Final Exam Practice Questions
Psych 3101

Vocabulary

The real exam will ask for definitions of 10-15 terms taken from the final vocabulary list.

Conceptual questions

1. What is the alternative to the repeated-measures ANOVA when data are defined on an ordinal scale?

2. A Mann-Whitney test is used in place of _____ when the dependent variable is defined on _____, or when its distribution is _____ and _____.

3. For a two-tailed paired-samples t-test, there are three equivalent conditions that lead to the conclusion that the two measurements have different means. These conditions are based on the p-value, the t statistic, and the confidence interval for the mean difference score. Write each condition.
p:
t:
CI:

4. What is the difference between a ratio scale and an interval scale?

5. In a regression predicting weight in pounds, the coefficient for hamburgers eaten per week equals 16. Write a sentence describing what this means.

6. 200 new depression drugs are tested, each in a separate experiment. 100 of these drugs work and 100 have no effect, but of course the researchers don't know this in advance. Assuming all experiments are tested using an alpha level of .05, about how many of the 100 drugs that actually work will be concluded to have an effect? Of the 100 drugs that don't work, about how many will be concluded to have an effect? Answer each question with a number or "impossible to determine."

7. A bag contains 24 red marbles and 31 blue marbles, which are all identical except for their color. If a single marble is removed, what is the probability it will be red?

8. As the sizes of the groups in an independent-samples t-test are increased, what will tend to happen to the values of the t statistic and Cohen's *d* computed from the data? Assume the difference between the population means is non-zero (i.e., the alternative hypothesis is correct) and that *t* and *d* are positive.

t:	increase	no change	decrease
d:	increase	no change	decrease

9. When a roulette wheel hits either of the "house" numbers, 0 or 00, every player loses. A fair wheel will have a 1/19 probability of doing this. You play 30 spins and a house number comes up 4 times. You calculate an exact binomial test and come up with a one-tailed p-value of .019. Write a sentence stating as precisely as you can the relationship between the count of 4 house numbers out of 30 spins, the question of whether the wheel is fair or crooked, and the number .019.

10. If variable X_1 has an effect on X_2, and X_2 has an effect on Y, but X_1 has no direct effect on Y, what can you say about the regression coefficient for X_1 in each of the following two situations?

Regress Y on X_1 alone:	near-zero	non-zero
Regress Y on X_1 and X_2:	near-zero	non-zero

Math questions

You don't need to show your work, but I will give partial credit for partial answers.

1. Calculate the correlation between X = [4, 8, 2, 3, 5, 8] and Y = [48, 52, 63, 57, 56, 60]. Keep in mind that these are raw scores.

2. Calculate the t statistic for a one-sample t-test with null hypothesis $\mu = 100$ and data $X = [106, 109, 98, 95, 112]$.

3. 82 people are asked their favorite colors from a list of six options. The observed frequencies are shown below. Calculate the chi-square statistic for the test of whether all five colors are equally popular.

Red	Orange	Yellow	Green	Blue	Purple
20	7	11	18	17	9

4. Calculate the chi-square statistic for the test of whether men and women differ in their distributions of political affiliations, based on the observed frequencies shown below.

Party	Men	Women
Democrat	52	29
Republican	56	21
Green	17	20
Libertarian	15	8

5. Calculate the F statistic for the test of whether there are any reliable differences among the means of the groups [17, 23, 20, 24], [16, 13, 21, 18, 17], and [24, 21, 26, 17, 27]. The degrees of freedom are $df_{\text{treatment}} = 2$ and $df_{\text{residual}} = 11$.

6. Using the data below, a regression of digit span results in regression coefficients of .20 for age in years and -.72 for days of sleep deprivation, with an intercept of 6.18. Calculate the F statistic for the test of whether the regression explains any meaningful variation in the outcome. The degrees of freedom are $df_{\text{regression}} = 2$ and $df_{\text{residual}} = 2$.

Age	Sleep Dep.	Span
20	6	4
15	2	8
17	7	6
31	0	13
22	1	9

7. Calculate a 95% confidence interval for the mean of the sample [4096, 3158, 5520, 5193, 3984, 4272, 3819]. The critical t-value for $\alpha = .05$ with 6 degrees of freedom is 2.45.

