Final Exam
Psych 3101, Fall 14

Vocabulary

1. Random assignment: A process of assigning values of an independent variable, in which every subject has an equal chance of ending up in any group

2. Estimator: A statistic that is used to produce an estimate of some population parameter

3. Mode: The most frequent value in a distribution

4. Histogram: Graph showing frequencies from a distribution

5. Reliability: The degree to which a pattern in the sample is likely to be true in the population

6. Outlier: Score that is different from the rest of a distribution

7. Type I error: Rejection of a true null hypothesis

8. Replication: A process of (actually or hypothetically) repeating an experiment exactly, using a different sample from the same population

9. Independent variable: Variable that's manipulated by the experimenter

10. Grand mean: The average of all scores in all groups in a study

Conceptual Questions

1. Subjects are randomly assigned to two groups, one of which undergoes a week of meditation practice, while the other group spends an equal amount of time in a control task. Afterward, everyone's blood cortisol level is measured.

Is this an experiment or a non-experimental study? Experiment

What is the independent variable? Meditation vs. control

What is the dependent variable? Blood cortisol

2. Give an example of a ratio-scale variable.

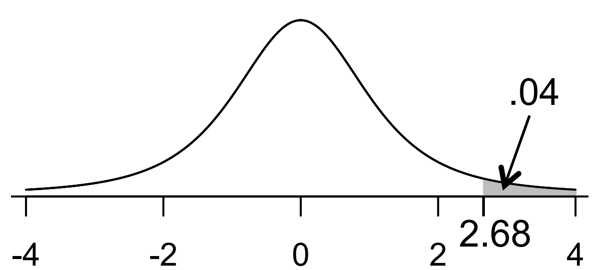
Time, height, number of siblings, etc.

3. Imagine we run a t-test to decide whether the mean of a sample is reliably different from zero. If the confidence interval for the mean is [0.2, 4.8], what can you say about the relationship between p and α?

$p < \alpha$

4. A t-test gives a result of $t = 2.68$. Under a t distribution with the appropriate df, the probability of a value greater than 2.68 is .04. What is the p-value for a <u>two-tailed</u> test?

.08



5. Normal male subjects average 85% on a memory test, and normal female subjects average 89%. Male subjects given midazolam average 65%. If there were no interaction between sex and midazolam administration, what would the average score be for females given midazolam?

69%

6. 40% of house mice prefer cheddar cheese. If species and cheese preference are independent, what percentage of field mice prefer cheddar?

40%

7. Sixty subjects are measured on a nominal variable with 4 possible values (e.g., apple, pumpkin, pecan, blueberry). According to the null hypothesis that all values are equally probable, what is the expected frequency (f^{exp}) for each?

15

8. What kind of hypothesis test should you use to test whether the average (median) of an ordinal-scale variable differs across five groups of subjects?

Kruskal-Wallis

9. If you want to do a single-sample t-test, but the scores are non-normally distributed and your sample size is small, what test should you use instead?

Wilcoxon

Math questions

1. Consider the sample [38,32,35,39]. The critical value you will need below is $t_{\text{crit}} = 3.18$.

(a) Find the sample mean (M) and sample standard deviation (s).

$$M = \frac{\sum X}{n} = \frac{38+32+35+39}{4} = \frac{144}{4} = \underline{36}$$

$$\begin{aligned} s &= \sqrt{\frac{\sum (X-M)^2}{n-1}} \\ &= \sqrt{\frac{(38-36)^2+(32-36)^2+(35-36)^2+(39-36)^2}{3}} \\ &= \sqrt{\frac{4+16+1+9}{3}} \\ &= \sqrt{10} \\ &\approx \underline{3.16} \end{aligned}$$

(b) Calculate the z-score for the first subject.

$$z = \frac{X-M}{s} \approx \frac{38-36}{3.16} \approx \underline{0.63}$$

(c) Calculate the standard error of the mean (σ_M).

$$\sigma_M = \frac{s}{\sqrt{n}} \approx \frac{3.16}{\sqrt{4}} = \underline{1.58}$$

(d) Calculate a 95% confidence interval for the population mean.

$$CI = M \pm t_{\text{crit}} \cdot SE \approx 36 \pm 3.18 \cdot 1.58 \approx \underline{[30.98, 41.02]}$$

(e) Calculate a t statistic for testing the null hypothesis that the population mean (μ) is equal to 33.

$$t = \frac{M - \mu_0}{\sigma_M} \approx \frac{36 - 33}{1.58} \approx \underline{1.90}$$

(f) Do you retain or reject the null hypothesis that μ = 33? Give two (brief) reasons for your answer, one based on your answer to part d and one based on your answer to part e.

Retain it, because $t < t_{\text{crit}}$ and because 33 is in the confidence interval.

