ANOVA Midterm Review Problems ANSWER KEY

<u>Interpreting Results:</u>
Example 1

1) One-way ANOVA

1a) Homogeneity of variance is not violated according to the non-significant Levene's test, $F(2, 12) = 0.61$, $p = .560$

1b) Yes, skew and kurtosis for all three groups are within the cutoffs: skew is between -3 and +3 and kurtosis is between -10 and +10. Also the histograms, don't quite look normal, but since ANOVA is robust, it doesn't seem like it will be a major issue (especially given the skew and kurtosis values)

1c) There was a significant difference somewhere between the three groups, $F(2, 12) = 11.10$, $p = .002$, $\eta^2 = .65$

1d) There were significant differences between two of the three pairwise comparisons:
$M_{\text{Gryffindor}} - M_{\text{Hufflepuff}} = 7.00$, $p = .002$
$M_{\text{Gryffindor}} - M_{\text{Slytherin}} = 5.00$, $p = .017$

1e) There was a significant difference somewhere between the three houses (Gryffindor, Hufflepuff, and Slytherin) on the number of transfigurations they performed, $F(2, 12) = 11.10$, $p = .002$, $\eta^2 = .65$. This is a large effect with house assignment (i.e., Gryffindor, Hufflepuff, and Slytherin) accounting for 65% of the variance in number of transfigurations performed. Upon further post-hoc analyses utilizing Tukey HSD, the significant differences were between Gryffindor and Hufflepuff ($M_{\text{Gryffindor}} - M_{\text{Hufflepuff}} = 7.00$, $p = .002$) and Gryffindor and Slytherin ($M_{\text{Gryffindor}} - M_{\text{Slytherin}} = 5.00$, $p = .017$) such that Gryffindor students ($M = 11.20$, $SD = 3.03$) were able to conjure significantly more transfigurations than both Hufflepuff ($M = 4.20$, $SD = 2.39$) and Slytherin ($M = 6.20$, $SD = 1.64$). This suggests that individuals sorted into Gryffindor are more likely than either of the other two houses to practice their transfiguration.

1f) Individuals that were sorted into the Gryffindor group ended up casting a large amount more transfigurations than those in the other two houses. Individuals in Gryffindor cast around 11 spells on average, whereas those in Hufflepuff cast around 4 and those in Slytherin cast around 6. Although Slytherin cast more spells than Hufflepuff, they didn't cast a whole lot more. This suggests there is something going on with Gryffindors that is causing them to practice more than the other houses. Shall we do more studies to try to figure out what that something is??! (Maybe some qualitative this time, eh??)

2) Factorial ANOVA

2a) The assumption of homogeneity of variance is not violated according to the non-significant Levene's test, $F(8, 170) = 1.04$, $p = .405$

2b) The assumptions of the dependent variable being normally distributed for each group was not violated as skew and kurtosis for each of the six groups were within the necessary cut-off values (i.e., skew between -3 and +3, kurtosis between -10 and +10). The histograms were all normally distributed as well.

2c) There was a significant main effect on the number of transfigurations students cast based on the type of house they were in, $F(2, 170) = 17.40$, $p < .001$, partial $\eta^2 = .17$. This is a large effect size with student house accounting for 17% of the variance in transfiguration spells.

2d) There was a significant main effect on the number of transfigurations students cast for the year in school they were in, $F(2, 170) = 24.30$, $p < .001$, partial $\eta^2 = .22$. This is a large effect size with school standing accounting for 22% of the variance in transfiguration spells.

2e) There was an interaction between the type of house individuals were in and the year in school they were in on the number of transfigurations they cast, $F(4, 170) = 12.00$, $p < .001$, partial $\eta^2 = .22$.

2f) There was a simple effect between the different houses that first year students were in on the number of transfiguration spells they cast, $F(2, 54) = 4.61$, $p = .014$, $\eta^2 = .15$.

There was a simple effect between the different houses that second year students were in on the number of transfiguration spells they cast, $F(2, 59) = 16.10$, $p < .001$, $\eta^2 = .35$.

There was a simple effect between the different houses that third year students were in on the number of transfiguration spells they cast, $F(2, 57) = 19.00$, $p < .001$, $\eta^2 = .40$.

