Review of Inference Procedures

Hypothesis Test

- Setting up and testing hypotheses is an essential part of statistics.
- In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved, for example, claiming that a new drug is better than the current drug for treatment of the same symptoms.
- In each problem considered, the question of interest is simplified into two competing hypotheses between which we have a choice; the null hypothesis, denoted H_0, against the alternative hypothesis, denoted H_1.

Hypothesis Tests

- These two competing hypotheses are not however treated on an equal basis: special consideration is given to the null hypothesis.
- We have two common situations with regard to formulating hypothesis tests.
- In the first case, the experiment has been carried out in an attempt to disprove or reject a particular hypothesis, the null hypothesis, thus we give that one priority so it cannot be rejected unless the evidence against it is sufficiently strong.

For example,

H_0: there is no difference in taste between Coca Cola and Pepsi

H_1: there is a difference in taste between both.

Occam's Razor

- If one of the two hypotheses is 'simpler' we give it priority so that a more 'complicated' theory is not adopted unless there is sufficient evidence against the simpler one.
- See Occam's Razor
- For example, it is 'simpler' to claim that there is no difference in flavour between Coca Cola and Pepsi than it is to say that there is a difference.
- You would only prefer the more complex claim if there is sufficient evidence to back it up

Testing Parameter values: The hypotheses are often statements about population parameters like expected value and variance; for example H_0 might be that the expected value of the measurements taken by one clinical method are different from those of a similar clinical method (assuming both methods have the same purpose).

- A hypothesis might also be a statement about the distributional form of a characteristic of interest, for example that clinical measurements of a certain quantity are normally distributed.
- The outcome of a hypothesis test is "Reject H_0 in favour of H_1" or "Do not reject H_0".

Null Hypothesis

The null hypothesis, H_0, represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. We would write

H_0: there is no difference between the two drugs on average.

We give special consideration to the null hypothesis. This is due to the fact that the null hypothesis relates to the statement being tested, whereas the alternative hypothesis relates to the statement to be accepted if / when the null is rejected. The final conclusion once the test has been carried out is always given in terms of the null hypothesis.

Important:

- We either "Reject H_0 in favour of H_1" or "Do not reject H_0". We never conclude "Reject H_1", or even "Accept H_1".
- If we conclude "Do not reject H_0", this does not necessarily mean that the null hypothesis is true, it only suggests that there is not sufficient evidence against H_0 in favour of H_1.
- Rejecting the null hypothesis then, suggests that the alternative hypothesis may be true.

Alternative Hypothesis

- The alternative hypothesis, H_1, is a statement of what a statistical hypothesis test is set up to establish.
- For example, in a clinical trial of a new drug, the alternative hypothesis might be that the new drug has a different effect, on average, compared to that of the current drug.
- We would write H_1: the two drugs have different effects, on average.

The alternative hypothesis might also be that the new drug is better, on average, than the current drug. In this case we would write H_1: the new drug is better than the current drug, on average. The final conclusion once the test has been carried out is always given in terms of the null hypothesis. As before, we either "Reject H_0 in favour of H_1" or "Do not reject H_0". We never conclude "Reject H_1", or even "Accept H_1". If we conclude "Do not reject H_0", this does not necessarily mean that the null hypothesis is true, it only suggests that there is not sufficient evidence against H_0 in favour of H_1. Rejecting the null hypothesis then, suggests that the alternative hypothesis may be true.

Confidence Intervals

Confidence Interval

- A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.
- If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter.
- Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.

- The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter (recall precision).
- A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.
- Confidence intervals are more informative than the simple results of hypothesis tests (where we decide "reject H_0" or "don't reject H_0") since they provide a range of plausible values for the unknown parameter.

Confidence Limits

Confidence limits are the lower and upper boundaries / values of a confidence interval, that is, the values which define the range of a confidence interval. The upper and lower bounds of a 95% confidence interval are the 95% confidence limits. These limits may be taken for other confidence levels, for example, 90%, 99%, 99.9%.

- The confidence level is the probability value $(1 - \alpha)$ associated with a confidence interval. It is often expressed as a percentage. For example, say $\alpha = 0.05 = 5\%$, then the confidence level is equal to (1-0.05) $= 0.95$, i.e. a 95% confidence level.

