ANOVA Midterm Review Problems

Interpreting Results:
Example 1

Professor McGonagall wants to know if being sorted into a house has an effect on the amount that first years practice transfiguration. Fifteen first years were randomly assigned to 3 Hogwarts houses (Gryffindor, Slytherin, Hufflepuff). Each of the 15 participants reported on the number of transfiguration spells they cast in their first week. Professor McGonagall hypothesized that there would be a difference in the number of spells cast by members of the three houses.

(1) Professor McGonagall is many things (an Animagus, a talented teacher, and a skilled legilimens), but she does not know what analyses she needs to run. What test would you suggest she use?

The output for the analyses is below the set of questions.

1a) Is the assumption of homogeneity of variance violated? Provide evidence by reporting all relevant statistics according to APA style.

1b) Is your data normally distributed? How do you know?

1c) Overall, what did you find? Report all relevant statistics according to APA format.

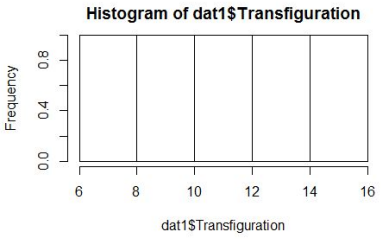
1d) Which, if any post-hoc analyses were significant? Report any relevant statistics in an orderly manner.

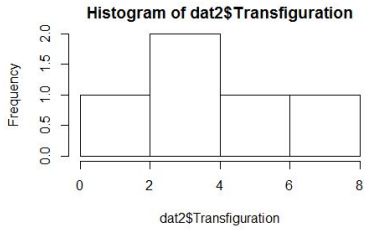
1e) Interpret the output for Professor McGonagall. Although she didn't know what test to run, she does understand statistical language.

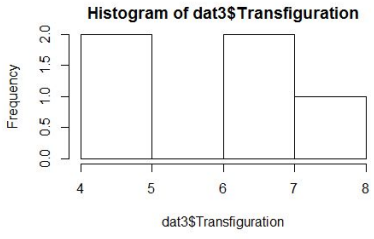
1f) Your dear friend Hagrid is interested in what you found. While Hagrid has many skills with raising insane creatures, he has little knowledge about stats. Statistical jargon and complicate analyses will be hard for him to understand

Output

```
 Descriptive statistics by group
group: Gryffindor
                vars n mean   sd median trimmed  mad min  max range  skew kurtosis   se
House*             1 5  NaN   NA     NA     NaN   NA Inf -Inf  -Inf    NA       NA   NA
Transfiguration    2 5 11.2 3.03     11    11.2 2.97   7   15     8 -0.11    -1.74 1.36
------------------------------------------------------------------------------------------------
group: Hufflepuff
                vars n mean   sd median trimmed  mad min  max range skew kurtosis   se
House*             1 5  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
Transfiguration    2 5  4.2 2.39      4     4.2 2.97   1    7     6 -0.1     -1.9 1.07
------------------------------------------------------------------------------------------------
group: Slytherin
                vars n mean   sd median trimmed  mad min  max range  skew kurtosis   se
House*             1 5  NaN   NA     NA     NaN   NA Inf -Inf  -Inf    NA       NA   NA
Transfiguration    2 5  6.2 1.64      7     6.2 1.48   4    8     4 -0.25    -1.99 0.73
>
> dat1 <- subset(dat, dat$House == "Gryffindor")
> dat2 <- subset(dat, dat$House == "Hufflepuff")
> dat3 <- subset(dat, dat$House == "Slytherin")
>
>
> hist(dat1$Transfiguration)
> hist(dat2$Transfiguration)
> hist(dat3$Transfiguration)
```







```
> jmv::anova(data = dat, dep = 'Transfiguration', factors = c('House'), postHoc = 'House',
postHocCorr = 'tukey', effectSize = 'partEta', homo = TRUE)
```

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2p
House	130.0	2	65.00	11.1	0.002	0.649
Residuals	70.4	12	5.87			

