Name:__________________ Section: __________

Statistics 311 **Final Exam** 11 December 2007

Please read the following instructions carefully. **Do not turn the page and start the exam until you are told to begin.**

You may have the formula card that comes with the textbook (or a copy of the same). You may also have one sheet (double-sided) of 8.5 x 11 paper with notes (handwritten or typed). You may use a calculator, although be sure to show your work.

There are 3 multi-part short answer problems worth a total of 98 points; you get 1 point for your name and 1 point for your quiz section. In addition, there are 5 multiple choice problems and 5 true/false problems. The multiple choice and true/false problems are worth one point each for up to 10 points of extra credit. Point values for each question or part of a question are designated in parentheses at the beginning of the problem.

GOOD LUCK!

Problem 1 (6+10+4+6): Researchers of the Marine Science Institute of the University of Texas at Austin noticed a difference in the seagrass beds in Corpus Christi Bay (CCB) versus those in Lower Laguna Madre (LLM). Researchers know that ammonium is essential to seagrass growth. As such, they decided to compare sediment (pore water) ammonium concentrations (in μM) at the two locations. Summary statistics based on collected samples are given below*.

$$\bar{x}_{\text{CCB}} = 115.1 \quad s_{\text{CCB}} = 79.4 \quad n_{\text{CCB}} = 51$$
$$\bar{x}_{\text{LLM}} = 24.3 \quad s_{\text{LLM}} = 10.5 \quad n_{\text{LLM}} = 19$$

a) Write down the null and alternative hypotheses to test whether, on average, ammonium concentrations at CCB are greater than those at LLM. Be sure to appropriate label your parameters.

$H_0: \mu_{CCB} = \mu_{LLM}$ or $\mu_{CCB} - \mu_{LLM} = 0$ versus $H_1: \mu_{CCB} > \mu_{LLM}$ or $\mu_{CCB} - \mu_{LLM} > 0$

b) Construct the appropriate test statistic for your test in part a) using an un-pooled procedure.

$$t_{obs} = \frac{(\bar{x}_{CCB} - \bar{x}_{LLM}) - 0}{\sqrt{\frac{s_{CCB}^2}{n_{CCB}} + \frac{s_{LLM}^2}{n_{LLM}}}} = \frac{(115.1 - 24.3) - 0}{\sqrt{\frac{6304.36}{51} + \frac{110.25}{19}}} = \frac{90.8}{\sqrt{123.615 + 5.803}} = \frac{90.8}{\sqrt{129.418}} = \frac{90.8}{11.376} = 7.98$$

c) Using a 5% significance level, what is the critical value for this test?

$t_{.05(1),18} = 1.734$

d) Using your test statistic and critical value, state the conclusion from your hypothesis test in relation to the original question.

Since 7.98 > 1.734, reject the null and conclude that sample evidence supports that the mean ammonium concentration at CCB is greater than that the mean concentration at LLM.

*The data set is from exercise 10.64 in Neil Weiss' *Introductory Statistics*, 8th edition, 2008, Pearson Education, Inc. Data are from a paper by Lee and Dunton, *Marine Ecology Progress Series*, 2000, Vol. 196, pp. 39-48.

Problem 2 (5+4+4+8+4+2+2+4+3+8+4 points): Many relationships can be characterized by spatial proximity. One such relationship is average monthly temperature and latitude. Average monthly temperatures for January, April, and August were calculated from daily temperature readings for 20 major cities throughout the United States. A plot of average August temperatures in degrees Fahrenheit (*y*) versus degrees latitude (*x*) shows that a linear model may be an appropriate method for describing the relationship between the two variables. A simple linear regression analysis was performed using Excel. The regression summary output is shown below. In addition, $\bar{x} = 37.95$ and $SS_{xx} = 592.95$.

SUMMARY OUTPUT

Regression Statistics	
R Square	**0.610**
Standard Error	**4.622**
Observations	**20**

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	**113.584**	**7.277**	**15.608**	**6.615E-12**
Latitude	**-1.006**	**0.190**	**-5.301**	**4.864E-05**

a) Write down the equation for the regression line.

$\hat{y} = 113.584 - 1.006x$

b) Interpret the estimated slope value in the context of the problem.

The estimated slope is -1.006 degrees F/degree latitude. For each one degree increase in latitude, the average August temperature in U.S. cities is expected to decrease by 1.006 degrees F, on average.

c) Does the estimated *y*-intercept have any practical meaning in the context of the problem? **Briefly explain.**

$x = 0 \Rightarrow 0^\circ$ latitude, which is located at the equator. While many think that temperatures at the equator are high, the *y*-intercept really has no practical meaning in the context of this problem because the data were collected across U.S. cities, which does not include any observations at the equator.

d) Calculate a 95% confidence interval for the estimated slope in the regression equation. ***Interpret the CI in the context of the problem.***

$$-1.006 \pm t_{.05(2),18}(0.190) \Rightarrow -1.006 \pm 2.101(0.190) \Rightarrow -1.006 \pm 0.399$$

$$\Rightarrow (-1.405, -0.607)$$

The average change in mean Aug. temperatures per degree Lat falls between -1.405 and -0.607 degrees F. The interval does not include 0 ➔ slope is statistically different than 0 at the 5% significance level.

e) Write down the null and alternative hypotheses to test that the slope is less than zero.

