Exam 3 PS 217, Spring 2008

1. [Sprinthall] Assume that a researcher is interested in whether or not there is a relationship between a politician's physical attractiveness and her or his years in office. The researcher takes a random sample of eight members of Congress from throughout the country and ranks them on attractiveness (1 = "lowest attractiveness" and 8 = "highest attractiveness"). These rankings are paired with the number of years each has served in the House of Representatives. Use these data to provide an answer to the question. [15 pts]

Member	Attractiveness	Years Served	Rank Years	Att x Rank Yr
A	1	6	1	1
B	2	8	2	4
C	3	12	3.5	10.5
D	4	14	5	20
E	5	12	3.5	17.5
F	6	20	6.5	39
G	7	20	6.5	45.5
H	8	24	8	64
Σ	36		36	201.5
Σ Squared	204		203	

Of course, because attractiveness scores are ranked, this problem requires Spearman's *rho*.

$$SP = 201.5 - \frac{(36)(36)}{8} = 39.5$$

$$SS_X = 204 - \frac{36^2}{8} = 42$$

$$SS_Y = 203 - \frac{36^2}{8} = 41$$

$$rho = \frac{39.5}{\sqrt{42 \times 41}} = .95$$

$rho_{\text{Crit}} = .738$

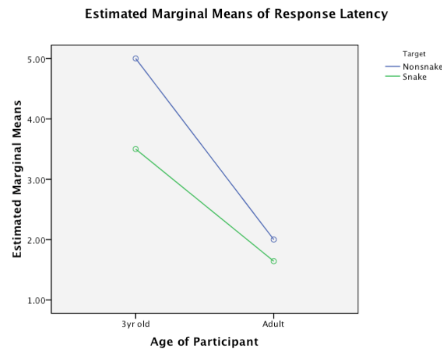
There is a significant positive correlation between attractiveness (ranks) and years served (ranked), r_s (8) = .95. (If you'd ranked the years served from 24 = rank1 to 1 = rank 8, then the *rho* would have been negative.) Thus, the more attractive a politician, the more likely he or she is to have served longer. Keep in mind causal claims. It could be that attractive people are more likely to continue to be (re)elected. It might also be that people who have more power (and have been in the public view more often) are found to be more attractive. It may also be that people with a lot of resources are able to fund campaigns and to hire assistants to make them look more attractive (third variable).

2a. In a study of early ability to detect a fear-relevant stimulus (a snake), LoBue and DeLoache (2008) presented 3-year-old children and adults (Age: 3-year old vs. adult) a series of 3x3 matrices of pictures. The subject's task was to point out a target by touching one of the nine pictures on a touch-screen (Target: either a *snake* among eight non-snake distractors or a non-snake animal, such as a *caterpillar*, among eight snake distractors). Thus, we can think of this study as a 2x2 independent groups design. Below is a partially completed source table that is consistent with their results (Experiment 3). Complete the source table and interpret the results as completely as you can. [15 pts]

Descriptive Statistics

Dependent Variable:Response Latency

Age of Parti...	Target	Mean	Std. Deviation	N
3yr old	Nonsnake	5.0000	.95346	12
	Snake	3.5000	.79772	12
	Total	4.2500	1.15156	24
Adult	Nonsnake	2.0000	.82572	12
	Snake	1.6000	.76396	12
	Total	1.8000	.80434	24
Total	Nonsnake	3.5000	1.76315	24
	Snake	2.5500	1.23500	24
	Total	3.0250	1.58053	48



Dependent Variable:Response Latency

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
Age	72.030	1	72.03	102.9	.000	.700	1.000
Target	10.830	1	10.83	15.5	.000	.259	.970
Age * Target	3.630	1	3.63	5.2	.028	.105	.604
Error	30.920	44	.70				
Corrected Total	117.410	47					

$F_{Max} = \frac{.95346^2}{.76396^2} = \frac{.909}{.584} = 1.56$ **and $F_{\text{Max Crit}}$ = ~5 Thus, you should be not be concerned about heterogeneity of variance and use α = .05.**

Thus, the interaction and both main effects are significant ($p < .05$). To interpret the interaction, you'd first need to compute Tukey's HSD:

$HSD = 3.77\sqrt{\frac{.7}{12}} = .91$ **You might then conclude:**

Three-year olds responded faster to snakes ($M = 3.5$) than non-snakes ($M = 5.0$). However, adults did not respond faster to snakes ($M = 1.6$) than to non-snakes ($M = 2.0$).

2b. Looking at the SPSS output, what are the effect sizes for the three *F*s? How are the effect sizes related to the *F*-ratios and to the power estimates?

