing the UK smoking rate in 1948 and 1970, using circles each representing 1,000,000 people. For 1948, 40 circles are shown, representing the total population of 40 million adults, with 26 coloured in, to represent 26 million smokers. For 1970, 45 million adults, of which 25 million are smokers, is represented.

Audio:

Percentages can be tricky things to understand and communicate.

Have a go at this exercise to see what you know.

Text:

What is the difference between a percentage and percentage points?

Have a go at these questions which illustrate the difference.

Text for question 1

Between 1948 and 1970, the number of smokers in Britain decreased by 1m from 26m smokers to 25m smokers. Select the statement that is correct.

- This a decline of 4 percent [correct]
- This is a decline of 4 percentage points.

Response for question 1

This is a decline of 4 per cent. Whenever you are using the percentages to refer to a change in the 'number', e.g. 'number' of smokers,' 'number' of students, 'number' of voters, then the correct term to refer to the change is 'per cent.'

Text for question 2

The smoking rate (the proportion of the total population that smoked) went from 65% in 1948 to 55% in 1970. Select the statement that is correct.

- This is a decline of 10 percent.
- This is a decline of 10 percentage points.

Response for question 2

This is a decline of 10 percentage points.

Whenever you compare the difference between two percentages, the correct term is 'percentage point'. To simplify this, always ensure you clarify the actual number of the decline. For example, 'the number of smokers in Britain decreased by 1m from 26m smokers to 25m smokers.'

Text displayed at end of exercise

Helpful tip:

Percentages are effective ways of scaling numbers for comparison. We choose something to set as 100, and then work everything else out relative to that. That might be a proportion (X is 30% of Y) or a change (X is 30% greater than Y).

When referring to changes in 'numbers' the correct term is percentage; when referring to changes or differences between 'percentages', you say percentage points. In the example here, this will help show the true changes in the numbers of smokers and the proportion of the population that smoke.

For clarity, avoid small percentages like 0.03% – it's better to use something like 3 in 10,000. Percentage increases greater than 100% can be confusing too. 'Tripled' is more intuitive than 'increased by 200%'.

Resources:

None.

Section 4: Surveys and polls

Screen 6: Surveys and polls

Learning outcomes:

Users will:

- understand the importance of surveys and polls, as well as getting to the heart of questions asked
- gain an awareness of the context of surveys and how to exercise judgement when interpreting results.

Screen type:

Text with pop up question answer.

Screen content:

Graphics:

None.

Audio:

When using surveys and polls in reporting, the wording of a question can hugely influence the answer you get. Take a look at this question, as an example, to find out more.

Text:

What number of people would need four chairs? Or have five heartbeats? Or two ballot papers? Or leave three sets of footprints?

Click Answer to find out the answer: Answer.

When reporting on poll results, it is important to know what questions are being asked, and what exactly is being counted. With most polls it is not always practical to do a full census (surveying everyone in the population.). In most cases, a proxy measure is used.

Text displayed when Answer is selected

One answer is this family of five – a pregnant mother, infant son, young daughter and father. They use four chairs, have five heartbeats, receive two ballot papers and leave three footprints (as the infant son is carried by the dad). Here, the measure of four chairs, or two ballot papers, is a proxy

for measuring the family. Here's the scientific method broken down a little. Shuffle these steps into the correct order.

As a journalist, you'll be sent the results of surveys and polls on an almost daily basis. In this section, we'll look at what surveys measure, the uncertainty that's introduced as a result, and bias.

Surveys and polls try to measure some aspect of a group – eg opinions of UK voters or favourite foods of cats. Almost all surveys will include some amount of estimation, approximation, and indirect measurement. Why?

Firstly, it is usually too difficult for everyone in a group to be asked, or is too costly or impractical.

A second issue is that it is often impossible to directly measure a value. Take homelessness or poverty, for example. 'Proxy measures' are used instead – shelter visitor numbers, benefit claimants. Proxy measures are indirect measures.