2. An experiment using a repeated-measures design tests five subjects in three conditions. These are the data:

	Condition		
Subject	A	B	C
1	18	21	24
2	23	23	29
3	30	28	29
4	20	27	25
5	24	21	33

(a) Calculate the variability explainable by differences among conditions, $SS_{\text{treatment}}$.

$$M_{\text{A}} = 23 \quad M_{\text{B}} = 24 \quad M_{\text{C}} = 28 \quad \overline{M} = 25$$

$$SS_{\text{treatment}} = \sum n\left(M_i - \overline{M}\right)^2 = 5(23-25)^2 + 5(24-25)^2 + 5(28-25)^2 = 20 + 5 + 45 = \underline{70}$$

(b) The variability explainable by individual differences is $SS_{\text{subject}} = 102$, and the total variability is $SS_{\text{total}} = 250$. The relevant degrees of freedom are $df_{\text{treatment}} = 2$ and $df_{\text{residual}} = 8$. Calculate an F statistic for the repeated-measures ANOVA to test for differences among the conditions.

$$SS_{\text{residual}} = SS_{\text{total}} - SS_{\text{treatment}} - SS_{\text{subject}} = 250 - 70 - 102 = 78$$

$$MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{78}{8} = 9.75$$

$$MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} = \frac{70}{2} = 35$$

$$F = \frac{MS_{\text{treatment}}}{MS_{\text{residual}}} = \frac{35}{9.75} \approx \underline{3.59}$$

3. A regression using 4 predictors measured on 30 subjects yields the following sums of squares:
$SS_{\text{total}} = 1000$, $SS_{\text{regression}} = 600$, $SS_{\text{residual}} = 400$.

(a) Calculate R^2 for this regression.

$$R^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}} = \frac{600}{1000} = \underline{0.6}$$

(b) The degrees of freedom are $df_{\text{regression}} = 4$ and $df_{\text{residual}} = 25$. Calculate an F statistic for testing whether the regression explains meaningful variability in the outcome.

$$MS_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{600}{4} = 150$$

$$MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{400}{25} = 16$$

$$F = \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{150}{16} = \underline{9.375}$$

4. Two nominal variables, X and Y, are measured on 100 subjects. Here are the observed frequencies:

	y_1	y_2	y_3
x_1	3	2	5
x_2	12	10	4
x_3	20	8	6
x_4	15	11	4

Find the expected frequency (f^{exp}) for the upper-left cell, i.e. for the combination x_1 & y_1.

$$f^{\text{exp}}_{x_1 \& y_1} = \frac{f^{\text{obs}}_{x_1} \cdot f^{\text{obs}}_{y_1}}{n} = \frac{10 \cdot 50}{100} = \underline{5}$$

The sum of the x_1 row is 10, meaning 10 subjects are x_1. The sum of the y_1 column is 50, meaning 50 subjects are y_1. Overall there are 100 subjects. This means half the subjects are y_1, and so if the variables are independent then half of the 10 x_1 subjects should be y_1, meaning 5 subjects are both x_1 & y_1.

R questions

1. Based on the following output, which hypothesis should you believe (null: $\mu=0$; alternative: $\mu\neq0$), using an alpha level of 5%? There are two parts of the output that you could use to decide between the hypotheses. Describe them both.

```
> t.test(x,mu=0)

      One Sample t-test

data:  x
t = 1.2523, df = 299, p-value = 0.2114
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.04026887  0.18120764
sample estimates:
 mean of x
0.07046938
```

Conclusion (circle): Null Alternative

Reason 1: $p > 5\%$

Reason 2: zero is in the confidence interval

2. Fill in the missing values in the following output from the `anova()` function. (It's easiest to do them in order.)

```
Response: memory
            Df  Sum Sq Mean Sq F value    Pr(>F)
study        1     0.8     0.8       D  0.502537
test         B    48.4    24.2       C  0.000281
study:test   2    57.6    28.8 16.9412  0.000112
Residuals   16    27.2       A
```

A: A = 27.2/16 = 1.7

B: 48.4/B = 24.2 → B = 48.4/24.2 = 2

C: C = 24.2/1.7 ≈ 14.24

D: D = 0.8/1.7 ≈ 0.47

3. Describe in words what this command produces, or when you would use it.

```
> qt(.05, 11, lower.tail=FALSE)
```

The command returns the 95th percentile of a t distribution with 11 degrees of freedom, i.e. the value that has a .05 probability of being exceeded. You would use this as the critical value in a one-tailed t-test with α=5%.

4. Use the regression output below to predict how long an alien species will take to achieve space flight if they have 8 fingers and their planet has a surface area of 10,000,000 square miles.

```
Call:
lm(formula = years.to.space.flight ~ fingers + area.square.miles)

Coefficients:
(Intercept)     fingers     area.square.miles
     35,000        -750               -0.0001
```

$$\hat{Y} = b_0 + b_{\text{fingers}} X_{\text{fingers}} + b_{\text{area}} X_{\text{area}} = 35000 - 750 \cdot 8 - 0.0001 \cdot 10000000 = 35000 - 6000 - 1000 = \underline{28000}$$

5. What statistic is computed by the following commands?

```
> Yhat = b0 + b1*X1 + b2*X2 + b3*X3
> answer = sum((Y-Yhat)^2)
```

SS_{residual} for a regression

6. What statistic is computed by the following command? What kind of hypothesis test might you use it for?

```
> answer = sum((f.obs-f.exp)^2/f.exp)
```

χ^2 for a goodness-of-fit test, such as a multinomial test or test for independence

7. What is the result of the following command?

```
> median(1:5)
```

3