PSYCHOLOGY 300B (A01)

Additional Problems and Answers (answers appear at the end of the document)

1. What is the difference between a rejection criterion and a region of rejection?

2. A researcher is planning to test the effectiveness of a memory enhancing drug on lab animals. One group of animals is to receive the drug and another is to receive a placebo. The researcher decides to use a nondirectional alternative hypothesis and plans to set $\alpha$ at .05.
   (a) How many animals must be used in each group to have power of .80 to detect an effect size of half a standard deviation?
   (b) How many animals would be needed if the researcher were to change $\alpha$ to .10?

3. A researcher plans to test the idea that there is a correlation between self-rated physical attractiveness and attitude ratings of self-confidence. He wants to measure enough subjects so that the power to detect a true correlation coefficient of .5 will be .80, using $\alpha = .05$ and a nondirectional alternative hypothesis. How many subjects will he need?

4. A researcher interested in the effect of alcohol on performance of a perceptual-motor task randomly assigned a sample of 40 subjects to two groups of 20. One group received a drink of alcohol and the other received a drink of a neutral substance prior to performing the task. The following data were obtained. Higher scores indicate better performance.

<table>
<thead>
<tr>
<th></th>
<th>Alcohol</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>22.8</td>
<td>26.3</td>
</tr>
<tr>
<td>$s$</td>
<td>9.2</td>
<td>8.8</td>
</tr>
<tr>
<td>$n$</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

(a) Compute an independent samples $t$ test and verify that the result is not significant with $\alpha = .05$ and a nondirectional alternative hypothesis.

(b) Compute the observed effect size ($d$) using the pooled estimate of variance to get a value for population standard deviation.

(c) Using Cohen's terms, how would you describe this effect size?

(d) Assuming that the observed effect size computed in part (b) is the effect size that truly exists in the population, what is the power of this study to obtain a significant $t$ ratio?

(e) Given the power of the study to detect the effect size computed in part (b), is it safe to conclude that such an effect size does not exist in the population? Explain.

(f) On the basis of this study's result, is it safe to conclude that an effect size ($d$) of 1.0 does not exist in the population? Explain.

5. Which population parameter is estimated by $MS_{error}$ in a between-subjects analysis of variance?

6. The following data are error scores for two groups of subjects on a typing test. One group was tested under conditions of high noise and one was tested under conditions of normal office noise. Carry out an independent samples $t$ test and an ANOVA to test the null hypothesis that noise has
no effect on errors. Use an $\alpha$ of .05 in each case. What is the relationship between the obtained $t$ ratio and the obtained $F$ ratio? What value in the ANOVA computation is equal to the pooled estimate of variance obtained when computing the $t$ ratio?

Noise:  6  9  12  6  7
Normal:  4  7  2  3  4

7. Three groups of subjects were formed by random assignment. Subjects were asked to read a story about a social event, then to describe it to another person. Each subject in group 1 described the event to someone younger than herself or himself, subjects in group 2 each described the event to someone of approximately the same age, and subjects in group 3 each described the event to someone older. The experimenter measured the number of nonverbal behaviors that each subject used while describing the event. The raw data for each group are shown below. Compute an analysis of variance. Decide whether the null hypothesis can be rejected with $\alpha$ at .05. Conduct a Fisher LSD test to compare each pair of groups. Based on the outcome of these comparisons, what conclusions can you draw regarding the relationship between age of audience relative to story teller and the number of nonverbal behaviors used by the story teller?

Group 1:  6  4  3  3
Group 2:  8  5  9  10
Group 3: 10 12 8 14

8. Provide the missing values in the following ANOVA summary table which is based on an experiment involving four groups.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>12.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>45</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>84.12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Here are the means and ANOVA summary table from a study that had two independent variables, A and B. Use the information below to compute a simple effects test of the effect of factor A at each level of factor B. Also compute a $t$ test to examine the main effect of factor A, specifically testing whether the first and second levels of factor A are significantly different. Note that there is an equal number of subjects in each cell of the design. You can figure out the value of $n$.

Condition means:

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>48.6</td>
<td>53.5</td>
</tr>
<tr>
<td>A2</td>
<td>48.3</td>
<td>45.5</td>
</tr>
<tr>
<td>A3</td>
<td>50.0</td>
<td>42.2</td>
</tr>
</tbody>
</table>
10. In a study using a repeated-measures design, a researcher administered a pretest (measuring current level of mental stress) to a group of subjects prior to their receiving a treatment to relieve chronic stress. At each of three time points after treatment was complete (one week, one month, six months), a post-test was given to the subjects, resulting in four stress scores per subject. A repeated-measures ANOVA was computed to determine whether stress levels were affected by the treatment and how long the effect lasted. The mean stress level at each time point and the ANOVA summary table are shown below. It is clear that stress levels were different at the different time points. Use pairwise t tests to answer the following questions:
   (a) Were stress levels lower one week after treatment was completed than they were before treatment began?
   (b) Is there any evidence of a treatment effect 6 months after treatment was completed?
   (c) Is there any evidence that the effect of the treatment declined between one week and six months after treatment?

