

Author: Charles S. Leger
Course: Biostatistics 5081
Instructor: Dr. Joel Shore

5081 Seminar: ANOVA

This document includes examples of independent one-way, two-way and repeated-measures ANOVA. All analyses were conducted in R (R Core Team (2017). R: A language and environment for statistical computing). For each test, data are initially explored for relative normality and variance assessment.

Using the same data, one-way, two-way ANOVA analyses were also conducted in SAS®. The SAS examples (which have the results identical to R) can be found on a separate document.

Prior to conducting an analysis, it is important to know the number per group (n) required to achieve a large effect size with at least 80% power. This can be calculated using the `pwr()` function:

```
library(pwr)
```

Important variables are k = number of groups; n = number of observations per group; f = effect size (Cohen); `sig.level`; and `power`, with most studies aiming for 80% power to reject the null hypothesis. Any one of the variables (k , f , `power`, or n) can be removed to arrive at the relevant estimate.

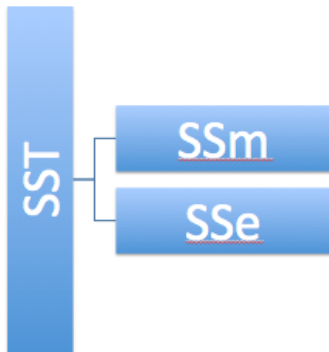
For example for these variables, and as configured below, the requirement is $n = 21$ (per group).

```
pwr.anova.test(k=3, sig.level = .05, f=.40, power=.80);
```

Our sample, data has only an n of 7 ($n=7$), so it would be too under powered (~31% power) for an actual study, though this would not make it invalid:

```
pwr.anova.test(k=3, sig.level = .05, f=.40, n=7) power = 0.3067576
```

Conceptually, and as depicted in the diagram below, one-way ANOVA can be viewed as partitioning the total sum of squares (SST, also referred to as the total variance) into variance accounted for by the explanatory variable (SSm) and the unsystematic, unexplained variance (SSe) (Field, 2012).



These are the “libraries” that R requires for one-way ANOVA

```
library(car)
library(compute.es)
library(pastecs)
library(multcomp)
library(ggplot2)
```

Generating a fictitious dataset

The lines below were executed to generate a simulated dataset using the `runif()` function:

```
R4S<-runif(7, min= 2, max=10)
RW<-runif(7, min= 5, max=13)
EC<-runif(7, min=6, max=12)
```

This dataset will be used in the one-way ANOVA example. It will also be used in the repeated-measures example, though the repeated-measures ANOVA dataset will be augmented by the addition of single column for individual subjects.

This fictitious scenario and data involves 3 different reading compression programs. The dependent (response) variable is a single score that represents a composite of correct answer response and speed of response (in minutes). The independent (explanatory) variable is reading program method. The method is made up of reading programs named "Read 4 Speed" (R4S), "Read Well" (RW), and "Eye Candy" (EC). Fictitious subjects were randomly allocated to one of the 3 different reading programs. Each program involved 2 a week course, with 3 hours of daily reading practice following program guidelines.

Here is the simulated data:

```
R4S<-c(8.249930, 3.851183, 5.166481, 5.819707, 8.554169, 2.519100,
4.331016)
length(R4S)
RW<-c(10.466106, 6.934875, 5.192835, 11.630756, 9.662392, 7.690234,
7.267778)
length(RW)
EC<-c(8.949562, 9.539441, 8.415509, 9.850214, 9.303608, 10.499970,
10.221738)
length(EC)
```

Testing for normality: a p-value > .05 indicating relatively normal distribution

```
shapiro.test(R4S)
Shapiro-Wilk normality test
data: R4S
W = 0.93525, p-value = 0.5964
```

```
shapiro.test(RW)
Shapiro-Wilk normality test
data: RW
W = 0.96241, p-value = 0.8391
```

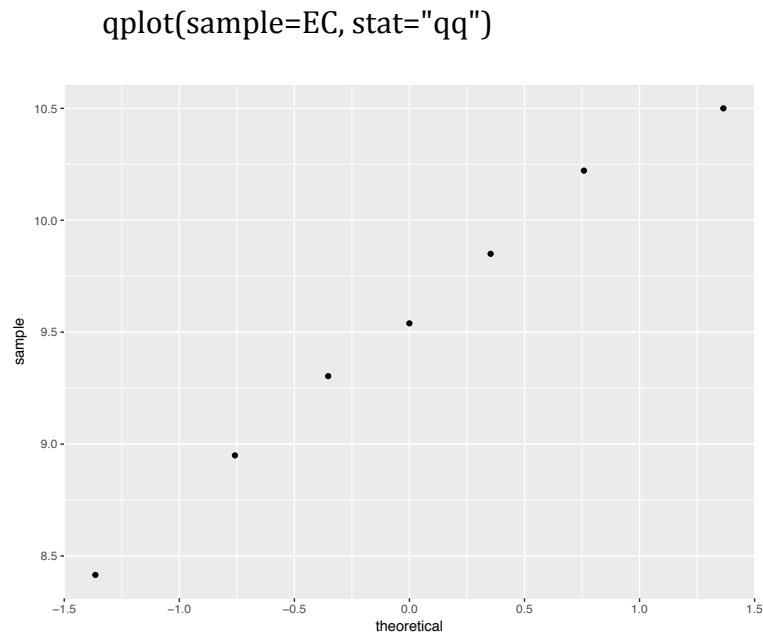
```
shapiro.test(EC)
Shapiro-Wilk normality test
data: EC
W = 0.98454, p-value = 0.9786
```

Residuals should be normally distributed for each group (here R4S, RW, and EC). These are small groups, so we expect greater variability than in the actual population and the residuals may not be as well behaved as in groups with a larger n . In the plots below the residuals (dots) should approximately adhere to a diagonal pattern; off-diagonal points (residuals) suggest some non-normal distribution of residuals. A more elaborate test of relative normality of residuals will be conducted using the ANOVA model that is created after we have completed after this initial exploration of data to assess relative normality (and testing of the assumption of homogeneity of variance).