ASSUMPTION CHECKS

Test for Homogeneity of Variances (Levene's)

F	df1	df2	p
0.609	2	12	0.560

POST HOC TESTS

Post Hoc Comparisons - House

House		House	Mean Difference	SE	df	t	p-tukey
Gryffindor	-	Hufflepuff	7.00	1.53	12.0	4.57	0.002
	-	Slytherin	5.00	1.53	12.0	3.26	0.017
Hufflepuff	-	Slytherin	-2.00	1.53	12.0	-1.31	0.419

```
describeBy(dat, dat$House)
```

```
 Descriptive statistics by group
group: Gryffindor
                vars n mean   sd median trimmed  mad min  max range  skew kurtosis   se
House*             1 5  NaN   NA     NA     NaN   NA Inf -Inf  -Inf    NA       NA   NA
Transfiguration    2 5 11.2 3.03     11    11.2 2.97   7   15     8 -0.11    -1.74 1.36
------------------------------------------------------------------------------------------
group: Hufflepuff
                vars n mean   sd median trimmed  mad min  max range skew kurtosis   se
House*             1 5  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
Transfiguration    2 5  4.2 2.39      4     4.2 2.97   1    7     6 -0.1     -1.9 1.07
------------------------------------------------------------------------------------------
group: Slytherin
                vars n mean   sd median trimmed  mad min  max range  skew kurtosis   se
House*             1 5  NaN   NA     NA     NaN   NA Inf -Inf  -Inf    NA       NA   NA
Transfiguration    2 5  6.2 1.64      7     6.2 1.48   4    8     4 -0.25    -1.99 0.73
>
```

Interpreting Results:

Example 2

Next, Professor McGonagall wanted to examine the effect of house and year on the amount of transfiguration practiced. She collected data from 180 students in 3 houses (Gryffindor, Hufflepuff, and Slytherin) across 3 years (first, second, and third). Professor McGonagall hypothesized that there would be an interaction in the number of spells cast by members of the three houses and standing at Hogwarts (i.e., year in school).

2) What test would you suggest she use for this?

The output for the analyses is below the set of questions.

2a) Is the assumption of homogeneity of variance violated? Provide evidence by reporting all relevant statistics according to APA style.

2b) Is your data normally distributed? How do you know?

2c) Was there a main effect for type of house? Report all relevant statistics according to APA format.

2d) Was there a main effect for year in school? Report all relevant statistics according to APA format.

2e) Was there indeed an interaction between the type of house and year in school like Professor McGonagall hypothesized? Report all relevant statistics according to APA format.

2f) Which, if any, simple effect analyses were significant? Report any relevant statistics according to APA format.

2g) Which, if any, of the follow up analyses from the simple effects were significant (i.e., the post-hoc Tukeys from following up the simple effects)? Report all relevant statistics in an orderly manner.

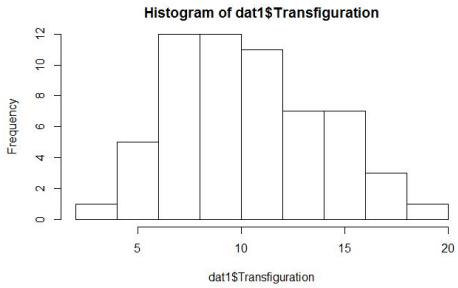
2h) Interpret the output for Professor McGonagall. Although she didn't know what test to run, she does understand statistical language.

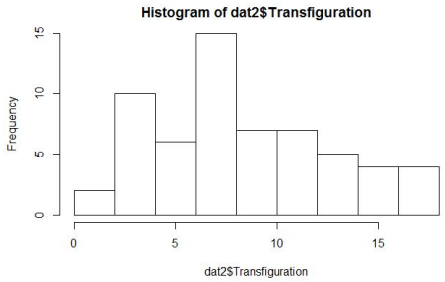
2i) Your dear friend Hagrid is interested in what you found. While Hagrid has many skills with raising insane creatures, he has little knowledge about stats.

