*T*-tests

(& hypothesis testing)

# Slide 1 – Title

Hey everyone! Josh, here. In this lesson, we are going to learn about *t*-tests – specifically, two sample *t*-tests – and how they are used to test simple hypotheses about group differences in a statistical framework.

# Slide 2 – Lesson objectives

By the end of this lesson, you should be able to:

1. Explain how *t*-statistics are used to test hypotheses about group differences
2. Apply *t*-statistic calculations to real data and interpret the significance of results
3. Evaluate how sample size, effect size, and standard deviation affect the ability to detect (truly) significant differences in *t*-tests
4. Use R to simulate and analyse data

# Slide 3 – *T*-tests: A difference in means

The purpose of a *t*-test is to determine whether the mean of a group of observations is significantly different from an expected value. This might involve sampling a single group for some property to determine if the mean of the group differs from some predefined expectation. Alternatively, we can sample two groups for the same property and ask whether the means of those groups are significantly different. The first example is called a “One sample *t*-test”, the other is called a “Two sample *t*-test” – and it is this second example that we will focus on in this lesson.

# Slide 4 – Hypotheses

Okidokey – I have just told you that a two sample *t­*-test can be used to determine if the means between two sampled groups differ.

\*\* Take a moment to pause this video a write down the two competing hypotheses that you think are being tested in a two sample *t*-test. \*\*

Welcome back! The two hypotheses being tested in a two sample *t*-test are referred to as the “null” and “alternate” hypothesis.

A “null” is something that is associated with a zero value, so the null hypothesis is that “there is no difference in means between the two groups”.

Therefore, the “alternate” hypothesis is the oppose, in that “there is a difference in means between the two groups”.

We can formulate these written hypotheses with mathematical notation. If the symbol *μ* represents a group’s mean, and the numbers 1 and 2 represent groups 1 and 2, then the null can be phrased as:

And the alternate can be phrased as:

# Slide 5 – Data properties

We now have two formal hypotheses to test: equal means, or different means. However, we cannot simply calculate the means of two groups and say they are significantly different. Imagine if we had one group with a mean of 5 and another group with a mean of 6 – do you think these two groups significantly differ?

The answer is that it “depends”. Determining significance depends on several properties of our data, which include:

1. The number of samples we collect in each group
2. The true difference, or “effect size” between group means
3. The variability, or spread, of values around their group means

Statistical frameworks help us assess whether observations are significant given the properties of our data – because making false conclusions are bad! However, typically, concluding there is a difference when there truly isn’t one is worse than concluding there isn’t a different when there truly is one.

# Slide 6 – The *t*-statistic (two samples)

To test between the null and alternate hypothesis, we need a statistic that captures the difference in means between the two groups. We call this the *t*-statistic.

The *t*-statistic is separated into two components. The numerator (top part of the fraction) is simply the difference in means of group 1 and group 2. The denominator (bottom part of the fraction) is the standard error (SE) of the sample.

For a two sample *t*-test, the *t*-statistic is a “standardised difference in means”.

Dividing the difference in means (*μ*1 − *μ*2) by the standard error (SE) controls for the inherent variability in data and the respective sample sizes of each group.

# Slide 7 – Assumptions

There are some assumptions that our data must adhere to before we conduct a *t*-test.

1. Firstly, the data should be (approximately) normally distributed.
2. Secondly, to use this formulation of the *t*-statistic for two samples, the variance of both groups must be equal.

# Slide 8 – Simulation: Data properties

We are now going to use R code to explore how *t*-statistic calculations work, and how properties of a dataset affect the conclusions drawn from a *t*-test.

Open up R and load the *StatsModules\_2sample\_t.R* script. Alternatively, you can simply cut and paste the code from these slides into R.

Code Snippet 1 sets up the parameters for our simulations.

**library(ggplot2)**

**#### SNIPPET 1 ####**

**# Params:**

**# n = sample size (both groups)**

**# eff = effect size (negative or positive number)**

**# s = standard deviation**

**n <- 50**

**eff <- 1**

**s <- 1**

To start the simulations, we will work with a sample size of *n* = 50 for both groups in our sample. We will simulate two scenarios: one where the null hypothesis is true, and another where it is false (that is, the alternate hypothesis is true). When the alternate is true, the group means must differ, and we will simulate this difference using *eff* = 1, an “effect size” of 1 unit difference. Finally, each group will have a standard deviation of *s* = 1 around their respective means.

# Slide 9 – Simulation: Group distributions

We will now use a function called *rnorm* to simulate data following the parameters we specified. The code for the simulated data is in Code Snippet 2.

