Exam 2 PS 217, Fall 2008

1. Suppose that you are interested in whether the type of mood reflected by a word (happy, neutral, or sad) affects how well the word is remembered. You construct a list of 10 happy words (e.g., joyful, bright), 10 neutral words (e.g., derive, convey), and 10 sad words (e.g., gloomy, lonely). You present the list repeatedly to 8 participants until they can recite the entire list correctly twice in a row. You repeat this procedure for each word list, for a total of 24 participants. One week later, each participant attempts to recall the entire list. The number of items correctly recalled (the DV) is seen below as a function of the type of word (the IV). Analyze the results of this experiment as completely as possible, including whatever interpretation seems appropriate. [20 pts]

	Happy	Neutral	Sad	
	5	4	3	
	6	3	4	
	4	5	2	
	5	3	1	
	3	1	2	
	6	3	4	
	2	2	3	
	5	3	1	
T (ΣX)	36	24	20	G = 80
Mean	4.5	3.0	2.5	$\Sigma X^2 = 318$
SS	14	10	10	

Source	*SS*	*df*	*MS*	*F*
Between (Mood)	**17.33**	**2**	**8.67**	**5.35**
Within (Error)	**34.0**	**21**	**1.62**	
Total	**51.33**	**23**		

Hartley's F_{Max} = 1.4, so with $F_{Max\ Crit}$ = 6.94 you wouldn't be concerned about heterogeneity of variance, so you'd use α = .05. With $F_{Crit}(2,21) = 3.47$, you would reject H_0, because $F_{Obt} \geq F_{Crit}$. Having done so, you would next need to compute HSD to determine which means differed. With q = 3.57, HSD = 1.61. Thus, Happy mood leads to significantly more words recalled (*M* = 4.5) than Sad (*M* = 2.5), but no other differences were significant.

2. Suppose that you were interested in computing an ANOVA on 3 sets of data from 30 different people, for which summary statistics are shown below. [5 pts]

	Condition 1	Condition 2	Condition 3
Mean	15	20	40
Variance (s^2)	5	15	70
n	10	10	10

a. What parameter is MS_{Within} intended to estimate? **σ^2**

b. In the ANOVA computed on the data above, what would MS_{Within} be? **30 (mean of variances)**

c. What F_{Crit} would you use to evaluate the F_{Obt} for the overall ANOVA?

Here you need to be careful. F_{Max} = 35 and $F_{MaxCrit}$ = 5.34, so you'd want to use α = .01 for the *F* in the overall ANOVA. Thus, $F_{Crit}(2,27) = 5.49$.

d. If the three means had been *identical* (e.g., 15, 15, 15), what would F_{Obt} be? **0**

e. If the same study were conducted as a repeated measures design, how many total participants would you need to produce at least as many pieces of data (i.e., at least 30) as seen above?

Again, be careful. You'd want to counterbalance! Without counterbalancing, you'd need only 10 participants. However, with three conditions, you should use complete counterbalancing, which means that you'd need six orders. Thus, to get at least 30 pieces of data, you'd want to run 12 participants (generating 36 pieces of data).

3. People like Jacob Cohen suggest that we should conduct experiments with power of at least .80. What level of Type II error are they willing to tolerate? [2 pts]

.20 (Power and Type II Error are complementary)

4. Briefly explain why it is essential to develop an error term (MS_{Error}) for the repeated measures design that includes only variability due to random factors? (In other words, what's going on in the $MS_{Between}$?) [5 pts]

In a repeated measures design, the variability among the condition means is attributable to any effects of treatment as well as random variability. However, there is no contribution of individual differences. As a result, using an error term that includes individual differences would be inappropriate. It is for that reason that we need to develop an error term (MS_{Error}) for which the only source of variability is random effects.

5. In the repeated measures lab, we never got around to computing the final repeated measures ANOVA. (You may have done so on your own.) Below is a SPSS analysis of the data from the lab. As you may recall, the lab dealt with the notion of bats flying through a maze comprised of wires of varying thickness (Thin, Medium, Thick). We first considered an analysis of an independent groups design, and then the repeated measures design. Analyze and interpret the results below as completely as you can. [13 pts]

Descriptive Statistics

	Mean	Std. Deviation	N
Thin	3.0556	1.21133	18
Medium	1.6111	.97853	18
Thick	.2778	.46089	18

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Wire	Sphericity Assumed	69.0	2	34.5	71.1	.000	.808	143.013	1.000
Error(Wire)	Sphericity Assumed	16.5	34	.485					

a. Computed using alpha = .05

H_0: $\mu_{Thin} = \mu_{Medium} = \mu_{Thick}$ H_1: Not H_0

Decision: Reject H_0, because $p < .05$.

To determine which conditions differ, you'd need to compute a post hoc test (Tukey's

HSD):

$$HSD = 3.47\sqrt{\frac{.485}{18}} = .57$$

All three means differ by at least .57, so we would conclude that bats hit the thin wire significantly more often (*M* = 3.06) than the medium wire (*M* = 1.61) or the thick wire (*M* = .28). The bats also hit the medium wire more often than the thick wire.
SPSS provides an estimate of effect size (partial eta squared), which is quite high (.808).

6. Dr. Thomas Tomm is interested in the impact of distraction on driving ability. To that end, he has people drive a car on a closed course. The independent variable is the type of distraction (carrying on a cell-phone conversation, sending a text message, or eating a slice of pizza). Each of the twelve participants first drives the course in the cell-phone condition, then in the text-message condition, and finally in the pizza condition. The dependent variable is the number of times the driver strays from the course. Interpret these data as completely as you can, telling Dr. Tomm what he should do next. [10 pts]

Descriptive Statistics

	Mean	Std. Deviation	N
cellphone	13.3333	1.49747	12
textmsg	13.4167	1.16450	12
pizza	13.0000	1.20605	12

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
distractor	Sphericity Assumed	1.167	2	.583	.562	.578	.049	1.124	.131
	Greenhouse-Geisser	1.167	1.777	.656	.562	.559	.049	.999	.126
	Huynh-Feldt	1.167	2.000	.583	.562	.578	.049	1.124	.131
	Lower-bound	1.167	1.000	1.167	.562	.469	.049	.562	.105
Error(distractor)	Sphericity Assumed	22.833	22	1.038					
	Greenhouse-Geisser	22.833	19.549	1.168					
	Huynh-Feldt	22.833	22.000	1.038					
	Lower-bound	22.833	11.000	2.076					

a. Computed using alpha = .05

For starters, you should note that the results are not significant, with *p* > .05. Next, you should think about reasons that the study produced results that were not significant. You should note, for example, that there is no control group (no distractions). These three distraction conditions may not differ from one another, but that doesn't mean that the distractions have no impact on driving. That would be one way to increase the treatment effect. You might think of other ways to increase the treatment effect (e.g., not just pizza, but pizza *and* soda!). You might also think about ways to decrease random variability (e.g., be sure that the driving conditions are as close to identical as you can make them). Finally, you should tell the good Dr. Tomm how important it is to counterbalance a repeated measures study. As described above, all participants go through the conditions in the same order (cell-phone, then text-messaging, then pizza). That's a big no-no. In other words, Dr. Tomm has an appropriate number of participants (a multiple of 6), but he didn't counterbalance!