Review for Final

Tables Provided (names are from the web site): Normal distribution (like Table A.3), t-distribution (like Table A.5), t curve table (Table A.8), F tables (like Table A.9), Studentized Range (Table A.10).
Note: MC stands for Multiple Choice/True False question.

Chapter 8

1. Be able to determine if a certain error is Type I or Type II.

2. Be able to interpret the meaning of the P-value. (MC)

3. Be able to determine which test to perform (upper tailed, lower tailed, two-tailed and mean or proportion).

4. Be able to use the 7 steps in hypothesis testing.
   1) Identify the parameter of interest and describe it in the context of the problem situation.
   2) Determine the null value and state the null and alternative hypothesis.
   3) Select the significance level α.
   4) Give the formula for the computed value of the test statistic and substitute in the values and calculate P (if required)
   5) Determine the rejection region.
   6) Decide whether H_0 should be rejected and why.
   7) State the conclusion in the problem context.

5. The following are the test statistics:

μ	$Case\ I: z = \frac{\overline{x} - \mu_o}{\sigma/\sqrt{n}}$	$Case\ III: t = \frac{\overline{x} - \mu_o}{s/\sqrt{n}}$
p	$large\ sample\ size: z = \frac{\hat{p} - p_o}{\sqrt{p_0(1 - p_0)/n}}$	

6. The following are the rejection regions for the appropriate tests

	Alternative Hypothesis	z	t
upper tailed	H_a: μ (p) > μ_0 (p_0)	$z \geq z_\alpha$	$t \geq t_{\alpha,n-1}$
lower tailed	H_a: μ (p) < μ_0 (p_0)	$z \leq -z_\alpha$	$t \leq -t_{\alpha,n-1}$
two-tailed	H_a: μ (p) ≠ μ_0 (p_0)	$z \geq z_{\alpha/2}$ OR $z \leq -z_{\alpha/2}$	$t \geq t_{\alpha/2,n-1}$ OR $t \leq -t_{\alpha/2,n-1}$

7. The following are the P-values for the appropriate tests:

	Alternative Hypothesis	z	t
upper tailed	H_a: μ (p) > μ_0 (p_0)	$1 - \Phi(z)$	$t_{n-1}(t)$
lower tailed	H_a: μ (p) < μ_0 (p_0)	$\Phi(z)$	$t_{n-1}(-t)$
two-tailed	H_a: μ (p) ≠ μ_0 (p_0)	$2[1 - \Phi(\|z\|)]$ or $2\Phi(-\|z\|)$	$2t_{n-1}(\|t\|)$

8. Be able to calculate the sample size and Type II error probabilities for Case I upper tail/lower-tail ONLY (no proportions, no two-tailed)

	upper tail	lower tail
β(μ')	$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\frac{\sigma}{\sqrt{n}}}\right)$	$1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu'}{\frac{\sigma}{\sqrt{n}}}\right)$
n	$\left[\frac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2$	

9. Be able to explain the differences between statistical significance and practical significance and their practical meanings.

Chapter 9

10. Be able to determine which test to perform: 2-sample independent, 2-sample paired or proportion (large population). Note: limited questions will be asked concerning the 2-sample proportions.

11. The following are the test statistics: Note: I am only including the test statistic for Case III for the 2 - sample tests for the population mean though you are required to know for Case I also.

μ	$2 - sample^* = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$paired = \frac{\bar{d} - \Delta_o}{s_D/\sqrt{n}}$
p	$large\ sample\ size: z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $where\ \hat{p} = \frac{X + Y}{n_1 + n_2}$	

*ν for the t test will be given if required on the exam.
The rejection regions are the same as for Chapter 8.

12. The following are for the CIs: Note: I am only including the interval for Case III for the 2 - sample tests for the population mean though you are still required to know what the interval is for Case I. In addition, be able to calculate the upper and lower bounds of the appropriate CIs.

μ	$2 - sample: \bar{x} - \bar{y} \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$paired: \bar{d} \pm t_{\alpha/2,n-1}\frac{s_D}{\sqrt{n}}$
p	$large\ sample\ size: \tilde{p}_1 - \tilde{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}_{1_1}\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}}$	

Note: There will be no questions on sample sizes or the probability of Type II errors in the 2 - sample situation.

Chapter 10

13. Be able to determine the factor, the levels of the factor and the response variable.

14. Be able to explain the general procedure of ANOVA and why it works including the derivation of the test statistic. (MC)

15. State the assumptions required for ANOVA to be valid.
   a) Normality.
   b) Constant variance.

16. Be able to determine the one-way ANOVA table via hand (fill in the boxes and know equations). No explicit calculations necessary from the double sums.

