Homework 9: Factorial & Repeated-Measures ANOVAs

Homework 9:
One-way between-subjects ANOVA Questions 1-2
Factorial and Repeated Measures ANOVA Question 3-4

SPSS Part 1:

Factorial ANOVA - Two-way ANOVA

1. Enter the following data into SPSS:
   Patients with two kinds of diagnoses were assigned randomly to one of three types of therapy. The effectiveness of the therapy was measured on a 1-to-15 scale (with a higher number indicating greater effectiveness). There were two patients per cell.

	Therapy A	**Therapy B**	**Therapy C**
Diagnosis I	6	3	2
	2	1	4
Diagnosis II	11	7	8
	9	9	10

2. Conduct a two-way ANOVA using **Analyze → General Linear Model → Univariate**. Under **Options** select **Descriptive Statistics**, **Estimates of Effect Size**, and **Homogeneity Tests → Continue.** Conduct post-hoc tests on each of the main effects (if they have more than 2 levels). Under **Post Hoc** select **Scheffé → Continue → OK**.

3. On the output, you will notice that there is no *p*-value for Levene's Test. Just assume that it is greater than .05. Make a table of cell and marginal means and describe the pattern of means. Circle and label the *p*-values for the effects (p_{rows} , $p_{columns}$, $p_{\text{interaction}}$), indicate which ones are significant at the .05 level, and describe the results in terms of the variables in the study.

4. Also, circle and label the **effect sizes** for each main and interaction effect ($R^2_{Rows}, R^2_{Columns}, R^2_{\text{interaction}}$) and indicate what they mean in terms of the variables in the study.

5. Circle the p -values for the post hoc tests, indicate which ones are significant at the .05 level, and describe the results of the tests in terms of the variables in the study. Finally, label the values for SS_{Rows}, $SS_{Columns}$, $SS_{Interaction}$, SS_{Within}, SS_{Total}, df_{Rows}, $df_{Columns}$, $df_{Interaction}$, df_{Within}, df_{Total}, MS_{Rows}, $MS_{Columns}$, $MS_{Interaction}$, MS_{Within}, F_{Rows}, $F_{Columns}$, *and* $F_{Interaction}$.

<u>SPSS Assignment 2</u>:

<u>One-way Repeated-Measures ANOVA</u>

1. Data are in the "Data" folder on your McKenna drive saved as "3RM_group_ANOVA_beer.sav". Open the data file so that you can analyze them using SPSS. These data represent participants who tasted 3 types of beer (e.g., PBR, Natural Ice, and Milwaukee's Best) and rated on a scale from 1 – 9 how much they like them. 1=Not at all and 9=Very much.

2. Conduct a one-way Repeated-Measures ANOVA using **Analyze → General Linear Model → Repeated Measures**. Define the repeated measures factor by renaming it from "factor 1" to "beer" so that the name of your variable is meaningful in terms of your variable of interest. Then indicate the number of levels of the beer variable. Because there are 3 types of beer, enter a 3 into that box. Then click **"Add"** and the box will read "beer(3)". This simply means that you are comparing 3 levels of the beer variable. Now click the "Define" button and move the 3 types of beer into the Within-Subjects Variables box. There is **no** between-subjects variable because all 7 people tasted all 3 types of beer.

3. Under **Options** select **Descriptives** and **Estimates of Effect Size.** The **Post-Hoc** tests for repeated measures cannot be found by clicking on post hoc tests. To do so, take your "beer" factor that is in the Marginal means box and put it in the box labeled **Display Means For:** and underneath that box check **Compare Main Effects**. Because your familywise error rate will be inflated (bigger than you intend) if you are comparing 3 groups or more, you must correct your alpha level, so select **Bonferroni → Continue.** If you are interested in examining whether different *p* values correspond to different pairwise-comparison methods, run the ANOVA again using either the LSD or Sidak confidence interval adjustment measure. Compare the Sidak and LSD those *p* values to those from the Bonferroni procedure. You should see that some or all *p* values resulting from the Bonferroni procedure are greater than the LSD and Sidak procedures, thus illustrating the fact that the Bonferroni procedure is more conservative than several other measures.