The residuals can be calculated by subtracting the outcome from the mean of the outcome (outcome - mean(outcome)) and then plotted.

```
res_EC<-(EC-mean(EC))
library(ggplot2)
```

Here is an example using just the EC (Eye Candy condition)



Testing for homogeneity of variance with the Levene Test

Creating a grouping factor (aka explanatory) variable for testing of variance with the `leveneTest()` function, though this function will also coerce a factor. The `leveneTest()` assesses variance between levels of a given predictor (explanatory variable); such variances should be similar, and we should see p-values $> .05$ which would indicate that the assumption of homogeneity of variance has not been violated.

Note, the `leveneTest` requires the `car` package, which was called at the outset with `library(car)`

Making grouping/factor "method" which includes the reading programs ("Read 4 Speed", "Read Well", and "Eye Candy". This is the explanatory variable ("method"); the response/dependent variable is now the actual score values in "score". Note, the `table()` function below is not necessary but shows that "method" has three levels and the number of subjects in each level.

```
score<-c(R4S, RW, EC)
method<-rep(c(1,2,3), c(7,7,7))
method<-factor(method, levels = c(1:3), labels = c("Read 4 Speed", "Read
Well", "Eye Candy"))
dat1<-data.frame(method, score)
```

```
table(dat1$method)
```

The Levene test, conducted below has a p-value $> .05$ ($p = .11$) indicating the assumption of homogeneity of variance was not violated; variances among groups are not extreme or non-normal.

```
leveneTest(dat1$score, dat1$method)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	2	2.4518	0.1144
	18		

Always look at the dataset, or least the first 6 rows of the dataset; this can be indispensable for a general grasp of the column and row labels and perhaps even some insight to patterns in the data.

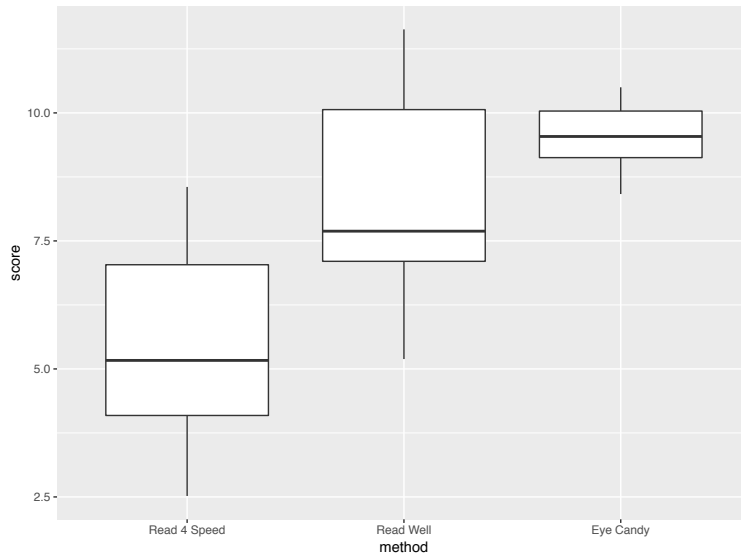
```
dat1
  method      score
1 Read 4 Speed 8.249930
2 Read 4 Speed 3.851183
3 Read 4 Speed 5.166481
4 Read 4 Speed 5.819707
5 Read 4 Speed 8.554169
6 Read 4 Speed 2.519100
7 Read 4 Speed 4.331016
8  Read Well 10.466106
9  Read Well  6.934875
10 Read Well  5.192835
11 Read Well 11.630756
12 Read Well  9.662392
13 Read Well  7.690234
14 Read Well  7.267778
15 Eye Candy  8.949562
16 Eye Candy  9.539441
17 Eye Candy  8.415509
18 Eye Candy  9.850214
19 Eye Candy  9.303608
20 Eye Candy 10.499970
21 Eye Candy 10.221738
```

Graphs

The box plot below shows the quartiles, including the medians (middle lines in the boxes). Extending from the bottom of each box is whisker to first line is first

quartile(25%); the 2nd quartile is from first line to median line (50%); the 3rd quartile is from median line to the top of the upper box; the upper quartile is from the top of a box to the end of the upper whisker. The length of the whiskers reflects variation within the group. Note that the highest median score occurred in the Eye Candy reading group.

```
ggplot(dat1, aes(method, score)) + geom_boxplot()
```



Here are the actual quantiles for “Eye Candy”, the best performing reading programme:

```
quantile(dat1$score[dat1$method=="Eye Candy"], c(.25, .50, .75))
 25%      50%      75%
9.126585  9.539441 10.035976
```

The means, variance and standard deviation can be easily extracted using a few methods. One approach is to use the `tapply()` function:

```
tapply(dat1$score, dat1$method, mean)
Read 4 Speed  Read Well  Eye Candy
5.498798     8.406425     9.540006
tapply(dat1$score, dat1$method, var)
Read 4 Speed  Read Well  Eye Candy
5.0123978    5.0863527    0.5252534
tapply(dat1$score, dat1$method, sd)
Read 4 Speed  Read Well  Eye Candy
2.2388385    2.2552944    0.7247437
```

Note that “Eye Candy” has the highest mean.

One-way ANOVA

Since we have determined that the dataset is relatively normal, a standard ANOVA procedure can be conducted. The test below shows a significant effect of Method, indicating there is a significant difference among the means of the 3 reading methods ($p=.0024$)

