STAT 201: **Final Exam Review**

Please read!

- The comprehensive final exam will be given on Friday, May 1 from 8:30AM - 10:30AM.
- The final exam will have a similar layout as the previous exams. However, it will be 1.5 to 2 times longer and out of 150 points. Thus, there will be 18 multiple choice questions and 6 free response problems (you will be required to complete) on the exam. On the free response portion of the exam there will be two sections: a mandatory problem section of three problems and a non-mandatory problem section with a total of five problems and you will chose three to complete. If you decide to complete all five, **ONLY** the first three will be graded. You will need to put a big X on the two problems that you do not want graded.
- You should only bring pencils and a calculator to the exam. All tables and a formula sheet will be provided. A SAMPLE formula sheet is attached for your review.
- **IN ADDITION** to the concepts and material covered on previous exams, the final will (ALSO) include the new concepts learned since the third exam which are:
  - The one and two-sample t-tests and confidence intervals (Section 9.4)
  - ANOVA
- For extra practice, there are 20 comprehensive practice problems attached to this review. It is also recommended that you work through your old exams, quizzes, and homework!!!!
- GOOD LUCK!!!!!!

Final Exam Formulae:

- $b = r\left(\frac{s_y}{s_x}\right)$
- $a = \bar{y} - b\bar{x}$
- $Z_c = \frac{\widehat{p} - p_0}{SE}$
- $t_c = \frac{\bar{x} - \mu_0}{SE}$
- $Z_c = \frac{\widehat{p}_1 - \widehat{p}_2}{\text{SE}}$
- $t_c = \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}}$
- $SE = \frac{s}{\sqrt{n}}$
- $SE = \frac{\sigma}{\sqrt{n}}$
- $SE = \sqrt{\frac{p(1-p)}{n}}$
- $SE = \sqrt{\frac{p_0(1-p_0)}{n}}$
- $SE = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$
- $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- $SE = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$
- $SE = \sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

STAT 201 - Final Exam Practice Problems

The following are some extra problems to help you prepare for the Final Exam. Other than these problems, I strongly suggest that you also look over all your homework problems, all your old exams, and the examples we have done in class. You should also review and understand all concepts that we have covered in class.

1. The manager of a local Target store would like to know if costumers spend more money when paying with a credit card than when paying with a debit card. In order to determine if a difference exists, random samples of 20 debit purchases and 20 credit card purchases are taken. Summary information for the purchases is given in the following table.

	Payment Amount		
Payment Type	n	Mean	Standard Deviation
Debit	20	\$47.5214	\$21.4785
Credit	20	\$54.2872	\$36.5472

   (a) Is there enough evidence to support the manager's claim? State the null and alternative hypotheses for this problem.

   (b) Test the hypotheses you stated in (a) by calculating a test statistic and p-value. Make you decision using a significance level of $\alpha = 0.1$. Be sure to draw a conclusion in terms of the problem.

2. The ages of 40 randomly selected college seniors were obtained. The mean age was found to be 23.1 years with a standard deviation of 1.85 years. Calculate and interpret a 99% confidence interval for the true mean age of college seniors.

3. Three sets of five mice were randomly selected to be placed in a standard maze but with different color doors. The different colors are red, green, and black. The response in the time required to complete the maze for each mouse was recorded. The resulting data may be found below. The researcher thinks that color of the door does not affect the mean time it takes to complete the maze. At the 0.05 level of significance, is there sufficient evidence to support the researchers claim?

Color of Door		
Red	Green	Black
9	20	6
11	21	5
10	23	8
9	17	14
15	30	7

   (a) State the null and alternative hypotheses.

   (b) Are the assumptions roughly satisfied?

   (c) Using your calculator, what is the value of the test statistic? What is the p-value?

   (d) Based on the ANOVA results above, is there enough evidence to support the researchers claim?
   If you conclude that the researchers claim is false, state which door colors have different mean completion times and why.

4. Suppose that the length of cockroaches have a mean of 20 millimeters and a standard deviation of 2 millimeters.

   (a) What is the probability that the length of one cockroach exceeds 22 millimeters?

   (b) What is the probability that the average length of a sample of 40 cockroaches exceeds 22 millimeters?

