Introductory Statistics with Excel

Andrew Parnell

University College Dublin

Class 3 - Hypothesis testing

Learning outcomes

In this class we will cover

- Sampling distributions
- Setting up hypothesis test
- The working of a hypothesis test
- Interpreting the output of a hypothesis test

Statistical inference and sampling distributions

Why are we doing statistical inference? Consider the following examples:

- We have a new drug which might be used to treat those suffering with cancer. We trial the drug in an RCT. How do we know whether the drug works or not?
- The mean northern hemisphere temperature has risen from around 13°C (1980-1990 average) to around 19°C (2000-2010 average). How do we know this difference is real and not just from random chance?
- A new particle, purported to be the Higgs Boson, has been found at mass 125GeV after an experiment at CERN. How do we know whether this particle is really there or whether it was just experimental chance?

Parameters

Reminder

We are interested in estimating **parameters**: characteristics of interest from the **population**. We assume that the parameter is an unchanging, fixed value

Examples of parameters:

- The proportion of the population that consider themselves left-handed
- The probability that a baby who sleeps with a night light suffers from myopia at age 20
- The income difference between arable and pastural farmers

Statistics

Reminder

A **statistic** is an estimate computed from a sample of data. These data may have been collected from a trial, an experiment, an observational study, or a survey. We often try to choose statistics that are **good estimators** of the parameters

Examples of statistics:

- The sample proportion of 100 people who consider themselves left-handed
- The sample proportion of adults aged 20 with myopia from a sample of adults aged 20 who had night lights as babies
- The sample mean income difference between arable and pastural farmers

Key ideas

- These statistics are just **estimates** of the parameters. If we ran the experiment again we would get different values of the statistics. These estimators are **uncertain** and the statistics are **random variables**
- If we can make some assumptions about the behaviour of the statistic over different samples we can create the **sampling distribution** of the statistic
- The practice of estimating parameters from statistics taking into account their uncertainty is called **statistical inference**
- We will cover two parts of statistical inference: **confidence intervals** and **hypothesis testing**

Answering questions about parameters

One of the great things about statistics (and maths in general) is that the **methods** used in all sorts of different applications turn out to be the same. However, it can be quite difficult to translate from the questions of scientific interest into questions about parameters which are amenable to statistical analysis.

Example

Do pigs on fodder produce higher quality meat than those who forage? Possible parameters of interest:

- proportion of pigs on fodder with higher quality meat than those who forage
- average meat quality increase between fodder/foraging pigs

We could then run a study of pigs measuring their meat quality on each of the different feed types

Sampling distributions of statistics

Statistics as random variables

Important

We can think of taking a sample or conducting an experiment as one big random circumstance. The value of the statistic is then one possible outcome of that random circumstance. Thus a statistic is just a **random variable** and has an associated **probability distribution** (or probability density). We call the probability distribution of a statistic the **sampling distribution**

If we know the probability distribution then we can make statements about the parameter from the probability distribution. We can then work out, for example, the probability that the statistic lies in the range (a, b) or the probability that the statistic is bigger than a chosen value.

Features of sampling distributions

In the situations we will cover:

- The sampling distribution is **approximately normal**. Remarkably this occurs all the time (see associated app – `https://gallery.shinyapps.io/CLT_mean/`).
- The mean of the sampling distribution is the **value of the parameter**. For example, the mean of the sampling distribution for the sample mean is the true parameter value
- The standard deviation of the sampling distribution measures how the value of the statistics **vary** over across different samples

Remember that the normal distribution is **completely defined** by its mean and standard deviation (or variance). So to specify the sampling distribution we only need to calculate the mean and the standard deviation

Standard error

Definition

The sample standard deviation for a statistic is known as the **standard error**.

Note that this is different to the standard deviation of the data (again see app). To be clear, we append the name of the statistic on to the end of the name, so that the standard deviation of the sample mean is written as the **standard error of the mean**

Hypothesis tests

Example

Reminder of the milk yield data:

169.6	142	103.3	111.6	123.4	143.5	155.1
101.7	170.7	113.2	130.9	146.1	169.3	155.5

1. Want to investigate the farmer's claim that the mean weekly milk yield for the herd is 120kg.
2. What error could the we make when making their decision?
3. Want to assess uncertainty in mean weekly milk yield – construct and interpret a 95% confidence interval for the mean milk yield.