   Under **Plots** move your beer factor into the Horizontal (X) axis box and click **"Add" → Continue**. Click **Descriptives** and **Estimates of Effect Size → Continue → OK.**

4. On the output, make a table of cell and marginal means and describe the pattern of means. Circle and label the *p*-value corresponding to the omnibus ANOVA and indicate if it is significant at the .05 alpha level. Then circle and label the *p*-value(s) corresponding to the post-hoc contrasts for which you used the Bonferroni procedure and indicate which ones are significant at the .05 alpha level. Finally, describe the statistical results in terms of beer preferences. Finally, locate and label the values for $SS_{Beer}, SS_{error}, MS_{Beer}, MS_{error}, F_{Beer}$

One-Way ANOVA

1. For each of the following studies, test whether a comparison in which the researcher figures an *F* value of 17.21 would be significant using the Scheffé method.

	Number of Groups	Participants in Each Group	Significance Level
A	5	10	0.05
B	6	10	0.05

2. Describe the following comparisons for ANOVA, and indicate when you would want to use each one.
    a. Bonferroni
    b. Dunnet's
    c. Tukey's
    d. Scheffe

Factorial and Repeated Measures ANOVA

3. The following table of means represents the results of a study using a *factorial* research design. **Assuming that any differences are statistically significant**:

		Age	
		Young	*Old*
Class	*Lower*	20	35
	Upper	25	100

IV 1= Class (Lower or Upper)
IV 2= Age (Young or Old)
DV= Income (in thousands of dollars)

a. Make one bar graph that groups one of the two IVs on the abscissa; your DV should always appear on the ordinate.

b. Make a second, complementary bar graph that groups the other IV on the abscissa. The reason for visualizing the data using these two graphs is because one visual interpretation often affords itself better to interpretation than another.

c. Indicate which potential significant effects (main effects and interactions) might exist in the data set.

d. Describe in words what the pattern of means represents (that is, explain any main or interaction effects or lack thereof). You can describe this meaning aloud to yourself, but make sure you explain how the variables in the study account for the IV-DV relationship.

4. The following table of means represents the results of a study using a factorial research design. Assuming that any differences are statistically significant:

Group		**Gender**	
		Females	*Males*
	Exercisers	2.0	2.5
	Controls	3.1	3.6

IV 1= Group (Exercisers or Control)
IV 2= Gender (Female or Male)
DV= Days sick per month

a. Make one bar graph that groups one of the two IVs on the abscissa; your DV should always appear on the ordinate.

b. Make a second, complementary bar graph that groups the other IV on the abscissa. The reason for visualizing the data using these two graphs is because one visual interpretation often affords itself better to interpretation than another.

c. Indicate which potential significant effects (main effects and interactions), might exist in the data set.

d. Describe in words what the pattern of means represents (that is, explain any main or interaction effects or lack thereof). You can describe this meaning aloud to yourself, but make sure you explain how the variables in the study account for the IV-DV relationship.

Homework 9: Factorial & Repeated-Measures ANOVAS
Answers

1. Is the test significant with a Scheffe test?

a. Overall study's df_{BG} = 5-1= 4
Overall study's df_{WG} = 9+9+9+9+9= 45
Scheffé corrected F = 17.21/4 = 4.30
Overall study's cutoff F (4, 45) = 2.58
4.30 > 2.58, so this comparison is significant.

b. Overall study's df_{BG} = 6-1 = 5
Overall study's df_{WG} = 9+9+9+9+9+9 = 54
Scheffé corrected F = 17.21/5 = 3.44
Overall study's cutoff F (5, 50) = 2.58
(Our df_{BG} = 5 and df_{WG} = 54, but we use df_{WG} = 50 when we look it up in the table)
3.44 > 2.58, so this comparison is significant.