2g) *Within First year:*
Gryffindor – Slytherin: $M_{\text{Gryffindor}} - M_{\text{Slytherin}} = 2.78$, $p = .010$

Within Second Year:
Gryffindor – Hufflepuff: $M_{\text{Gryiffindor}} - M_{\text{Hufflepuff}} = 5.11$, $p < .001$
Hufflepuff – Slytherin: $M_{\text{Hufflepuff}} - M_{\text{Slytherin}} = -3.09$, $p = .003$

Within Third Year:
Gryffindor – Slytherin: $M_{\text{Gryffindor}} - M_{\text{Slytherin}} = 4.89$, $p < .001$

Hufflepuff – Slytherin: $M_{\text{Hufflepuff}} - M_{\text{Slytherin}} = 5.89$, $p < .001$

2h) A factorial ANOVA test was used to examine the effects of house, student standing, and the interaction between the two independent variables on the number of transfiguration spells casted. The omnibus ANOVA found a significant interaction, $F(4, 170) = 12.00$, $p < .001$, $\eta^2 = .22$. This is a large effect size with the interaction between the two independent variables accounting for 22% of the variance in number of transfiguration spells. We followed up this omnibus test (i.e., blob test) with simple effects and found that there were differences across the house students were in at each year of school. At the first year in Hogwarts, Gryffindor students cast significantly more transfiguration spells than Slytherin ones, $M_{\text{Gryffindor}} - M_{\text{Slytherin}} = 2.78$, $p = .010$. At the second year, both Gryffindor cast significantly more transfiguration spells than Hufflepuff, $M_{\text{Gryiffindor}} - M_{\text{Hufflepuff}} = 5.11$, $p < .001$. Additionally, at the first year of Hogwarts, Slytherin students cast significantly more transfiguration spells than Hufflepuff, $M_{\text{Hufflepuff}} - M_{\text{Slytherin}} = -3.09$, $p = .003$. Finally, at the third year at Hogwarts, Gryffindor cast significantly more spells than Slytherin, $M_{\text{Gryffindor}} - M_{\text{Slytherin}} = 4.89$, $p < .001$, and Hufflepuff cast significantly more spells than Slytherin, $M_{\text{Hufflepuff}} - M_{\text{Slytherin}} = 5.89$, $p < .001$.

2i) Overall, all three groups start off very similar to each other in the number of transfiguration spells they are casting. The only difference was that Gryffindor first years were casting more spells than Slytherin first years. By the second year, Gryffindor are still casting a good amount of spells, but Slytherin students have now upp-ed their game (i.e., they're getting better at spells). Now both Gryffindor students and Slytherin students are casting more spells than Hufflepuff ones. By the third year, though, Hufflepuff has really pulled through and improved. Both Hufflepuff and Gryffindor are casting a whole lot more spells than Slytherin students. That is a big rise in spells cast for Hufflepuff between the second to third year and a rather large drop for Slytherin students during the same time period. Those slackers!

Conceptual Questions:

1) The assumption of homogeneity of variance is the assumption that all groups have the same or similar variance. You can test this assumption using Levene's test, which provides an *F* value and a *p* value. A nonsignificant result indicates that the assumption of homogeneity of variance is satisfied, and you can proceed with the inferential statistics test (e.g., ANOVA).

The assumption of homogeneity of variance is important for *t* tests and *F* tests because comparing means of groups with significantly different variances can result in invalid data. As long as group sizes are (relatively) equal, the *F* test is pretty robust, but if group sizes are different, it can wreak havoc on your results. For instance, when the assumption is violated, and the smallest size group has the highest variance, you are inflating your alpha.

2) **Main effects** are the effects of one IV on the DV ignoring all other IVs (or in other words, averaging across the levels of the other IV). **Interaction effects** are when the effect one IV depends depending on the other IV. That is, the effects of one IV change across the levels of the other IV. **Simple effects** are the effect of one IV only looking at a single level of the other IV. (Also see Andrew's slides)

3) If you ran an ANOVA comparing, for example, 3 sets of <u>*identical*</u> data in which <u>*no*</u> score deviated from the mean, you would have 0 between group variance. This would lead to the <u>*very rare*</u> instance of an *F* value of 0. Generally, any *F* value less than 1 is bad news because it means your error variance is higher than your between groups variance.

4) The largest possible *F* value is technically limited only by the largest deviation score from the mean in your data. According to many theologists, if you obtained an infinite *F* value it means you accidentally sampled God. Good luck explaining that to IRB.

5) A *p* value of .30 indicates my results were likely due to chance (i.e., there is a 30% chance I could have obtained my results simply from the randomness of life). A *p* value of .01 means that it is highly unlikely that my obtained results were due to chance (i.e. there is a 1% chance my results were obtained due to the randomness of life).

Hi Grandma, thanks for the cookies. We use *p* values to describe how likely something is to happen. For instance, if you told me you rolled across all the lanes of the I-5 Freeway in your wheelchair and were hit by a car, we would not be surprised. The I-5 is a freeway traveled by millions of speeding cars each day. That means there's a high probability you're gonna get hit…

Let's say you told me you rolled across the I-5 and were NOT hit by a car. We would be pretty surprised, right? Given what we assume about how busy a freeway is, you NOT being hit while crossing is pretty unlikely (the null hypothesis). We COULD suppose, however, with the right amount of dumb luck, you might be able to cross the freeway ONE time unscathed. Let's say dumb luck could get you safely across 1 in 10 times. The *p* value for your story about crossing the freeway one time unharmed is 0.10 assuming that the null hypothesis is true.