<table>
<thead>
<tr>
<th>Condition means:</th>
<th>Time of test</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>1 week</td>
<td>1 month</td>
<td>6 months</td>
</tr>
<tr>
<td></td>
<td>34.0</td>
<td>30.6</td>
<td>32.3</td>
<td>35.4</td>
</tr>
</tbody>
</table>

11. An environmental psychologist tests two methods for encouraging people to switch from driving to work each day to taking public transportation. A random sample of 90 commuters is randomly divided into two groups of 45. One group is exposed to method 1 and the other group is exposed to method 2. After 6 months, the psychologist asks each person in the sample whether he or she has taken public transportation to work an average of at least once per week over the past six months. The number of people in each group who responded yes and no is shown below. Conduct a χ² test to determine whether one program was more effective than the other.

<table>
<thead>
<tr>
<th>Response</th>
<th>Method 1</th>
<th>Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>No</td>
<td>33</td>
<td>28</td>
</tr>
</tbody>
</table>

ANOVA summary table: | Source | SS  | df  | MS   | F     | p   |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1,128</td>
<td>2</td>
<td>564.1</td>
<td>7.71</td>
<td>&lt; .05</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>215</td>
<td>1</td>
<td>214.7</td>
<td>2.94</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>AxB</td>
<td>1,634</td>
<td>2</td>
<td>817.1</td>
<td>11.17</td>
<td>&lt; .05</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>17,112</td>
<td>234</td>
<td>73.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20,089</td>
<td>239</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12. Given the following output from R which shows the results of a Bayesian \( t \) test based on an independent-samples design, compute the posterior probability that the alternative hypothesis is correct rather than the null hypothesis. Does the Bayesian method indicate at least positive evidence in favor of the alternative hypothesis?

Bayes factor analysis

[1] Alt., r=0.707 : 18.1500 ±0%

Against denominator:
  Null, mu1-mu2 = 0

Bayes factor type: BFindepSample, JZS
Answers

1. A rejection criterion is the least extreme observed outcome that will allow rejection of the null hypothesis. A region of rejection represents the area of a distribution of outcomes that will allow rejection of the null.

2. (a) The relevant δ value for this case, taken from the power table, is 2.80. So we have
\[ n = \frac{(2.80^2)}{.5^2} = 62.72 \]
which means the researcher needs 63 subjects per group.

(b) Here, δ = 2.50 instead, so we have
\[ n = \frac{(2.50^2)}{.5^2} = 50 \]
which means that 50 subjects per group are needed.

3. From the power table, the required value of δ for power to be .80 in this situation is 2.80. So we have \[ n = \frac{(2.80^2)}{.5^2} + 1 = 31.4 + 1 = 32.4, \]
which means 32 or 33 subjects would be needed.

4. (a) \( t = -1.23 \), not significant.
(b) \( d = \frac{(22.8 - 26.3)}{(9.00)} = -0.39 \) (or .39--it is arbitrary which mean is subtracted from which).
(c) This effect size is between small and medium.
(d) For \( d = .39 \), the statistical effect size for this independent samples \( t \) test is:
\[ \delta = .39 \sqrt{\frac{20}{2}} = 1.23 \]
and from the power table with \( \delta = 1.23 \), power = .22.
(e) One cannot conclude from this study that an effect size of .39 does not exist in the population. This study had very low power to detect such an effect, so even if the effect existed, a study such as this one would be very unlikely to produce a significant \( t \)-ratio.
(f) This study had better power to detect an effect size of one standard deviation: for \( d = 1, \delta = 3.16 \) and power = .88. Therefore, it is reasonable to conclude that an effect of that size likely does not exist in the population, because if such an effect existed, this study most likely would have detected it.

5. Variance of scores in the population (\( \sigma^2 \)).

6. \( t = (8 - 4)/1.414 = 2.83 \); reject the null hypothesis.
ANOVA: \( MS_{\text{group}} = 40, \text{df} = 1; MS_{\text{error}} = 5, \text{df} = 8; F(1, 8) = 8.00 \), reject the null hypothesis
\( t^2 = F; (2.83)^2 = 8 \)
\( MS_{\text{error}} \) = pooled variance estimate in the \( t \)-test.