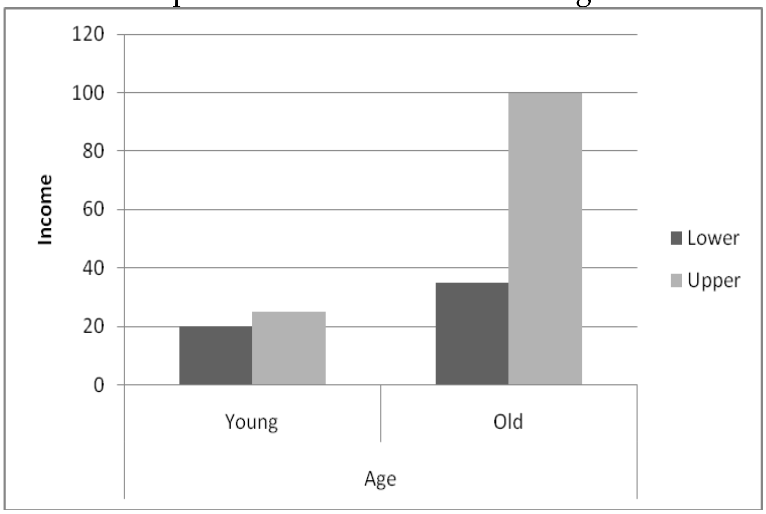
2. ANOVA comparisons question

	Describe Comparison	When to use?
a. Bonferroni	Divide original alpha by number of comparisons you plan to make. The goal is to use a more stringent significance level for each contrast, to keep the overall chance for any one of the contrasts being mistakenly significant low.	Very Conservative The power to detect significant differences for any one comparison is reduced. Use when you have a small number of comparisons.
b. Dunnet's	Simple or complex pairwise comparisons. Compares a control group to the other groups without comparing the other groups to each other	Less conservative Useful when you want to test significance of contrasts compared to a control group mean.
c. Tukey's HSD	Compare all pairs of means, computes a single minimal difference value between means; any comparison between sample means that is greater than that value is significant.	Conservative Useful when you want to examine all possible pairwise comparisons between many means.
d. Scheffé	Calculates a new F value that is used for all comparisons made.	Conservative Useful for many post hoc comparisons, especially complex comparisons.

Factorial ANOVA

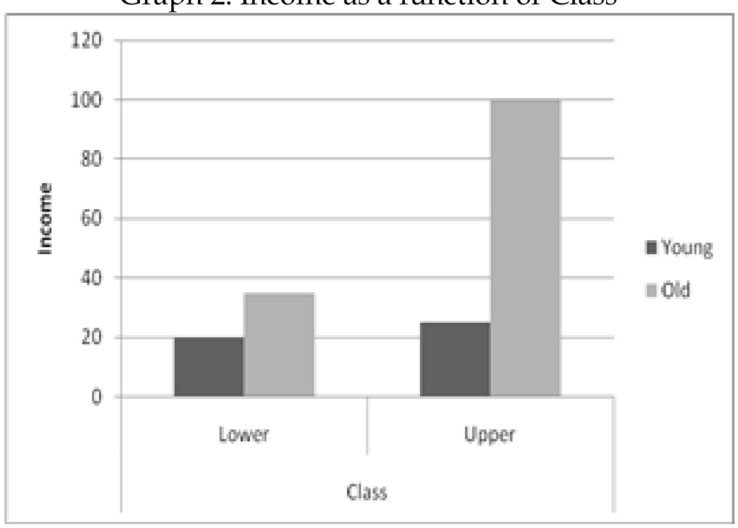
a. Make one bar graph

Graph 1: Income as a function of Age



b. Make a second, complementary bar graph that groups the other IV on the abscissa.

Graph 2: Income as a function of Class



c. Indicate which potential significant effects), might exist in the data set.

In Graph 1 and Graph 2, a *main effect of Age* and a *main effect of Class* appear to exist. An *interaction* between Class and Age effect also seems present.

