Exam 2 PS 217, Spring 2011

1a. In each of the chapters we've covered for this exam, there has been some discussion of effect size. First of all, tell me how you would measure effect size (i.e., tell me the statistic you would use). Then, tell me how effect size and power are related. [3 pts]

For ANOVA, we've been using eta squared (η^2). For an independent groups ANOVA, $\eta^2 = \frac{SS_{Between}}{SS_{Total}}$. For a repeated measures ANOVA, $\eta^2 = \frac{SS_{Between}}{SS_{Total} - SS_{Subjects}}$. If you're dealing with a small effect size, you'll need a lot of power to detect the effect. If, on the other hand, your effect size is large, you'll have a lot of power.

1b. If researchers work to achieve a level of power of .80, what does that say about their tolerance for Type II Error relative to Type I Error, given the usual α-level used in research? [2 pts]

Because they're complementary, the probability of a Type II error would be .20 (or four times larger than the probability of Type I error).

2. Dr. Rick Call is interested in whether the type of mood reflected by a word (happy, neutral, or sad) affects how well the word is remembered. He constructs a list of 18 words: 6 happy words (e.g., joyful, bright), 6 neutral words (e.g., derive, convey), and 6 sad words (e.g., gloomy, lonely). He then presents the list repeatedly to 30 participants until they can recite the entire list correctly twice in a row. One week later, each participant attempts to recall the entire list. The number of items correctly recalled as a function of the type of word is analyzed.

If you were Dr. Call's research assistant, how might you suggest that he construct the list of words? [3 pts]

I would probably generate more than one list of words. However, what I'd be sure to do in any list is to ensure that among the first trials (e.g., the first six) there would be a random order of two happy, two neutral, and two sad words. Thus, that first block of six words might be SHNHSN. The last block of six words might be NNSHSH. Doing so ensures that subjects will see happy, neutral, and sad words roughly equally often at the beginning (primacy effects?) and at the end of the list (recency effects?).

Complete the source table below and interpret the results of this study as completely as you can. [12 pts]

Descriptive Statistics

	Mean	Std. Deviation	N
Happy	4.3333	.80230	30
Neutral	2.0000	.78784	30
Sad	3.6667	1.09334	30

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[a]
Emotion	Sphericity Assumed	86.67	2	43.33	50.27	.000	.634	1.000
Error(Emotion)	Sphericity Assumed	50.00	58	.862				

Decision: Reject H_0, because $p < .05$. To determine which means differed, I would generate Tukey's HSD.

$$\boldsymbol{HSD = 3.4\sqrt{\frac{.862}{30}} = .58}$$

Thus, all the means differ. I would conclude:

There was a significant effect of word type on subsequent recall, $F(2,58) = 50.27$, $MSE = .862$, $p < .001$, $\eta^2 = .634$. Post hoc tests using Tukey's HSD indicate that people recalled significantly more happy words ($M = 4.33$) than sad words ($M = 3.67$) or neutral words ($M = 2.0$). They also recalled more sad words than neutral words.

3. Eskine, et al. (2011) recently published an article in *Psychological Science* with the intriguing title "A bad taste in the mouth: Gustatory disgust influences moral judgment." Their abstract reads, in part…

Can sweet-tasting substances trigger kind, favorable judgments about other people? What about substances that are disgusting and bitter? Various studies have linked physical disgust to moral disgust, but despite the rich and sometimes striking findings these studies have yielded, no research has explored morality in conjunction with taste, which can vary greatly and may differentially affect cognition. The research reported here tested the effects of taste perception on moral judgments. After consuming a sweet beverage, a bitter beverage, or water, participants rated a variety of moral transgressions.

From their method section:

> [Participants] were told that we were exploring the effects of motor interference (specifically arm-hand movements) on cognitive processing, and we therefore directed them to drink a beverage during a moral-judgment task to instantiate this movement in a natural way. Participants in the sweet condition were given Minute Maid Berry Punch, those in the bitter condition received Swedish Bitters, and control participants were given water. They were not told the identity of the beverages, although an ingredient list was provided so they could check for potential allergens. Beverages were administered in two 1-teaspoon doses in a small cup; the first dose was given at the onset of the moral-judgment task, and the second one was administered at the halfway point to ensure that the taste lingered throughout the task. Participants were instructed to drink each dose in its entirety in a single swift motion, "as if you were drinking a shot."
>
> Moral judgments were assessed using Wheatley and Haidt's (2005) moral vignettes, which portray various moral transgressions (second cousins engaging in consensual incest, a man eating his already-dead dog, a congressman accepting bribes, a lawyer prowling hospitals for victims, a person shoplifting, and a student stealing library books). All participants received the same six moral vignettes, in counterbalanced order. After each vignette, participants rated "how morally wrong" the offense was on a scale consisting of a 14-cm line representing a continuum from *not at all morally wrong* to *extremely morally wrong*. Participants were asked to make a slash at the point on the continuum corresponding to their impressions. These marks were then converted to scores ranging from 0 to 100, with higher scores indicating harsher moral judgments.