7. \( MS_{\text{group}} = 49.33, \text{df} = 2; MS_{\text{error}} = 4.44, \text{df} = 9; F(2, 9) = 11.10 \), reject the null hypothesis
Planned comparisons (critical \( t(9) = 2.262 \))
group 1 vs 2:
\[ t = \frac{4 - 8}{4.44 + 4.44} = \frac{-4}{1.49} = -2.68 \]

(significant)

group 1 vs 3:
\[ t = \frac{4 - 11}{1.49} = -4.70 \]

(significant)

group 2 vs 3:
\[ t = \frac{8 - 11}{1.49} = -2.01 \]

(not significant)

When telling a story to someone younger than themselves, subjects use less nonverbal behavior than when telling a story to someone older than or similar in age to themselves.

8. **Source** | **SS** | **df** | **MS** | **F**
--- | --- | --- | --- | ---
Group | 12.05 | 3 | 4.02 | 2.51
Error | 72.07 | 45 | 1.60 |
Total | 84.12 | 48 |

9. Simple effects test for A at each level of B. Critical F ratio for each effect test has 2 and 234 degrees of freedom; from the \( \alpha = .05 \) F table, critical \( F = 3.04 \) (based on 2 and 200 df).

Mean of the three A means at B1 = 49.0; mean of the three A means at B2 = 47.1.

\( SS_{A@B1} = 40[(48.6 - 49.0)^2 + (48.3 - 49.0)^2 + (50.0 - 49.0)^2] = 40(1.65) = 66.0 \)

\( MS_{A@B1} = 66.0/2 = 33.0 \ F_{A@B1} = 33.0/73.1 = 0.45 \) (not significant)

\( SS_{A@B2} = 40[(53.5 - 47.1)^2 + (45.5 - 47.1)^2 + (42.2 - 47.1)^2] = 40(67.53) = 2,701.2 \)

\( MS_{A@B2} = 2,701.2/2 = 1,350.6 \ F_{A@B2} = 1,350.6/73.1 = 18.48 \ p < .05 \)

Conclusion: Factor A has an effect at B2 but not at B1.

\( t \) test for comparing A1 vs. A2, averaging across levels of B. The critical \( t \) ratio has \( df = 234 \), but the closest \( df \) available in the \( t \) table is 100, so critical \( t \) with \( \alpha = .05 \) is \( \pm 1.984 \).

\[ t = \frac{51.1 - 46.9}{\sqrt{73.1 + 73.1}} = \frac{4.2}{1.352} = 3.11 \]

\( p < .05 \)

Conclusion: A1 differs from A2, averaging across levels of B.

10. (a) Compare the pretest and 1-week conditions. The \( df \) for this \( t \) test is the same as the \( df \) for the \( MS_{error} \) in the ANOVA: 57. With \( \alpha = .05 \), the critical \( t \) is 2.009, based on \( df = 50 \) (because 50 is as close as the \( t \) table comes to 57).

\[ t = \frac{34.0 - 30.6}{\sqrt{15.59 + 15.59}} = \frac{3.4}{1.249} = 2.72 \]

\( p < .05 \)

Conclusion: Stress levels were lower one week after treatment than they were in the pretest.

(b) Compare the pretest and 6-month conditions.
Conclusion: There is no evidence for a treatment effect after six months.

(c) Compare the 1-week and 6-month conditions.

\[
t = \frac{30.6 - 35.4}{\sqrt{\frac{15.59}{20} + \frac{15.59}{20}}} = -3.84
\]

\[p < .05\]

Conclusion: There is a treatment effect declined between one week and six months after treatment.

11. The expected frequency for each cell, computed using RC/N is shown in the table in parentheses beside the corresponding observed frequency. The total for each row (R) and for each column (C) are also shown.

<table>
<thead>
<tr>
<th>Response</th>
<th>Method 1</th>
<th>Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>12 (14.5)</td>
<td>17 (14.5)</td>
</tr>
<tr>
<td>No</td>
<td>33 (30.5)</td>
<td>28 (30.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>45</th>
<th>45</th>
</tr>
</thead>
</table>

\[
\chi^2 = \frac{(12-14.5)^2}{14.5} + \frac{(17-14.5)^2}{14.5} + \frac{(33-30.5)^2}{30.5} + \frac{(28-30.5)^2}{30.5} = 1.27
\]

With \(df = 1\) and \(\alpha = .05\), the critical \(\chi^2 = 3.84\), so the test is not significant. Conclusion: there is no evidence that one method is more effective than the other.

12. Given that \(BF_{10} = 18.15\). As shown in the table discussed in class, positive evidence is considered to be a Bayes factor at least as large as 3, so in this case the Bayes factor favoring the alternative hypothesis clearly qualifies as positive evidence.