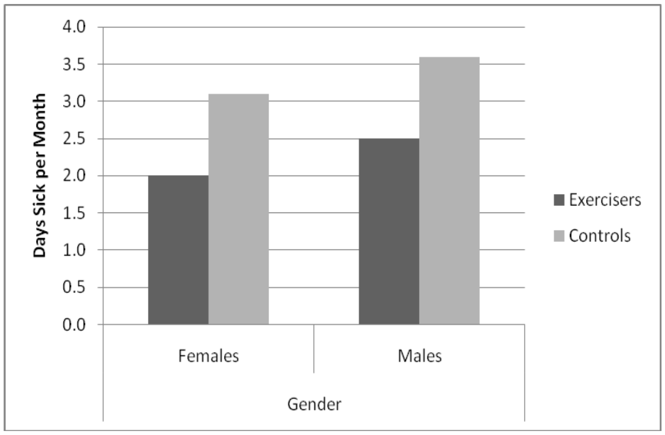
d. Describe in words what the pattern of means represents (that is, explain any main or interaction effects or lack thereof).

There is a main effect of age: income is greater in general for older adults than younger adults. There is a main effect of class: income is greater in general for upper-class than lower class. However, there is an interaction that best explains these data: individuals who are older and in the upper class have incomes higher than what would be expected from the effects of either variable alone.

4. Factorial ANOVA

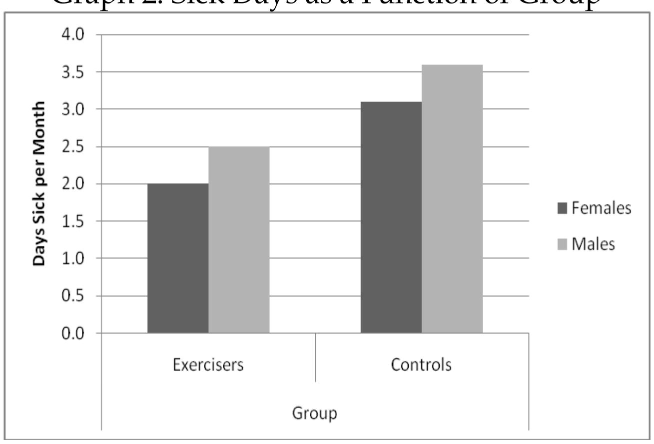
a. Make one bar graph.

Graph 1: Sick Days as a Function of Gender



b. Make a second, complementary bar graph that groups the other IV on the abscissa.

Graph 2: Sick Days as a Function of Group



c. Indicate which potential significant effects (main effects and interactions), might exist in the data set.

In Graph 1 there is a *main effect of Gender* and in Graph 2 there is a *main effect of Group*. However, Group and Gender do not seem to *interact*.

d. Describe in words what the pattern of means represents.

There is a main effect of gender: females generally miss fewer days per month than do males. There is a main effect of group: those who exercise generally miss fewer days per month than controls. Finally, there is no interaction: The effect of exercise does not depend on the gender, and the effect of gender does not depend on the group.

SPSS 1

Factorial ANOVA: Two-way ANOVA

Syntax:

```
UNIANOVA Effectiveness BY Therapy Diagnosis
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /POSTHOC=Therapy(SCHEFFE)
  /PRINT=ETASQ HOMOGENEITY DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=Therapy Diagnosis Therapy*Diagnosis.
```

Table of Cell and Marginal Means

	Therapy A	Therapy B	Therapy C	*Diagnosis Factor Marginal Means*
Diagnosis I	4	2	3	3
Diagnosis II	10	8	9	9
Therapy Factor Marginal Means	7	5	6	*GM* = 6