R questions

1. What is the result of the following commands?

```
> X = 3:9
> X[4:5]
```

2. What statistic is computed by the following commands?

```
> step1 = sqrt(var(X))
> step2 = step1/sqrt(length(X))
> answer = mean(X)/step2
```

3. What can you say about the result of the following commands?

```
> X = rnorm(100)
> var(X)
```

4. What is the result of the following commands?

```
> t = qt(.05,7,lower.tail=FALSE)
> pt(t,7,lower.tail=FALSE)
```

5. The output below contains three p-values. Write the hypothesis that each leads you to endorse. (Write the meaning of each hypothesis, not just null or alternative.)

```
Analysis of Variance Table

Response: mazeTime
                              Df Sum Sq Mean Sq F value   Pr(>F)
amphetamine                    1  8.566   8.566  8.7789 0.004467
hippocampusRemoved             1  0.547   0.547  0.5604 0.457240
amphetamine:hippocampusRemoved 1  0.992   0.992  1.0169 0.317603
Residuals                     56 54.640   0.976
```

6. Based on the output below, what would you conclude about the null hypothesis that the population mean is 10, using a two-tailed test and an alpha level of .05? Explain your answer.

```
> t.test(memory.score)

   One Sample t-test

data:  memory.score
t = 5.6937, df = 9, p-value = 0.0002967
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  7.468779 17.315885
sample estimates:
mean of x
 12.39233
```

Answers

Conceptual

1. Friedman Test

2. an independent-samples t-test; an ordinal scale; not normal; the sample size is small

3. $p < \alpha$; $|t| > t_{crit}$; confidence interval does not contain 0

4. On a ratio scale, zero is a well-defined value.

5. For every additional hamburger per week, the average person is 16 pounds heavier, assuming all other predictors are held constant.

6. About 5 of the useless drugs will be concluded to have an effect. We can't say how many of the useful drugs will be concluded to have an effect.

7. 24/55 or 43.6%

8. t should increase and d should stay about the same.

9. If the wheel were fair, there would have been a 1.9% chance for a house number to come up four or more times.

10. When X_1 is the only predictor, its coefficient will be non-zero, but when X_2 is included the coefficient for X_1 will be close to zero (i.e., zero except for sampling error).

Math

1.

$M_X = 5$; $M_Y = 56$

$$s_X = \sqrt{\frac{\sum\left(X - M_X\right)^2}{n-1}} = \sqrt{\frac{(4-5)^2 + (8-5)^2 + (2-5)^2 + (3-5)^2 + (5-5)^2 + (8-5)^2}{6-1}} = \sqrt{\frac{32}{5}} = 2.53$$

$$s_Y = \sqrt{\frac{\sum\left(X - M_Y\right)^2}{n-1}} = \sqrt{\frac{(48-56)^2 + (52-56)^2 + (63-56)^2 + (57-56)^2 + (56-56)^2 + (60-56)^2}{6-1}} = \sqrt{\frac{146}{5}} = 5.40$$

$$z_X = \frac{X - M_X}{s_X} = [-.40, 1.19, -1.19, -.79, 0, 1.19]$$

$$z_Y = \frac{Y - M_Y}{s_Y} = [-1.48, -.74, 1.30, .19, 0, .74]$$

$$r = \frac{\sum z_X z_Y}{n-1} = \frac{-.40(-1.48) + 1.19(-.74) - 1.19 \cdot 1.30 - .79 \cdot .19 + 0 \cdot 0 + 1.19 \cdot .74}{5} = \underline{-.22}$$

2.

$M = 104$

$$s = \sqrt{\frac{\sum\left(X - M\right)^2}{n-1}} = \sqrt{\frac{(106-104)^2 + (109-104)^2 + (98-104)^2 + (95-104)^2 + (112-104)^2}{5-1}} = \sqrt{\frac{210}{4}} = 7.25$$

$$\sigma_M = \frac{s}{\sqrt{n}} = \frac{7.25}{\sqrt{5}} = 3.24$$

$$t = \frac{M - \mu_0}{\sigma_M} = \frac{104 - 100}{3.24} = \underline{1.23}$$

3.

$$f^{\text{exp}} = \frac{82}{6} = 13.67$$

$$\chi^2 = \sum \frac{\left(f^{\text{obs}} - f^{\text{exp}}\right)^2}{f^{\text{exp}}} = \frac{(20-13.67)^2}{13.67} + \frac{(7-13.67)^2}{13.67} + \frac{(11-13.67)^2}{13.67} + \frac{(18-13.67)^2}{13.67} + \frac{(17-13.67)^2}{13.67} + \frac{(9-13.67)^2}{13.67}$$

$$= 2.93 + 3.25 + .52 + 1.37 + .81 + 1.60 = \underline{10.48}$$

4.