Hypothesis Testing

Hypothesis testing involves two contradictory hypotheses:

- The Null Hypothesis (H_0), initially assumed to be true.
- The Alternative Hypothesis (H_A).

These hypotheses are equivalent to the innocent (H_0) until proven guilty (H_A) in a court of law.

We examine the evidence (the **data**) to see if it proves guilt; we do not try to prove innocence as the defendant is presumed innocent until proven guilty.

Hence, when we choose the hypotheses, H_A should be the hypothesis for which we seek evidence.

NOTE: The same hypothesised value must be used in H_A as in H_0.

Hypothesis Testing

The decision whether or not to reject H_0 is based on the evidence contained in the sample **data**.

The result of the hypothesis test will be either:

- Reject H_0 in favour of H_A: when the data provide sufficient evidence that H_A is true. The result is said to be **statistically significant**.

- Fail to reject H_0: when we have insufficient evidence that H_A is true. The result is not **statistically significant**.

Note: Failing to reject H_0 does not mean that we have evidence that it is true - we simply do not have sufficient evidence to reject it. Failing to prove guilt does not prove innocence!!

Incorrect Decisions

	The Truth	
The Decision	H_0 True	H_A True
Reject H_0	Type I error	:)
Do not reject H_0	:)	Type II error

Type I error: Reject the Null Hypothesis when it is true. Type II errors: Do not reject the Null Hypothesis when it is false.

LEGAL ANALOGY:

H_0: Innocent and H_A: Guilty

Type I error: Innocent person found guilty.

Type II error: Guilty person found innocent.

Incorrect Decisions

The advantage of using statistical hypothesis testing instead of any other decision making procedure is that we can quantify and control the probability of making a type I or type II error.

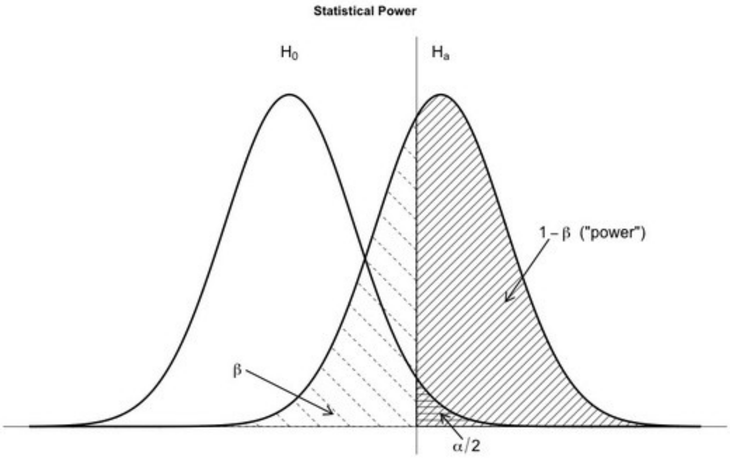
The Probability of a **Type 1** error is called the **significance level** of the test.
The Probability of a **Type 2** error is one minus the **power** of the test.

Reducing the significance level increases the power and vice versa. The only way to improve both is to increase the sample size.

The value of the significance level is decided by the statistician and the values usually used are 0.05 or 0.01.

(These values have statistical motivation – they're not just randomly chosen values!)

Power and significance



Taken from `https://qph.ec.quoracdn.net/`

Test Statistics and Rejection Regions

Assuming H_0 is true, we compute a **test statistic** which follows a known distribution (e.g. the Normal distribution).

$$\textbf{test statistic} = \frac{\text{sample statistic - null value}}{\text{standard error}}$$

When we have a large sample (n$\geq$30) the Central Limit Theorem tells us that, irrespective of the underlying population distribution, a mean has approximately a **Normal** distribution with:

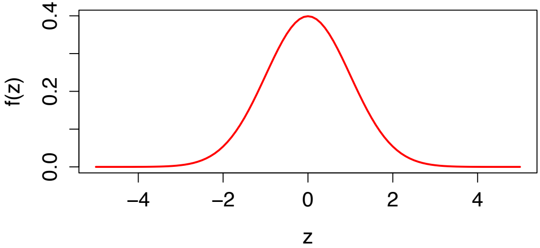
- mean $=$ population mean (the value in H_0).
- standard error $=$ sample standard deviation$/\sqrt{n}$.
  (the **standard error** - see earlier app)

Test Statistics and Rejection Regions

Hence, our test statistic

$$\frac{\text{sample statistic - null value}}{\text{standard error}}$$

has a **standard Normal** distribution (i.e. mean 0 and standard deviation 1).



