



Best practice guide for using statistics in communications

These guidelines will cover:

- Introduction
- Why statistics are used
- What statistical information should include in PR activities
- Analysis
- Reporting on survey methodology
- What common statistical terms mean – and pitfalls to watch out for
- Common pitfalls that can undermine your message
- Statistics and the CIPR Code of Conduct
- Sources of statistical advice
- Acknowledgements.

1. Introduction

Using statistics can help to achieve public relations goals whether raising awareness, informing decision making or influencing behaviour.

This guide is about using statistics effectively, both to achieve day-to-day goals and to contribute over time to enhanced reputation. Good and accurate use of statistics can help to establish credibility and to increase influence. Poor use of statistics can lead to loss of trust and reduced authority.

The guide has been produced to help PR professionals know what should be included in their work, and to spot when they might want to obtain expert statistical advice. It is not meant to be an instruction manual in statistical techniques.

A number of PR exercises incorporate results from survey research. This is not a comprehensive guide to the use of survey results in PR, but does cover some of the key issues. For more in-depth information, consult the research agency commissioned to undertake the survey or refer to the guidelines produced by The Market Research Society or its Code of Conduct. (See *Sources of statistical advice* below).

2. Why are statistics used?

Statistics and survey results help to provide a better understanding of the way things are, to back up claims being made, to measure impact, to provide valuable contextual or background information or to make forecasts.

When estimating the proportions of a given audience or producing forecasts, survey research can help to improve understanding of the level of certainty of estimates, findings or projections.

3. What statistical information should be included in public relations activities?

As a basic guide, the following information should be made available to the public as this will allow audiences to understand the context and limitations of your work:

- the client commissioning the survey
- the purpose of the investigation (reasons for undertaking the survey, experiment etc)
- the 'universe' effectively represented (all adults, voters etc)
- how survey participants were selected and method of interview (eg face-to-face interviews with British adults aged 16 years and over or Managing Directors in SME firms across the UK)
- the sample size, geographic coverage, fieldwork period and response rate
- what was measured and how data were weighted, if relevant

Best practice guide for using statistics in communications

- the survey results, taking into consideration the statistical margins of error and overall reliability of the findings
- the percentages or proportions upon which conclusions are based, including the numbers from which they were calculated.

This information will help audiences to make an assessment of how authoritative the survey is and how reliable the results are. It is also important to make sure that audiences are not misled or do not inadvertently misunderstand what is being reported.

4. Analysis

An analysis of the survey findings should consider:

- the distribution of the results

- whether the data are evenly spread or clustered around a particular value
- if the data are spread symmetrically or skewed with more measurements to one end or the other
- the range of the results, ie the lowest and highest values
- if an average has been calculated, whether it was the mean, median or mode, and why it was chosen
- the certainty of the results – confidence intervals or levels of statistical significance (which give an assessment of how likely the result was, if chance alone were operating).

5. Reporting on survey methodology

Any news release should include a technical specification outlining the way the survey was conducted.

Methods or terms should be clearly defined, if not in the main text then in notes to editors in press releases or footnotes to a briefing.

Typically, a news release might include:

- purpose of the survey and what was measured or asked
- audience or subjects represented, eg people, businesses, journeys
- sample size and method of sampling to show appropriate representation
- response rates
- fieldwork dates
- whether or not the survey data have been weighted
- results with accompanying commentary on key findings (and ensuring that any comparisons are based on results that are statistically different from each other)
- Information on source data and any assumptions to accompany any forecasts.

Note how the use of words can be as important as using numbers. It is important that any description or interpretation of the survey results is accurate and supported by the statistics.

Also consider the extent to which a conclusion is expressed reasonably and responsibly. For example, if there is an extreme result which is uncommon or unlikely (and this is not spelled out) then it could be alarmist to focus on it. This might provide a good short term PR advantage but can undermine credibility in the long term.

Above all, transparency is critical. If there are issues that could have affected the accuracy of your results, spell them out – otherwise there is the risk of exposure from a third party. If a survey is being reported, it may be advantageous, therefore, to make the raw survey data available for others to inspect and analyse.