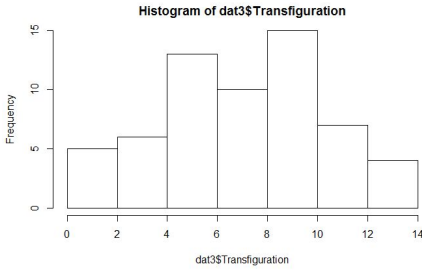
Output

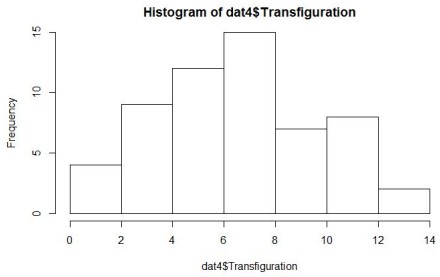
```
> View(dat)
> dat1 <- subset(dat, dat$House == "Gryffindor")
> dat2 <- subset(dat, dat$House == "Hufflepuff")
> dat3 <- subset(dat, dat$House == "Slytherin")
>
> describeBy(dat, dat$House)

 Descriptive statistics by group
group: Gryffindor
                vars  n  mean   sd median trimmed  mad min  max range skew kurtosis
YearInSchool*      1 59   NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA
House*             2 59   NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA
Transfiguration    3 59 10.25 3.58     10   10.15 4.15 2.4 18.2  15.8 0.25    -0.68
                 se
YearInSchool*    NA
House*           NA
Transfiguration 0.47
------------------------------------------------------------
group: Hufflepuff
                vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
YearInSchool*      1 60  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
House*             2 60  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
Transfiguration    3 60 8.31 4.25    7.6     8.1 4.15 0.3 16.9  16.6 0.36    -0.78 0.55
------------------------------------------------------------
group: Slytherin
                vars  n mean   sd median trimmed  mad min  max range  skew kurtosis
YearInSchool*      1 60  NaN   NA     NA     NaN   NA Inf -Inf  -Inf    NA       NA
House*             2 60  NaN   NA     NA     NaN   NA Inf -Inf  -Inf    NA       NA
Transfiguration    3 60 7.05 3.16    7.2    7.08 3.34 1.1 13.2  12.1 -0.03    -0.87
                 se
YearInSchool*    NA
House*           NA
Transfiguration 0.41
>
> hist(dat1$Transfiguration)
> hist(dat2$Transfiguration)
> hist(dat3$Transfiguration)
>
>
> dat4 <- subset(dat, dat$YearInSchool == "First")
> dat5 <- subset(dat, dat$YearInSchool == "Second")
> dat6 <- subset(dat, dat$YearInSchool == "Third")
>
```

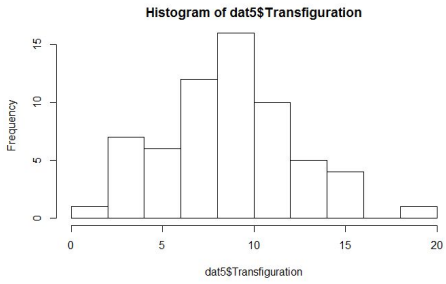
```
> describeBy(dat, dat$YearInSchool)

 Descriptive statistics by group
group: First
                vars  n mean   sd median trimmed  mad min  max range skew kurtosis  se
YearInSchool*      1 57  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA  NA
House*             2 57  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA  NA
Transfiguration    3 57 6.51 3.01    6.6    6.46 3.11 1.1 13.2  12.1 0.14    -0.78 0.4
------------------------------------------------------------
group: Second
                vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
YearInSchool*      1 62  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
House*             2 62  NaN   NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
Transfiguration    3 62  8.6 3.55    8.6    8.51 3.56 0.3 18.2  17.9 0.22     -0.1 0.45
------------------------------------------------------------
group: Third
                vars  n  mean   sd median trimmed mad min  max range skew kurtosis   se
YearInSchool*      1 60   NaN   NA     NA     NaN  NA Inf -Inf  -Inf   NA       NA   NA
House*             2 60   NaN   NA     NA     NaN  NA Inf -Inf  -Inf   NA       NA   NA
Transfiguration    3 60 10.37 4.1   10.2   10.45 4.6 1.4 17.5  16.1 -0.1    -0.89 0.53
>
> hist(dat4$Transfiguration)
> hist(dat5$Transfiguration)
> hist(dat6$Transfiguration)
```