Test Statistics and Rejection Regions

The **critical value(s)** divide the possible values of the test statistic into a **rejection region** and a **nonrejection region**.

The critical values depend on the significance level of the test α.

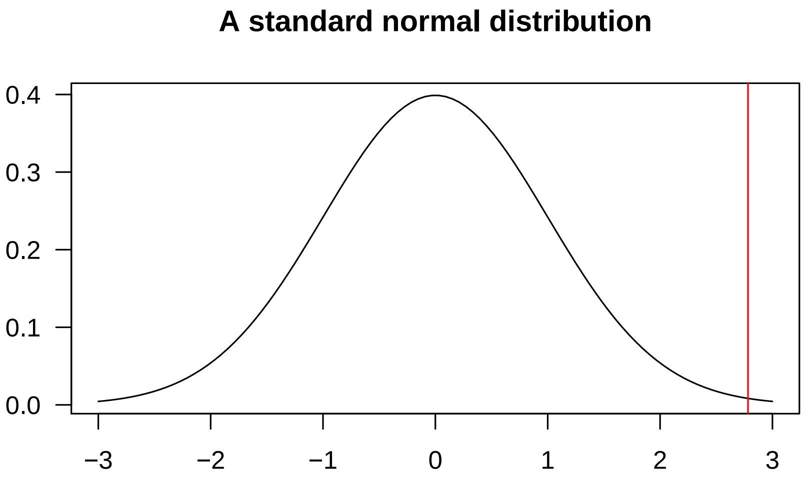
Why?

- If H_0 is true, the sample statistic is unlikely to be very different from the hypothesized value.
- Or, a test statistic very different from 0 is unlikely.
- The significance level is the value we assign to the word ‘unlikely’.
- If the test statistic lies in the rejection region we reject H_0 as the probability it lies there is below what we defined as ‘unlikely’.

Back to the cows data

- We want to test whether the mean value is 120 or not so H_0: population mean $=$ 120, vs H_A: population mean $\neq 120$.
- We already know the sample mean is 138.3kg and the sample standard deviation is 24.6kg
- We have 14 observations so that standard error of the mean is $\frac{24.6}{\sqrt{14}} = 6.56$.
- Our test-statistic is then: $\frac{\text{sample statistic - null value}}{\text{standard error}} = \frac{138.3-120}{6.56} = 2.78$.
- The 97.5th percentile of the standard normal distribution is 1.96 (Excel - `NORM.INV(0.975,0,1)`). We are thus in the **rejection region** so we might reject the null hypothesis

Plot of the test statistic



Interpreting the test statistic value

- Often, rather than specifically choosing a significance level, the area under the curve more extreme than the red line is calculated. This is known as the **p-value**
- Since the area under the curve is 1, the p-value will be less than 1
- Interpreting the p-value is hard. A small p-value (e.g. less than the type 1 error level of say 0.05) is often called a **statistically significant** result
- This is using the word 'significance' to mean that it 'signifies' something, not that the effect is large or important
- The mis-use of p-values pervades science, and many statisticians do not use them
- It is much harder to calculate the power of the test - you need to assume a probability distribution for the alternative hypothesis

Class 3 summary

- Whatever the shape of the probability distribution of the population, it's usually safe to assume that the sample mean has a sampling distribution that is **normally distributed**
- A hypothesis test is setup with a **null and alternative hypothesis**. We create a test statistic and, if it lies in the **rejection region** (or the p-value is small) we reject the null hypothesis
- We might make a **Type 1** or **Type 2** error
- Be careful when interpreting the output of a 'statistically significant' result