Complete the source table below. In light of their hypotheses, interpret the results as completely as you can (as you might in a discussion). [15 pts]

Descriptives

Moral

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Bitter	17	78.2353	10.29884	2.49784	72.9401	83.5305	53.00	93.00
Sweet	17	59.5882	15.72044	3.81277	51.5055	67.6709	34.00	89.00
Water	17	61.5294	12.99576	3.15193	54.8476	68.2112	28.00	93.00
Total	51	66.4510	15.44450	2.16266	62.1071	70.7948	28.00	93.00

Test of Homogeneity of Variances

Moral

Levene Statistic	df1	df2	Sig.
1.553	2	48	.222

ANOVA

Moral

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3574.58	2	1787.29	10.27	.000
Within Groups	8353.44	48	174.03		
Total	11928.02	50			

Because the Levene test is not significant ($p > .05$), there appears to be no concern about heterogeneity of variance. You could compute F_{Max} if you were so inclined, but you'd arrive at the same conclusion. However, because the ANOVA results in an F-ratio that is so large ($p < .01$), even if the data led you to use a more conservative α, you would still reject H_0.

Decision: Reject H_0, $p < .01$.

$$HSD = 3.42\sqrt{\frac{174.03}{17}} = 10.94$$

There was a significant effect of flavor of beverage on harshness of judgments, $F(2,48) = 10.27$, $MSE = 174.03$, $p < .001$, $\eta^2 = .30$. Post hoc tests using Tukey's HSD indicate that people who ingested the bitter beverage made significantly harsher judgments ($M = 78.24$) than those who ingested a sweet beverage ($M = 59.59$) or water ($M = 61.53$).

4. Generally speaking, what does a post hoc test (such as Tukey's HSD) do for you? You should know, of course, that it allows you to compare two means to determine if they are drawn from populations with different means. But you could test such a hypothesis with a *t*-test or an ANOVA. So, what does a post hoc test do for you *beyond* allowing you to test for the difference between two conditions? [5 pts]

Post hoc tests such as Tukey's HSD adjust the per-comparison α to control for an inflated chance of Type I error. Thus, such tests are more conservative than a simple *t*-test or ANOVA.

5. A psychologist would like to examine the relative effectiveness of three therapy techniques for treating mild phobias. A group of 15 individuals who display a moderate fear of spiders is obtained. These individuals are randomly assigned to one of the three therapies with $n = 5$. The dependent variable is a measure of the reported fear of spiders, with higher numbers indicating greater fear. Analyze these data as completely as you can. [20 pts]

	Therapy A	Therapy B	Therapy C	
	8	3	1	
	5	3	0	
	5	0	1	
	7	2	2	
	5	2	1	
(ΣX) or T	30	10	5	45
ΣX^2	188	26	7	221
SS	8	6	2	

Source	*SS*	*df*	*MS*	*F*
Between	70	2	35	26.25
Within	16	12	1.33	
Total	86	14		

$$SS_{Total} = \sum X^2 - \frac{G^2}{N} = 221 - \frac{45^2}{15} = 86$$

$$SS_{Between} = \sum \frac{T^2}{n} - \frac{G^2}{N} = 205 - \frac{45^2}{15} = 70$$

$F_{Max} = \frac{2}{.5} = 4$ and $F_{MaxCritical} = 15.5$, so there's no concern about violating the homogeneity of variance assumption. I would then us $\alpha = .05$ to determine the $F_{Critical}$ for the overall ANOVA. $F_{Critical}(2,12) = 3.88$

Decision: Reject H_0, because $F_{\text{Obtained}} \geq F_{\text{Critical}}$.

$$HSD = 3.77\sqrt{\frac{1.33}{5}} = 1.94$$

There was a significant effect of therapy on reported fear of spiders, $F(2,12) = 26.25$, $MSE = 1.33$, $p < .05$, $\eta^2 = .81$. Post hoc tests using Tukey's HSD revealed that Therapy C led to significantly less fear of spiders ($M = 2$) than Therapy A ($M = 8$) or Therapy B ($M = 6$).

6. Two researchers were interested in studying the effects of reward magnitude (the IV) on performance. Both researchers used the same independent variable, with three levels of reward magnitudes ($1, $5, $20). They used the same pool of introductory psychology students as participants, the same total number of participants (24), the same apparatus, the same task, and the same performance measure (DV). One researcher used an independent groups design and the other researcher used a repeated measures design. Assume that neither study has a major flaw (e.g., repeated measures design is properly counterbalanced, random assignment to conditions). In each case, the researchers used $\alpha = .05$. Complete the source tables for the two experimenters seen below, then make a decision about the H_0: $\mu_{\$1} = \mu_{\$5} = \mu_{\$20}$. Do the results surprise you? Why? How might you explain the potentially unexpected outcome? [10 pts]

Independent Groups Design:

Source	*SS*	*df*	*MS*	*F*
Between	16	2	8	4
Within	42	21	2	
Total	58	23		

Using $\alpha = .05$, $F_{\text{Critical}}(2,21) = 3.47$

Thus, you'd reject H_0, because $F_{\text{Obtained}} \geq F_{\text{Critical}}$.

Repeated Measures Design:

Source	*SS*	*df*	*MS*	*F*
Between	12	2	6	3
Within	94	69		
Subject	2	23		
Error (Subj x Treat)	92	46	2	
Total	106	71		

Using $\alpha = .05$, $F_{\text{Critical}}(2,46) = 3.2$

Thus, you would retain H_0, because $F_{\text{Obtained}} < F_{\text{Critical}}$.

Because you should expect a repeated measures analysis to be more powerful, you should then be surprised to find that you reject H_0 for the independent groups ANOVA, but not for the repeated measures ANOVA. However, if you look at the source table for the repeated measures ANOVA, you should notice that the SS_{Subjects} was actually quite small, so that the error term wasn't reduced as much as you'd like to be the case.