For Age, partial eta squared is .7. For Target, partial eta squared is .26. For the interaction, partial eta squared is .11. The *n* for the means in the main effects is larger (24)

than for the interaction (12), so power should be lower for the interaction. Generally speaking, as the effect size goes down, the *F* would decrease unless the *n* increased. In this case, with equivalent *n*, lower effect size for target results in a lower *F* and less power, compared to age. With smaller *n* for the interaction, the decrease in effect size results in an even lower *F* and less power.

2c. The *F*-ratio for the interaction may be significant, but it is smaller than the *F*-ratios for the main effects. Why? How might you increase the *F*-ratio for the interaction?

As noted above, the *F* is lower because the effect size is lower, as is the *n* contributing to each mean. To offset that lower effect size, you might consider increasing sample size (*n*).

2d. Suppose that you re-computed the ANOVA as a single-factor analysis on the Age factor. What would that source table look like?

Source	*SS*	*df*	*MS*	*F*
Age	72.03	1	72.03	73.0
Within (Error)	45.38	46	.99	
Total	117.41	47		

2e. Test the null hypothesis that for the adults, H_0: $\mu = 2.0$. That is, using the information in the Descriptive Statistics table for the problem, compute the appropriate statistic to test the hypothesis that the mean for the sample of adults was drawn from a population with $\mu = 2.0$.

H_0: $\mu = 2.0$ and H_1: $\mu \neq 2.0$
$t_{Crit}(23) = 2.069$

$$t = \frac{1.8 - 2}{.804/\sqrt{24}} = \frac{-.2}{.164} = -1.22$$

Decision: Retain H_0, because $t_{Obt} < t_{Crit}$.
Conclude: It's possible that adults could have come from a population with $\mu = 2.0$.

3. In the first lab in PS 306 this semester (something to look forward to), we collected a number of different academic measures. Below are the results from a correlation analysis of two different SAT scores (Math and Verbal/Critical Reading). First of all, tell me what you could conclude from these results. Then, given an SAT-V score of 600, what SAT-M score would you predict using the regression equation? Given the observed correlation, if a person studied only for the SAT-V and raised her or his SAT-V score, would you expect that person's SAT-M score to increase as well? What would you propose as the most likely source of the observed relationship? [10 pts]

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.438[a]	.192	.173	59.47261

a. Predictors: (Constant), satv

ANOVA[b]

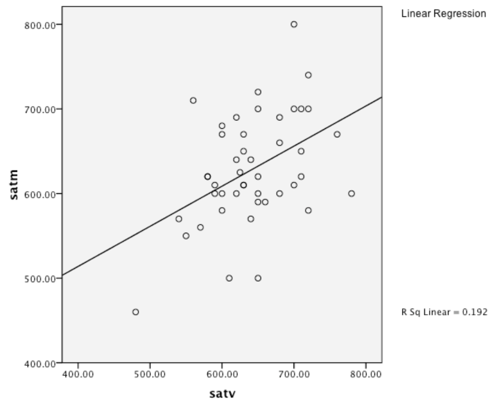
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	36109.377	1	36109.377	10.209	.003[a]
	Residual	152090.623	43	3536.991		
	Total	188200.000	44			

a. Predictors: (Constant), satv
b. Dependent Variable: satm

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	324.301	95.566		3.393	.001
	satv	.474	.148	.438	3.195	.003

a. Dependent Variable: satm



There is a significant positive linear relationship between SAT-M and SAT-V, *r*(43) = .438, *p* = .003.

The regression equation would be: Y = .474X + 324.3. Thus, with SAT-V = 600, I would predict SAT-M = 608.7.

If a student focused on one component (SAT-V), there is no reason to think that the other component (SAT-M) would show a similar increase, even though there is a positive relationship between the two. To think otherwise is to presume a causal relationship between the two.

One reason that the two tests might be positively correlated is due to the native intelligence of the people taking the tests.

4. A variety of research results suggest that visual images interfere with visual perception. In one study, Segal and Fusella (1970) had seven subjects watch a screen, looking for occasional brief presentations of a small blue arrow. On some trials, the subjects were also asked to form a mental image (e.g., imagine a volcano). The results show that subjects made more errors on trials on which they were forming images compared to trials on which they were not forming images. Data similar to Segal and Fusella are as seen below. Analyze the data to see if the trials differed, interpreting the results as completely as possible. [15 pts]

	Errors With Image	Errors Without Image	P
	13	4	17
	9	2	11
	12	10	22
	7	8	15
	10	6	16
	8	6	14
	9	4	13
Sum	68	40	108
Mean	9.71	5.71	
SS	27.43	43.42	

H_0: $\mu_{Image} = \mu_{NoImage}$ H_1: Not H_0

Source	*SS*	*df*	*MS*	*F*
Between	56	1	56	9.9
Within	70.85	12		
Subject	36.9	6		
Error	33.95	6	5.66	
Total	126.85	13		

$F_{Crit}(1,6) = 5.99$ Decision: Reject H_0, $F_{Obt} \geq F_{Crit}$.