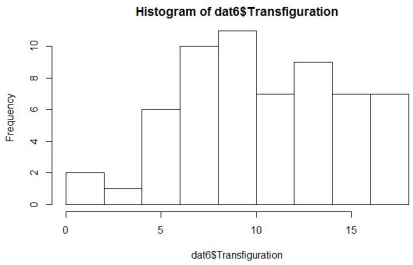













```
> jmv::anova(data = dat, dep = 'Transfiguration', factors = c('House', 'YearInSchool'), homo = TRUE, qq = TRUE)
```

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p
House	313	2	156.30	17.4	< .001
YearInSchool	437	2	218.73	24.3	< .001
House:YearInSchool	434	4	108.38	12.0	< .001
Residuals	1530	170	9.00		

ASSUMPTION CHECKS

Test for Homogeneity of Variances (Levene's)

F	df1	df2	p
1.04	8	170	0.405

```
> jmv::anova(data = dat, dep = 'Transfiguration', factors = c('House', 'YearInSchool'), effectSize = 'partEta',
 postHoc = c('House', 'YearInSchool'), postHocCorr = 'tukey')
NOTE: Results may be misleading due to involvement in interactions
NOTE: Results may be misleading due to involvement in interactions
```

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2p
House	313	2	156.30	17.4	< .001	0.170
YearInSchool	437	2	218.73	24.3	< .001	0.222
House:YearInSchool	434	4	108.38	12.0	< .001	0.221
Residuals	1530	170	9.00			

POST HOC TESTS

Post Hoc Comparisons - House

House		House	Mean Difference	SE	df	t	p-tukey
Gryffindor	-	Hufflepuff	1.90	0.550	170	3.45	0.002
	-	Slytherin	3.23	0.550	170	5.87	< .001
Hufflepuff	-	Slytherin	1.33	0.548	170	2.42	0.043

Post Hoc Comparisons - YearInSchool

YearInSchool		YearInSchool	Mean Difference	SE	df	t	p-tukey
First	-	Second	-2.13	0.551	170	-3.87	< .001
	-	Third	-3.86	0.555	170	-6.96	< .001
Second	-	Third	-1.74	0.543	170	-3.19	0.005

```
>
> jmv::anova(data = dat4, dep = 'Transfiguration', factors = c('House'), effectSize = 'eta',
 postHoc = 'House', postHocCorr = 'tukey')
```

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2
House	74.2	2	37.09	4.61	0.014	0.146
Residuals	434.8	54	8.05			

POST HOC TESTS

Post Hoc Comparisons - House

House		House	Mean Difference	SE	df	t	p-tukey
Gryffindor	-	Hufflepuff	1.60	0.921	54.0	1.74	0.201
	-	Slytherin	2.78	0.921	54.0	3.02	0.010
Hufflepuff	-	Slytherin	1.18	0.921	54.0	1.29	0.409

```
>
> jmv::anova(data = dat5, dep = 'Transfiguration', factors = c('House'), effectSize = 'eta',
 postHoc = 'House', postHocCorr = 'tukey')
```

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2
House	272	2	136.00	16.1	< .001	0.353
Residuals	499	59	8.46			

POST HOC TESTS

Post Hoc Comparisons - House

House		House	Mean Difference	SE	df	t	p-tukey
Gryffindor	-	Hufflepuff	5.11	0.909	59.0	5.62	< .001
	-	Slytherin	2.01	0.909	59.0	2.22	0.076
Hufflepuff	-	Slytherin	-3.09	0.897	59.0	-3.44	0.003

```
>
> jmv::anova(data = dat6, dep = 'Transfiguration', factors = c('House'), effectSize = 'eta'
 postHoc = 'House', postHocCorr = 'tukey')
```