Significance Level

The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true.

It is the probability of a type I error and is set by the investigator in relation to the consequences of such an error. That is, we want to make the significance level as small as possible in order to protect the null hypothesis and to prevent, as far as possible, the investigator from inadvertently making false claims. The significance level is usually denoted by α

$$SignificanceLevel = P(typeIerror) = \alpha$$

Usually, the significance level is chosen to be 0.05 (i.e. 5

Test Statistic

A test statistic is a quantity calculated from our sample of data. Its value is used to decide whether or not the null hypothesis should be rejected in our hypothesis test. The choice of a test statistic will depend on the assumed probability model and the hypotheses under question.

P-Value

The probability value (p-value) of a statistical hypothesis test is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H_0, is true. It is the probability of wrongly rejecting the null hypothesis if it is in fact true.

It is equal to the significance level of the test for which we would only just reject the null hypothesis. The p-value is compared with the actual significance level of our test and, if it is smaller, the result is significant. That is, if the null hypothesis were to be rejected at the 5% significance level, this would be reported as ”p ¡ 0.05”.

Small p-values suggest that the null hypothesis is unlikely to be true. The smaller it is, the more convincing is the rejection of the null hypothesis. It indicates the strength of evidence for say, rejecting the null hypothesis H_0, rather than simply concluding "Reject H_0' or "Do not reject H_0". When using R, we will see a classification structure for various levels of p-values.

Type I and II Error

- In a hypothesis test, a type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H0 is wrongly rejected. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; i.e.

  H0: there is no difference between the two drugs on average.

Type I and II Error

Type I Errors

- A type I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them.
- The hypothesis test procedure is therefore adjusted so that there is a guaranteed 'low' probability of rejecting the null hypothesis wrongly; this probability is never 0.
- This probability of a type I error can be precisely computed as

$$\Pr(\text{type I error}) = \text{significance level} = \alpha$$

Type I and II Error

- For any given set of data, type I and type II errors are inversely related; the smaller the risk of one, the higher the risk of the other.
- A type I error can also be referred to as an error of the first kind.

Type II Error

- In a hypothesis test, a type II error occurs when the null hypothesis H0, is not rejected when it is in fact false.
- For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; i.e. H0: there is no difference between the two drugs on average.

Type II Error

- A type II error would occur if it was concluded that the two drugs produced the same effect, i.e. there is no difference between the two drugs on average, when in fact they produced different ones. A type II error is frequently due to sample sizes being too small.
- The probability of a type II error is generally unknown, but is symbolised by and written

$$P(\text{type II error}) = \beta$$

- A type II error can also be referred to as an error of the second kind.
- The exact probability of a type II error is generally unknown. If we do not reject the null hypothesis, it may still be false (a type II error) as the sample may not be big enough to identify the falseness of the null hypothesis (especially if the truth is very close to hypothesis).

Review of Inference Procedures

Summary:
The following table gives a summary of possible results of any hypothesis test:

	H_0 True	H_0 False
Reject H_0	Type I Error	Correct Rejection
Fail to Reject H_0	Correct Decision	Type II Error

Type I and II Error

- A type I error is often considered to be more serious, and therefore more important to avoid, than a type II error.

Power

- The power of a statistical hypothesis test measures the test's ability to reject the null hypothesis when it is actually false - that is, to make a correct decision.

Review of Inference Procedures

- In other words, the power of a hypothesis test is the probability of not committing a type II error.
- It is calculated by subtracting the probability of a type II error from 1, usually expressed as:

$$\textit{Power} = 1 - P(\text{type II error}) = 1 - \beta$$

- The maximum power a test can have is 1, the minimum is 0. Ideally we want a test to have high power, close to 1.

One Tailed Testing

- A one-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H0 are located entirely in one tail of the probability distribution.
- In other words, the critical region for a one-sided test is the set of values less than the critical value of the test, or the set of values greater than the critical value of the test.

Review of Inference Procedures

- We generally use it to formally test whether a parameter value is greater or less than some specified value, or for the case of two samples, if the parameter values from sample is greater or less than the corresponding parameter value from the other sample.
- A one-sided test is also referred to as a one-tailed test of significance.
- Remark: For the sake of brevity, we will mostly work on the basis of two-tailed tests throughout this module.
- The choice between a one-sided and a two-sided test is determined by the purpose of the investigation or prior reasons for using a onesided test.