Now let's say you told me you crossed the I-5 freeway every day for the last year and you weren't hit a single time. Could we chalk that up to random luck? The thought of you successfully avoiding the unpredictable pattern of cars screaming by 365 times in a row would require an almost unthinkable amount of dumb luck - let's call those chances 1 in a million. Our *p* value for your story about crossing the freeway 365 times unharmed would be 0.000001.

That *p* value is so low that I would be forced to conclude SOMETHING STRANGE AND INTERESTING must be going on to explain your amazing feat. I would probably ask you the follow-up question: "What time of day did you cross?". Of course, you would have a completely reasonable explanation: "I crossed every day at 5PM rush hour. The freeway was at a standstill."

In science, our data tell the story. If the story can be explained by dumb luck, chance, or the randomness of life, it has a high *p* value. As the the stories become less and less explainable by chance, the *p* value drops. Once the *p* value drops below .05, scientists formally believe there is some explanation besides "chance" for the story. The scientist doing the research often has a presupposed alternate explanation for the story (we call that an alternative hypothesis), but that's a talk for another day. The bottom line is that to scientists, a *p* value less than .05 means there's very likely SOMETHING STRANGE AND INTERESTING besides luck that explains what is going on.

6) Your sample size is used to calculate df which in turn is used as the denominator in your calculation of MS_{Error}. Therefore, larger df will lead to smaller MS_{Error}. Because MS_{Error} is in the denominator of your F ratio, a small error value will make for a large F value. So, when between group differences are held constant, larger sample sizes will make for higher F values (and more likely trigger significant results). *(Study Tip: You can demonstrate this mathematically by changing the sample size, and consequently degrees of freedom, but maintaining all other values in the calculation)*

7) An ANOVA is a "blob test" which is not designed to calculate which individual group is different from another; rather, an ANOVA is only sensitive to whether there is a statistical difference SOMEWHERE among your groups. Once the ANOVA has returned a significant result, it is appropriate to run post-hoc t-tests to determine EXACTLY which specific group is different from another specific group.

8) Firstly, doing this makes you come across as a total lunatic. Secondly, every time you run a t-test you are inviting the possibility that you will get a significant result by chance (the size of this possibility is determined by your alpha level). For this reason, running 100 t-tests "inflates" your alpha level 100x and will lead to completely inaccurate statistical findings. To correct for inflated alpha issues, I can use Tukey's HSD or Bonferroni's method, both of which make adjustments in calculation to accommodate the number of t-tests being run on the data. *(Study Tip: Refer to the Familywise Error Rate formula to mathematically determine the probability of making at least one Type 1 Error if you don't control for alpha inflation during repeated t-tests. HINT: Kathy discussed this at the end of module 1 - CLASSES BUILD UPON EACH OTHER?!?! SAY WHAAAA?!? Also, you can Google this information)*

9) **Complete eta-squared** = $SS_{treatment} / SS_{total}$
Basically, complete eta-squared looks at the variance accounted for by a given variable and divides it by the total variance in the model. Therefore, if you have a complete eta-squared value of 0.35, you can say the IV explains 35% of the variance in the DV. Complete eta-squared is appropriate to report for *one-way ANOVAs* and for *simple effects of two-way ANOVAs* because the variance of the treatment is only

accounted for by one IV. In these situations, the complete eta-squared = partial eta-squared! *mindblown*

Partial eta-squared = $SS_{treatment}$ / ($SS_{treatment}$ + SS_{error})

As noted above, partial eta-squared values will always be identical to complete eta-squared values for one-way ANOVAs or simple effects of two-way ANOVAs. While complete eta-squared takes into account all the variance in the model, partial eta-squared is only concerned with the *variance unique to a particular factor*. Therefore, in a factorial design with Factor A, Factor B, and an interaction term (AxB), you would only consider the variance associated with Factor A and the error variance in the denominator for the calculation of the effect size of Factor A. It doesn't make sense to include the unique variance accounted for by Factor B when you want to discuss the effect size of Factor A.

ANOVA Summary Table Completion

1)

Source	SS	df	MS	F
Between Groups	18	2	9	3
Within Groups	99	33	3	
Total	117	35		

2)

Source	SS	df	MS	F
Between Groups	30	2	15	7.50
Within Groups	54	27	2	
Total	84	29		

3)

Source	SS	*df*	MS	F
Between Groups (Treatment)	56	2	28	21.05
Within Groups (Error)	12	9	1.33	
Total	68	11		

4)

Source	**SS**	***df***	**MS**	**F**
Between Groups (Treatment)	100	20	5	2.5
Within Groups (Error)	60	30	2	
Total	160	50		

5)

Source	**SS**	**df**	**MS**	***F***
Factor A	0	1	0	0
Factor B	12	1	12	16
Interaction AxB	12	1	12	16
Within Groups (Error)	6	8	.75	
Total	30	11		