Never confuse a proxy measure with the value it was trying to discern. Understand what is being

Resources:

None.

Screen 7: Surveys and polls – margins of error

Learning outcomes:

Users will:

- understand the importance of surveys and polls, as well as getting to the heart of questions asked
- gain an awareness of the context of surveys and how to exercise judgement when interpreting results.

Screen type:

Text with two hotspots for pop-up text.

Screen content:

Graphics:

Two calendars are displayed. One for August 2013 with 28 August 2013 circled, one for August 2014 with 28 August 2014 circled.

Audio:

When exploring how surveys work, you need to consider other factors when interpreting them.

And remember; there is always a measure of uncertainty about a result.

Let's look at the opinion polls leading up to the 2014 vote on Scottish independence to explore this further.

Text:

A lot of the numbers in the news comes from polls and surveys. Let's look at opinion polls leading up to the 2014 vote on Scottish independence to explore this further. Respondents were asked the following question: 'Should Scotland be an independent country?'

Click each of the calendars to see the poll results for each year.

Text shown for each calendar

- 28 August 2013: Yes - 51% No - 49%
- 24 August 2014: Yes - 47% No - 53%

Text shown after both calendars clicked:

Click on the +Confidence Interval button to find out more.

Text shown after +Confidence Interval button clicked:

What a difference a year makes! But, when we then factor in confidence intervals and a margin of error, a different picture emerges. Polls of 1,000 people typically mention (in small print) that the figures are 'plus or minus 3%.' This is called a '95% confidence interval.' This means that at least 95% of the time we are sure that the true value of what is being measured (all the people of Scotland, not just those asked by the pollsters) lies in the margin of error.

+With Cconfidence iIntervals (+/- 3%)

28th August 2013

Between 48% and 54% said 'Yes' and between 46% and 52% said 'No.'

288th August 2014

Between 44% and 50% said 'Yes' and between 50% and 56% said 'No.'

The 'yes' and the 'no' results overlap on each day, and overlap when comparing 2013 to 2014. It becomes increasingly difficult to determine which vote is most popular or least popular, and also difficult to determine if there has been a real increase or decrease in results from 2013 to -2014.

Helpful tip: be careful about over-interpreting small shifts in polls - we'd expect the numbers to bounce around by a few % just because of the sampling. Looking at longer-term trends and comparing multiple polls can help.

Resources:

- British Polling Council: A Journalist's Guide To Opinion Polls http://www.britishpollingcouncil.org/a-journalists-guide-to-opinion-polls/
- National Council on Public Polls: 20 Questions a journalist should ask about poll results: http://www.ncpp.org/?q=node/4
- What Scotland Thinks http://whatscotlandthinks.org/questions/should-scotland-be-an-independent-country-1#bar
  Look at the "Notes and Methodology for this question" link.
  Scroll down to the date of the poll, 28 August 2013. Who conducted the poll? Who commissioned the poll? Who and how many people were sampled? (e.g. 16 and over, or 18 or over?) How do these results compare to other polls, e.g. the poll on 28 August 2014? Have we compared like with like in the exercise? What was the exact question asked – do

you think that’s been fairly summarised? What other factors could affect the results, or how the information from the polls is reported to you?

Screen 8: Surveys and polls – understanding bias

Learning outcomes:

Users will :

- understand the importance of surveys and polls, as well as getting to the heart of questions asked
- gain an awareness of the context of surveys and how to exercise judgement when interpreting results.
- understand bias in surveys

Screen type:

Text and graphics with pop-up results text.

Screen content:

Graphics:

- Two chess boards (black and white square grids) with white counters scattered across each board.
- Background image “My Go Board”– used under CC-BY licence by Flickr/Matt Ryall.

Audio:

Unwanted bias is one of the most important issues in opinion surveys, but also one of the hardest to solve.

Have a go at this exercise to find out more. Try to keep it under one minute!

