ST 372 Final Exam ~ Practice

Printed Name: __________________________

Signature: ____________________________ (Honor Code: By signing, I agree to neither give help to nor receive help from another student on this exam.)

Instructions:

1. There are 17 questions. Please make sure you have received all 12 exam pages.

2. Questions 1 - 13 are Multiple Choice or Short Answer. Circle the correct answer, or fill in the blank.

3. The remaining questions involve problem solving. Show all work to receive full credit. Use $\alpha = 0.05$ unless otherwise specified.

4. Read each question carefully. Make all work legible.

5. Turn in exam. You may keep your note sheet.

Multiple Choice / Short Answer (2 points each)

1. Suppose that $X$ is a random variable with mean 20 and variance 10 and that $Y$ is a random variable with mean 30 and variance 15. Assume also that $X$ and $Y$ are independent. Then the difference $X - Y$ has mean and variance
   a. 10 and -5
   b. 10 and 25
   c. -10 and -5
   d. -10 and 25

2. For simple linear regression, we assume that the random error term $e$
   a. is normally distributed
   b. has an expected value equal to 0
   c. has a constant variance $\sigma^2$
   d. All of the above

3. For simple linear regression, what method is used to estimate the intercept and slope parameters?
   a. Least squares
   b. Central limit theorem
c. Method of moments
d. None of the above

4. Suppose that $H_0 : \mu = 90$ and $H_a : \mu > 90$ for $n = 90$. The test statistic value is $z = 1.37$. What is the p-value?
a. 0.085
b. 0.171
c. 0.915
d. None of the above.

5. In a statistical test of hypothesis, how is a Type II error made?
a. By rejecting the null hypothesis when the alternative hypothesis is true
b. By rejecting the null hypothesis when the null hypothesis is true
c. By failing to reject the null hypothesis when the alternative hypothesis is true
d. By failing to reject the null hypothesis when the null hypothesis is true

6. If the form for a large-sample confidence interval for $\mu$ is given by $\bar{x} \pm 1.88(s/\sqrt{n})$, what is the level of confidence?
a. 92%
b. 96%
c. 90%
d. 94%
e. 98%

7. Suppose that the probability of a type II of a statistical test is 20%. Then, the probability of a type I error is
a. .80
b. .60
c. .40
d. .20
e. Can not tell
8. Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size $n$ from a distribution with parameter $\theta$. If the Maximum Likelihood Estimator (MLE) of $\theta$ is $3\bar{X}$, what is the MLE of $(\theta + 5)^2$?
   a. $3\bar{X}$
   b. $3\bar{X} + 5$
   c. $(3\bar{X} + 5)^2$
   d. None of the above

9. In simple linear regression analysis, the vertical deviations $y_1 - \hat{y}_1, y_2 - \hat{y}_2, \ldots, y_n - \hat{y}_n$ from the estimated regression line are called the
   a. Residuals
   b. Fitted values
   c. Predicted values
   d. Sums of Squares
   e. None of the above

10. Value of the sample correlation will always be between _______ and _______

11. In simple linear regression, the slope $\beta_1$ represents the expected change in ________
    for one unit _________ in predictor.

12. In a specific regression analysis, we obtain two intervals for a specific value of $x = 25$: (159.31, 172.78) and (164.28, 167.81). Among these, ____________ is the confidence interval for $E(y|x = 25)$ and ____________ is the prediction interval for a new $y$.

13. In a single factor ANOVA, suppose you found that the means of the groups you are comparing are different. Now, in a further analysis, which method would you use to find out exactly which two means are different from each other? ________________
    [Hint: not a t-test]

Turn to next page
14. To detect the presence of harmful insects in farm fields, one researcher put up boards covered with a sticky material and examined the insects trapped on the boards. The researcher wants to know which colors best attract insects. She placed 6 boards of each of the 4 colors (blue, green, white and yellow) at random locations in a field of oats and measured the number of beetles trapped.

<table>
<thead>
<tr>
<th>color</th>
<th>n</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue</td>
<td>6</td>
<td>14.833</td>
<td>2.182</td>
</tr>
<tr>
<td>green</td>
<td>6</td>
<td>31.167</td>
<td>2.574</td>
</tr>
<tr>
<td>white</td>
<td>6</td>
<td>16.167</td>
<td>1.537</td>
</tr>
<tr>
<td>yellow</td>
<td>6</td>
<td>47.167</td>
<td>2.774</td>
</tr>
</tbody>
</table>

We want to perform a ANOVA on this data set.

- What is the factor and how many levels does it have?

- Clearly define the parameters of interest and write down $H_0$ and $H_a$ in terms of these parameters.

- Compute mean square for treatments (MSTr).

- Compute mean square for error (MSE).
• Perform an F-test (compute the test statistic, specify degrees of freedoms and your decision)

• Write down the ANOVA table for this problem.
15. Data were collected on $x = $ ulna length (cm) and $y = $ height (cm) for males. Answer the following questions based on the accompanying output and summary statistic.

(a) Identify the dependent and independent variables.

(b) Obtain the equation of the estimated regression line, and interpret its slope in the context of the problem.

(c) Calculate the coefficient of determination ($R^2$) for this model. Interpret this value.
(d) Note from the output that SST = S_{yy} = 626.00. Based on your answer in (c), give an estimate of the error variance $\sigma^2$.

(e) Compute the standard error for $\hat{\beta}_1$ and construct a 95% confidence interval. Interpret.

(f) Perform a model utility test.

(g) Write down the ANOVA table for this model.
(h) What assumption are you making to perform this test in (f) and to construct the interval in (e)?

(i) Based on the output given above, are these assumptions reasonable? Clearly explain why or why not.
16. Seven people measured their heart rate (HR) before using a treadmill and after walking on a treadmill for 10 minutes. The mean HR before and after treadmill use are 83 bpm and 93 bpm, respectively. The standard deviation of the difference (among these individuals) was measured to be 6 bpm. Assume that the heart rates are normally distributed.

(a) Let \( \alpha = 0.01 \). Perform a hypothesis test to determine whether there is a significant difference HR before and after treadmill use by completing the following steps.

(i) Parameter of interest:

(ii) Null hypothesis:

(iii) Alternative hypothesis:

(iv) Name of the test and formula of the test statistic:

(v) Value of the test statistic (show calculation):

(vi) Rejection region and p-value:

(vii) Conclusion:
(b) Construct a 90% confidence interval for the true mean difference in HR.

(c) Interpret the interval.

(d) Now use the confidence interval to determine whether the mean difference in HR is statistically significant at $\alpha = 0.1$ or not. Carefully explain how you reached your conclusion.
17. Suppose $X_1, \ldots, X_n$ constitute a random sample of size $n$ from a Poisson distribution with pmf

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}.$$ 

(a) Find the maximum likelihood estimator of $\lambda$. 

(b) Calculate the variance of this estimator. You may use the fact that the variance of Poisson random variable is \( \lambda \).

(c) Construct the MLE of the variance in part (b).