Conclusion: People made significantly more errors with image ($M = 9.71$) and with no image ($M = 5.71$).

5. A researcher is interested in investigating the impact of the drug magnesium pemoline (MgPe) on retention of learned material. A group of 16 people is randomly selected from the population of students at a large university. The people are then randomly assigned to one of four conditions: Placebo, 10cc of MgPe, 20cc of MgPe, or 30cc of MgPe. All the people are given the same material to read and four hours later are tested for retention (higher scores indicate greater retention). Analyze the data as completely as you can, including very clear advice about what to do next. [20 pts]

	Placebo	10cc MgPe	20cc MgPe	30cc MgPe
	8	10	11	10
	6	7	6	8
	7	8	8	7
	5	6	9	9
Sum	26	31	34	34
ΣX^2	174	249	302	294
SS	5	8.75	13	5

$F_{Max} = 13/5 = 2.6$ $F_{Max\ Crit} > 20.6$ Thus, there is no concern about heterogeneity of variance, so use $\alpha = .05$. $F_{Crit}(3,12) = 3.49$

Source	*SS*	*df*	*MS*	*F*
Between (Drug)	10.69	3	3.56	1.35
Within (Error)	31.75	12	2.646	
Total	42.44	15		

Decision: Retain H_0, $F_{Obt} < F_{Crit}$.

I would then turn my attention to increasing the power of the study. I might consider increasing the treatment effect (MORE DRUGS!, use 40cc, 50cc, etc.). I might also consider decreasing individual differences (e.g., using only men, using a very specific type of to-be-learned information, using only people of similar weight). I might also consider decreasing random variability by improving the precision of measuring the drug levels, ensuring that the study conditions were similar (and quiet), etc.

6. Researchers are interested in a phenomenon called *hindsight bias*. When you hear people talk about being a Monday Morning Quarterback, they're referring to this hindsight bias. Thus, knowing the results of professional football games on Sunday leads people to be more confident that they would have been able to predict the outcomes before the games (i.e., on Saturday). Psychologists have studied this phenomenon in a number of ways, but one way

is through anagram solving (turning scrambled letters into words). That is, we can present some people with anagrams to solve and measure the time (in minutes) to solve the anagrams (the Worksight Condition). For other people, we can present the anagrams and the solutions simultaneously. Their task is to estimate the time they think it would take someone to solve the anagram (the Hindsight Condition). Thus, to the extent that hindsight bias is present, the estimated times in the Hindsight Condition will be less than the actual solution times in the Worksight Condition. In addition to this factor, we might also examine the extent to which anagram length has an impact on hindsight bias. Thus, the two independent variables would be: TASK (actually solve anagrams vs. with solution present, estimate time to solve) and LENGTH (anagrams would be 4, 6, or 8 letters long). Complete the analysis below, and interpret the results as completely as you can. [20 pts]

Descriptive Statistics

Dependent Variable:"Solution" Time

Task	Anagram ...	Mean	Std. Deviation	N
Hindsight	4-Letter	.6690	.08239	10
	6-Letter	1.3600	.10530	10
	8-Letter	2.0190	.11377	10
	Total	1.3493	.56908	30
Worksight	4-Letter	1.0920	.21049	10
	6-Letter	1.9480	.11545	10
	8-Letter	2.5950	.50322	10
	Total	1.8783	.69891	30
Total	4-Letter	.8805	.26700	20
	6-Letter	1.6540	.32024	20
	8-Letter	2.3070	.46194	20
	Total	1.6138	.68587	60

Dependent Variable:"Solution" Time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[b]
Task	4.2	1	4.2	73.71	.000	.577	1.000
Length	20.42	2	10.21	179.09	.000	.869	1.000
Task * Length	.08	2	.04	.74	.480	.027	.170
Error	3.08	54	.057				
Corrected Total	27.755	59					

$F_{Max} = \frac{.50322^2}{.08239^2} = \frac{.253}{.007} = 35.7$ **and $F_{\text{Max Crit}}$ = 7.8, so you should be concerned about heterogeneity of variance and use α = .01.**

Thus there is no significant interaction and two significant main effects ($p < .01$).

Conclusion:

For Task, people with Worksight have significantly higher scores (M = 1.88) than those with Hindsight (M = 1.35). For Length, you would first need to compute Tukey's HSD.

$$HSD = 3.42\sqrt{\frac{.057}{20}} = .18$$

Thus, 8-letter anagrams had significantly longer (actual and predicted) times (M = 2.31) than times for 6-letter anagrams (M = 1.65) and 4-letter anagrams (M = .88). Six-letter anagrams had longer times than 4-letter anagrams.