```
mod1<-aov(score~method, data= dat1)
summary(mod1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Method	2	60.83	30.416	8.589	0.0024 **
Residuals	18	63.74	3.541		

This is a is an alternative ANOVA calculation approach using just a single line:

```
summary(aov(dat1$score~factor(dat1$method)))
              Df    Sum Sq   Mean Sq  F value  Pr(>F)
factor(dat1$method) 2     60.83    30.416   8.589   0.0024 **
Residuals           18     63.74     3.541
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect size:

An eta-squared (SS between groups/ SST) effect size can be easily derived using the `summary.lm()` function. This will provide an r^2 value (labeled multiple R^2 as this is a regression function). An adjusted version is also provided that conveys the extent to which this model generalizes to the population. The adjusted r^2 value is .42

```
summary.lm(mod1)
.42
```

Conducting pairwise tests (which method significantly differ):

```
pairwise.t.test(dat1$score, dat1$method, p.adjust.method = "fdr" )
```

Pairwise comparisons using t tests with pooled SD

data: dat1\$score and dat1\$method

Pairwise comparisons using t tests with pooled SD

data: dat1\$score and dat1\$method

	Read 4 Speed	Read Well
Read Well	0.0292	-
Eye Candy	0.0024	0.8237

Effect size; using the `mes()` function, which takes the form `mes(mean group1, mean group2, sd group1, sd group2, n group1, n group2)`

```
library(compute.es)
```

Read 4 Speed vs. Read Well

```
mes(5.498798, 8.406425, 2.2388385, 2.2552944, 7, 7)
```

using effect size of $r = .54$;

$r^2 (-.54)^2 = .29$

Read 4 Speed vs. Eye Candy

```
mes(5.498798, 9.540006, 2.2388385, 0.7247437, 7, 7)
```

$r = -0.75$

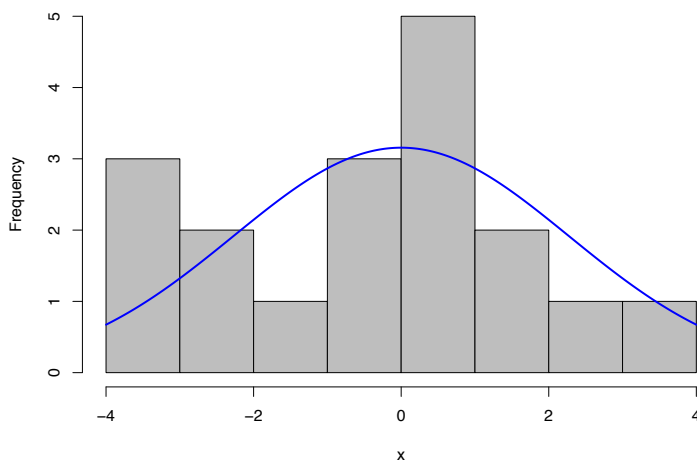
$r^2 (-.77)^2 = .59$

Additional test of normality of residuals using the completed model

```
library(rcompanion)
```

```
x=residuals(mod1)
```

```
plotNormalHistogram(x)
```



Statement:

There was not a significant difference between the Read Well and Eye Candy programs. However, the means of Read 4 Speed ($M=5.50$, $SD=2.24$), and Read Well

($M= 8.4, SD=2.25$) differed; this difference proved to be significant, $p =.03, r^2=.29$. Method accounted for a 29% of the variance in performance (the composite score outcome). The means of Read 4 Speed ($M=5.50, SD=2.24$), and Eye Candy also differed ($M= 9.54, SD=.72$) differed; this difference proved to be significant, $p =.002, r^2=.59$. Here, method accounted for 59% of the variance in performance. Overall, Eye Candy had the highest mean result and contributed to 59% of the variance in reading performance.

Non-normal dataset: non-parametric ANOVA

The code below was used to generate a non-normal data set appropriate for a non-parametric or robust ANOVA. Here a non-parametric ANOVA Kruskal-Wallis test was used.

Simulating non-normal data:

```
nuts=rchisq(30, df=3)
c1<-sample(nuts, size=10)
c2<-sample(nuts, size=10) - 10
c3<-sample(nuts, size=10) + 7
```

Making a grouping variable and factor, here called groupc:

```
groupc<-rep(c(1,2,3), c(10,10,10))
groupc<-factor(groupc, levels = c(1:3), labels = c("c1", "c2", "c3"))
```

Putting all generated values into a single dataset (actually a vector)

```
reading<-c(c1,c2,c3)
```

Combining the grouping variable and data in a data frame:

```
dat2<-data.frame(method1=groupc, reading)
```

```
head(dat2)
  method1 reading
1    c1 6.8267049
2    c1 3.8972707
3    c1 2.2951466
4    c1 0.3787745
5    c1 0.9271258
6    c1 5.3481738
```

Tests of normality and variance; the Levene test of normality as conducted first followed by the Shapiro-Wilk test;