Best practice guide for using statistics in communications

The British Polling Council requires that all data and research findings made on the basis of surveys conducted in the United Kingdom by member organisations that enter the public domain must include reference to many of the above criteria. (See *Sources of statistical advice* below).

6. What common statistical terms mean – and pitfalls to watch out for

Having a good understanding of a few common statistical terms can make all the difference. The following goes into a little more depth on these and highlights some of the pitfalls that can undermine the credibility of good public relations.

- ***Distributions***

If lots of things are being measured, it is very likely that differences will occur. Collectively we have different eye colour, height and incomes. Different people respond differently in a drugs trial. One shoe selling business or electricity company will have bigger profits (or losses) than another. And so on.

In other words, things come in distributions of values or attributes. That is why statistics are beneficial.

They allow basic features of distributions to be summarised in order that decisions can be made or conclusions reached.

- ***Averages***

There are actually several types of average. For example; mean, median and mode – each summarising lots of figures in one number.

When people think of the 'average' mostly they are thinking of the mean, which involves adding up all the figures and dividing by the total number of them.

Where measurements and results have a basically symmetrical spread, like height, then the mean is usually used.

For skewed distributions, the mean may not be so useful. For example, in the case of incomes, most people have moderate salaries. Pop stars, premier league footballers or chief executives of big businesses earn a large wage. In these cases, the mean is higher than what most of us earn, so using the median could be more appropriate. If the incomes of all adults were lined up from smallest to largest then the median is the one halfway through.

The third kind of average tells us about what is most common for example, a bestselling fiction book or ice cream flavour. This average is the mode.

It's important to know, though, that the choice of average depends on the question that you are trying to answer.

- Potential pitfall – using the 'wrong' average:

When a distribution is particularly skewed the mean and the median can be very different. You might overstate or understate your case – or others might use your choice to undermine your message.

Choosing carefully and explaining your choice will help overcome this.

Example

Consider the following 25 numbers, in no particular order:

6 4 1 0 2 3 4 5 35 6 3 10 2 3 9 5 20 5 7 8 1 0 1 3 4

To find the **mean** all the numbers are added up and divided by how many there are:

Best practice guide for using statistics in communications

4

$6+4+1+0+2+3+4+5+35+6+3+10+2+3+9+5+20+5+7+8+1+0+1+3+4=147$

And $147/25 = 5.88$

To find the **median**, the numbers are put in order lowest to highest, including any repeated numbers, and the middle one is the median:

0 0 1 1 1 2 2 3 3 3 3 4 4 4 5 5 5 6 6 7 8 9 10 20 35

In this case the median number is 4

The **mode** is the number that appears more than any other one. The brackets show how many times each number occurs:

0 (2) 1 (3) 2 (2) 3 (4) 4 (3) 5 (3) 6 (2) 7 (1) 8 (1) 9 (1) 10 (1) 20 (1) 35 (1)

The number 3 appears more times than any other number.

So the mean is 5.88, the median is 4 and the mode is 3. The numbers are different because they are different averages, which shows how important it is to be precise about the type of average that has been calculated.

Remember, the choice you make will depend on the question that you are trying to answer.

- **Sample size**

Reasons for taking a sample include, where it is impossible or too expensive to undertake a survey of a whole population or group of businesses or other subjects. If the sample is drawn in an appropriate way and it is representative of the intended audience then it will provide a truly representative response.

The aim is that the sample is good enough for the results to closely reflect what you would have got if you could have measured everything. Statistical techniques applied to the results obtained from questioning those in the sample can allow a calculation to be made of a range of values in which the true figure for a quantity lies, together with its certainty.

This certainty is related to the sample size. The larger the sample size, the smaller the confidence interval (see below) for any particular level of certainty, assuming that the samples under consideration are equally representative of the population being surveyed. Opinion polls of members of the general public, for example, are quite often undertaken with a sample size of about 1,000 respondents. This allows a survey result to be calculated within about three per cent either way.

- Potential pitfall using a subsample to squeeze out too much.

