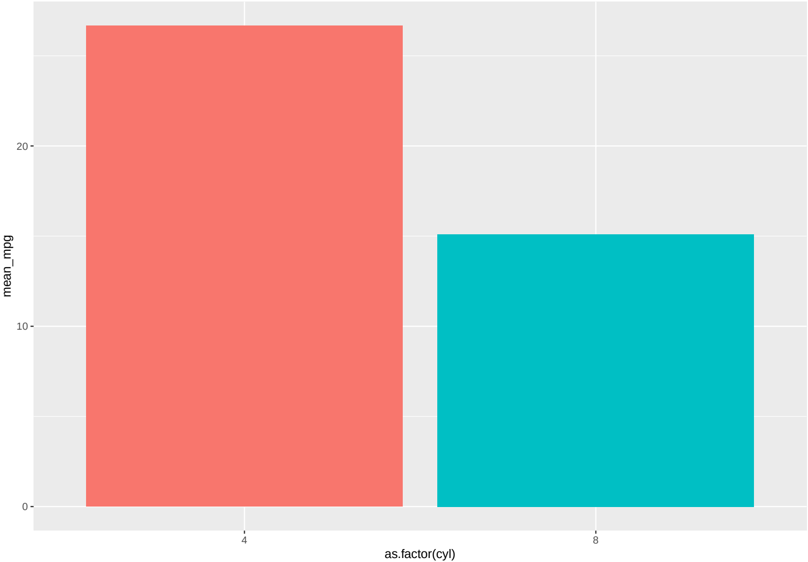
Independent Samples t-tests

Jonathan A. Pedroza, PhD

Looking at Differences

- the easiest form of an experiment is to compare two groups on an outcome
  - Ex: treatment vs control on depression rates
- you can also just compare two groups on an outcome that do not need to be in an experimental design
  - preliminary analyses - comparing the ages of male and female
- two different ways of comparing means
  - **between-groups/subjects or independent design/independent-samples**
    - comparing two *separate* groups together
  - **repeated-measures or within-subjects design**
    - comparing two time points

Technically Categorical Predictors in the Linear Model



Technically Categorical Predictors in the Linear Model

- the t-test is comparing the means between our two groups
  - it also fits in the linear model that we have been focusing on
  - Ex: comparing 4 and 8 cylinder cars on miles per gallon (MPG)

$$outcome_i = (model) + error_i$$

- Let's look at how these variables fit in our linear model

$$Y_i = (b_0 + b_1 X_{1i}) + e_i$$

$$MPG_i = (b_0 + b_1 8cyl_i) + e_i$$

Technically Categorical Predictors in the Linear Model

- since cylinder is a nominal variable, we convert this variable into numbers
  - we refer to these as **dummy variables**, where 0 means the value is not representative of that column and 1 means the value is representative of that column
  - the first case, **8 cylinder = 0, 4 cylinder = 1** states that the car in question is not a 8 cylinder and is a 4 cylinder

```
     8 4
[1,] 0 1
[2,] 1 0
[3,] 1 0
[4,] 0 1
[5,] 0 1
[6,] 1 0
```

Technically Categorical Predictors in the Linear Model

- Since the outcome is the same as the group mean for one group, we can make some changes to the equation

$$MPG_i = b_0 + b_1 8cyl_i$$

$$\overline{X}_{4cyl} = b_0 + (b1x0)$$

$$\overline{X}_{4cyl} = b_0$$

Technically Categorical Predictors in the Linear Model

- this gives us the intercept for our model, which is our b0, which is also known as our y-intercept

$$\overline{X}_{4cyl} = \overline{X}_{8cyl} + b_1$$

$$b_1 = \overline{X}_{4cyl} - \overline{X}_{8cyl}$$

- so b1 is just difference between the means of the two groups
  - what is the 4 cylinder average MPG minus the 8 cylinder average MPG
- since this is a linear model, you can technically run this in SPSS as a regression and it will give you the constant (b0) and the difference between the two groups (b1) because what is working underneath the hood is a t-test

Technically Categorical Predictors in the Linear Model

```
    Two Sample t-test

data:  mpg by cyl
t = 8.1024, df = 23, p-value = 3.446e-08
alternative hypothesis: true difference in means is not equ
95 percent confidence interval:
  8.611261 14.516012
sample estimates:
mean in group 4 mean in group 8
       26.66364        15.10000
```

Technically Categorical Predictors in the Linear Model

```
Call:
lm(formula = mpg ~ as.factor(cyl), data = mtcars)

Residuals:
   Min     1Q Median     3Q    Max
-5.264 -2.264  0.100  2.200  7.236

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       26.664      1.068  24.966  < 2e-16 ***
as.factor(cyl)8  -11.564      1.427  -8.102 3.45e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 3.542 on 23 degrees of freedom
Multiple R-squared:  0.7405,	Adjusted R-squared:  0.7293
F-statistic: 65.65 on 1 and 23 DF,  p-value: 3.446e-08
```

The t-test

- there are three (two are more useful) types of t-test
  - **independent t-test** is used to compare two means that come from different conditions for two separate groups (between)
  - **paired-samples t-test** or **dependent t-test** is when you want to compare two means that come from different conditions from the same participants (within)
  - **one-sample t-test**

Rationale for the t-test

- we want to see if the means of two different samples are different from one another
- if the samples come from the same population, they would be roughly equal (no statistical significance found)
  - while we could have difference due to sampling variation, significant differences would occur infrequently (5%)
  - we are interested in the difference between the sample means
    - if there is = statistical significance = difference between groups
    - if no difference = not statistically significant = groups are roughly equal

