This review sheet is intended to cover everything that could be on the exam; however, it is possible that I will have accidentally left something off. You are still responsible for everything in the chapters covered except anything that I explicitly say you are not responsible for. Therefore, if I left something off of this sheet, it can still be on the exam. There will be no multiple-choice questions. Most of the questions will be like the ones in the homework assignments, and possibly a few definition questions, but I am more likely to ask questions that make you use the definitions rather than recite them.

The review session will be at a time to be determined in class, probably Sunday 4/19.
I will provide the following as on the previous exam. The number before each refers to the lab. Words in small letters, P, s, E, and n, refer to data you will provide and/or a cell. Bold equations were ones where I gave an incorrect version or too long a version in class because I was using Excel 2007 not Excel 2013.
6 NORM.INV(probability, mean, standard_dev) 6FREQUENCY(data_arrays, bins_arrays)crtl-shift-enter
6 RAND() 6 mean $\pm$Z*standard deviation/SQRT(n) where Z = 2.58 (99%), 1.96 (95%), 1.645 (90%)
6 n = (Z*standard deviation/desired error)^2 6 P$\pm$Z*SQRT(P*(1-P)/n) 6 n = π(1-π)*(Z/E)^2
7 left critical value = NORM.S.INV(probability) 7 two tail critical value = ABS(NORM.S.INV(prob/2))
7 P-value left tail test = NORM.S.DIST(Z,cumulative) 7 P-value right tail test = 1- left tail P-value
7 P-value for the two-tail test = IF(Z>0,2*(1-NORM.S.DIST(Z,cumulative)),2*NORM.S.DIST(Z,cum.))
7 left critical value for t is T.INV(probabilty,df) 7 right critical value = T.INV.RT(probability,df)
7 Two tail critical value is T.INV.2T(probabilty,df) 7 left P-value T.DIST(t,df)
7 right P-value T.DIST.RT(t,df) 7 Two tail P-value T.DIST.2T(t,df) 8 $Z=(\bar{X}_1-\bar{X}_2)/SQRT(s_1^2/n_1+s_2^2/n_2)$

Chapter 17: What is meant by **statistical significance, null hypothesis (H_0), alternative hypothesis (H_A or H_1), two-tail test, one-tail test, left-tail test, right-tail test,** and **critical value**? *The null hypothesis is often called the status quo or no difference hypothesis. Therefore, the alternative is things have changed or there is a difference. For a two-tail test, the alternative is X ≠ Y. For a left-tail test, the alternative is X<Y.* What is **α** and what does it tell us? *α is the level of significance we want, which means it is also the probability of making a Type I Error.* What are **Type I** and **Type II Errors**? Be able to explain what they are given a null hypothesis. For example, if the null hypothesis is that the two tests are equally difficult, what is the Type I Error? *Since α is measuring the probability we reject the null hypothesis by pure luck, Type I Errors are we wrongly reject th null hypothesis. Remember, α is the first letter of the Greek alphabet, so it is Type I. Type II errors are the other wrong decision – we fail to reject a false null hypothesis. Since β is the second letter in the Greek alphabet, it is Type II error.* How does getting a bigger sample affect α and β? How else can you reduce one of them? Page 437 has a graph which helps with that. What do the variables in $Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ mean? The formula for **t-test** is basically the same except that you have s_X in the denominator. What is meant by the **P-value** and how is it used? *P is for probability and is the probability of a Type I error. So, if P-value is less than α, then reject the null hypothesis because there is a lower probability than we need.* What are **parametric tests**? They assume the **observations** are **independent, normally distributed**, and the **populations** have **equal variance**. Note, this gives you a **Z distribution**. If you have only one sample, and it is small (under 30) use a t-test. You also use a t-test if the **standard deviation** is not known, or if the population is not normally distributed. However, if the sample size is over 120, you can use a Z-test. If you have **related samples (paired samples)**, then you must use a t-test. If there are **proportions** (percent of the population fitting a criteria), you must use Z. Do not worry about non-parametric tests. See Labs for

more information.

