Review for Final

For a detailed review of Chapters 1 – 7, please see the review sheets for exam 1 and 2. The following only briefly covers these sections. The final exam could contain problems that are included in this study guide from Chapters 1 – 7. In addition, review the discussions that I made in class after each midterm.

All tables that are required will be provided. To save size on copying, I may choose NOT to include tables that are not required, though some may still be included. Tables that are provided will be out of the following list: Z table, t table, part of Table 5, Table 9, Part of Table 10, Table 11, Studentized Range Distribution.

Chapter 1

Type of studies: anecdotal, observational, experimental
Random sampling

Chapter 2

Types of variables: quantitative, continuous, discrete, categorical, ordinal, nominal
Histogram and Boxplot: shape description
Measurement of center and dispersion: mean, median, range, quartiles, IQR, variance, and standard deviation; resistant to outliers
Determine outliers
Transformation

Chapter 3

Probability: Frequency interpretation, probability tree
Calculate union (or) and intersection (and) of two probabilities
Conditional probability
Independent and dependent events
Mean and variance (standard deviation) of random variables
Binomial distribution

Chapter 4

Normal distribution
Standardization
Using Z-table to find out percentages and percentiles
Determine if a distribution has a Normal distribution

Chapter 5

Sampling distributions for quantitative variables
Central limited theorem: Normal approximation
Normal approximation to the binomial distribution using the continuity correction

Chapter 6

Difference between SE (standard error) and SD (standard deviation).
Confidence interval for a single population mean, a difference of population means, effect of other variables, approximations and validity conditions
T-distribution: using t-table
Determination of sample sizes.

Chapter 7

Interpretation of the P value.
Perform the hypothesis test using the 9 (8) steps for one- and two-sample t-test and using a directional or nondirectional alternative hypothesis and know the assumptions that are required.
Differentiate between Type I and Type II errors.
Compare the results of a hypothesis test and confidence interval.
The use of effect size to determine sample size.
Wilcoxon-Mann-Whitney Test

For all sections:

1. Be able to determine if the conclusion of the test is an association or causation (MC)

2. Is the method appropriate for the situation (validity of conditions) (MC)

3. State the assumptions that are required for each test.

4. Be able to perform the 9 (8)-step process for all hypothesis tests.

	Steps	Comments
1	State the scientific question to be answered	Use a complete sentence with no symbols
2	Define the parameters of interest	μ, μ_1, etc.
3	State the null hypothesis H_0 and the alternative hypothesis H_A in terms of parameters	The null hypothesis is usually an equality The alternative hypothesis can be $\neq$, > or <
4	State the significance level α.	If not given, then $\alpha = 0.05$.
5	Calculate the test statistic from the data.	The distribution can be z test, t test, χ^2 test, F tests, or WMW. Include other parameters, like df, for the distribution if appropriate
6	Calculate the critical value, rejection region (and P-value).	Find the critical value assuming that the distribution of the test statistic assuming that H_0 is true. This defines the rejection region.
7	Compare the test static to the rejection region or compare the P-value to α.	Is test statistic ≤ or ≥ the critical value? Is the P-value ≤ α?
8	Make a decision about the null hypothesis a) if the test statistic is in the rejection region, or the P-value is smaller than α, state "reject H_0" b) if not, state "do not reject H_0"	The word 'accept' should not be used
9	Form a scientific conclusion based on that decision. if 8a) (reject H_0), then start with "This study provides evidence … " if 8b) (fail to reject H_0), then start with "This study does not provide evidence …" followed by "[(P=x)] at the ___ significance level that " followed by the verbal statement of H_A	Use complete sentences with no symbols (except possibly P) Use similar working in step 3 You must not affirm H_0 It is important to NOT be creative here.

