$\bar{X} = 77$

$M = 78$

10	2
9	11
8	9
7	14
6	7
5	6

ST 305: Exam 3

Spring 2019

By handing in this completed exam, I state that I have neither given nor received assistance from another person during the exam period. I have used no resources other than the exam itself and the basic mathematical functions of a calculator (ie, no notes, electronic communication, notes stored in calculator memory, etc.) I have not copied from another person's paper. I understand that the penalty if I am found guilty of any such cheating will include failure of the course and a report to the NCSU Office of Student Conduct. **I understand that I must show all work/calculations, even if they seem trivial, to get credit for my answers.**

Using your calculator for values from probability distributions like the normal, binomial, or t is OK; however, if you are doing that type of calculation, show your work all the way to the point of plugging the final entries into your calculator.

Name: KEY

ID#: ____________________

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$Z = \frac{X - \mu}{\sigma}$$

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}{n-1}$$

$$b_1 = r\frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$\text{residual} = y - \hat{y}$$

$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A^C) = 1 - P(A)$$

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$\mu_X = \sum x_i p_i$$

$$\mu_{a+bX} = a + b\mu_X$$

$$\mu_{X+Y} = \mu_X + \mu_Y$$

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$

$$\sigma_{a+bX}^2 = b^2\sigma_X^2$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$$

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A \text{ and } B) = P(A)P(B \mid A)$$

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

$$\mu_X = np$$

$$\sigma_X = \sqrt{np(1-p)}$$

$$\hat{p} = X / n$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$P(X = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$

$$m = z^*\sigma/\sqrt{n}$$

$$\bar{x} \pm m$$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$n = \left(\frac{z^*\sigma}{m}\right)^2$$

Simple Linear Regression

$$b_1 = r\frac{s_y}{s_x}; \quad b_0 = \bar{y} - b_1\bar{x}$$

$$e_i = y_i - \hat{y}_i; \quad s^2 = \frac{\sum e_i^2}{n-2}$$

$$b_j \pm t^* SE_{b_j}; \quad t = \frac{b_j}{SE_{b_j}} \quad \text{df} = n - 2$$

$$\hat{\mu}_y \pm t^* SE_{\hat{\mu}}; \quad \hat{y} \pm t^* SE_{\hat{y}}$$

$$\text{SST} = \sum (y_i - \bar{y})^2$$

$$\text{SSM} = \sum (\hat{y}_i - \bar{y})^2$$

$$\text{SSE} = \sum (y_i - \hat{y}_i)^2$$

$$SE_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$SE_{b_0} = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}$$

$$SE_{\hat{\mu}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

$$SE_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Multiple regression changes:

$$s^2 = \frac{\sum e_i^2}{n - p - 1}$$

$$\text{df} = n - p - 1$$

Chapter 7 Stuff

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}, \ \text{df} = n - 1$$

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \ \text{df} = \min(n_1, n_2) - 1$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \ \text{df} = n - 1$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \ \text{df} = \min(n_1, n_2) - 1$$

$$(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \ \text{df} = n_1 + n_2 - 2$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \ \text{df} = n_1 + n_2 - 2$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Definitions. (5 points each) **Clearly define each of the following terms.**

1. p-value:

2. standard error:

3. significance level:

What is wrong? (4 points each)

If a statement is incorrect, edit it to make it correct by replacing the underlined word or phrase with another word or short phrase. If the statement is correct, simply state that it is correct.

