UNIVERSITY OF BATH

Data Science and Statistics in Research: unlocking the power of your data

Session 2.2:

Hypothesis testing I

Outline

Sampling Variability

Sampling Variability

Sample statistics

Sample statistics can be used to estimate the characteristics of an **underlying population or process** from which the samples are drawn.

There are two potential problems:

- **bias**: the sample may not be representative of the population as a whole
- **chance**: the sample may differ to a greater or lesser extent from the population by chance alone.

Sampling Variability

- Bias is important in the planning of a study, where we must be clear about any possible selection effects that may bias a sample.
- For example, public opinion surveys on the street on weekdays will miss out many different types of people, in particular the working population, who may have strongly differing views on various issues.
- If we have a **well-designed** and **carefully executed** study bias should not pose a major problem.
- Chance is unavoidable. Where variability is present, the sample statistics calculated from any particular sample will be different to those calculated from another independent sample.

SAMPLING VARIABILITY

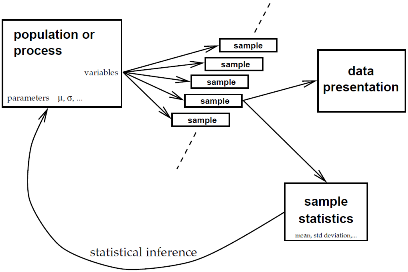
- Suppose we are interested in the height of the population of Bath and we record the heights of all the students at the University of Bath.
- The mean and standard deviation of the population and the sample are shown

	Population	Sample
Mean	μ	$\bar{x}$
Standard deviation (SD)	σ	s

- The sample mean, $\bar{x}$, is an estimate of the true population mean, μ of the population of Bath.
- The sample standard deviation, s, is an estimate of the true standard deviation, σ of the population of Bath.

Sampling Variability

- If we repeatedly take random samples from the overall population and each time record the mean and the standard deviation we would find that values
  - $\bar{x}$ and s would vary from sample to sample
  - $\bar{x}$ would be distributed symmetrically around the true population mean, μ, the value we are interested in
  - near μ would occur more frequently than those far from μ.



Confidence Intervals

Confidence Intervals

What is a confidence interval?

- Given a sample of data, the sample mean gives an estimate of the true mean, but there is **uncertainty** associated with this.
- A **confidence interval** (CI) quantifies this uncertainty by giving a range in which we are 'confident' the true mean will lie.

When is it used?

- It is use to indicate the level of precision of an estimate from a sample, with larger samples giving more precision.
- A wider interval means that there is more uncertainty.
- Often they are used to assess whether a particular value is likely or not.

Confidence Intervals

What is the output?

- Usually a CI is symmetric, centered around the sample mean.

What is the output?

- It is most common to use 95% CIs.
- In turn, we often center it around the sample mean

$$24.9 \quad (24.5, 25.3).$$

Confidence Intervals

How do we interpret it?

- The CI will be calculated based on a specified level of significance where we are ‘confident’ that the true mean lies within the given interval.
- A wider interval means that there is more uncertainty.

Are there any restrictions on its use?

- The calculation usually relies on the fact that data is not heavily skewed and does not have heavy tails and in certain cases this will not be a good assumption.
- We can use CIs to make certain assumptions about population means using the observed values from a sample.
- The procedure is to construct CIs and then to see whether or not the values of interest lie in those intervals.

EXAMPLE: MOTOR TREND CAR ROAD TESTS

- The `mtcars` dataset in `R` contains fuel consumption and 10 other aspects of 32 cars recorded from 1973-74.
- Let's create a 95% CI for the fuel consumption (in miles per gallon) to understand where the true mean may lie.
- There are 32 samples, with the fuel consumption having sample mean 20.09 and sample standard deviation 6.03, so we get the CI of

$$20.09 \quad (17.92, 22.26).$$

- Therefore we are 95% 'confident' that the true mean fuel consumption is between 17.91 and 22.26 miles per gallon.

Hypothesis Tests

Hypothesis Tests

- Research involves the development and testing of hypotheses.
- Statistical tests help us to evaluate these hypotheses.
- To do this, two alternative conclusions (hypotheses) are set up, and we assess if the data generated by an experiment is more consistent with one hypothesis than the other.
- This this is assessed using a pre-specified levels of confidence.

Hypothesis Tests

- There are two hypotheses constructed, the null and alternative.
- The **null hypothesis**, H_0: we hypothesise that there is no difference between
  - your data and a known mean
  - the mean of two groups.
- The **alternative hypothesis**, H_1: we hypothesise that there *is* a difference between
  - your data and a known mean
  - the mean of two groups.