```
leveneTest(dat2$reading, dat2$method1)
Levene's Test for Homogeneity of Variance (center = median)
      Df    F value    Pr(>F)
group 2     0.6238    0.5435
      27
```

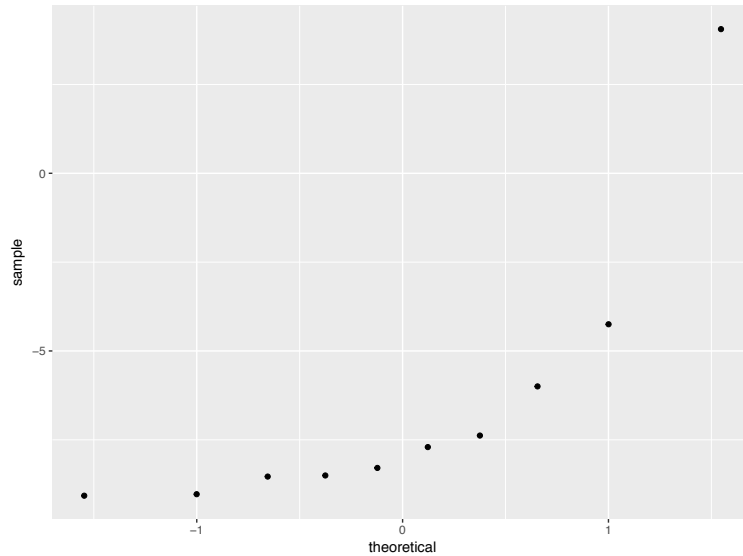
Assumption of homogeneity of variance was not violated ($p > .05$).

Just testing one group, here c2, with the Shapiro-Wilk test, and this data is significantly non-normal. This is also conveyed by the `qqplot()` of residuals that follows the Shapiro-Wilk test. The `qqplot()` shows lots of residuals off the diagonal.

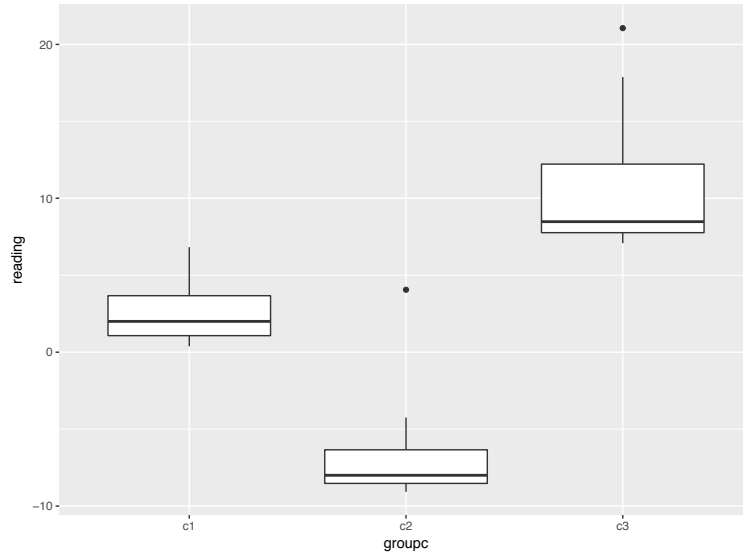
```
shapiro.test(c2)
```

Shapiro-Wilk normality test

```
data: dat2 c2
W = 0.67139, p-value = 0.0003972
```



Here is a box plot of the non-normal dataset; it clearly shows outliers.
`ggplot(dat2, aes(groupc, reading)) + geom_boxplot()`



Conducting the Kruskal-Wallis test

`(mod2<-kruskal.test(reading~method1, data=dat2)) p > .05, ns`

Kruskal-Wallis rank sum test

data: reading by method

Kruskal-Wallis $X^2 = 19.771$, $df = 2$, $p\text{-value} = 5.09e-05$

As with the preceding one-way ANOVA example, a statement for the omnibus ANOVA outcome and pairwise tests would be required (but we will skip this to save time).

Two-way ANOVA

Conceptually, two-way independent ANOVA is similar to one-way with the SST partitioned in to the sum of squares explained by the model SSm and the error (SSe). However, the SSm or model is partitioned in to the two main effects (here named SSA and SSB) as well as the interaction of $SSA \times SSB$.



This study also used simulated data. The aim was to assess if 1st year female and male university students had differing anxiety levels during lectures relative to exams. Factor A is gender: Male, Female (two levels); Factor B is Anxiety level during lectures vs. exams. Each factor therefore had 2 levels. The Beck Anxiety Inventory (BAI) scale was used to scale anxiety: 0-9 is minimal; 10-18 mild 19-29 moderate; 30-63 is severe.

Here is a schematic of this two-way ANOVA

		Anxiety	
		Lecture	Exam
Gender	Male		
	Female		

This time we import the dataset that is in an Excel csv format:

```
two_way<-read.csv("5081_TwoWay.csv", header=TRUE)
```

Here we can verify that R automatically set the correct variable formats, where the gender and status (we will change this name to Event) are factors.

```
is.factor(two_way$gender) TRUE
is.factor(two_way$status) TRUE
is.factor(two_way$Anxiety_level) FALSE
```

Changing column names:

```
colnames(two_way)<-c("Gender", "Event", "Anxiety")
```

```
(two_way)
```

```
  Gender Event Anxiety
1 female lecture    5
2 female lecture    4
3 female lecture   11
4 female lecture   13
5 female  exam   14
6 female  exam   19
7 female  exam   11
8 female  exam    8
9  male lecture   16
10 male lecture    5
11 male lecture    3
```

12	male	lecture	7
13	male	exam	20
14	male	exam	12
15	male	exam	13
16	male	exam	17

Testing data for data relative normality and variance

Here will extract both the Anxiety outcome across gender groups as well as create subsets of the data by gender to allow for more specific assessments, including separate assessments of anxiety variance and relative normality for females and males.

Extracting a subject of the dataset by gender using the `subset()` function, which employs a logical operator to extract specific elements of the data. To extract study data for female participants:

```
female<-subset(two_way, Gender=="female", select=c("Event", "Anxiety"))
female
Event Anxiety
1 lecture 5
2 lecture 4
3 lecture 11
4 lecture 13
5 exam 14
6 exam 19
7 exam 11
8 exam 8
```

Here we do the same for male participants:

```
male<-subset(two_way, Gender=="male", select=c("Event", "Anxiety"))
male
Event Anxiety
9 lecture 16
10 lecture 5
11 lecture 3
12 lecture 7
13 exam 20
14 exam 12
15 exam 13
16 exam 17
```

Testing normality using the Shapiro-Wilk test: first for females, then males, then test across groups without differentiating gender.