ANOVA

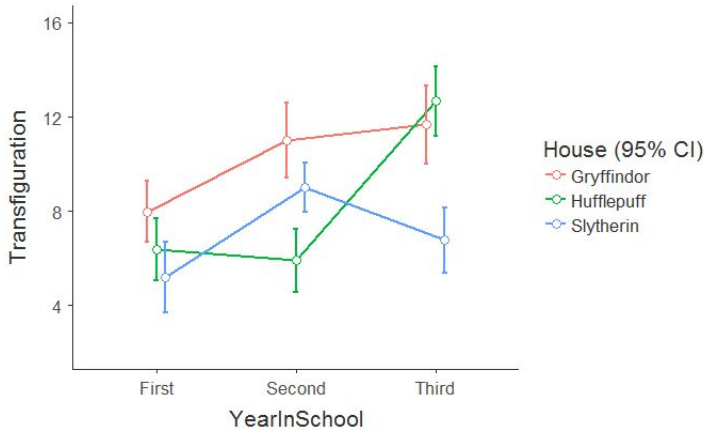
ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2
House	397	2	198.6	19.0	< .001	0.400
Residuals	596	57	10.5			

POST HOC TESTS

Post Hoc Comparisons - House

House		House	Mean Difference	SE	df	t	p-tukey
Gryffindor	-	Hufflepuff	-1.01	1.02	57.0	-0.982	0.591
	-	Slytherin	4.89	1.02	57.0	4.775	< .001
Hufflepuff	-	Slytherin	5.89	1.02	57.0	5.758	< .001



Conceptual Questions:

1. What is homogeneity of variance? Why is it an assumption for independent *t* tests and ANOVAs?

2. Explain the difference between main effects, interaction effects, and simple effects.

3. What is the smallest possible *F* value? What would it mean if you were to obtain this *F* ratio?

4. What is the largest possible *F* value? What would it mean if you were to obtain this *F* ratio?

5. Explain what it means to obtain a *p* value of 0.30. Explain what it means to obtain a *p* value of 0.01. Explain to your grandmother (assuming she isn't a statistician) what these values mean in the context of null hypothesis significance testing. (HINT: explain this to a lay audience, people without statistical training)

6. How does sample size affect *df* when pooling error variance in an ANOVA? How might this lead to a higher *F* value?

7. Why can't an ANOVA tell me which specific group in my sample is different from another specific group in my sample? What sort of analysis do I need to perform instead?

8. What's wrong with running a bunch of regular *t* tests on my data to figure out individual group differences? What can I do to correct for this issue?

9. What is the difference between complete and partial eta-squared? When would you use each?

ANOVA Summary Table Completion

1. Professor Trelawney conducts a one-way ANOVA to determine if there is a difference in the number of times a bad omen (such as the Grim!) occurs for students of different houses. She reads the tea leaves of 12 students from each of 3 houses: Gryffindor, Slytherin, and Hufflepuff. Solve for each missing value of the ANOVA summary table below.

Source	SS	df	MS	F
Between Groups	?	?	9	?
Within Groups	?	?	?	
Total	117	?		

2. A one-way ANOVA was conducted. The independent variable represents the different departments (psychology, education, theology). Each group contains 10 students. Solve for each missing value of the ANOVA summary table below.

Source	SS	df	MS	F
Between Groups	?	?	15	7.50
Within Groups	?	?	?	
Total	?	?		

3)

Source	SS	df	MS	F
Between Groups	56	?	?	?
Within Groups	?	9	?	
Total	68	11		

4)

Source	SS	df	MS	F
Between Groups	?	20	?	2.5
Within Groups	?	?	2	
Total	160	?		

5)

Source	SS	df	MS	F
Factor A	0	1	?	0
Factor B	12	?	?	?
Interaction AxB	?	1	?	16
Within Groups	?	?	.75	
Total	30	?		