STAT 205 **Name:________________________________**
Fall 2006
Final Exam

$$\tilde{p} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+Z^2_{\alpha/2}}} \quad \text{where} \quad \tilde{p} = \frac{Y + \frac{1}{2}Z^2_{\alpha/2}}{n+Z^2_{\alpha/2}}$$

$$(\tilde{p}_1 - \tilde{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}} \quad \text{where} \quad \tilde{p}_1 - \tilde{p}_2 = \frac{Y_1+1}{n_1+2} - \frac{Y_2+1}{n_2+2}$$

$$\sum_{i=1}^{n}\frac{(O_i - E_i)^2}{E_i}$$

$$\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-y)}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$\overline{y} - b_1\overline{x}$$

$$\sqrt{\frac{SS(resid)}{n-2}}$$

$$\frac{S_{Y|X}}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}}$$

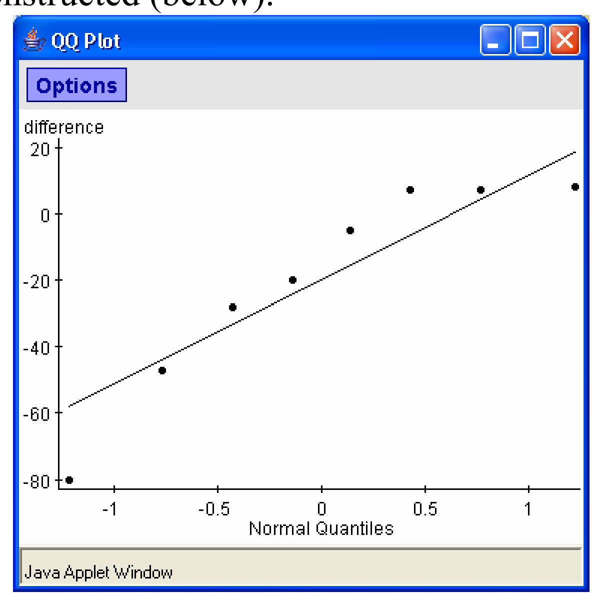
$$\frac{[\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})]^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}$$

This exam is worth a total of 120 points.
Part I: Answer eight of the following nine questions. If you complete more than eight, I will grade only the first eight. Five points each.

1) State the definition of a P-value.

2) The Central Limit Theorem says that for any i.i.d. random sample, $Y_1, Y_2, \ldots, Y_n$ where $E[Y_i] = \mu$ and $E[(Y_i-\mu)^2] = \sigma^2$, then as $n\rightarrow\infty$ the distribution of the sample mean is _______________ with mean, ____, and variance, _______(note, I'm asking for variance here – not standard deviation).

3) The compound *m*-chlorophenylpiperazine (mCPP) is thought to affect appetite and food intake in humans. In a study of the effect of mCPP on weight-loss, eight moderately obese men were given mCPP in a double-blind, placebo controlled experiment. Some of the men took mCPP for two weeks, then took nothing for two weeks (a "washout period"), and then took a placebo for two weeks. The rest of the men took the placebo during the first two weeks, then had a two week washout period, then took mCPP for the final two weeks. The The men were asked to rate how hungry they were at the end of each two-week period (hunger rating for mCPP period – hunger rating for placebo period). A QQplot of the differences was constructed (below).



(Circle the correct answer.) The use of the **<u>independent samples t-test</u> / <u>dependent samples t-test</u> / <u>signtest</u> / <u>Wilcoxon-Mann-Whitney test</u>** would be appropriate here.

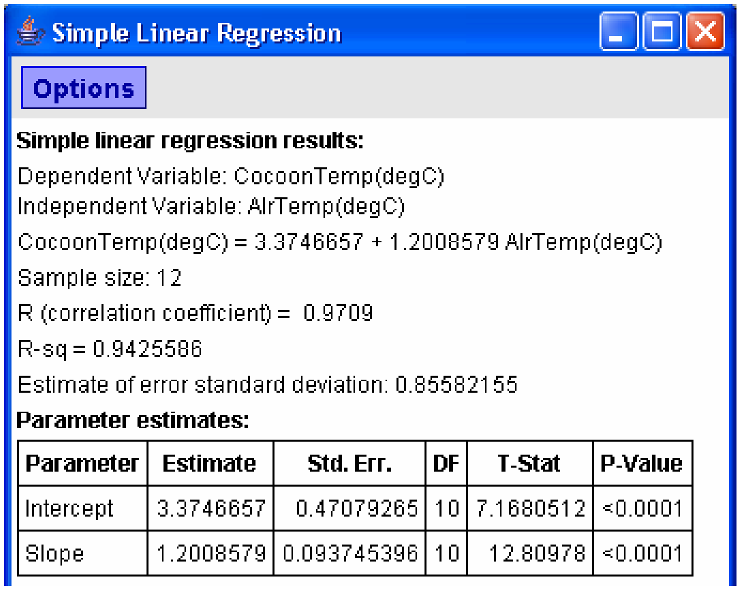
4) State the assumptions we need to check (in terms of the errors) for simple linear regression before using the regression model for inference.