Hypothesis Tests

- Hypothesis tests are always stated so that either one or the other (but not both) can be true.
- On the basis of the hypotheses a statistical decision rule is constructed
  - **IF** the observed value of a statistic takes on a certain range of values
  - **THEN** we have enough evidence to reject the null hypothesis
  - **OTHERWISE** there is not enough evidence to reject the null hypothesis.
- The range of values for the statistic will be determined by the desired levels of statistical significance.

Hypothesis Tests

- Given a sample of data, we calculate a summary statistic
  - for example, the mean.
- We compare this with the value we have chosen for our null hypothesis
  - for example we want to test that the mean fuel consumption (in miles per gallon) in the `mtcars` dataset is 22.5.
  - the difference will be the sample mean minus 22.5.
- Larger differences would suggest that the null hypothesis is not true.

HYPOTHESIS TESTS

- We need to know how large the difference needs to be to have a **statistically significant difference**.
- For this we calculate a **test statistic** which is a standardised version of the difference.
- We use this to obtain a **p-value** to assess how significant the difference is.
- Smaller p-values provide evidence to reject the null hypothesis.

One-sample t-test

Example: Motor Trend Car Road Tests

- Let's return to the `mtcars` dataset.
- Suppose a car company has tested 32 cars and thinks that the true mean fuel consumption (in miles per gallon) is 22.5.
- We set up the following hypotheses
  - null: the true mean fuel consumption is equal to 22.5
  - alternative: the true mean fuel consumption is not equal to 22.5.
- How do we test this hypothesis?

One-sample t-test

What is it for?

- A one sample t-test is used to determine whether the mean of a sample significantly differs from a known population mean.

What does it do?

- It tells you whether the mean of a sample differs from a hypothesised mean.

What is the output?

- A p-value which indicates the probability that the data are consistent with the null hypothesis that there is no statistical difference between the sample mean and the hypothesised mean.

One-sample t-test

How do you interpret the output?

- If the p-value is small, typically <0.05, then there is enough evidence to reject the null hypothesis.

What restrictions are there on its use?

- If the data are severely skewed or sample sizes are small, then a non-parametric test may be more suitable.

Example: Motor Trend Car Road Tests

- We want to test the following hypotheses
  - null: the true mean fuel consumption is equal to 22.5
  - alternative: the true mean fuel consumption is not equal to 22.5.
- We choose a significance level of 0.05 for our test and construct the statistical decision rule
  - **IF** the p-value is less than 0.05
  - **THEN** we have enough evidence to reject the null hypothesis
  - **OTHERWISE** there is not enough evidence to reject the null hypothesis.

Example: Motor Trend Car Road Tests

- There are 32 samples in our dataset, with sample mean 20.09 and standard deviation 6.03.
- Performing a one-sample t-test gives us a p-value of 0.0309.
- This is less than the chosen significance level.
- We have enough evidence to reject the null hypothesis.
- We conclude the true population mean is not 22.5.

One-tailed and two-tailed Hypothesis tests

- There are two types of hypothesis test; **one-tailed** and **two-tailed**.
- One-tailed tests allow for the possibility of an effect in just one direction.
- In other words would like to test whether true population mean is significantly *greater than* or *less than* than a particular value.
- In two-tailed tests, you are testing for the possibility of an effect in either direction.

Example: Motor Trend Car Road Tests

- We previously tested the following hypotheses
  - null: the true mean fuel consumption is equal to 22.5
  - alternative: the true mean fuel consumption is not equal to 22.5.
- This is an example of a two-tailed test as we are not testing to see if the true mean is greater than or less than 22.5.
- Suppose we want to change to the following hypotheses
  - null: the true mean fuel consumption is 22.5
  - alternative: the true mean fuel consumption is less than 22.5.
- This is an example of a one-tailed test.

Example: Motor Trend Car Road Tests

- We choose a significance level of 0.05 for our test and construct the statistical decision rule
  - **IF** the p-value is less than 0.05
  - **THEN** we have enough evidence to reject the null hypothesis
  - **OTHERWISE** there is not enough evidence to reject the null hypothesis.
- Performing a one-sample t-test gives us a p-value of 0.0155.
- This is less than the chosen significance level.
- We have enough evidence to reject the null hypothesis.
- We conclude the true population mean is less than 22.5.

Any Questions?