```
Testing female Anxiety by level: lecture
shapiro.test(female$Anxiety[female$Event=="lecture"])
Shapiro-Wilk normality test
```

```
data: female$Anxiety[female$Event == "lecture"]
W = 0.87832, p-value = 0.3315
```

```
Testing female Anxiety by level: exam
shapiro.test(female$Anxiety[female$Event=="exam"])
Shapiro-Wilk normality test
```

```
data: female$Anxiety[female$Event == "exam"]
W = 0.98403, p-value = 0.9252
```

```
Testing male Anxiety by level: lecture
shapiro.test(male$Anxiety[male$Event=="lecture"])
Shapiro-Wilk normality test
```

```
data: male$Anxiety[male$Event == "lecture"]
W = 0.86966, p-value = 0.2964
```

```
Testing male Anxiety by level: exam
shapiro.test(male$Anxiety[male$Event=="exam"])
Shapiro-Wilk normality test
```

```
data: male$Anxiety[male$Event == "exam"]
W = 0.92654, p-value = 0.5742
```

```
Testing across groups without differencing gender
shapiro.test(two_way$Anxiety)
```

```
Shapiro-Wilk normality test
```

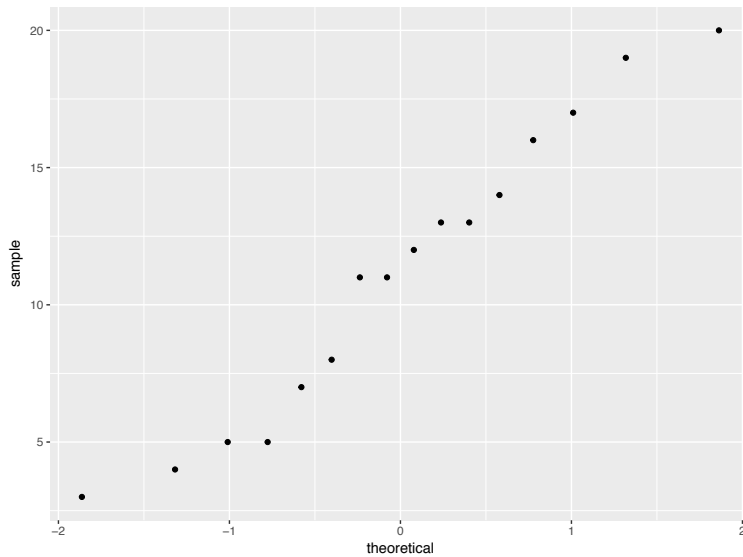
```
data: two_way$Anxiety
W = 0.95404, p-value = 0.5563
```

None of the p-values were $< .05$ suggesting these data are relatively normal.

The final test of normality will again entail using the `qqplot()` to test relative normality of residuals. For the sake of brevity we will use `qqplot()` to look just at total

Anxiety without differentiating Anxiety by specific levels (lecture, exam) or Anxiety or Gender. In this qqplot the residuals look mostly aligned on the diagonal suggesting that the dataset is relatively normal.

```
qqplot(sample=two_way$Anxiety, stat="qq")
```



Conducting the Levene Tests, first for Anxiety by Gender, then Anxiety by Event, and finally for the interaction of Gender and Event.

```
leveneTest(two_way$Anxiety, two_way$Gender)
Levene's Test for Homogeneity of Variance (center = median)
  Df    F value    Pr(>F)
group 1      0.6173      0.4452
  14