5) (Circle the correct answer.) If the assumptions of a regression model for predicting *y* from *x* are met, and we do not reject the null hypothesis that $\beta_1=0$, then we conclude that *x* **can** / **cannot** be used to predict *y*.

6) (Circle the correct answer.) If the assumptions of a linear regression model for predicting *y* from *x* are met and we do reject the null hypothesis that $\beta_1=0$, then we **may** / **may not** conclude that *x* causes *y*.

Questions 7,8, and 9 are on the next page!

A simple linear regression was performed to relate the cocoon temperature (Y in ^{o}C) to the outside air temperature (X in ^{o}C). Use this portion of DoStat's output to answer the following 3 questions.



Simple Linear Regression

Options

Simple linear regression results:

Dependent Variable: CocoonTemp(degC)
Independent Variable: AirTemp(degC)

CocoonTemp(degC) = 3.3746657 + 1.2008579 AirTemp(degC)

Sample size: 12

R (correlation coefficient) = 0.9709

R-sq = 0.9425586

Estimate of error standard deviation: 0.85582155

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	T-Stat	P-Value
Intercept	3.3746657	0.47079265	10	7.1680512	<0.0001
Slope	1.2008579	0.093745396	10	12.80978	<0.0001

7) What is the r^2 for this regression? __________

8) Interpret the value of the coefficient of determination (r^2) in the context of the setting.

9) Referring to the default test DoStat performs for the β_1, fill in the blanks:

H_0: ____

H_A: ____

P-value: ____________________________________

Part II: Answer every part of the next three problems. Read each question carefully, and show your work for full credit.

1) It is common folk wisdom that drinking cranberry juice can prevent urinary tract infection in women. In 2001, the British Medical Journal reported the results of a Finnish study in which two groups of 50 women were monitored for these infections over 6 months. One group drank cranberry juice every day and the other group did not drink cranberry juice. At the end of the study, the number of women who had urinary tract infections (one or more) was 8 for the cranberry juice group and 18 for the group that did not drink cranberry juice.

1a) (20 points) Construct a 95% Agresti-Caffo confidence interval for the difference in the proportion of urinary tract infections for these two groups.

1b) (5 points) Interpret the interval you just computed in part (a).

2) (20 points) Research has indicated that the stress produced by today's lifestyles results in health problems for a large proportion of society. An article in the *International Journal of Sports Psychology* (July – Sept. 1990) evaluated the relationship between physical fitness and stress. 549 employees of companies participating in the Health Examination Program offered by Health Advancement Services (HAS) were classified into three groups of fitness levels: good, average, and poor. Each person was tested for signs of stress. The following table reports the results.

	Signs of Stress	***No Sign of Stress***	Total
Poor Fitness	38	204	**242**
Average Fitness	28	184	**212**
Good Fitness	10	85	**95**
Total	**76**	**473**	**549**

Test whether there is a significant association between fitness level and stress at the 0.05 significance level. I've numbered the steps for you, please write the appropriate step next to the appropriate number.

(1)

(2)

(3)

(4)

(5)

(6)

3) Twenty plots, each of equal area, were randomly chosen in a large field of corn. For each plot, the plant density (number of plants in the plot) and the mean cob weight (g of grain per cob) were observed. The results are given in the table.

Plant Density *X*	**Cob Weight *Y***	**Plant Density *X***	**Cob Weight *Y***
137	212	173	194
107	241	124	241
132	215	157	196
135	225	184	193
115	250	112	224
103	241	80	257
102	237	165	200
65	282	160	190
149	206	157	208
85	246	119	224

Preliminary calculations yield the following results:

$$\bar{x} = 128.05 \qquad \bar{y} = 224.1$$

$$\Sigma(x_i - \bar{x})^2 = 20{,}209.0 \qquad \Sigma(y_i - \bar{y})^2 = 11{,}831.8$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -14{,}563.1$$

$$SS(\text{resid}) = 1{,}337.3$$

3a) (7 points) Calculate the least-squares regression line using X=plant density as the predictor variable and Y=cob weight as the response.

3b) (7 points) Calculate the residual standard deviation ($S_{Y|X}$).

3c) (7 points) Give an estimate of the mean and standard deviation of cob weight at a plant density of 145.

3d) (7 points) Calculate a 95% confidence interval for β_1 (slope of the true line).

3e) (7 points) Interpret the interval you just computed in part (d).