MA4605 Lecture 2B Part 2

(Theory for Inference Procedures)

- **Types I and II error**
- **Power**
- **One Tailed Testing**
- **Example: Two Sample Testing**
- **Formal Test for Equality of Variances**
- **Paired t-test**

Type I Error

Important: In a hypothesis test, a type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H_0 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.

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

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

P(type I error) = significance level = α

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).

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 H_0, 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.

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

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 β and written

P(type II error) = β

A type II error can also be referred to as an error of the second kind.

The rate of the type II error is related to the power of a test (which equals 1−β).

Summary

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

		Decision	
		Reject H0	**Don't reject H0**
Truth	**H0**	Type I Error	Right decision
	H1	Right decision	Type 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.

In other words, the power of a hypothesis test is the probability of not committing a ***type II error***.0

It is calculated by subtracting the probability of a type II error from 1, usually expressed as:

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

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.

Power analysis can be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size.

Power analysis can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size.

One Tailed Testing

- A one-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H_0 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.
- 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.
- 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 one-sided test.

Recall the titration experiments from the previous week:

A	10.08	10.11	10.09	10.10	10.12
B	9.88	10.14	10.02	9.80	10.21
C	10.19	9.79	9.69	10.05	9.78
D	10.04	9.98	10.02	9.97	10.04

We shall perform a series of one sample and two sample tests.

Recall the true value of the titration experiment in each case is supposed to be 10.

First we will consider the case of A's measurements (as we did in the previous lecture).

The mean and standard deviation of A's measurements are as follows:

```
> mean(X.A)
[1] 10.1
> sd(X.A)
[1] 0.01581139
```

We will perform two one-tailed tests.

To recap, we performed a two tailed test in the previous lecture (i.e. the true mean is equal to zero). To contrast with the one-tailed tests, here is it again, with the alternative specified.

```
> t.test(X.A, mu=10, alternative = "two.sided")

        One Sample t-test

data:  X.A
t = 14.1421, df = 4, p-value = 0.0001451
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
 10.08037 10.11963
sample estimates:
mean of x
     10.1
```

We will perform a "greater than" test. The null and alternative are specified as follows:

H_0: $\mu_A \leq 10$ True value of population mean is no more than 10

H_1: $\mu_A > 10$ True value of population mean is greater than 10.

```
> t.test(X.A, mu=10, alternative = "greater")

        One Sample t-test

data:  X.A
t = 14.1421, df = 4, p-value = 7.256e-05
alternative hypothesis: true mean is greater than 10
95 percent confidence interval:
 10.08493      Inf
sample estimates:
mean of x
     10.1
```

In this case, we would reject the null hypothesis, based on the extremely low p-value. There is very convincing evidence to say that the true mean of A's measurements is greater than 10.

(Furthermore there is a systematic upward bias in A's measurements)

Now we will perform a "less than" test. The null and alternative are specified as follows:

H_0: $\mu_A \geq 10$ True value of population mean is at least 10

H_1: $\mu_A < 10$ True value of population mean is less than 10.

```
> t.test(X.A, mu=10, alternative = "less")

        One Sample t-test

data:  X.A
t = 14.1421, df = 4, p-value = 0.9999
alternative hypothesis: true mean is less than 10
95 percent confidence interval:
     -Inf 10.11507
sample estimates:
mean of x
     10.1
```

In this case, we would fail to reject the null hypothesis, based on the extremely high p-value.

Example: Two Sample Testing

In the previous class, we discussed the test of equality of population mean for two independent samples.

Let us use the measurements from Students A and B. The mean and standard deviation of B's measurements are as follows:

```
> mean(X.B)
[1] 10.01
> sd(X.B)
[1] 0.1717556
```

The hypotheses can be stated as follows:

H_0: $\mu_A = \mu_B$ Population means are equal for students A and B

H_1: $\mu_A \neq \mu_B$ Population means are not equal for A and B

To implement such as test in ***R***, we simply specify both data sets.

```
> t.test(X.A, X.B)

        Welch Two Sample t-test

data:  X.A and X.B
t = 1.1668, df = 4.068, p-value = 0.3071
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1227643  0.3027643
sample estimates:
mean of x mean of y
    10.10     10.01
```

Based on the p-value, we fail to reject the null hypothesis.

Note the name given to the output of the procedure " ***Welch Two Sample t-test***".

There are in fact two different two-sample t-tests that can be implemented in R.

The test that we have just completed, i.e. the Welch test, does not make the assumption that both data sets have equal variance.

Conversely, the ***"Student Two Sample t-test"*** does make that assumption when performing the test.

To perform a Student Two Sample t-test using ***R***, we must additionally specify that the variances are assumed to be equal

```
> t.test(X.A, X.B, var.equal =TRUE)

        Two Sample t-test

data:  X.A and X.B
t = 1.1668, df = 8, p-value = 0.2769
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0878765  0.2678765
sample estimates:
mean of x mean of y
    10.10     10.01
```

In this instance, we would come to the same conclusion whether or not we specify the assumption of equal variances.
Notice, however, the p-value and confidence intervals are quite different.
In some instances, it is possible that a null hypothesis would be rejected by one test, but not by the other.

Formal Test for Equality of Variances

R provides us with a formal test for equality of variances for two samples.

The hypotheses can be stated as follows:

$H_0: \sigma^2_A = \sigma^2_B$ Population variances are equal for A and B

$H_1: \sigma^2_A \neq \sigma^2_B$ Population variances are not equal for A and B

Specifically, R considers the ratio of the variances. If two populations have equal variance, then this variance ratio is one.

The hypotheses can be re-stated as follows:

$H_0: \sigma^2_A / \sigma^2_B = 1$ Population variances ratio is 1

$H_1: \sigma^2_A / \sigma^2_B \neq 1$ Population variances ratio is not 1

To recap – the variances of A's and B's measurements are as follows. Also computed is the ratio of these variances.

```
> var(X.A)

[1] 0.00025

> var(X.B)

[1] 0.0295

> var(X.A)/var(X.B)

[1] 0.008474576

> var(X.B)/var(X.A)

[1] 118
```

The test is carried out using the `var.test()` command.

```
> var.test(X.A, X.B)

        F test to compare two variances

data:  X.A and X.B
F = 0.0085, num df = 4, denom df = 4, p-value = 0.0004213
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.000882352 0.081394321
sample estimates:
ratio of variances
       0.008474576
```

Here we reject the null hypotheses. The p-value is extremely small.

It doesn't matter which order the data sets are specified, other than the fact that it will reciprocate the variance ratio.

```
> var.test(X.B, X.A)

        F test to compare two variances

data:  X.B and X.A
F = 118, num df = 4, denom df = 4, p-value = 0.0004213
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
   12.28587 1133.33453
sample estimates:
ratio of variances
               118
```

Notice the 95% confidence interval for the variance ratio. We are 95% confident that this interval contains the true variance ratio (i.e. we are 95% confident that the true variance ratio is between 12.28 and 1133.33).

We can reject the null hypothesis on the basis that the value of 1 is not contained in that interval.

Referring back to the two-sample t test; the first test i.e. the Welch test, did not rely on the equality of variances, so it is the preferred procedure in that case.

Summary: before performing a two-sample t test, check to see if the assumption of equal variance is valid.

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. patient's 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).

The hypotheses can be stated as follows:

H_0: $\mu_{diff} = 0$ Population mean of case-wise differences is zero

H_1: $\mu_{diff} \neq 0$ 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.