4. Decreasing the sample size ~~increases~~ the confidence level of a confidence interval.

   X Does not change

5. A small p-value suggests that the null hypothesis is false.

   CORRECT

6. Increasing the sample size leads to an increase in the ~~significance level~~ of a test.

   X Power

7. Prediction intervals for an individual value are ~~narrower~~ than confidence intervals for mean response.

   X Wider

8. In order to reduce the probability of a Type I error, I might change my significance level from 0.05 to 0.01.

   CORRECT

9. Answer each of the following **in 15 words or less**: (5 points each)

a) Why do we compute a confidence interval (CI)?

to estimate an unknown parameter value

b) Why do we carry out a hypothesis test (HT)?

to answer a yes/no question about the value of a parameter

c) What fact about the HT and CI procedures studied in Chapter 6 makes them unlikely to be useful in most real-life situations?

they assume σ is known, which is unrealistic

10. Provide a mathematical expression for the **general form** of the equations for:

a) computing a confidence interval for an unknown parameter (5 points)

$$\text{estimate} \pm t^* \, SE_{\text{estimate}}$$

b) computing the t-statistic used to test the null hypothesis that a parameter has a specific value of interest (5 points)

$$t = \frac{\text{estimate} - H_0 \text{ value}}{SE_{\text{estimate}}}$$

For the remaining questions on the exam:

- **Show all work, even if the math is trivial**
- **Use confidence level 0.95 or significance level 0.05 unless stated otherwise**
- **If you conduct a hypothesis test, clearly show your null and alternative hypotheses, and provide the p-value for the test.**

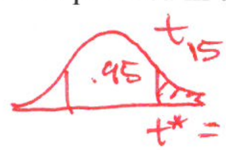
11. A recent study compared the numbers of foxes infected with rabies in two regions of Germany. The scientists did this by dividing the North and South regions into "areas" of the same size, and then taking a SRS of size 16 from the North and another SRS of size 16 from the South. The numbers of infected foxes in each of those 16x2=32 areas were counted. The data on the number of infected foxes per area is summarized in the table below. Use this data to answer parts (a) through (c)

Summary statistics:

Column	n	Mean	Variance	Std. dev.	Std. err.	Median	Range	Min	Max	Q1	Q3
North	16	4.1875	6.4291667	2.5355801	0.63389504	4	10	0	10	2.5	5.5
South	16	2.5	4.9333333	2.2211108	0.55527771	2	9	0	9	1	3.5

a) Find a 95% confidence interval for the mean number of infected foxes per area in the South region. (5 points)

$\bar{x} \pm t^* s/\sqrt{n}$



$2.5 \pm 2.131(.555)$

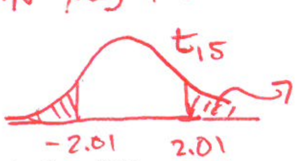
b) Does this sample provide strong evidence that the mean number of infected foxes per area differ between the North and South regions? Carry out an appropriate statistical procedure to answer this question. (5 points)

$H_0: \mu_N - \mu_S = 0$

$H_A: \mu_N - \mu_S \neq 0$

$t = \frac{\bar{x}_N - \bar{x}_S}{\sqrt{\frac{s_N^2}{16} + \frac{s_S^2}{16}}} = \frac{4.1875 - 2.5}{\sqrt{\frac{6.43}{16} + \frac{4.93}{16}}} = \frac{1.6875}{\sqrt{.71}} = 2.01$