Source	df	SS (Sum of Squares)	MS (Mean Square)
Model	I – 1	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\bar{x}_{i.} - \bar{x}_{..})^2$	$\frac{SSM}{dfm}$
Error	n – I	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2$	$\frac{SSE}{dfe}$
Total	n – 1	$\sum_{i=1}^{k}\sum_{j=1}^{n_k}(x_{ij} - \bar{\bar{x}})^2$	

17. Be able to recognize the physical meanings of SSM, SSE, SST, MSM and MSE in ANOVA. (MC)

18. Be able to calculate the test statistic:

$$F_s = \frac{MSM}{MSE}$$

19. Be able to perform the 7-step hypothesis test for one-way ANOVA.
   Rejection region is $F \geq F_{\alpha,\, dfm,\, dfe}$

20. Be able to explain what the problems are with performing a multiple comparison test. (MC)

21. Be able to perform Tukey's multiple comparison analysis indicating the factors that are different, presenting a visual display and determining the final answer.

$$width = q_a \sqrt{\frac{MSE}{n_i}} \; if \; n_i = n_j$$

22. Be able to state the differences and similarities between the 2-sample independent t test and ANOVA and when each would be used and why (MC).

Chapter 12.

23. Be able to state the problems of causation versus association. (MC)

24. Be able to determine the relationship between y and x visually using a scatter plot.

25. Be able to write down the simple linear regression model including the distribution of the error term.

26. Be able to recognize where the randomness error occurs in linear regression analysis. (MC)

27. State the assumptions required for Linear Regression to be valid.
   a) Linearity
   b) Normality of the residuals
   c) Constant variance of the residuals

28. For linear regression calculations, you will be provided with the following: S_{xy}, S_{xx}, SSE, SST, $\bar{x}$ and $\bar{y}$. You do not need to know the ANOVA table, but you do need to know how to calculate MSE and the value of dfe = n – 2.

29. Be able to recognize the physical meanings of SSM, SSE, SST, MSM and MSE in linear regression. You do not need to know the ANOVA table, but you do need to know the value of dfe. (MC)

30. Be able to calculate least squares regression line and determine the point estimates of the parameters

$$\hat{\beta}_1 = b_1 = \frac{S_{xy}}{S_{xx}}, \qquad \hat{\beta}_0 = b_0 = \bar{y} - b_1\bar{x}, \qquad \hat{\sigma}^2 = s^2 = MSE$$

31. Be able to interpret the slope using the rise over run concept.

32. Be able to calculate a point estimate for the response variable and state whether the prediction should be made or not and why.

33. Be able to calculate the residual.

34. Be able to calculate the coefficient of determination, R^2 given the appropriate values

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST}$$

35. Be able to state the interpretation of R^2 and what it doesn't mean.

36. Be able to explain how association can be determined via the slope. (MC)

37. Be able to calculate the confidence intervals for β_1

$$b_1 \pm t_{\alpha/2,n-2}\frac{\sqrt{MSE}}{\sqrt{S_{xx}}}$$

38. Be able to perform hypothesis tests for β_1 with a test statistic of

$$t = \frac{b_1 - \beta_{10}}{\sqrt{MSE}/\sqrt{S_{xx}}}$$

40. State the differences between performing inference on the correlation vs. the slope.
   a) assumptions
   b) confidence interval
   c) null value

Additional Objectives

41. Be able to determine which methodology would be best for a particular specific example. These will be multichapter questions. Possibilities are one-sample mean (z or t test), one-sample proportion, two-sample mean independent (z or t), two sample mean paired (z or t), ANOVA, linear regression. (MC)

Information about material from the past Midterms:

If not otherwise mentioned, the objectives for the chapters are the same from what has been previously posted.

Chapter 1: The only material included is simple questions involving the equations for the sample mean and sample standard deviations. If the sample mean and/or sample deviation are required for another problem (like hypothesis testing or CI), then those values will be provided. There might be questions about the trimmed means which was not included in Midterm 1.

Chapter 3: You will not need to calculate anything from the pmf using summations. That is, you will not have to calculate any expectations or variances by using the definitions from the pmf. You still might need to know the properties of expectations and variances 29b) c), 30 and 31 a) b). You do not need to have the equations for the expectations and variances for the hypergeometric on your cheat sheet, 33 c). Objectives 33a) and b) still might be on the exam.

Chapter 4: Nothing involving QQ plots (normal probability plots). You do not need to have the equations for the uniform distribution on your cheat sheet.

Chapter 6: As in Midterm 2, only possible questions will be T/F or Multiple choice questions. These will be conceptual questions.

Chapter 7: You need to know all of this.