Initial interpretation of the marginal means in this table suggests that there are main effects of both *type of therapy* and *type of diagnosis*. **If the main effects were statistically significant**, they would suggest that, *overall*, patients receiving Therapy A experienced the greatest effectiveness in therapy, followed by those receiving Therapy C and Therapy B (as seen by marginal means). Patients who received Diagnosis II experienced greater therapy effectiveness than those who received Diagnosis I (as seen by marginal means). Beyond the main effects there is no apparent interaction effect because the differences between the cell means (A: 4 - 10 = -6; B: 2 - 8 = -6; C: 3 - 9 = -6) are all equal. Thus, the effect that the type of diagnosis has on therapy effectiveness does not seem to depend on the type of therapy received. In other words, based on these data, one type of therapy does not seem to be more appropriate for one diagnosis than for the other. An interaction would suggest that a certain therapy is more effective for treating a specific diagnosis.

Here we assume $p > .05$, and that the Levene's Test is non-significant. Actually, sample sizes for this example are too small for calculating this statistic, which is why the analyses have missing values. NOTE: If there are few participants, the F and p values might be missing. Remember, ANOVA requires larger sample sizes. The small sample size used here was only for practice.

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Effectiveness

F	df1	df2	Sig.
.	5	6	.

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Therapy + Diagnosis + Therapy * Diagnosis

$P_{columns}$ P_{rows} $P_{interaction}$

Tests of Between-Subjects Effects

Dependent Variable: Effectiveness

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Pa Squared
Corrected Model	116.000[a]	5	23.200	7.733	.014	.866
Intercept	432.000	1	432.000	144.000	.000	.960
Therapy	$SS_{Columns}$ 8.000	$df_{Columns}$ 2	$MS_{Columns}$ 4.000	$F_{Columns}$ 1.333	.332	.308
Diagnosis	SS_{Rows} 108.000	df_{Rows} 1	MS_{Rows} 108.000	F_{Rows} 36.000	.001	.857
Therapy * Diagnosis	$SS_{Interaction}$ .000	$df_{Interaction}$ 2	$MS_{Interaction}$ .000	$F_{Interaction}$ .000	1.000	.000
Error	SS_{Within} 18.000	df_{Within} 6	MS_{Within} 3.000			
Total	SS_{Total} 566.000	df_{Total} 12				
Corrected Total	134.000	11				

a. R Squared = .866 (Adjusted R Squared = .754)

The above *p*-values interpretation:

$P_{columns}$ is non-significant at the .05 level. We thus fail to reject the null hypothesis that there is no main effect of type of therapy on the effectiveness of the therapy.

P_{rows} is significant at the .05 level. We reject the null hypothesis that there is no difference in the effectiveness of the therapy depending upon the diagnosis. Because we only have two levels of type of diagnosis, we conclude that the patients with Diagnosis II (M = 3) experience the therapy as more effective than the patients with Diagnosis I (M = 3).

$P_{interaction}$ is not significant at the .05 level. This confirms what we found when observing the previous table of marginal means. The effect of the type of diagnosis on therapy effectiveness is not dependent on the type of therapy.

Effect size interpretation:

Partial Eta Squared:

Therapy Partial Eta squared, $\frac{SS_{BG}}{SS_{BG} + SS_{Error}} = \frac{8}{8+18} = \eta_p^2 = .308$. This suggests that 30.8% of the variance in therapy effectiveness (the DV) can be accounted for by type of therapy. The rest of the variability must be due to other variables, systematic or random.

Diagnosis Partial Eta squared, $\eta_p^2 = .857$. This suggests that 85.7% of the variance in therapy effectiveness can be accounted for by diagnosis type. The rest of the variability must be due to other variables, systematic or random.

Interaction Partial Eta squared, $\eta_p^2 = .0$. 0% of the variance in therapy effectiveness is accounted for by the interaction between therapy type and diagnosis type.

However, note that "partial" eta-squared as a measure of effect size is not based on the total variability in your dataset (SS total = 556), but rather part of the variability. This is why adding the effect size measures together will not sum to 1.0 and will actually exceed 1.0.