.84



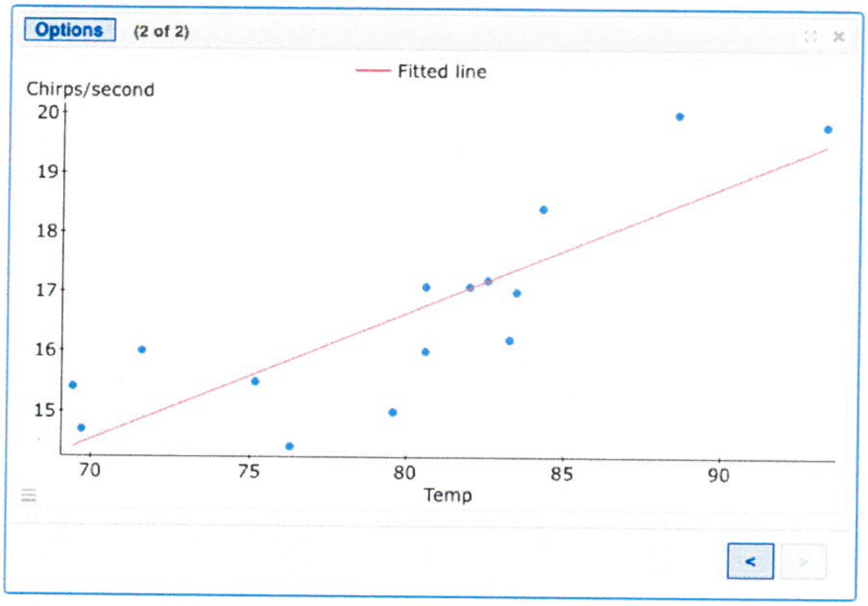
b/w .025 & .05

$\Rightarrow$ p-val is between .05 & .1. Fail to reject H_0.

c) Estimate the difference between the mean number of infected foxes per area in the North and the mean number of infected foxes per area in the South using a 95% confidence interval. (5 points)

$(\bar{x}_N - \bar{x}_S) \pm t^* \sqrt{\frac{s_N^2}{16} + \frac{s_S^2}{16}} \Rightarrow 1.6875 \pm .84\, t^*$

12. In *The Song of Insects* Dr. GW Pierce investigated the relationship between the speed of chirps made by the striped ground cricket and temperature. The following StatCrunch output summarizes a statistical analysis of 15 crickets from that study. Use it to answer the questions on the next page (you can tear this page out if that is helpful). Some values have been replaced with XXXX, and you will need to find those values to answer some questions. You may use any values from the output without showing how they were computed.



****

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	95% L. Limit	95% U. Limit
Intercept	-0.3091419	3.1085839	13	-7.0248291	6.4065453
Slope	0.21192498	0.038711226	13	0.12829446	0.2955555

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	28.287325	28.287325	XXXXXX	0.0001
Error	13	12.270006	XXXXXXX		
Total	14	40.557331			

Predicted values:

X value	Pred. Y	s.e.(Pred. y)	95% C.I. for mean	95% P.I. for new
74	15.373306	0.34291827	(14.632477, 16.114136)	(13.147561, 17.599052)

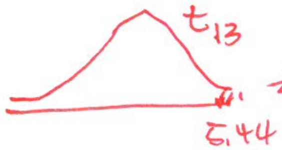
****

a) **Using a t-test**, test to see if there is a positive linear relationship between temperature and the number of chirps/second. (5 points)

$H_0: \beta_1 = 0$

$H_A: \beta_1 > 0$

$t = \dfrac{b_1 - 0}{SE_{b_1}} = \dfrac{.212}{.039} = 5.44$



$< .0005 \Rightarrow$ Reject H_0 & conclude $\beta_1 > 0$

b) **Using an F-test**, test to see if there is a linear relationship (either positive or negative) between temperature and the number of chirps/second. (5 points)

$H_0: \beta_1 = 0$

$H_A: \beta_1 \neq 0$

$F = MSM/MSE = 28.29/.944 = \underline{29.97}$

$= \dfrac{SSE}{DFE} = \dfrac{12.27}{13} = .944$

p-val $\approx .0001$

$\Rightarrow$ Reject H_0, conclude $\beta_1 \neq 0$

c) Write an equation for a line that can be used to predict the number of chirps/second for a cricket when the temperature is 80 degrees. (3 points)

$\hat{y} = b_0 + b_1$ $\qquad \hat{y} = -.309 + .212x$

d) Provide a 95% confidence interval for the slope of the line relating the number of chirps/second to temperature. (4 points)

$(.128, .296)$

e) You want to find a pet cricket for your apartment, which you keep at 74 degrees. Give a 95% interval predicting the speed of its chirps. (4 points)

$(13.15, 17.6)$

f) What is the value of the correlation coefficient between temperature and chirps/second? (4 points)

$r^2 = \dfrac{SSM}{SST} = \dfrac{28.29}{40.56} = .697 \Rightarrow r = .835$

(positive b/c b_1 is positive)