A sample of a 1,000 adults across the UK is sufficiently robust to provide a good estimate of opinion upon this audience. Reporting on a specific sub-sample group of around 100 respondents needs more care to ensure that the results are statistically significant. (ie how likely that the result is not down to chance).

- Potential pitfall confusing the mid-point of the range with the 'true' value.

Although a confidence interval is often calculated by taking a number minus, say, three per cent, and the same number plus that three per cent, doesn't mean that the 'true' value is the number.

All we know is the 'true' value lies in that range, to a certain degree of certainty.

7. Common pitfalls that can undermine your message

- **Using a sample that is not representative of the audience being represented**

The way in which a sample of the required audience is selected is of key importance to how representative the survey results are. A common mistake is to think that a particular group is representative of the target community, if many people are asked. People with similar views tend to have similar behaviour, for example going to work on the early morning trains, or attending certain sporting events, or being in the town centre in mid-morning. A random sample of a lot of these people will clearly not be representative of a more general population.

- ***Getting bias through self-selection***

Getting people in a community to 'have their say' is commonly used to gather views. This can be useful information but is very unlikely to be representative of the whole community as individuals have been able to opt in to the survey. This is because people with strong views on an issue are more likely to respond than those without strong views.

Some newspapers and magazines use this kind of survey and then say that these are the views of their readership or the country. As a 'bit of fun' these sorts of poll can be okay, will not have any statistical validity as summaries of the overall population.

- ***Getting bias through leading questions***

Questions need to be phrased and ordered in a balanced way to ensure that they are not misleading in any way.

Although the response may be relevant to the client, it may not be useful as an indicator of a person's true behaviour. It actually takes considerable statistical expertise to design questionnaires, even going as far as randomising the order of questions to different people to minimise any bias that a particular order might bring.

Even the choice of words can be very important. For example, people will tend to give different answers depending on if they are asked how satisfied they are with a product or service or how happy they are. An open question at the end of a survey might be helpful so that there is a chance for people to raise any issues that your survey hasn't properly covered.

- ***Quoting percentages without providing the sample size on which the survey is based***

Reporting that a certain percentage of the audience surveyed responded in a particular way is only half the story. It will be important to know how many people overall were interviewed. This is related to the issue of sample sizes.

- ***Quoting trend data incorrectly***

This is about knowing how to express a change in percentage result from one time period to another. For example, if 28 per cent of adults now like a particular spread but it used to be 20 per cent, then the improvement is eight percentage points – not eight per cent (or a 40 percentage point improvement).

Again, the number from which the percentage has been calculated should be included.

- ***Not spelling out the whole story on risk***

Best practice guide for using statistics in communications

Care needs to be taken when reporting on risk, and changes in it. This often comes up in reporting health-related stories.

There are two types of risk. The absolute risk of a disease, for example, is the chance that a person will get it over a certain time period. Relative risk compares the absolute risk between two different groups.

For example, generally we may have an absolute risk of five per cent of getting a disease. For some people that risk might be six per cent because of their lifestyles. Although this is only one percentage point higher, the relative risk is increased by 20 per cent.

In reporting on relative risks, the absolute risks should always be quoted.

- ***Misplacing emphasis on extreme, but uncommon, values***

It can be tempting to focus on an extreme value that arises from an investigation or a projection.

However, if it is uncommon or unlikely it may lead to subsequent activity being inappropriately focused, or raise an undue sense of alarm.

- ***Two things that appear related don't necessarily mean one is caused by the other***

Correlation is where two things seem to appear together in a particularly frequent way. It can be tempting to think one causes the other. This may be the case but there may be a third, unknown factor that is actually at play.

- ***Providing unwarranted precision***

The more precise a figure the more trust people can have in it. But often the precision is unwarranted. For example, a figure for home ownership of 65,748 in a town is too precise if it comes from scaling up from a sample of 900 people. The margins of error would at best only allow the figure to be quoted as 65,700 or even 66,000.