Chapter 8 Comparison of Paired Samples

5. Be able to perform a (1 - α)100% C.I and interpret the result.
   equations

$$\overline{d} = \overline{y}_1 - \overline{y}_2 \qquad SE_{\overline{d}} = \frac{s_d}{\sqrt{n_d}} \qquad \overline{y}_1 - \overline{y}_2 \pm t(df)_{\alpha/2} SE_{\overline{d}}$$

6. Be able to perform a paired sample hypothesis test
   Equations

$$\overline{d} = \overline{y}_1 - \overline{y}_2 \qquad SE_{\overline{d}} = \frac{s_d}{\sqrt{n_d}} \qquad t_s = \frac{\overline{d}}{SE_{\overline{d}}} \qquad df = n_d - 1$$

7. Be able to determine if a paired analysis can be applied or if the independent samples is appropriate (MC).

8. Be able to determine whether to perform a paired analysis or a sign test (MC).

Chapter 9/10: Categorical Data: One -Sample Distributions/Categorical Data: Relationships

9. Be able to calculate, $\tilde{p}$

$$\tilde{p} = \frac{y + 2}{n + 4}$$

10. Be able to use $\tilde{p}$ to calculate the sampling distribution using the binomial theorem with Y. You will need to know the formula to calculate Pr(Y = y) if Y ~ Binomial(n,p)

11. Be able to calculate and interpret the confidence interval for p:

$$\tilde{p} \pm z_{\alpha/2} SE_{\tilde{p}} \; where \; \tilde{p} = \frac{y + 0.5z_{\alpha/2}^2}{n + z_{\alpha/2}^2} \; and \; SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + z_{\alpha/2}^2}} \; (general \; C.I.)$$

$$\tilde{p} \pm 1.96 SE_{\tilde{p}} \; where \; \tilde{p} = \frac{y + 2}{n + 4} \; and \; SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + 4}} \; (95\% \; C.I.)$$

12. Given a guess for the standard deviation, be able to determine an appropriate sample size for a given the half-width, w ($\overline{y} \pm$ w) for the categorical case.

$$n = \frac{z_{\alpha/2}^2 \tilde{p}(1-\tilde{p})}{w^2} - z_{\alpha/2}^2 \; (general \; C.I.)$$

$$n = \frac{4\tilde{p}(1-\tilde{p})}{w^2} - 4 \; (95\% \; C.I.)$$

Note: if you do not have a guess for $\tilde{p}$, use a value of 0.5.

13. Be able to calculate the C.I. for two variables: $\tilde{p}_1 - \tilde{p}_2 \pm z_{\alpha/2} \, SE_{\tilde{p}_1 - \tilde{p}_2}$

$$\tilde{p}_i = \frac{y_i + 1}{n_i + 2} \qquad SE_{(\tilde{p}_1 - \tilde{p}_1)} = \sqrt{SE_1^2 + SE_2^2} = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2 + 2}}$$

14. Be able to perform hypothesis tests for categorical variables (χ^2 tests)
    a) Test statistic:

$$\chi_s^2 = \sum_{i=i}^{k} \frac{(O-E)^2}{E}$$

    Here, k is the total number of cells in the contingency table
    b) Degrees of Freedom
    one-sample: df = categories - 1, $E_i = n(p_i)$
    two-samples: df = (rows - 1)(columns - 1), $E_i = \frac{(row\ total\ )(column\ \ total\ )}{grand\ \ total}$
    c) Note: In table 9: directional: α, non-directional: 2α
    d) For dichotomous variables, be able to determine if the test is directional or non-directional.
    e) Be able to determine if the normality assumption is valid.

15. Be able to determine if a χ^2 test is appropriate for a particular situation (MC)

16. Be able to calculate the relative risk for a particular situation.

Chapter 11 ANOVA

17. Describe why multiple t tests are not appropriate when you have more than 2 samples (MC).

18. Be able to create an ANOVA table (one-way ANOVA only):

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between	I - 1	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	$\frac{SS}{df}$
Within	n. - I	$\sum_{i=1}^{I}(n_i - 1)s_i^2$	
Total	n. - 1	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$	

Note: The actual calculations using the sums of the SS's will not be asked. However, given some of the SS's, be able to calculate the rest of them.