5. Is there a difference in the mean scores of Exam 1 and Exam 2? A random sample of 35 students is selected, and both exam scores are recorded for each student. Consider the following summary information:

	Grade		
Exam	n	Mean	Standard Deviation
Exam 1	35	81.9714	17.1832
Exam 2	35	82.0286	16.5498

Assuming grade differences are normally distributed, calculate and interpret a 90% confidence interval for the true mean difference in exam scores.

6. A professor believes that if a class is allowed to work on an examination as long as desired, the times spent by the students would be approximately normally distributed with mean 45 minutes and standard deviation 6 minutes.

   (a) What is the probability that it takes a student between 50 and 60 minutes to finish the exam?

   (b) The professor would like the end the exam when 96% of the students have turned in their exam. Given this information, how much time should he allot for the exam?

   (c) What is the probability that it takes a random sample of 10 students an average of more than 47.5 minutes to finish the exam?

7. Three brands of batteries are under study. It is suspected that the lives (in weeks) of the three brands are different. Five batteries of each brand are tested with the weeks of life recorded for each battery.

   (a) Using the information given in the problem, complete the following partial ANOVA table.

Source	df	SS	MS	F	p-value
Group		204.10			0.015
Error					
Total		403.30			

   (b) *Are the lives of these brands of batteries different?* Give the null and alternative hypotheses, along with the test statistic, p-value, decision, and conclusion in terms of the problem. (Use $\alpha = 0.01$)

   (c) Based on your decision in part (b), does it make sense to perform multiple comparisons to see which brands are different?

8. After the exam is over, you decide to go to a bar and throw darts (probably with a picture of your professor's head as the bullseye ☺). The probability that you hit the bullseye is 0.4. You find out that they are having a dart-throwing contest. You are given 10 darts and will win \$5 for each bullseye that you hit.

   (a) Define the random variable of interest, the number of trials, and the probability of success.

   (b) How much money do you expect to win in this contest?

   (c) Find the standard deviation.

   (d) What is the probability that you win only \$5 in the contest?

   (e) What is the probability that you win at least \$40 in the contest?

9. During *March Madness*, many people participate in bracket pools to pick the NCAA national champion. In a random sample of 136 brackets, 116 people believe that one of the four number-one seeds will win the NCAA basketball tournament.

   (a) Compute the point estimate of the true proportion of people that believe that a number-one seeded team will win the NCAA men's basketball tournament.

   (b) Find the margin of error for a 90% confidence interval for the true proportion of people that believe that a number-one seeded team will win the NCAA men's basketball tournament.

   (c) Compute the 90% confidence interval for the true proportion of people that believe that a number-one seeded team will win the NCAA men's basketball tournament.

   (d) Interpret the confidence interval you computed in the previous part.

10. An experiment was conducted to compare the mean lengths of time required for the bodily absorption of two drugs, A and B. One hundred people signed up for the study, and were randomly assigned to take one of the two drugs. Each person received an oral dosage of the assigned drug and the length of time (in minutes) for the drug to reach a specified level in the blood was recorded. Summary statistics are in the following table.

	Time		
Drug	n	Mean	Standard Deviation
A	50	27.2	16.36
B	50	33.5	16.92

Can you conclude that there is a difference in mean absorption times for the two drugs? (Use $\alpha = 0.05$)

11. A consumer protection agency believes that the weight of packages of Golden Grahams differs from the advertised amount of 24 ounces. A random sample of 101 boxes is chosen. The sampled boxes had a mean weight of 23.94 ounces with a standard deviation of 0.73 ounces. Does the sample evidence support the agency's claim. (Use $\alpha = 0.05$)

12. An agricultural researcher wishes to see if a kelp extract prevent frost damage on tomato plants. Two similar small plots are planted with the same variety of tomato. Plants in both plots are treated identically, except the plants on plot 1 are sprayed weekly with a kelp extract while the plants in a plot 2 are not. After the first frost in the autumn, the percentage of damaged fruit is determined. For plants in plot 1, 20 of 100 tomatoes on the vine exhibited damage. For plants in plot 2, 36 out of 100 tomatoes on the vine exhibited damage. Let p_1 and p_2 be the actual proportion of all tomatoes of this variety that would experience crop damage under the kelp and no kelp treatments, respectively, when grown under conditions similar to those in the experiment.

    (a) What is the margin of error for a 99% confidence interval?

    (b) What is the 99% confidence interval to estimate the mean difference in proportion of all tomatoes that would experience crop damage under the kelp and no kelp treatments.