leveneTest(two_way$Anxiety, two_way$Event) OK
Levene's Test for Homogeneity of Variance (center = median)
  Df    F value    Pr(>F)
group 1      0.1197      0.7345
  14
```

Do the combinations of the factors (Gender and Event) have significant interactions such that significant differences in Anxiety variances for Gender x Event might result?

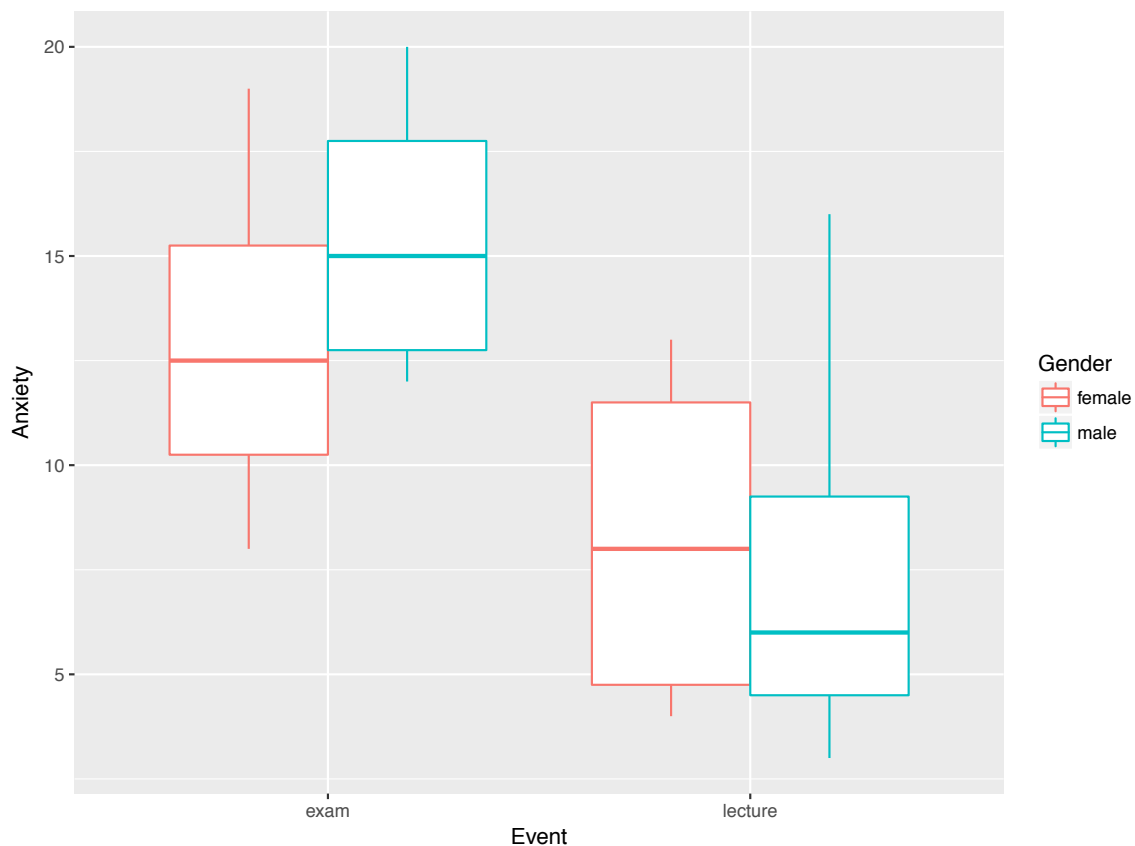
```
leveneTest(two_way$Anxiety, interaction(two_way$Gender,
two_way$Event))
Levene's Test for Homogeneity of Variance (center = median)
```

	Df	F value	Pr(>F)
group	3	0.0736	0.973
	12		

There was no indication of significantly different variances by group or resulting from an interaction: all p-values were > .05

Graphs

This graph combines both female and male with Anxiety levels for lecture and exam. Interestingly, males show greater exam anxiety but less lecture anxiety (at least in our simulated data).



Looking specifically at the Anxiety means, variance and standard deviation for females and males, it is apparent that males do indeed have higher exam Anxiety relative to females. We will see if this is a significant difference with the two-way ANOVA.

```
tapply(female$Anxiety, female$Event, mean)
Event
exam lecture
```



```
13.00 8.25
tapply(female$Anxiety, female$Event, var)
Event
exam lecture
22.00 19.58
tapply(female$Anxiety, female$Event, sd) shows mean anxiety based on
Event
exam lecture
4.690416 4.425306
```

Male Anxiety levels

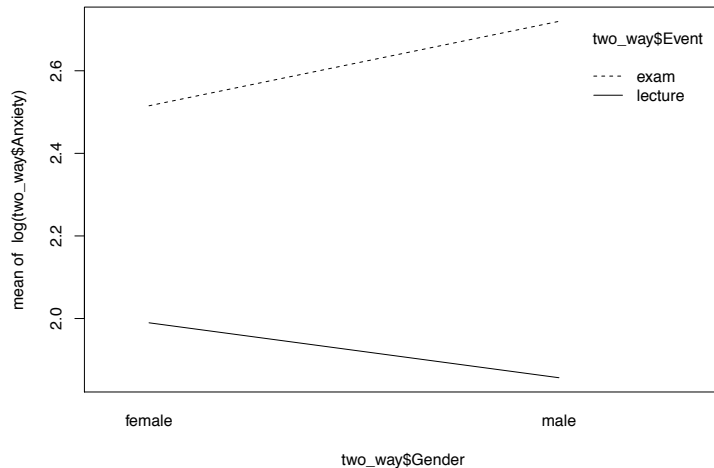
```
tapply(male$Anxiety, male$Event, mean)
exam lecture
15.50 7.75
tapply(male$Anxiety, male$Event, var)
exam lecture
13.67 32.92
tapply(male$Anxiety, male$Event, sd)
exam lecture
3.69 5.74
```

This is the mean, variance and standard deviation of anxiety across groups without differentiating by gender:

```
tapply(two_way$Anxiety, two_way$Event, mean)
exam lecture
14.25 8.00
tapply(two_way$Anxiety, two_way$Event, var)
exam lecture
17.07 22.57143
tapply(two_way$Anxiety, two_way$Event, sd)
exam lecture
4.13 4.75
```

Here we test for an indication of a possible ***interaction***; this provides a critical early insight to a potential interaction effect (Gender x Event); parallel lines for males and females would suggest no interaction while intersecting or non-parallel lines suggest there is an interaction. The interaction plot below clearly displays non-parallel lines, suggesting a potential interaction. We will need to test if this interaction is significant with a two-way ANOVA.