**#### SNIPPET 2 ####**

**# Two groups, x or y.**

**# Group y0 is Ho TRUE, and group y1 is Ho FALSE.**

**x <- rnorm(n=n, sd=s, mean=50)**

**y0 <- rnorm(n=n, sd=s, mean=50)**

**y1 <- rnorm(n=n, sd=s, mean=50+eff)**

**# Print object contents to screen, they contain**

**# simulated values.**

**x**

**y0**

**y1**

In this simulation, we have created 3 objects, x, y0, and y1, which represent different groups that have a sample size of *n* and a standard deviation of *s*. The mean for group X is 50. Group Y0 is a second group, which we have simulated to have the same mean as group X. In contrast, group Y1 is an alternate version of the second group, where the mean differs by the effect size that we specified as *eff*.

# Slide 10 – Simulations: Visualise distributions

We can more easily visualise these differences between the simulated groups using histograms to graph their frequency distributions – as per Code Snippet 3.

**#### SNIPPET 3 ####**

**# Plotting**

**plot(ggplot(data.frame(vals=c(x,y0,y1)**

**, group=c(rep('x', n), rep('y0', n), rep('y1', n)))**

**, aes(x=vals, fill=group))**

**+ theme(legend.position='none'**

**, strip.text.x = element\_text(face='bold'))**

**+ geom\_histogram(colour='black')**

**+ geom\_vline(xintercept=50, linetype='dashed', colour='black')**

**+ scale\_fill\_manual(values=c(x='#737373', y0='#d65cd4', y1='#51d5e1'))**

**+ facet\_wrap(~ group, ncol=1)**

**+ labs(x='Values', y='Frequency')**

**)**

****

This graph will look different each time due to the randomness simulations. What you should be able to see more clearly is the different distribution of the values in each group. The black dashed line indicates a value of 50. The distribution of group X and Y0 values will be roughly centred on 50, whereas the distribution of Y1 will sit slightly higher than 50 because we simulated an effect size of 1.

# Slide 11 – Simulations: *t*-statistics by hand

We will now calculate the *t*-statistics for two simulated scenarios: First, group X with Y0, where the null hypothesis is true (*i.e.* no difference in group means); and second, group X with Y1, where the null hypothesis is false (*i.e.* there is a difference in group means).

In Code Snippet 4, we first calculate the sample standard error for each of the respective scenarios. Next, we calculate the group means for each of X, Y0, and Y1. Finally, we calculate the *t*-statistic from each scenario by dividing the difference in group means by the sample standard error.

**#### SNIPPET 4 ####**

**# Functions:**

**# sqrt() = calcualtes square root**

**# sd() = calculates the standard deviation**

**# Remember that SD^2 = variance**

**se0 <- sqrt((sd(x)^2/n)+(sd(y0)^2/n))**

**se1 <- sqrt((sd(x)^2/n)+(sd(y1)^2/n))**

**# Means**

**x.mean <- mean(x)**

**y0.mean <- mean(y0)**

**y1.mean <- mean(y1)**

**# Observed t-scores**

**t0 <- (x.mean-y0.mean-0)/se0**

**t1 <- (x.mean-y1.mean-0)/se1**

# Slide 12 – Simulations: Significance testing

We will now conduct a two-sided test to determine whether or not the calculated statistics from each scenario suggest statistically significant differences between group means. See Code Snippet 5.

**#### SNIPPET 5 ####**

**# Take a look at the simulated t-statistics**

**t0**

**t1**

**# P-value: Probability of observing t if Ho = TRUE,**

**# given the number of degree of freedom (df).**

**# Two-tailed test**

**# p-value: simulated difference is 0 (Ho = TRUE).**

**2\*(1-pt(abs(t0), df=n+n-2, lower.tail=TRUE))**

**# p-value: simulated difference isn't 0 (Ho = FALSE).**

**2\*(1-pt(abs(t1), df=n+n-2, lower.tail=TRUE))**

There are two important concepts two take away from Code Snippet 5:

1. The expected value of *t*-statistics calculated in each simulated scenario.
2. The probability of observing each statistic, if the null hypothesis was true.

We will address these points in the next two slides.

# Slide 13 – Observed (vs expected) *t*-statistics

Let’s first reflect on the *t*-statistics you calculated from each scenario where the null was true (t0) or false (t1).

**# Take a look at the simulated t-statistics**

**t0**

**t1**

\*\* Stop and think for a moment about what values of t0 and t1 you were expecting, given the scenarios they were simulated under. \*\*

In theory, t0 should be close to zero: X and Y0 were simulated to have the same mean of 50. In contrast, t1 should be different from zero: X and Y1 were simulated to have different means, X with a mean of 50, and Y1 with a mean of 50 + effect size.

The actual values of these observed *t*-statistics will vary with each simulation. So have a go a rerunning the code from Code Snippet 1 🡪 5, and see how the values of t0 and t1 change.

# Slide 14 – Probability of *t*, if HO is true

To test whether our data is in alignment with the null or alternate hypothesis, we test whether our observed *t*-statistic is likely to arise from a scenario where the null hypothesis is true – that is, a scenario where there is truly no difference between group means. This probability value is called the “*p­­*-value”.