Chapter 18: What is **Pearson's Correlation Coefficient, r**? *It tells you how correlated the variables are and ranges from -1 to 1.* Know how to interpret the value. Note that r assumes a **linear relationship,** and a **bivariate normal distribution**. To tell if r is significant, you run a t-test, $t=r/\text{sqrt}((1-r^2)/(n-2))$. Understand the difference between "X causes Y", "Y causes X", "X causes Y and vice verse", and "some other variable causes both." Can the correlation coefficient tell them apart? Understand the problem of **artificial correlations**. Know how to interpret a **regression** result of $\hat{Y} = \beta_0 + \beta_1 X$ gotten via the **method of least squares** and how to use it to predict a value. Why is the **confidence band** bow tie shaped? Note the confidence band = $\underline{+}t\ \alpha_{/2}s*\text{sqrt}(1+1/10+(X-\bar{X})^2/SS_X)$. Understand the difference between the **explained deviation $\hat{Y}$-$\bar{Y}$, unexplained deviation Y - $\hat{Y}$** and **total deviation Y - $\bar{Y}$.** Similarly, **SS_r (sum of squares of regression**) is the sum of the squares of the explained deviations, **SS_e** (errors) is the sum of the squares of the unexplained deviations, and **SS_t** (total) is the sum of the squares of the total deviation. *However, unlike the deviations, $SS_r + SS_e$ does not = SS_t.* Know how the **degrees of freedom, df,** are found for the regression and the error. How is the **mean square for both regression** and **error (MS_r** and **MS_e** are found. Know that their ratio gives us **F-ratio**.

Chapter 19: Know the differences between the following types of reports, **formal, informal, long, short, technical, management, paper, electronic,** and both paper and electronic. Why does the book recommend you figure out the type of paper, the access to the paper, what ways you are going to depict the data, and your recommendations before you write the report? Why should a short report include: Why it is researched, state the question asked of you (if applicable) otherwise your question, and an answer to the question? Why should it be brief and direct? Why should details be in an **appendix**? For each of the following parts of a long report, know what it includes and why it is in the order listed: **letter of transmittal, title page, authorization letter** if going to a **public organization, executive summary, table of contents, introduction, methodology** (if applicable)**, findings, conclusions** including **recommendations, appendices** (if applicable), and **bibliography.** Know why the introduction includes **problem statement, research objectives,** and **background**? Why does the methodology section (if applicable) include **sampling design, research design, data collection methods, data analysis,** and **limitations** in that order? Why should your write an **outline** before writing the paper? Why should your sentences be written in a positive tone and why should you wait a day or more before submitting it? ***Note that if you want the writing center to read your paper for this class, you must have the rough draft to them 48 hours before they return it to you. Therefore, since the paper is due Friday, you should get it to them on Tuesday so you can revise it Thursday night.***

Labs: I cannot ask you to do the computer work. Read each lab for the types of questions I could ask.

Lab 6 (Workbook Chapters 8 - 9): Know what the **Law of Large Numbers** is. *The bigger the sample, the closer the mean will be to the actual mean.* Know the **Central Limit Theorem**. *The means of samples will approximate normal distribution even if the distribution is not normal.* Know how to interpret the results of the **sampling function** in the data analysis tool. Know what highlighting a bunch of empty cells beside a column called brackets or bin? Then typing "=FREQUENCY(A2:A20,B2:B5) crtl-shift-enter" will give you if your data is in A2 through A20 and the brackets are in B2 through B5. Why do we call **mean, median, mode, variance,** and **standard deviation** types of **point estimates**? What does **95% confidence interval** mean? The confidence interval is $\bar{X}\underline{+}Z*\sigma/\text{SQRT}(n)$ where Z = 2.58 (99%), 1.96 (95%), 1.645 (90%). To find the minimum sample size to get a certain error you need n=(Z*standard deviation/desired error)^2. The confidence interval for a population proportion is

$P \pm Z$*SQRT(P*(1-P)/n). The minimum sample size for a desired error E, is n = π(1-π)*(Z/E)^2. You will be given the equations at the top of the first page. In addition to asking what the terms in bold mean, I might describe what I typed into a spreadsheet and ask you what I am doing and what the result means. *For example, "If I typed in a cell: 5-1.96*3/SQRT(50), then what is this telling me?" The answer is: "With a sample of 50 elements and an average of 5 and a standard deviation of 3, there is a 95% probability (because Z = 1.96) that the real population mean is at least (because the second term is subtracted from the sample mean of 5) the final answer. That is the lower bound of the confidence interval".*

