T-test

Evaluation Methods & Statistics – Lecture 7

Benjamin Cowan & Andrew Howes

What we will cover

- Experiment Design
- Independent, Dependent Variables
- 2 samples t-test
  - Independent
  - Dependent
- Reporting the findings

A Study.....

Research Example

- Consequences of a secondary task on driving
- Does using a mobile phone to text cause driving quality to deteriorate?

Independent and dependent variables

- We systematically manipulate the IV

- We want to see how this systematic manipulation impacts an outcome (or DV)

Research Set-Up

Independent Variable (IV)

- **Secondary Driving Task**

This independent variable has **2 levels** or **factors**

1) Control Group

- Just driving (no secondary task)

2) Texting Group

- Participants asked to text whilst driving

Dependent Variable (DV)

- **Driving Quality Score**

Main task for the participants is to drive for an hour in a simulator

Hypotheses

Null Hypothesis

- There will be no significant difference between **secondary driving tasks** on **driving quality score**

Research Hypothesis

- There will be a significant difference between **secondary driving tasks** on **driving quality score**

Experiment Design

Methods of data collection

Between-subjects

Different groups of people take part in each driving task

Within-Subjects

Same group of people take part in each driving task

IV Level 1 Participants

IV Level 2 Participants

IV Level 1 Participants

IV Level 2 Participants

Methods of data collection

- This influences what statistical test you use

- It also influences how you design your experiment
  - Counterbalancing (In Within Subjects)
  - Sample matching (in Between Subjects)
  - To minimise confound effects on the variance

Two types of DV Variation

Unsystematic

Differences in DV due to unknown (or unmeasured) factors

e.g. personality, change in room environment between conditions 1 and 2

Systematic

Differences in DV due to the change in IV

e.g. Changes in driving score due to systematic manipulation of IV

Between-Subjects

Unsystematic variation due to sample differences

- naturally vary in IQ, Personality, attention span

We can try to control for this

- Include high risk variables as part of design
- E.g. high and low attention span as another IV

Randomly allocate participants to condition

- Spreads the variation across conditions

Within Subjects

- Unsystematic variation controlled
  - More power to identify effects
- Systematic confounds
  - Practice effects
  - Boredom effects
- Form a risk to our conclusions if we assume that systematic variation is all due to IV
- Counterbalance to reduce impact of this confound
  - 50% of participants- Condition 1 before 2
  - 50% of participants- Condition 2 before 1

T-Test

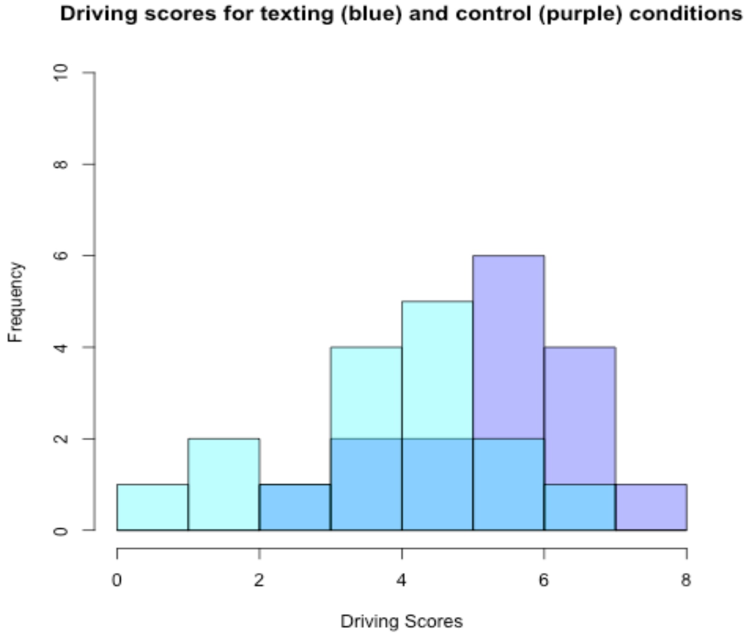
Guinness & t-test

- William Sealy Gosset (1876-1937)
- Worked for Guinness as a statistician
- They needed stats to examine which types of barley had the best yield
- In 1908 published a paper which introduced the t-distribution under the pseudonym "Student"

T-Test

- T-test is looking at **differences**
- It is testing to see if the two sample means gathered are from different populations
- Why might they be from different populations?
  - Due to the levels of our IV

T-test



T-Test: Rationale

If the sample mean difference is larger than we expect

- We have collected two sample by chance that are atypical of the population

OR

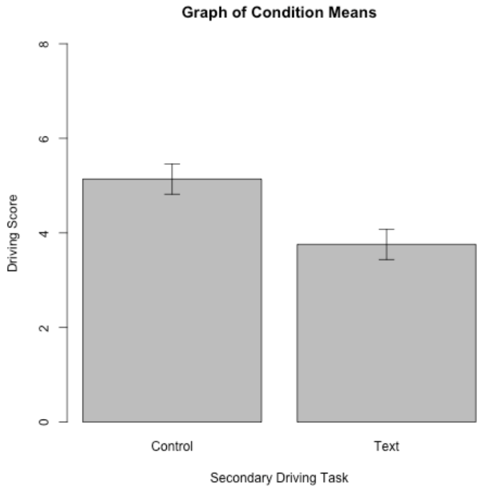
- The two samples are from different populations

T-Test: Rationale

If the sample comes from the same population (i.e. as assumed in H_0)

- we'd expect these means to be roughly equal
- **Large differences** between sample means should occur rarely

If they do, it is because sample means are from different populations



Types of t-test

- **Independent means t-test**
  - Different participants are allocated to each of the secondary driving task conditions
  - AKA independent measures & independent samples