Text:

You have one minute to add up all the counters that are on the black squares. If any counter overlaps a boundary, count it as if it is mainly on black.

When you have finished, type your answer in the space provided.

Now do the same for the counters on the white squares. If any counter overlaps a boundary, count it as if it is mainly on white.

When you have finished, type your answer in the space provided.

Try to limit yourself to 1 minute to answer the questions.

Results {button]

Text displayed when Results selected:

There are 15 counters mainly on white squares and 12 mainly on black squares.

How closely did your judgement match this?

Getting rid of unwanted bias is one of the most important issues in opinion surveys (and science in general!), but also one of the hardest to solve. There is a seemingly never-ending list of things that can unduly influence the result of a survey. Who is asked, how they are selected, the wording of the questions, how the data is collected; all of these demand careful consideration by the pollster and consequently by any journalist reporting the results of the survey.

Resources:

None.

Section 5: Risk – Absolute risk and relative risk

Screen 9: Absolute risk and relative risk

Learning outcomes:

Users will :

- be able to describe the distinction between absolute risk and relative risk
- understand the importance of communicating risk when presenting news stories

Screen type:

Text.

Screen content:

Graphics:

None.

Audio:

Risk is risky, and especially in reporting!

We'll further explore absolute and relative risk in the next few screens.

Text:

Risk is risky. A relative risk increase is meaningless without knowing the baseline absolute risk. Buying a second lottery ticket doubles your chance of winning, but the chance was so small to start with it's hardly a meaningful increase.

Absolute risk - the overall likelihood of an event (e.g. developing cancer) happening at all

Relative risk - the comparison of the risk of a particular event (e.g. developing cancer) for different groups of people. (e.g. those who drink alcohol and those who do not drink alcohol).

When reporting on risk, it is important to also include the absolute risk so as not to be misleading; it helps to look at both types of risks to understand how significant the change really is

For example, If substance x increases the risk of cancer, you should describe how many extra cases it would produce in 100 people, or 1,000 people.

Resources:

- Cancer Research- Absolute vs relative risk: making sense of media stories: http://scienceblog.cancerresearchuk.org/2013/03/15/absolute-versus-relative-risk-making-sense-of-media-stories/

Screen 10: Absolute risk and relative risk: Bacon case study

Learning outcomes:

Users will :

- be able to describe the distinction between absolute risk and relative risk
- understand the importance of communicating risk when presenting news stories

Screen type:

Text.

Screen content:

Graphics:

None.

Audio:

Take a look at how these two reports present their stories in such different ways.

Text:

To help further illustrate absolute risk and relative risk, take a look at these two reports:

A major report by the World Cancer Research Fund (2007) states that eating 50g of processed meat per day increases the risk of colorectal cancer by 20%.

The Sun reported **Careless Pork Costs Lives**.

But what does this actually mean?

Resources:

None.

Screen 11: Absolute risk and relative risk: Bacon case study - exercise

Learning outcomes:

Users will :

- be able to describe the distinction between absolute risk and relative risk
- understand the importance of communicating risk when presenting news stories

Screen type:

Text and graphics with text hotspots.

Screen content:

Graphics:

Two grids of 100 circles. One, captioned “Absolute risk of developing colorectal cancer without eating bacon”, has 95 grey circles and 5 blue circles The second grid, captioned “Relative risk of developing colorectal cancer if all 100 ate 3 extra rashers every day.”. has 94 grey circles, 5 blue circles and 1 red circle.

Audio:

Have a go at this exercise which helps illustrate the difference between absolute risk and relative risk.

Text:

Let’s explore this ‘20% difference.’

Click on each of the three statements to investigate their meaning.

Absolute risk of developing colorectal cancer without eating bacon:

- About 5 out of 100 people develop colorectal cancer, whether they eat bacon or not.