$f^{\text{obs}}_{\text{men}} = 140,\ f^{\text{obs}}_{\text{women}} = 78,\ f^{\text{obs}}_{\text{dem}} = 81,\ f^{\text{obs}}_{\text{rep}} = 77,\ f^{\text{obs}}_{\text{grn}} = 37,\ f^{\text{obs}}_{\text{lib}} = 23,\ n = 218$

$$f^{\text{exp}}_{\text{men-dem}} = \frac{f^{\text{obs}}_{\text{men}} \cdot f^{\text{obs}}_{\text{dem}}}{n} = \frac{140 \cdot 81}{218} = 52.0 \qquad f^{\text{exp}}_{\text{women-dem}} = \frac{f^{\text{obs}}_{\text{women}} \cdot f^{\text{obs}}_{\text{dem}}}{n} = \frac{78 \cdot 81}{218} = 29.0$$

$$f^{\text{exp}}_{\text{men-rep}} = \frac{f^{\text{obs}}_{\text{men}} \cdot f^{\text{obs}}_{\text{rep}}}{n} = \frac{140 \cdot 77}{218} = 49.4 \qquad f^{\text{exp}}_{\text{women-rep}} = \frac{f^{\text{obs}}_{\text{women}} \cdot f^{\text{obs}}_{\text{rep}}}{n} = \frac{78 \cdot 77}{218} = 27.6$$

$$f^{\text{exp}}_{\text{men-grn}} = \frac{f^{\text{obs}}_{\text{men}} \cdot f^{\text{obs}}_{\text{grn}}}{n} = \frac{140 \cdot 37}{218} = 23.8 \qquad f^{\text{exp}}_{\text{women-grn}} = \frac{f^{\text{obs}}_{\text{women}} \cdot f^{\text{obs}}_{\text{grn}}}{n} = \frac{78 \cdot 37}{218} = 13.2$$

$$f^{\text{exp}}_{\text{men-lib}} = \frac{f^{\text{obs}}_{\text{men}} \cdot f^{\text{obs}}_{\text{lib}}}{n} = \frac{140 \cdot 23}{218} = 14.8 \qquad f^{\text{exp}}_{\text{women-lib}} = \frac{f^{\text{obs}}_{\text{women}} \cdot f^{\text{obs}}_{\text{lib}}}{n} = \frac{78 \cdot 23}{218} = 8.2$$

$$\chi^2 = \sum \frac{\left(f^{\text{obs}} - f^{\text{exp}}\right)^2}{f^{\text{exp}}}$$

$$= \frac{(52-52.0)^2}{52.0} + \frac{(29-29.0)^2}{29.0} + \frac{(56-49.4)^2}{49.4} + \frac{(21-27.6)^2}{27.6} + \frac{(17-23.8)^2}{23.8} + \frac{(20-13.2)^2}{13.2} + \frac{(15-14.8)^2}{14.8} + \frac{(8-8.2)^2}{8.2}$$

$$= 0 + 0 + .9 + 1.6 + 1.9 + 3.5 + 0.0 = \underline{7.9}$$

5.

$$\overline{M} = \frac{17+23+20+24+16+13+21+18+17+24+21+26+17+27}{14} = 20.3$$

$$M_1 = \frac{17+23+20+24}{4} = 21 \qquad M_2 = \frac{16+13+21+18+17}{5} = 17 \qquad \overline{M} = \frac{24+21+26+17+27}{5} = 23$$

$$SS_{\text{total}} = (17-20.3)^2 + (23-20.3)^2 + (20-20.3)^2 + (24-20.3)^2$$

$$+(16-20.3)^2 + (13-20.3)^2 + (21-20.3)^2 + (18-20.3)^2 + (17-20.3)^2$$

$$+(24-20.3)^2 + (21-20.3)^2 + (26-20.3)^2 + (17-20.3)^2 + (27-20.3)^2$$

$$= 222.9$$

$$SS_{\text{residual}} = (17-21)^2 + (23-21)^2 + (20-21)^2 + (24-21)^2$$

$$+(16-17)^2 + (13-17)^2 + (21-17)^2 + (18-17)^2 + (17-17)^2$$

$$+(24-23)^2 + (21-23)^2 + (26-23)^2 + (17-23)^2 + (27-23)^2$$

$$= 130$$

$$SS_{\text{treatment}} = SS_{\text{total}} - SS_{\text{residual}} = 222.9 - 130 = 92.9$$