```
interaction.plot(two_way$Gender, two_way$Event, log(two_way$Anxiety))
```



Conducting two-way ANOVA

```
mod3<-aov(Anxiety~Event + Gender + Event:Gender, data = two_way)
```

```
summary(mod3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Event	1	156.2	156.25	7.089	0.0207 *
Gender	1	4.0	4.00	0.181	0.6777
Event:Gender	1	9.0	9.00	0.408	0.5348
Residuals	12	264.5	22.04		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Using the `summary.lm()` function just to extract r^2 , we obtain a value of r^2 (called multiple R^2) as well as an adjusted value. Note, as mention previously, the adjusted value reflects expected contribution to variance in the actual population: this value was .24.

```
summary(mod3)
```

```
.24
```

Statement:

The two-way ANOVA revealed that there was not a significant Gender x Event interaction effect, $p = .40$, nor was there a significant effect of Gender, $p = .18$. However, there were differing means between events for both females and males: event-lecture (females: $M=8.25$, $SD= 4.42$; males: $M=7.75$, $SD= 5.74$) and event-exam (females: $M=13.00$, $SD=4.69$; males: $M= 15.50$, $SD=3.69$). Without differentiating gender, the means for event-lecture ($M=8.00$, $SD=4.75$), and event-

exam ($M=14.25$, $SD= 4.13$) differed. This difference between events proved to be significant, with a significant main effect of Event, $p= .02$, $r^2=.24$, which indicates that event accounted for 24% of the variance in anxiety levels.

There was not a significant effect of Gender nor was there a significant interaction of Gender x Event. Because only the main effect of Event was significant, we can run a pairwise test for Event, which below shows a significant main effect of Event ($p =.014$)

Pairwise testing for Event

```
pairwise.t.test(two_way$Anxiety, two_way$Event, p.adjust.method = "fdr")
fdr= false discovery rate
```

Pairwise comparisons using t tests with pooled SD

```
data: two_way$Anxiety and two_way$Event
```

	exam
lecture	0.014

P value adjustment method: fdr

Effect size

Here we use a basic Welch t-test to facilitate arriving at an effect size using r^2 as the measure of effect strength; this t-test shows the same result as the pairwise test above, but running this t-test provides a very simple method of assigning a the t-test result to a variable (ta), then extracting the statistic and df values to calculate r^2 , which we find is .36.

```
ta<-t.test(two_way$Anxiety~two_way$Event,paired= FALSE, var.equal =
FALSE)
```

Welch Two Sample t-test

```
data: two_way$Anxiety by two_way$Event
```

```
t = 2.8076, df = 13.736, p-value = 0.01418
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
1.466933 11.033067
```

```
sample estimates:
```

```
mean in group exam mean in group lecture
```

```
14.25      8.00
```

```
t<-ta$statistic[[1]] -2.8
```

```
df<-ta$parameter[[1]] 13.73
```

$$r^2 = t^2 / (t^2 + df) = .36$$

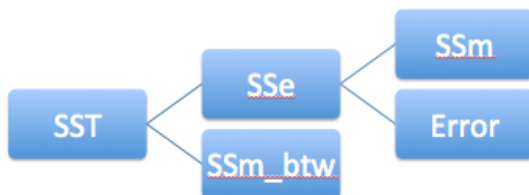
Statement:

As already reported, the means for event-lecture ($M=8.00$, $SD=4.75$), and event-exam ($M=14.25$, $SD= 4.13$) differed. This difference between events proved to be significant, with a significant main effect of Event lecture and Event exam, $p= .014$, $r^2=.36$, which indicates that event alone accounted for 36% of the variance in anxiety levels.

Repeated measures one-way ANOVA

In repeated-measures ANOVA the same subjects (or entities) participate in all treatments. A big advantage of repeated-measures ANOVA is the variance in the response variable due personal or individual differences is removed from the error term; difference between treatments then, can not be due to difference between participating subjects given the same subject is in each treatment.

The error term, then, changes in the repeated-measures design. As conveyed by the schematic below, the error in repeated measures ANOVA now includes these within-subject differences and any other variation not explained by the treatment effect. It must be underlined however, that within-subject individual differences are measured then removed from the denominator of the repeated-measures F-ratio. The removal of the variance of contributed by individual differences increases the power of repeated measures ANOVA relative to independent measures ANOVA



The dataset used in this repeated-measures ANOVA is, as noted at the outset, the same as dataset used in the one-way ANOVA example analysis. Details of the explanatory (reading program method) and response (reading composite score) variables can be found under *One-way ANOVA* (independent, unrelated measures) at the start of this document.

For the repeated-measures ANOVA however, a single column of individual subjects was added. Here is the dataset, imported for Excel as a csv.

```
rep_m<-read.csv("dat1rm.csv", header=TRUE)
```

	subject	method	score
1	s1	Read 4 Speed	8.249930
2	s2	Read 4 Speed	3.851183
3	s3	Read 4 Speed	5.166481
4	s4	Read 4 Speed	5.819707
5	s5	Read 4 Speed	8.554169
6	s6	Read 4 Speed	2.519100
7	s7	Read 4 Speed	4.331016
8	s1	Read Well	10.466106
9	s2	Read Well	6.934875
10	s3	Read Well	5.192835
11	s4	Read Well	11.630756
12	s5	Read Well	9.662392
13	s6	Read Well	7.690234
14	s7	Read Well	7.267778
15	s1	Eye Candy	8.949562
16	s2	Eye Candy	9.539441
17	s3	Eye Candy	8.415509
18	s4	Eye Candy	9.850214
19	s5	Eye Candy	9.303608
20	s6	Eye Candy	10.499970
21	s7	Eye Candy	10.221738

```
# Looking at data properties, to verify that R correctly formatted the data,  
distinguishing the factors (method and now subject as well).
```

```
is.factor(rep_m$subject)  TRUE  
is.factor(rep_m$method)  TRUE  
is.factor(rep_m$score)    FALSE
```

Here are the libraries used for repeated-measures:

```
library(TH.data)  
library(ez)  
library(pastecs)  
library(multcomp)  
library(nlme)  
library(ggplot2)  
library(reshape2)  
library(car)
```

If we chose to create the data rather than import it, we could do so as follows: (this is not a necessary step here, as we have imported the data, and R has automatically determined the correct data formatting):

```
7 sets (normally levels) of 3 scores
subj<-gl(7, 3, labels= c("s1", "s2", "s3", "s4", "s5", "s6", "s7"))
meth<-gl(3, 1, 21, labels = c("R4S", "RW", "EC"))
table(meth)
score1<-rep_m[, "score"]
dat5<-data.frame(subj, meth, score1)
dat5[order(dat5$meth),] just reorders by method
```

Note, that because this is the same dataset (same explanatory and response variables) as in the independent ANOVA example, the tests of normality and variance will have the same outcome; graphs will be the same; means and standard deviations will be the same across groups, and pairwise tests will be the same. As consequence none of these assessments are repeated here.