Relative risk of developing colorectal cancer if all 100 ate 3 extra rashers every day:

- The number would rise to six out of 100 people developing colorectal cancer

Bacon increases risk of colorectal cancer by 20%:

- This is the same as saying 'About 1 extra case per 100 people' - to the other 99 it will make no difference, at least to their colorectal cancer risk. This is like the percentages and percentage points we were talking about earlier. You could say that the absolute risk has

increased by one percentage point, from 5% to 6%.

Handy tip: So in short, when reading press releases or reporting on risk, be sure to check whether the findings mention relative risk and absolute risk, be prudent on how you present this to your audience, and be careful of headlines stating relative risk, rather than absolute risk.

Resources:

None.

Section 6: Correlation and causation

Screen 12: Correlation and causation

Learning outcomes:

Users will:

- be able to describe the distinction between correlation and causation
- understand the importance of communicating correlation and causation when presenting news stories.

Screen type:

Text with image.

Screen content:

Graphics:

Two graphs – one of pirates and global average temperature (Bobby Henderson) and one of cheese consumption in the USA and death from bedsheets (Tyler Vigen – Copyleft - http://tylervigen.com/view_correlation?id=7)

Audio:

Without a carefully controlled experiment, it's very difficult to state that A causes B. Instead, the numbers may show an association, or, correlation, between two things.

You should be aware of speculative links which may be explained by a third, or background factor.

Text:

Correlation: This is when two sets of data are strongly linked together. A pattern where something tends to happen in association with something else.

Causation: This goes further than correlation by stating that one event is the direct result of the occurrence of the other event; i.e. there is a causal relationship between the two events.

Select each of the graphs to find out more.

Text displayed when selecting first graph

This graph shows a relationship between the decrease in the number of pirates and the increase in the global average temperature. Has the decline in the number of pirates caused an increase in the global average temperature? Are pirates the missing link to global warming?

No, the decline in pirates has not caused global warming. Just because two things occur together, it does not imply one event directly caused the other.

Correlation: two things happen at the same time.

Causation: one of those things is actually making the other thing happen.

Text displayed when selecting second graph

This graph shows a correlation between the per capita consumption of cheese in the US, and the number of people who die by becoming tangled in their bed sheets. Does eating cheese cause an increase in the number of people who die by becoming tangled in their bedsheets?

No, this relationship is not causal, these are just two separate trends over time – an increase in cheese consumption and increase in number of deaths.

A significant *correlation* between two variables does not imply one *causes* the other.

Often there is a common cause for both variables, or, it's just a coincidence.

Correlation: two things happen at the same time.

Causation: one of those things is actually making the other thing happen.

Resources:

None.

Section 6: Regression to the mean

Screen 13: Regression to the mean

Learning outcomes:

Users will:

- be able to describe the distinction between correlation and causation
- understand how this applies to regression to the mean

Screen type:

Text with quiz.

Screen content:

Graphics:

None.

Audio:

Regression to the mean is not always easy to spot. In this scenario, you will see how a common situation of regression to the mean can be confused with a causative effect.

Start the exercise to see what happens.

Text:

Regression to the mean is another way of saying things return to normal. However, it's not always easy to spot. In this scenario you will see how a common situation of regression to the mean can be confused with a causative effect.

You have 6 streets to monitor, but only one speed camera; use your judgement to decide which street should have the speed camera.

Fatalities Report Year 1:

- Nightingale Crescent – 12
- Bayes Street – 8
- Cox Drive – 2
- Shewhart Lane – 5
- Blackwell Avenue – 4
- Chandra Road – 7

Text for incorrect response

That’s not quite right, please try again.

Text for second incorrect response

No, that’s not quite right. In this exercise it makes sense to place the speed camera on the road with the highest number of road deaths.

Text for correct response

Good choice! It makes sense to place the speed camera on the road with the highest number of road deaths.

Resources:

None.

Screen 14: Regression to the mean results

Learning outcomes:

Users will:

- be able to describe the distinction between correlation and causation
- understand how this applies to regression to the mean

Screen type:

Text with quiz.

Screen content:

Graphics:

None.