**# p-value: simulated difference is 0 (Ho = TRUE).**

**2\*(1-pt(abs(t0), df=n+n-2, lower.tail=TRUE))**

**# p-value: simulated difference isn't 0 (Ho = FALSE).**

**2\*(1-pt(abs(t1), df=n+n-2, lower.tail=TRUE))**

You do not need to worry about the specifics of this code, but you should appreciate the following:

* The degrees of freedom (DF) for assessing significance is the total sample minus 2, because we have two groups and each group independently has *n* – 1 DF.
* This is a two-tailed test, so we are not making explicit predictions about whether the first group’s mean is larger than the second, or vice versa.
* A probability < 0.05 suggests that the observed *t*-statistic has a very low probability of arising from a scenario where the null hypothesis is true.

Given the parameters of *n* = 50, *eff* = 1, and *s* = 1, the probability of t0 should be > 0.05 and t1 should be < 0.05.

Therefore, we would conclude non-significant difference in means between group X and Y0, and a significant difference in means between X and Y1.

# Slide 15 – Sample size, effect size, & SD (I)

In the final part of this lesson we are going to explore how properties of our data, that is, the sample size, effect size, and standard deviation, might affect our ability to identify significant differences between groups – even when the null hypothesis is false, and the alternate hypothesis is true.

Let us first consider the **effect size**. Remember this parameter determined how different the group means were when the null hypothesis was false.

Go back to Code Snippet 1 and change the value of *eff* to 0.025.

**n <- 50**

**eff <- 0.025**

**s <- 1**

Now rerun all the code from Code Snippets 1 🡪 5.

What you should notice is that the distribution of values in group Y1 is almost indistinguishable from that in group X, even though there is truly a difference between group means.

Indeed, when you look at the probability of t1, the value you get will almost certainly be greater than 0.05:

**2\*(1-pt(abs(t1), df=n+n-2, lower.tail=TRUE))**

**[1] 0.4494341E**

These results indicate that this effect size is too small to detect, given our sample, even though we know a difference in means between X and Y1 truly exists!

\*\* Before moving on, reflect on how the distribution of values, and observed *t*-statistics for the alternate scenario (t1) differ between this simulation (small effect size) and the original simulation (larger effect size) \*\*



# Slide 16 – Sample size, effect size, & SD (II)

When differences between groups is very small, we need a much larger **sample size** to detect those differences. Head back to Code Snippet 1 and edit the sample size to be *n* = 20,000.

**n <- 20000**

**eff <- 0.025**

**s <- 1**

Now rerun all the code from Code Snippets 1 🡪 5. Now note the probability for the test of the null hypothesis between group X and Y1:

**2\*(1-pt(abs(t1), df=n+n-2, lower.tail=TRUE))**

**[1] 0.004425624**

The value should be < 0.05. You should now be able to appreciate the importance of sample size for increasing our **power** to detect truly significant differences when those differences are small.



# Slide 16 – Sample size, effect size, & SD (III)

A small effect size is not the only thing that can obscure a detectable difference in means between groups. The **standard deviation** is the variance in the data, that is, how much it is spread out around the mean. When the standard deviation is large, the distributions between groups will potentially increase in their overlap, reducing the ability to detect significant differences when the null hypothesis is false.

Edit Code Snippet 1 to reset the sample size to *n* = 50 and effect size to *eff* = 1. Except this time, we are going to make the standard deviation ridiculously large, at *s* = 10.

**n <- 50**

**eff <- 1**

**s <- 10**

Rerun Code Snippets 1 🡪 5.

You should notice from the histogram that the range of values produced is huge, in comparison to our prior simulations that had a SD of 1. This huge spread of values also obscures the fact that group Y1 has a mean that is truly different from group X.

If you look at the probability of t1 assuming the null hypothesis, it is highly likely that all values you simulate will be > 0.05. This means that despite the true difference between X and Y1, the standard deviations are so large that we are unable to reject the possibility that the null hypothesis is true, even though it is false!

**2\*(1-pt(abs(t1), df=n+n-2, lower.tail=TRUE))**

**[1] 0.5921636**



# Slide 17 – Play time!

It is now your turn to play with the code and build your own understanding of *t*-tests and the effect of data properties on hypothesis testing.

Edit the simulation parameters *n*, *eff*, and *s* in Code Snippet 1, and see how they affect the outcomes of your analyses.

Some questions to ponder:

1. If the standard deviation is very large, what increases your ability to detect truly significant differences between X and Y1?
2. If the effect size is small, and the sample size is also small, what increases your ability to detect a truly significant difference between X and Y1?
3. How does a negative effect size (*eff* < 0) change (or not change) the calculated *t*-statistic and the test of significance?

# Slide 18 – Recap

Let us reflect once more on the learning objectives of this lesson. After going through this lesson material and experimenting with the code provided, you should now be able to:

1. Explain how *t*-statistics are used to test hypotheses about group differences
2. Apply *t*-statistic calculations to real data and interpret the significance of results
3. Evaluate how sample size, effect size, and standard deviation affect the ability to detect (truly) significant differences between groups
4. Use R to simulate and analyse data

Good job!