What will differ, however, is the power of the omnibus. Also, repeated-measures requires a change in assumptions. With repeated-measures ANOVA the same entities are in each treatment. But this violates the independent ANOVA assumption that scores are independent given scores across treatment conditions will be likely related by virtue of being derived from the same participant or entity. Consequently, a standard F-ratio will be inaccurate (Field, 2012).

The new assumption that arises out of repeated-measures version of ANOVA, is referred to as sphericity. Sphericity assumes that the extent of dependence between pairs of treatment conditions is approximately equal. Mauchly's test is often used to assess if the assumption of sphericity has been met. The repeated-measures assumption of sphericity is akin to the assumption of homogeneity of variance in independent ANOVA.

```
library(ez)
```

We will use the `ezANOVA()` function for repeated-measures initially, where `dv=` dependent/response variable; `wid=` subject, the unique case identifier; `within=` method, the explanatory/independent variable `DFn=` effect df; `DFd=` error df; `ges=` generalized eta-squared

```
mod5<-ezANOVA(data=rep_m, dv=(score), wid=.(subject), within=
.(method), detailed= TRUE, type=2)
mod5
$ANOVA
```

\$ANOVA

Effect	DFn	DFd	SSn	SSd	F	p p<.05	ges
1 (Intercept)	1	6	1282.5	30.31	253.83	3.881649e-06	* 0.9526534
2 method	2	12	60.83	33.43	10.919	1.989086e-03	* 0.4883104

\$`Mauchly's Test for Sphericity`

Effect	W	p	p<.05
2 method	0.887657		0.7423572

\$`Sphericity Corrections`

Effect	GGe	p[GG]	p[GG]<.05	HFe	p[HF]	p[HF]<.05
2 method	0.8990033	0.003023708		* 1.259646	0.001989086	*

Note that, using the very same data used for one-way independent ANOVA the repeated-measures ANOVA had a higher F-ratio: 10.92 for repeated-measures one-way compared to 8.59 for independent one-way. As noted at the start this section, this increase in power to reject the null hypothesis is because repeated-measures removes a portion of the error from the denominator, which shrinks the F-ratio error.

The SSn is the SS of the repeated-measures effect (SSm or the model) and this value is 60.83, and this is the same as in the independent one-way ANOVA SSm. The SSd is sum of squares for the residual (unsystematic, unexplained variance), which is 33.43. However, in the independent one-way ANOVA the sum of squares for the residuals was 63.74. Obviously, this is a large difference in total residual. Further look at the steps to the final outcome of the F-ratio for the repeated-measures and the independent ANOVA respectively as it illustrates the impact of the reduced error to result in a higher repeated-measures F-ratio:

Repeated-measures ANOVA outcome

SS mean square residual: $SSd/df: 33.42687/12 = 2.785$

SSn (or SSm) mean square: $60.83155/2 = 30.41$

F-ratio: $30.41/2.785573 = \mathbf{10.919}$

Independent ANOVA outcome

SS mean square residual: $SSr/df: 63.74/18 = 3.54$

SSm mean square: $60.83/2 = 30.415$

F-ratio: $30.415/3.54 = \mathbf{8.589}$

Mauchly's test is non-significant ($p = 0.74$) for sphericity indicating that sphericity is likely not a concern with this data. If Mauchly's test was significant we would need

to consult the Greenhouse and Geisser (1959) or Huynh and Feldt (1976) p-values which are corrected for sphericity.

Finally, a final omnibus test using a multilevel model was conducted mainly to provide additional perspective on the fit of the model (the SSm with the predictor) relative to baseline, where the predictor (explanatory variable) is not included. The dataset used is again the same one used for both one-way independent and repeated-measures ANOVA: the explanatory variable was reading "method", and the response variable was reading "score".

```
library(nlme)
```

The baseline result vs. the model with the predictor (mod6)

```
baseline<-lme(score~1, random= ~1|subject, data= rep_m, method= "ML")  
mod6<-lme(score~method, random= ~1|subject, data=rep_m,method="ML")
```

Comparing the baseline model with no predictor to the model the predictor method
anova(baseline, mod6)

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
baseline	1	3	102.98361	106.11718	-48.49181		
mod6	2	5	92.04028	97.26289	-41.02014	1 vs 2	14.94333 6e-04

The AIC (Akaike information criteria) measures how well a model fits the data, but also penalizes the model incrementally as predictors are added; with AIC more predictors will earn a higher penalty so parsimony with predictors is favored.

BIC (Baysian information criterion) and it too penalizes the model based on number of predictors.

The model (mod6) including the predictor (method) has lower AIC and BIC values than the baseline model without the predictor, and the L.Ratio (likelihood ratio; (which compares the two tests: baseline and mod6) indicate that the predictor model (mod6) is a significantly better fit ($p < .001$) relative to the baseline model.

End