19. Given the ANOVA table, be able to perform a hypothesis test for one-way ANOVA and two-way ANOVA.
    a) test statistic for one-way ANOVA:
    $F_s = \frac{MS(between)}{MS(within)}$, critical value: F_α(dfb, dfw)
    b) test statistics for two-way ANOVA:
    $F_s(interaction) = \frac{MS(interaction)}{MS(within)}$, critical value: F_α(df(interaction), dfw)
    $F_s(Factor\ A) = \frac{MS(between\ A)}{MS(within)}$, critical value: F_α(df(between A), dfw)
    $F_s(Factor\ B) = \frac{MS(between\ B)}{MS(within)}$, critical value: F_α(df(between B), dfw)

20. Be able to state (not perform) how to correct problems with the assumptions (MC).

21. Be able to explain why you would use a randomized block design (MC).

22. Be able to calculate, draw a line graph, interpret and compare the multiple comparison procedures from Bonferroni and Tukey methods (CI only)

$$Bonferroni\text{: } \overline{y}_{i\cdot} - \overline{y}_{j\cdot} \pm t(dfw)_{\alpha/2g}\sqrt{MSW\left(\frac{1}{n_i} - \frac{1}{n_j}\right)}$$

$$where\ g\ is\ the\ total\ number\ of\ comparisons\ = \frac{I(I-1)}{2}$$

$$Tukey\text{: } \overline{y}_{i\cdot} - \overline{y}_{j\cdot} \pm q(I, dfw)_{\alpha}\sqrt{\frac{MSW}{2}\left(\frac{1}{n_i} - \frac{1}{n_j}\right)}$$

Chapter 12 Linear Regression and Correlation

23. Be able to determine when linear regression is the appropriate method (MC).

24. Be able to determine if two samples are positively or negatively correlated given the graph, correlation coefficient or slope (MC).

25. Be able to state the difference between independent (not associated) and uncorrelated (r = 0) (MC).

26. Be able to calculate the correlation coefficient and interpret the meaning (both good and bad) for the number.

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right) = r = \frac{1}{n-1}\sum_{i=1}^{n} z_x z_y$$

where you will be given $\sum_{i=1}^{n} z_x z_y$

27. Be able to perform a hypothesis test for ρ
test statistic is:

$$t_s = r\sqrt{\frac{n-2}{1-r^2}}, df = n-2$$

28. Be able to recognize where the error occurs in linear regression analysis.
$Y = \beta_0 + \beta_1 X + \varepsilon$, $\varepsilon \sim N(0,\sigma)$

29. Be able to calculate the regression line and interpret what the variables mean

$$b_1 = r\left(\frac{s_y}{s_x}\right), \quad b_0 = \bar{y} - b_1\bar{x}$$

30. Be able to calculate the residuals: $e_i = y_i - \hat{y}_i$

31. Be able to state what the least squares lines minimize. (MC)

32. Be able to calculate s_e given SS(resid)

$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

33. Be able to calculate the coefficient of determination and interpret the meaning (both good and bad) for the number.

$$r^2 = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

34. Be able to calculate and interpret the C.I. for β_1

$$SE_{b_1} = \frac{s_e}{s_X\sqrt{n-1}}, \qquad b_1 \pm t(n-2)_{\alpha/2} SE_{b_1}$$

35. Know what factors decrease SE_{b_1} (MC).

36. Be able to perform a hypothesis test for β_1

$$t_s = \frac{b_1}{SE_{b_1}}, df = n - 2$$

37. Be able to state the assumptions for linear regression and know how they can be tested (MC).

Chapter 13: A Summary of Inference Methods

38. Given a research question, determine which method would be most appropriate to use (MC).