- **Dependent means t-test**
  - Each participant completes both of the secondary driving task conditions
  - AKA matched pairs & paired samples

T-Test: The Formula

- Keep in mind with t-test we are interested in **differences**

The Dependent t-test: Calculation

Mean difference between samples (Experiment effect)

Difference expected between population means (because of H0 then this is 0)

test statistic for the t-test

$$t = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}}$$

Standard error of the differences

Dependent t-test calculation

$$\overline{D}$$

score_control	score_text	D
3.85	5	-1.15
5.29	4.96	0.33
5.52	4.09	1.43
5.07	3.77	1.3
5.11	5.24	-0.13
4.04	4.6	-0.56
5.34	5.36	-0.02
6.66	6.41	0.25

- The mean of the D values

Standard Error of Differences

$$S_D / \sqrt{N}$$

Used as variability gauge between sample means

- If this is small -- most samples should have similar means, thus we would expect a small D
- If this is large – large D is likely

Independent t-test

The same premise as the dependent but some important differences

Instead of differences between pairs of scores it looks at differences between **overall means**

Because we don't have pairs of scores for each participant

Independent t-test

This means we cannot calculate the SE of differences by using the **differences in the sample**

- There are 2 independent samples

Calculated using the **variance sum law**

- The **variance of a difference** between two independent variables is equal to the sum of their variances
- Then square root this to get SE

Effectively this is doing the same thing as the denominator in previous equation

Independent t-test equation- with equal sample sizes

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{s_1^{\,2}}{N_1} + \frac{s_2^{\,2}}{N_2}\right)}}$$

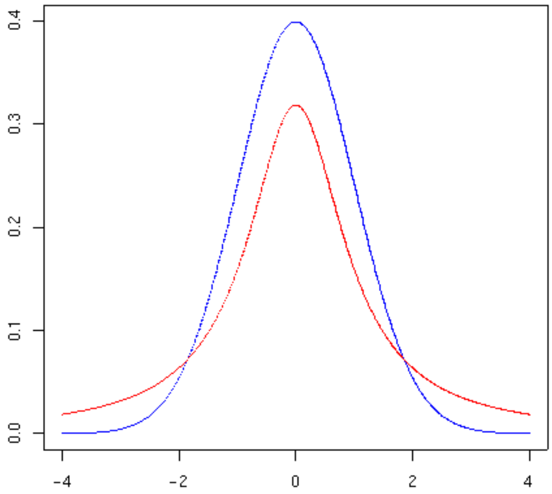
T-test: R output

```
> t.test (control, text, type=Student)

        Welch Two Sample t-test

data:  control and text
t = 2.673, df = 28.564, p-value = 0.01229
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3236936 2.4388064
sample estimates:
mean of x mean of y
 5.138125  3.756875
```

T Distribution



Reporting the t-test

- Need to report
  - The t statistic
  - The degrees of freedom (N-2)
  - The significance value (or full p statistic)

e.g. t (dof) = t statistic, pvalue

The Concept of Degrees of Freedom

- The number of observations that are free to vary
- Imagine a rugby team:
  - 15 people needed
  - If you arrive last then you have no choice in where to play
  - But the other 14 did
  - There are therefore 14 degrees of freedom (15-1)
  - In our t-test our degrees of freedom are the number of differences that are free to vary

The Concept of Degrees of Freedom

Degrees of freedom

- Dependent t-test
  - N-1
  - Because we are using 1 variable (mean of differences)
- Independent t-test
  - N-2
  - Because we are using 2 variables in the calculation

Parametric assumptions of t-test

Both tests

- Data is normally distributed (Shapiro Wilk test)
- Data is interval or ratio scale

Independent t-test only

- Variance in populations are roughly equal (Equality of variance)- Levene's test
- Scores are independent (i.e. They come from different people).

Shapiro-Wilk (Normality)

- Compares sample distribution to normal distribution
- If our sample varies significantly from this distribution, what would we expect?

```
> shapiro.test (control)

        Shapiro-Wilk normality test

data:  control
W = 0.9584, p-value = 0.6332
```

Shapiro-Wilk (Normality)

- Compares sample distribution to normal distribution
- If our sample varies significantly from this distribution, what would we expect?
- Test will be statistically significant
- We want it to be **non significant** ($p>.05$)

```
> shapiro.test (control)

        Shapiro-Wilk normality test

data:  control
W = 0.9584, p-value = 0.6332
```

Shapiro-Wilk

Levene's Test (Homogeinity of Variance)

- Compares variance in each sample to see if they are roughly equal
- We want both variances to be similar
- Do we want the test to be statistically significant or not?

```
> leveneTest(data$score, data$condition)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.9755 0.3312
      30
>
```

Levene's Test (Homogeinity of Variance)

- Compares variance in each sample to see if they are roughly equal
- We want both variances to be similar
- Do we want the test to be statistically significant or not?
- We want it to be **non significant** (p>.05)

```
> leveneTest(data$score, data$condition)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.9755 0.3312
      30
>
```

Levene's Test

Commands used in practical

Data Manipulation

- subset()

Graphing Data:

- barplot()
- boxplot()
- se.bar()
  - This is our own "homemade" function

Descriptives & Assumptions

- mysummary()
  - Again, a homemade function
- shapiro.test()
- leveneTest()

Analysis

- t.test()

Task for next week

- Complete the dependent and independent t-test manual calculation on Canvas
- Do this using a calculator (no R!)
- Solution will be posted online

Readings for this lecture

- Field, Miles & Field (2012) Chapter 9
- Howell (2010) Chapter 7 p.194-213