Lab 7 (Workbook Chapter 10): Know what the **null hypothesis H_0** and **alternative hypothesis H_1** or **H_A** are. Know whether you are doing a **one-tail** or **two-tail test,** the desired **level of significance**, and which test you are doing. (Much of that last part is the labs below.) To calculate Z we do ($\bar{X}$-μ)/(σ/SQRT(n)). The **left critical value** = NORM.S.INV(probability). The **right critical value** = -left critical value**.** The **two-tail critical value** = ABS(NORM.S.INV(probability/2)). The **P-value** for the left tail test = NORM.S.DIST(Z,cumulative). The P-value for the right tail test = 1- left tail P-value. The p-value for the two-tail test = IF(Z>0,2*(1-NORM.S.DIST(Z,cumulative)),2*NORM.S.DIST(Z,cumulative)). The t-value is calculated the same as Z except s replaces σ. The left critical value for t is T.INV(probabilty,df) and the right critical value is -left critical value. The two tail critical value is T.INV.2T(probability,df). The left P-value ***T.DIST(t,df)***. The right P-value ***T.DIST.RT(t,df)***. Two tail P-value ***T.DIST.2T(t,df)***, For sample proportion to determine if the sample proportion P is not = π, Z = (P-π)/SQRT(π(1-π)/n). As for questions I could ask, I could tell you what I typed and ask what the answer means as in the last part of the review for Lab 6, or I could give you a situation and ask what test you would do.

Lab 8 (Workbook Chapter 11): For **two independent populations** the Z used to determine if they are different is Z = ($\bar{X}_1$-$\bar{X}_2$)/SQRT($s_1^2/n_1+s_2^2/n_2$). For two independent populations to determine if the **proportions** are different Z = (P_1-P_2)/SQRT(P_C*(1-P_c)*($1/n_1+1/n_2$) where P_c = ($P_1n_1+P_2n_2$)/(n_1+n_2). For two populations where the standard deviations of the populations are not known, but assumed to be the same, you us the "**t-Test: two samples assuming equal variance**". Then for sample 1 and sample 2 you choose cells for those respective samples. For α, we normally use .05. What does that mean? For assumed difference, you put 0. For **paired samples**, use "**t-Test for paired samples.**" Otherwise, everything is the same. I could tell you what I want to test and ask you which test to use. I could also ask a question like the one I wrote at the end of the review for Lab 6.

Lab 9 (Workbook Chapter 12): Know that you do an **F-Test for two sample variances** to figure out if the variances are different. It will ask for Sample 1, Sample 2, *and alpha?* Know to use **ANOVA with a single factor** to determine if one factor affects the different samples and **ANOVA with two factors** if two factors affect the samples. I could ask when you use them or give you the results and ask you to interpret them.

Lab 10 (Workbook Chapter 13): Understand what **correlation analysis** means. On a **scatter diagram** with a **trend line**, **equation**, and **R^2**, be able to explain what the information is telling you. Which is the **dependent variable** and which is the **independent variable**? What does correlation coefficient tell us? (See Chapter 18.) What do **regression analysis** and **least square's regression** mean? (We do more of that in Lab 11. I will give you the scatter diagram with the line and R^2 and ask about it.

Lab 11 (Workbook Chapter 14): I will give you a printout of a multiple regression analysis. You should be able to tell me what the following statistics on the printout tell you: **adjusted R^2, standard error, observations, df, F, significance of F, coefficient, standard deviation, t-Statistic, P-value, Lower 95, Upper 95,** and **intercept.**

Non-graded Homework Assignment #7A to be reviewed with Assignment #7.

1) (20 points) Explain the difference in *audience* and *writing format* between a technical report and a managerial report. Why is the writing different? Which would you use if you were writing a press release? Explain your logic.

2) (5 points) Why do you think the book says to write the *executive summary* last?

3) (10 points) What is in an *authorization letter?* Why is it important to include the sponsor?

4) (5 points) Why should you want at least a day before sending the report?

5) (10 points) What is the difference between an *executive summary* and a *conclusion*?