    (c) Based on this confidence interval, would you reject the null hypothesis in a test of $H_0 : p_1 = p_2$ vs. $p_1 \neq p_2$?

13. A food company is developing a new breakfast drink, and their market analysts are currently working on preliminary taste-testing studies. To help with their marketing strategy, they were first interested in whether preference for the new product was related to a person's gender. There were 100 male and 100 female volunteers available for the taste-test. Both the males and females tasted the product and rated the flavor on a scale of 1 to 10, 1 being "very unpleasant" and 10 being "very pleasant". The mean rating for males was $\bar{x_1} = 6.4$, with a standard deviation $s_1 = 1.5$. The mean rating for females was $\bar{x_2} = 7.0$, with a standard deviation $s_2 = 2.0$. Let μ_1 and μ_2 represent the mean ratings we would observe for the populations of males and females, respectively, and assume our samples can be regarded as samples from these populations.

    (a) State the null and alternate hypothesis to test differences between mean ratings of men and women in the taste test.

    (b) What is the value of the test statistic?

    (c) What is the p-value for the above test?

    (d) What is your conclusion in terms of the problem for $\alpha = .01$? Write a **complete sentence** to state your conclusion.

    (e) Would a 99% confidence interval for the difference in mean ratings include 0? Explain why or why not.

14. **Is it wrong to cheat** Is it wrong to cheat on an exam or a quiz? This question was asked in a survey done in statistics classes at a large Northeastern University. The results are shown in the table below.

Gender	Wrong	Not wrong	Total
Female	389	60	449
Male	225	62	287
Total	614	122	736

(a) The professor believes that the proportion of female students who think it is wrong to cheat is larger than the proportion of male students who think it is wrong to cheat. Use appropriate statistical symbols to state the null and alternate hypotheses for this problem.

(b) Test the hypotheses you stated in part (a.). Use a significance level of $\alpha = 0.05$. Be sure to draw a conclusion.

15. **Resting Heart Rates** The distribution of resting heart rates is approximately Normal with mean $\mu = 70$ and standard deviation $\sigma = 8$ beats per minute. If we define X= resting heart rate, then $X \sim N(70, 8)$.

(a) Find the probability that the resting heart rate of an individual is less than 60.

(b) Find a value of resting heart rate, R, such that the probability of having a resting heart rate greater than R is 0.10.

(c) What is the probability that the mean resting heart rate for an individual measured over 16 days is above 71 beats per minute?

16. Many companies place advertisements to improve the image of their brand rather than to promote specific products. In a randomized comparative experiment, business students read advertisements that cited either the *Wall Street Journal* or the *National Enquirer* for important facts about a fictitious company. The students than rated the trustworthiness of the source on a 7 point scale. Here is a summary of their ratings. The research question is "Is the *Journal* regarded as the more trustworthy source as compared to the *Enquirer*?"

Source	n	$\bar{x}$	s
Wall Street Journal	66	4.77	1.50
National Enquirer	61	2.43	1.64

(a) Using statistical symbols, write out appropriate null and alternate hypotheses for this situation.

(b) Calculate the value of the appropriate test statistic.

(c) Calculate the corresponding p-value.

(d) State your conclusions to the hypothesis test using complete sentences. Be sure to answer the research question in your conclusion.

17. At a particular location, an environmentalist is interested is the effects of rainfall volume on runoff volume. The following table gives the 15 observations where each measurement is in m^3.

	Rainfall (m^3)	Runoff (m^3)
	5	4
	12	10
	14	13
	17	15
	23	15
	30	25
	40	27
	47	46
	55	38
	67	46
	72	53
	81	70
	96	82
	112	99
	127	100
Mean	53.2	42.867
Std. Dev.	38.347	32.111

(a) What is the value of the correlation coefficient r?

(b) Calculate the equation of the least squares regression line for predicting runoff volume based on rainfall volume. Make sure to write the equation in the terms of the problem.

(c) Interpret the slope of the regression line in terms of the problem.

(d) Use the regression equation to predict the runoff volume for rainfall of 70 m^3.

(e) What proportion of variation in the response variable is explained by the regression line? In view of this value, are you comfortable making the prediction in the previous question? Explain.

18. The data below is the point total (out of 200) for 10 students in a Physics course.

112	116	118	124	128	133	137	142	146	179

Calculate the five number summary. Check for outliers using the outlier test. Draw a modified boxplot.