R squared: As calculated before

Therapy: $R^2 = \frac{(SS_{BG})(df_{BG})}{(SS_{BG})(df_{BG}) + (SS_{WG})(df_{WG})} = \frac{MS_{BG}}{MS_{BG} + MS_{WG}} = \frac{4}{4+3} = .57$.

Diagnosis: $R^2 = \frac{108}{108+3} = .97$

Interaction: $R^2 = \frac{0}{0+3} = .00$

Of course, you really should not look at your post-hoc analyses if your ANOVA is not significant (you can if you decided to run planned comparisons), the table is shown below.

Multiple Comparisons

Effectiveness
Scheffe

(I) Therapy	(J) Therapy	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Therapy A	Therapy B	2.00	1.225	.332	-1.93	5.93
	Therapy C	1.00	1.225	.729	-2.93	4.93
Therapy B	Therapy A	-2.00	1.225	.332	-5.93	1.93
	Therapy C	-1.00	1.225	.729	-4.93	2.93
Therapy C	Therapy A	-1.00	1.225	.729	-4.93	2.93
	Therapy B	1.00	1.225	.729	-2.93	4.93

Based on observed means.
The error term is Mean Square(Error) = 3.000.

The above circled *p*-values for the post hoc comparisons of the different therapy groups are all non-significant. This is in line with our previous observation that there is no significant effect of type of therapy. The groups in Therapy A, Therapy B, and Therapy C, do not differ in their level of effectiveness of therapy.

SPSS
One-way Repeated-Measures ANOVA

Syntax:

```
GLM pbr natice beast
  /WSFACTOR=beer 3 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(beer)
  /EMMEANS=TABLES(beer) COMPARE ADJ(BONFERRONI)
  /PRINT=DESCRIPTIVE ETASQ
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=beer.
```

Selected Output:

Descriptive Statistics

	Mean	Std. Deviation	N
pbr	7.2857	.75593	7
natice	3.7143	1.60357	7
beast	4.7143	1.79947	7

Table of Means

PBR	Natural Ice	Milwaukee's Best
7.29	3.71	4.71

Based on the sample means in the table, PBR received the highest ratings. Second highest average rating was for Milwaukee’s Best. Natural Ice beer had the least favorable ratings. Were they statistically different from one another?

SS_{Beer} MS_{Beer} F_{Beer}

Omnibus ANOVA *p*-value (significant at the .05 alpha level).

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
beer	Sphericity Assumed	47.524	2	23.762	14.896	.001	.713
	Greenhouse-Geisser	47.524	1.964	24.192	14.896	.001	.713
	Huynh-Feldt	47.524	2.000	23.762	14.896	.001	.713
	Lower-bound	47.524	1.000	47.524	14.896	.008	.713
Error(beer)	Sphericity Assumed	19.143	12	1.595			
	Greenhouse-Geisser	19.143	11.786	1.624			
	Huynh-Feldt	19.143	12.000	1.595			
	Lower-bound	19.143	6.000	3.190			

SS_{Error} MS_{Error}

Overall we conclude that there the omnibus *F* tells us that ratings differed for the brands of beer when tested at the .05 alpha level. But what was the nature of the difference? Did people rate more than two type of beer different?

The mean rating of PBR beer is significantly higher than the mean rating of the Natural Ice.

Pairwise Comparisons

Measure:MEASURE_1

(I) beer	(J) beer	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	3.571*	.719	.008	1.208	5.935
	3	2.571*	.649	.022	.436	4.706
2	1	-3.571*	.719	.008	-5.935	-1.208
	3	-1.000	.655	.532	-3.152	1.152
3	1	-2.571*	.649	.022	-4.706	-.436
	2	1.000	.655	.532	-1.152	3.152

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

The mean rating of the PGR is significantly higher than the mean rating of the Milwaukee's Best.

The mean rating of the Natural Ice is not significantly different from the mean rating of the Milwaukee's Best.

Our post-hoc analyses indicate that people rated PBR significantly higher than the other two beers at the .05 level. People rated the other beers equal.