$$MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{130}{11} = 11.8$$

$$MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} = \frac{92.9}{2} = 46.5$$

$$F = \frac{MS_{\text{treatment}}}{MS_{\text{residual}}} = \frac{46.5}{11.8} = \underline{3.94}$$

6.

$$M_Y = \frac{4+8+6+13+9}{5} = 8$$

$$SS_{\text{total}} = \sum \left(Y - M_Y\right)^2 = (4-8)^2 + (8-8)^2 + (6-8)^2 + (13-8)^2 + (9-8)^2 = 46$$

$$\hat{Y} = b_0 + b_{\text{age}} X_{\text{sleep}} + b_{\text{age}} X_{\text{sleep}}$$

$$= [6.18 + .20\cdot 20 - .72\cdot 6, 6.18 + .20\cdot 15 - .72\cdot 2, 6.18 + .20\cdot 17 - .72\cdot 7, 6.18 + .20\cdot 31 - .72\cdot 0, 6.18 + .20\cdot 22 - .72\cdot 1]$$

$$= [5.86, 7.74, 4.54, 12.38, 9.86]$$

$$SS_{\text{residual}} = \sum \left(Y - \hat{Y}\right)^2 = (4-5.86)^2 + (8-7.74)^2 + (6-4.54)^2 + (13-12.38)^2 + (9-9.86)^2 = 6.78$$

$$SS_{\text{regression}} = SS_{\text{total}} - SS_{\text{residual}} = 46 - 6.78 = 39.22$$

$$MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{6.78}{2} = 3.39$$

$$MS_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{39.22}{2} = 19.61$$

$$F = \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{19.61}{3.39} = \underline{5.78}$$

7.

$M = 4291.72$

$$s = \sqrt{\frac{\sum (X - M)^2}{n-1}}$$

$$= \sqrt{\frac{(4096-4291.72)^2 + (3158-4291.72)^2 + (5520-4291.72)^2 + (5193-4291.72)^2 + (3984-4291.72)^2 + (4272-4291.72)^2 + (3819-4291.72)^2}{7-1}}$$

$$= \sqrt{\frac{3963149}{6}} = 812.73$$

$$\sigma_M = \frac{s}{\sqrt{n}} = \frac{812.73}{\sqrt{7}} = 307.18$$

$$CI_{\text{lower}} = M - t_{\text{crit}} \cdot \sigma_M = 4291.72 - 2.45 \cdot 307.18 = \underline{3539.13}$$

$$CI_{\text{upper}} = M + t_{\text{crit}} \cdot \sigma_M = 4291.72 + 2.45 \cdot 307.18 = \underline{5044.31}$$

<u>R</u>

1. `X` equals `[3 4 5 6 7 8 9]`, so the 4th and 5th components of `X` are <u>`[6 7]`</u>.

2. `step1` is sample standard deviation (s). `step2` is standard error of the mean (σ_M). `answer` is a <u>t statistic</u> (with $\mu_0 = 0$).

3. `X` is a large sample from the Standard Normal distribution. The second line computes the sample variance of `X`, which is an estimate of the population variance. The population variance for the Standard Normal is 1, so the result should be <u>close to 1</u>.

4. The first line defines `t` so that the probability greater than `t` according to a t distribution (with 7 df) is .05. The second line asks for the probability greater than `t` according to the same distribution, so the result is just <u>.05</u>. (It turns out that `t` equals 1.89, but there's no way for you to know that without a computer, and you don't need to know it to answer the question.)

5.
.004467: Amphetamines have an effect on average maze time.
.457240: Removing the hippocampus has no effect on average maze time.
.317603: The effect of amphetamines on average maze time does not depend on whether the hippocampus has been removed.

6. 10 is in the 95% confidence interval, so we would not reject the null hypothesis. (Remember that the p-value given by `t.test()` tests the null hypothesis $\mu = 0$, which is different from what we're testing here, but you can always use the confidence interval to evaluate any other null hypothesis.)