Psychology's Core Statistics Curriculum (More or less)

```
library(dplyr)
library(ggplot2)
```

Read in a data file

```
Df <- read.csv('../data/memorytest.csv')
```

Descriptive statistics

Now, we'll get the mean, median, maximum, minimum, standard deviation, sample size, of the memory scores for both conditions.

```
descriptives <- Df %>% group_by(condition) %>%
  summarize(mean = mean(memory),
            median = median(memory),
            max = max(memory),
            min = min(memory),
            sd = sd(memory),
            size = n())
```

	condition	mean	median	max	min	sd	size
1	coffee	23.87	23.40	34.52	14.65	4.56	20
2	control	23.98	25.11	33.45	13.27	5.12	30

Independent samples t-test

An independent samples t-test with the homogeneity of variances assumption is as follows:

```
results <- t.test(memory ~ condition, data=Df,
                  var.equal=T)
```

	Test	Results
1	Two Sample t-test:	t(48) = -0.08, p = .937, d = -0.02

Paired samples t-test

For the paired samples t-test, we'll make up some data. We'll make two paired samples of normally distributed random numbers of size $N = 25$.

```
x <- rnorm(N)
y <- rnorm(N)
results <- t.test(x,y, paired=T)
```

	Test	Results
1	Paired t-test:	t(24) = -0.43, p = .668, d = -0.12

Z-scores

```
scores <- rnorm(N, mean=100, sd=15)
z.scores <- (scores - mean(scores)) / sd(scores)
```

We get the area under the normal curve below each z-score as follows:

```
area.under.curve <- pnorm(z.scores)
```

χ^2 test

We'll make a 2×2 contingency table:

```
X <- matrix(c(120, 57, 75, 17), ncol = 2)
```

A χ^2 test (with no continuity correction) is as follows:

```
results <- chisq.test(X, correct=FALSE)
```

	Test	Results
1	Pearson's Chi-squared test:	chi^2(1) = 5.72, p = .017

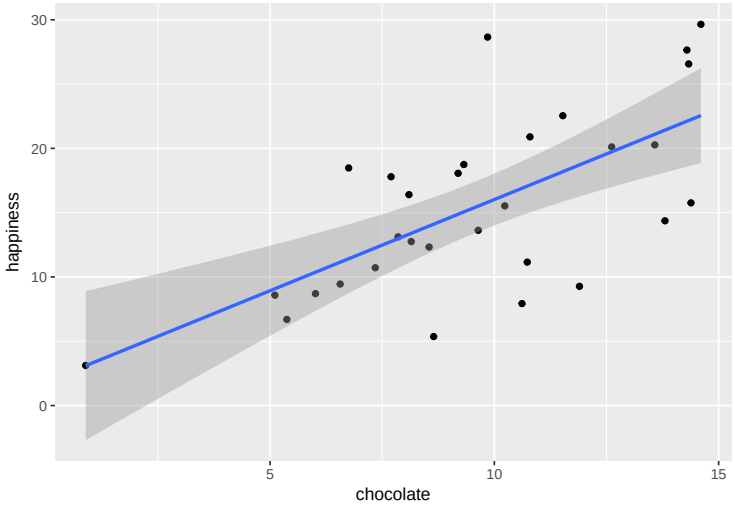
Simple linear regression

First, we'll read in our data from another csv file.

```
Df <- read.csv('../data/chocolate.csv')
```

Now, we'll do a scatter plot and line of best fit.

```
qplot(chocolate, happiness, data=Df) +
  geom_smooth(method='lm')
```



The regression model is fit with

```
M <- lm(happiness~chocolate, data=Df)
```

From this, we have $R^2 = 0.44$ (adjusted $R^2 = 0.42$), $F(1, 28) = 21.78$. The coefficients table is as follows:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.8387	3.0814	0.60	0.5555
chocolate	1.4183	0.3039	4.67	0.0001

One-way ANOVA

Again here, we'll read in a new data set from a csv file.

```
Df <- read.csv('../data/antidepressants.csv')
```

For this, we coul use the `lm` (linear model) command, but instead we'll use `aov`.

```
M <- aov(depression ~ drug, data=Df)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
drug	2	603.20	301.60	0.63	0.5345
Residuals	117	56029.04	478.88		

Mann Whitney U test:

```
x <- rnorm(10)
y <- rnorm(12)
results <- wilcox.test(x, y)
```

The U statistic is $U = 83$, $p = 0.14$.

Wilcoxon signed ranks test:

```
N <- 10
x <- rnorm(N)
y <- rnorm(N)
results <- wilcox.test(x, y, paired = TRUE)
```

The W statistic is $W = 45$, $p = 0.08$.

Multiple linear regression

```
Df <- read.csv('../data/stress.csv')
M <- lm(stress ~ social + extra + perfect,
        data=Df)
```

From this, we have $R^2 = 0.06$ (adjusted $R^2 = -0.004$), $F(3, 46) = 0.94$. The coefficients table is as follows:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	119.4799	20.2997	5.89	0.0000
social	-0.0618	0.3024	-0.20	0.8390
extra	0.1954	0.2506	0.78	0.4394
perfect	-0.4419	0.2752	-1.61	0.1152

Factorial Anova

Here, we look at the effect of two variables (drug type and therapy type) and their interaction on depression scores.

```
Df <- read.csv('../data/drugtherapy.csv')
M <- aov(depression ~ drug*therapy, data=Df)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
drug	2	129.93	64.96	0.17	0.8400
therapy	2	159.20	79.60	0.21	0.8077
drug:therapy	4	932.96	233.24	0.63	0.6443
Residuals	111	41296.17	372.04		

Repeated Measures Two Way Anova

```
Df <- read.csv('../data/movies.csv')
M <- aov(stress ~ music * image +
           Error(PID / (music * image)),
         data=Df)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	49	8344.18	170.29		
music	1	0.78	0.78	0.00	0.9543
Residuals1	49	11524.23	235.19		
image	1	61.11	61.11	0.30	0.5890
Residuals2	49	10126.52	206.66		
music:image	1	563.78	563.78	2.63	0.1115
Residuals	49	10519.86	214.69		

Repeated Measures Three Way Anova (One Between, Two Within) Anova

```
Df <- read.csv('../data/movies2.csv')
M <- aov(stress ~ time * music * image +
           Error(PID / (music * image)),
         data=Df)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
time	1	5.70	5.70	0.03	0.8571
Residuals	48	8338.48	173.72		
music	1	0.78	0.78	0.00	0.9548
time:music	1	21.96	21.96	0.09	0.7634
Residuals	48	11502.27	239.63		
image	1	61.11	61.11	0.29	0.5914
time:image	1	80.77	80.77	0.39	0.5374
Residuals 1	48	10045.76	209.29		
music:image	1	563.78	563.78	2.58	0.1149
time:music:image	1	23.74	23.74	0.11	0.7432
Residuals	48	10496.12	218.67		