- ***Regression to the mean***

This can be really subtle to spot. Sport is one area where it can be seen. Over time performances can tend to get better – that is a real effect often due to real improvements that come from practice and training.

However, from time to time, a really great score might be achieved, only to then have a run of less impressive ones. What is happening is that most of the time the scores will be about average, ie the mean, with a few extreme scores which are very good (or very bad). This is known as regression to the mean.

An extremely high or low value is likely to be followed by a value near the mean, simply because most values are nearer the mean than the extremely high or low values. So the fact that an extremely high value is followed by a smaller value does not mean that things are necessarily deteriorating. In the same way, the fact that an extremely low value is followed by a higher value does not mean that things are necessarily improving.

8. Statistics and the CIPR Code of Conduct

These best practice guidelines are designed as a point of reference for practitioners who are using statistics in the corporate world, in government and in the not for profit sector. Whilst these guidelines do not constitute a legal document, all CIPR members are bound by the Code of Conduct which is based

around three principles: integrity, competence, and confidentiality. The Code of Conduct should be adhered to when engaging in any public relations practice and there are elements within the Code's principles that are particularly relevant to any member using statistics.

Integrity: Integrity is essential when using statistics in PR, as indeed it is key to all elements of professionalism. It requires that members are honest and truthful when using statistics. Members should have regard for the public interest. They should be accurate when disseminating information.

Competence: Members should in this area as in others be aware of the limitations of their professional competence, and should therefore be willing to accept or delegate only that work for which they are suitably skilled and experienced. Members are advised that they should receive proper advice on gathering and interpreting statistics when necessary. The drafting of material by public relations professionals for their clients naturally has a role to play in many areas of professional endeavour. But all messages, statistics, and / or claims made in communications should be supported by proof, or a robust and reasonable rationale for the claim.

Confidentiality: Confidential information should not be disclosed unless specific permission has been granted by the parties concerned; or unless it is in the public interest; or unless required to do so by law. Members should bear in mind that failure to adhere to the CIPR Code of Conduct could lead to disciplinary action, with sanctions up to, and including, expulsion from the CIPR.

The comments above are intended only as an outline of some of the issues raised by the three Code principles; they do not in any way supersede the full text of the Code. This document is intended as a statement of best practice – not as an addition to the Code.

9. Sources of statistical advice

The Royal Statistical Society

- The RSS promote public understanding of statistics and provide professional support to users of statistics and statisticians. It has launched a new campaign, called getstats, to deliver this work.
- The RSS runs workshops for journalists and others to help them understand the basics of statistics.
- On the RSS web site is a directory of statistical consultants who can be engaged to carry out statistical work.

The Market Research Society

- With members in more than 70 countries, MRS is the world's largest association serving all those with professional equity in provision or use of market, social and opinion research, and in business intelligence, market analysis, customer insight and consultancy.
- In consultation with its individual members and Company Partners, MRS supports best practice by setting and enforcing industry standards. The commitment to uphold the MRS Code of Conduct is supported by the Codeline service and a wide range of specialist guidelines.
- MRS contributes significantly to the enhancement of skills and knowledge by offering various qualifications and membership grades, as well as training and professional development resources.

Straight Statistics

- A campaign established by journalists and statisticians to improve the understanding and use of statistics by government, politicians, companies, advertisers and the mass media.

British Polling Council

- The British Polling Council objectives are to ensure standards of disclosure designed to provide consumers of survey results that enter the public domain have an adequate basis for judging the reliability and validity of the results. Practitioners should be aware of the British Polling Council rules on disclosure which have made a material difference to the quality and quantity of material disclosed by PROs in recent years.

UK Statistics Authority

The Authority was set up by the Statistics and Registration Service Act 2007 and has the statutory objective to promote and safeguard the production and publication of official statistics that serve the public good. It is also required to promote and safeguard the quality and comprehensiveness of official statistics, and ensure good practice in relation to official statistics.

10. Acknowledgements

The CIPR would like to thank the guide's authors Andrew Garratt MCIPR and Adrian Penrose MCIPR and the Market Research Society and Royal Statistical Society for their input and endorsement.