One Sample t-test

A one sample t-test is a hypothesis test for answering questions about the mean where the data are a random sample of independent observations from an underlying normal distribution $N(\mu, \sigma^2)$, where is unknown. The null hypothesis for the one sample t-test is: H_0: $\mu = \mu_0$, where μ_0 is some specified number. That is, the sample has been drawn from a population of given mean and unknown variance (which therefore has to be estimated from the sample). This null hypothesis, H_0 is tested against one of the following alternative hypotheses, depending on the question posed:

H_1: μ is not equal to μ_0

H_1: $\mu > \mu_0$

H_1: $\mu < \mu_0$

Confidence Interval for a Mean

A confidence interval for a mean specifies a range of values within which the unknown population parameter, in this case the mean, may lie. These intervals may be calculated by, for example, a medical researcher who wishes to estimate the mean response by patients to a new drug; etc.

Confidence Interval for a Mean

The (two sided) confidence interval for a mean contains all the values of μ_0 (the true population mean) which would not be rejected in the two-sided hypothesis test of: H_0: $\mu = \mu_0$ H_1: μ not equal to μ_0 The width of the confidence interval gives us some idea about how uncertain we are about the unknown population parameter, in this case the mean.

Confidence Intervals

- A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.
  We calculate these intervals for different confidence levels, depending on how precise we want to be. We interpret an interval calculated at a 95% level as we are 95% confident that the interval contains the true population mean. We could also say that 95% of all confidence intervals formed in this manner (from different samples of the population) will include the true population mean.

Two Sample t-test

A two sample t-test is a hypothesis test for answering questions about the mean where the data are collected from two random samples of independent observations, each from an underlying normal distribution: Important: When carrying out a two sample t-test, it is common to assume that the variances for the two populations are equal, i.e. The null hypothesis for the two sample t-test is: H_0: $\mu_1 = \mu_2$ That is, the two samples have both been drawn from the same population.

Two Sample t-test

This null hypothesis is tested against one of the following alternative hypotheses, depending on the question posed. H_1: μ_1 is not equal to μ_2 H_1: $\mu_1 > \mu_2$ H_1: $\mu_1 < \mu_2$

Confidence Interval for the Difference Between Two Means

- A confidence interval for the difference between two means specifies a range of values within which the difference between the means of the two populations may lie.
- These intervals may be calculated by, for example, a medical researcher who wishes to estimate the difference in mean response by patients who are receiving two different drugs; etc.

The confidence interval for the difference between two means

- This interval contains all the values of $\mu_1 - \mu_2$ (the difference between the two population means) which would not be rejected in the two-sided hypothesis test of:

H$_0$: $\mu_1 = \mu_2$
against
H$_1$: μ_1 not equal to μ_2

Equivalently

H$_0$: $\mu_1 - \mu_2 = 0$
against
H$_1$: $\mu_1 - \mu_2$ not equal to 0

Important:

- If the confidence interval includes 0 we can say that there is no significant difference between the means of the two populations, at a given level of confidence. interpret an interval calculated at a 95% level as *we are 95% confident that the interval contains the true difference between the two population means*.
- We could also say that 95% of all confidence intervals formed in this manner (from different samples of the population) will include the true difference.

Review of Inference Procedures

Paired t-test A paired t-test is used to compare two population means where there are two samples in which observations in one sample can be paired with observations in the other sample.

Examples of where this might occur are:

- Before-and-after observations on the same subjects (e.g. patients diagnostic test results before and after a particular course of treatment).
- A comparison of two different methods of measurement or two different treatments where the measurements/treatments are applied to the same subjects (e.g. measurements made with Ultra Violet Spectroscopy and Near Infrared Reflectance spectroscopy).

Review of Inference Procedures

The hypotheses can be stated as follows:

H0: $mu_{diff} = 0$ Population mean of case-wise differences is zero

H1: $mu_{diff} = 0$ Population mean of case-wise differences is not zero

- The null hypothesis would articulate the argument that a course of treatment had no effect on the subjects, or for the second case, that there is no significant measurement bias between two methods of measurement.
- We will perform a case study of this in the Lab Classes.