$H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 < 0$

f) What is the *t*-score that is reported by Excel to test the hypotheses you specified in part e)?

t = -5.301

g) What is the *P*-value for the test you specified in part e)? NOTE: the *P*-value reported in the Excel output is for a two-tailed test.

$$P = \frac{4.864 \times 10^{-5}}{2} = 2.432 \times 10^{-5}$$

h) Based on the Excel output what do you conclude, in the context of the problem, for the test specified in part e)? Use $\alpha = 0.01$.

Since $P = 2.432 \times 10^{-5} < \alpha = 0.01$, reject the null. Conclude that there is sufficient sample evidence to suggest that the population slope for the rate of change in mean Aug temperatures/degree Lat is less than zero.

i) What is the expected value for average August temperature for U.S. cities located at 40 degrees latitude?

$$(\hat{y} | x = 40) = 113.584 - 1.006(40) = 113.584 - 40.24 = 73.34 \ ^\circ\text{F}$$

j) Calculate a 98% confidence interval for the average August temperature for all U.S. cities located at $x_0 = 40$ degrees latitude. ***Interpret the interval in the context of the problem.***

$$73.34 \pm t_{.02(2),18}(4.622)\sqrt{\frac{1}{20} + \frac{(40 - 37.95)^2}{592.95}} \Rightarrow 73.34 \pm 2.552(4.622)\left(\sqrt{.057}\right)$$

$$\Rightarrow 73.34 \pm 11.795(0.2387)$$

$$\Rightarrow 73.34 \pm 2.815 \Rightarrow (70.56, 76.16)$$

We are 98% confident that the mean of average Aug. temperatures for U.S. cities located at 40 degrees Lat will fall between 70.56 and 76.16 degrees F.

k) In the context of the problem, summarize the meaning of the R Square value reported in the Excel output?

$R^2 = 0.61 \Rightarrow$ 61% of the variation in average Aug. temperatures can b explained by knowing a city's latitude.

Problem 3 (15+5+4 points): In a survey of 190 college students, 134 students said that they believe there is extraterrestrial life.

a) Construct a 99% confidence interval for the proportion of all college students that believe there is extraterrestrial life.

$$\hat{p} = \frac{134}{190} = 0.705; \quad \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \Rightarrow 0.705 \pm (2.575)\sqrt{\frac{(.705)(.295)}{190}}$$
$$\Rightarrow 0.705 \pm (2.575)(.0331)$$
$$\Rightarrow 0.705 \pm (2.575)(.085) \Rightarrow (0.62, 0.79)$$

b) Does the confidence interval support the notion that the proportion of college students that believe there is extraterrestrial life exceeds 68%? Briefly explain.

No, since the lower bound of the CI includes 0.62 which is less than 68%.

c) If you had constructed a 90% interval instead of, would you expect this interval to be wider or narrower than the 99% interval? Briefly explain.

The 90% CI would be narrower since we are willing to be less confident. $z_{\alpha/2}$ would be smaller, making E smaller.

EXTRA CREDIT

EC1 (1 point each): Answer the following multiple choice questions by circling the letter with the correct response:

a) Which of the following quantities does NOT affect the width of a confidence interval for a population proportion?

A. The confidence level.

B. The sample proportion.

C. The sample size.

D. The population size.

b) When constructing a confidence interval for a population mean μ, for a sample of n = 29 observations and σ unknown, the correct formula to calculate the margin of error is

A. $E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

B. $E = t_{\alpha(2),28}\left(\frac{s}{\sqrt{n}}\right)$

C. None of the above.

e) The likelihood that a statistic would be as extreme or more extreme than what is observed based on the sample data is called a

A. statistically significant result.

B. test statistic.

C. significance level.

D. *P*-value.

f) The smaller the *P*-value, the

A. stronger the evidence against the alternative hypothesis.

B. stronger the evidence for the null hypothesis.

C. stronger the evidence against the null hypothesis.

D. None of the above.

g) Which statement is NOT true about hypothesis tests?

A. Hypothesis tests are only valid when the sample is representative of the population for the question of interest.

B. Hypotheses are statements about the population represented by the samples.

C. Hypotheses are statements about the sample (or samples) from the population.

D. Conclusions are statements about the population represented by the samples.

EC 2 (1 point each). Indicate whether each of the following statements is True or False.

False A confidence interval for the average value of *y* at a given value $x = x_0$, formed as part of a linear regression analysis, will always be wider than the corresponding prediction interval estimate for a new individual value of *y* at the same value $x = x_0$.

True One assumption for simple linear regression is that the epsilons are normally distributed with a mean of zero.

True Power is equal to $1 - \beta$.

False The level of significance associated with a significance test is the probability of committing a Type II error.

True A statistical hypothesis test is always conducted under the assumption that the null hypothesis is true.