Audio:

None.

Text:

A year has passed since the first report on the number of road accidents.

In light of this exercise, do you think it is proven that the speed camera reduced the number of road deaths on Nightingale Crescent?

Select your answer.

- Yes
- No

Fatalities Report Year 1:

- Nightingale Crescent – 12
- Bayes Street – 8
- Cox Drive – 2
- Shewhart Lane – 5
- Blackwell Avenue – 4
- Chandra Road – 7

Fatalities Report Year 2:

- Nightingale Crescent – 8
- Bayes Street – 7
- Cox Drive – 4
- Shewhart Lane – 2
- Blackwell Avenue – 10
- Chandra Road – 4

Text for incorrect response (Yes)

No. The number of accidents on Nightingale Crescent may have been unusually high before the speed camera was introduced, leading to a decrease not attributable to the speed camera, as the numbers return to average levels (“regression to the mean”). To further illustrate, if you placed a garden gnome at crash sites and find that the number of crashes falls, this does not mean the gnome has anything to do with the reduction of incidents. Good research takes account of regression to the mean and looks for longer-term effects – trends.

Text for correct response (No)

Correct. The number of accidents on Nightingale Crescent may have been unusually high before the speed camera was introduced, leading to a decrease not attributable to the speed camera, as the numbers return to average levels (“regression to the mean”). To further illustrate, if you placed a garden gnome at crash sites and find that the number of crashes falls, this does not mean the gnome has anything to do with the reduction of incidents. Good research takes account of regression to the mean and looks for longer-term effects – trends.

Resources:

None.

Section 7: Conclusion

Screen 15: Thank you for your time.

Learning outcomes:

- Summary of the course.

Screen type:

Text and audio.

Screen content:

Graphics:

None.

Audio:

You've now completed this module on Statistics for Journalists.

We hope this has helped you to understand the key points of how to approach statistics in your reporting. By following the basics in this module, you'll be able to ensure that your facts and figures have been investigated carefully, and presented accurately, to provide the best service to your audience.

You can always go to the Resources section to find out more about all of these topics in greater detail.

Thank you for your time on this course, and good luck with your reporting!

Text:

You've now competed this module on Statistics for Journalists, where we've looked at:

- Communicating numbers.
- Communicating percentages.
- Surveys and polls.
- Risk – absolute risk and relative risk.
- Correlation and causation.
- Regression to the mean.

By following the basics in this module, you'll be able to ensure that your facts and figures have been investigated carefully, and presented accurately, to provide the vest service to your audience.

You can go to the Resources section to find out more about all of these topics in greater detail.

Thank you for your time on this course, and good luck with your reporting!

Thank you for your time.

Resources:

- RSS Resources for journalists http://www.statslife.org.uk/resources/for-journalists
- Sense About Science – Making Sense of Statistics http://www.senseaboutscience.org/resources.php/1/making-sense-of-statistics
- Full Fact http://FullFact.org
- Understanding Uncertainty - http://understandinguncertainty.org/

Credits

Produced by the Royal Statistical Society as part of the RSS Science Journalism Training Programme, funded by the Department for Business, Innovation and Skills.

Online course developed by Marshall ACM from a storyboard produced by Vinet Campbell and Scott Keir (RSS) in consultation with Marshall ACM. Storyboard content primarily based on training materials produced for the RSS Science Journalism Training Programme, including ‘Statistics for journalists’ (Stories and Statistics) presentation material developed by Frank Swain, containing material prepared by Martin Griffiths, Sam Mugford, David Schley, Vinet Campbell, Paul Askew, Simon White. We thank the advisory board and volunteer trainers of the RSS Science Journalism Training Programme for their input, advice and feedback.

For further details, see statslife.org.uk/resources/for-journalists or contact Scott Keir, Head of Education and Statistical Literacy, Royal Statistical Society, 12 Errol Street, London, EC1Y 8LX, email: s.keir@rss.org.uk.

Published 3 August 2015.