Rationale for the t-test

- we use the standard error to gauge variability between sample means
- most test statistics are a signal-to-noise ratio, where the variance explained by the model is divided by what the model can't explain (error)

$$t = \frac{obs\ diff\ between\ sample\ means - expect\ diff\ between\ pop\ mea}{SE\ diff\ between\ 2\ sample\ means}$$

The Independent Samples t-test Equation

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{SE\ estimate}$$

- we are comparing the difference between the sample means and the population means
  - using the null hypothesis, we would assume that the populations would be same or that there would be no difference
  - so we can refer to that difference as zero

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{SE\ estimate}$$

The Independent Samples t-test Equation

- additionally we need to know what the SE estimate is
  - first, we'll need to get the variances, standard deviations for both groups
    - along with the n for each group

$$SS = \sum (X - \overline{X})^2$$

$$S^2 = \frac{SS}{df}$$

$$S = \sqrt{\frac{SS}{df}}$$

The Independent Samples t-test Equation

- then we can calculate the standard error for both groups

$$S_{\overline{X}} = \frac{S}{\sqrt{n}} \;\; OR \;\; S_{\overline{X}} = \frac{S^2}{n}$$

- we can use the **variance sum law**, which states that the variance of a difference between two two groups is equal to the sum of their variances
  - this means that we can estimate the variance of the sampling distribution of differences by adding together the variances

$$\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$$

The Independent Samples t-test Equation

- we could then get the standard error by taking the square root

$$S_{\overline{X_1}-\overline{X_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- if we don't have equal groups then we have an additional step

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

The Independent Samples t-test Equation

- then we can conduct a independent t-test

$$t_{obt} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

- the book doesn't cover this but our df changes slightly for our independent samples t-test

$$df = (n_1 - 1) + (n_2 - 1)$$

Steps to Independent Samples t-test

	group1	group2
1	1	5
2	4	6
3	6	3
4	7	4
5	3	8
6	6	7

Steps to Independent Samples t-test

- Get the means of both groups

```
(1 + 4 + 6 + 7 + 3 + 6)/6
```

```
[1] 4.5
```

```
# 4.5 group 1

(5 + 6 + 3 + 4 + 8 + 7)/6
```

```
[1] 5.5
```

```
# 5.5 group 2
```

Steps to Independent Samples t-test

- Get the n for each group
  - 6 per group

Steps to Independent Samples t-test

- Get the sum of squares

```
(1 - 4.5)^2 + (4 - 4.5)^2 + (6 - 4.5)^2 +
  (7 - 4.5)^2 + (3 - 4.5)^2 + (6 - 4.5)^2
```

```
[1] 25.5
```

```
# group 1 25.5 ss

(5 - 5.5)^2 + (6 - 5.5)^2 + (3 - 5.5)^2 +
  (4 - 5.5)^2 + (8 - 5.5)^2 + (7 - 5.5)^2
```

```
[1] 17.5
```

```
# group 2 17.5 ss
```

Steps to Independent Samples t-test

- df per group
  - 6 - 1 = 5

Steps to Independent Samples t-test

- get the variance and sd

```
25.5/5
```

```
[1] 5.1
```

```
sqrt(5.1)
```

```
[1] 2.258318
```

```
# group 1 variance = 5.1
# group 1 sd = 2.26
```

Steps to Independent Samples t-test

```
17.5/5
```

```
[1] 3.5
```

```
sqrt(3.5)
```

```
[1] 1.870829
```

```
# group 2 variance = 3.5
# group 2 sd = 1.87
```

Steps to Independent Samples t-test

- get the differences in standard error

$$S_{\overline{X_1}-\overline{X_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Steps to Independent Samples t-test

```
5.1/6
```

```
[1] 0.85
```

```
# group 1 part of the standard error = .85

3.5/6
```

```
[1] 0.5833333
```

```
# group 2 part of the standard error = .58

sqrt(.58 + .58)
```

```
[1] 1.077033
```

```
# standard error is 1.08
```

Steps to Independent Samples t-test

- get the t-obtained value

$$t_{obt} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

Steps to Independent Samples t-test

```
4.5 - 5.5
```

```
[1] -1
```

```
# numerator is -1

# denominator is 1.08 from previous step

-1/1.08
```

```
[1] -0.9259259
```

```
# t-test is -.93
```

Steps to Independent Samples t-test

- negative and positive values just indicate which group you're comparing to the other
  - since it is negative, we can conclude that the mean of group 1 is less than group 2's mean

Steps to Independent Samples t-test

- get df for the group comparisons
  - this is because we are looking at a sampling distribution for the differences between the two samples/groups

$$df = (n_1 - 1) + (n_2 - 1)$$

```
(6 - 1) + (6 - 1)
```

```
[1] 10
```

- df is 10
  - look at t-table for t-obtained value of -.93 and a df of 10

Reporting an Independent Samples t-test

- When reporting an independent samples t-test, we are comparing the means of the outcome between the two groups that are interested in comparing
  - Ex: Sex differences in statistics quiz scores
  - H0: There will be no differences in statistics quiz scores between male and female students
  - H1: There will be differences in statistics quiz scores between male and female students
  - Reporting: On average, when comparing male students ($M = 74.3$, $SD = 1.24$) to female students ($M = 94.7$, $SD = 4.40$), male students' scores were significant worse on their statistics quiz; $t(10) = 4.84$, $p = .03$.

Reporting an Independent Samples t-test

- the t-test includes:
  - an italicized lowercase t
  - df in the parentheses
  - the t-obtained value
  - the exact p value italicized (when using SPSS) or $p < .05$ (when using the t-table)