Math 2200
Final Exam
Spring 2009

Put your answers on the answer card. The answers offered have been rounded off; take that into account when looking for the correct answer. The true-false questions are two points each. The other eleven questions are eight points each. Two points for free.

All significance tests are to be done with $\alpha = .05$

PART I, TRUE-FALSE

1. Bernoulli trials are sequences of independent experiments with two outcomes, often called success and failure, with the probability of success being the same for each experiment.
   a. True
   b. False

2. The regression line for a set of points always goes through the point $(\bar{X}, \bar{Y})$. ($\bar{X}$ and $\bar{Y}$ are the average values of the $X$ and $Y$ coordinates of the data points.)
   a. True
   b. False

3. If you are studying data described by a Normal Model and you do not know $\sigma^2$, the variance of the model, you can use $s^2$, the estimate of $\sigma^2$ based on the data. However this commits you to using statistics based on the $t$-distributions.
   a. True
   b. False

4. Chi-square tests for goodness of fit can be one tailed or two tailed.
   a. True
   b. False

5. If the events $A$ and $B$ are independent then the conditional probabilities $P(B|A)$ and $P(A|B)$ will be equal.
   a. True
   b. False
6. Given the data

<table>
<thead>
<tr>
<th>x</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

Use linear regression of \( y \) on \( x \) to predict the \( y \) value that would be associated with a new observation which had \( x = 5 \).

a. 6.03
b. 6.06
\textbf{c. 6.11} 
\textbf{d. 6.15}
e. 6.21
f. 6.27
g. 6.33
h. 6.37
i. 6.43
j. 6.49

7. Suppose that in a group of pet owners 70% have a dog and 30% have a cat. Suppose further that 40% of the dogs have fleas and that 10% of the cats have fleas. Find the conditional probability that a pet is a dog given the information that it has fleas.

a. .375
b. .422
c. .491
d. .572
e. .648
f. .731
g. .856
\textbf{h. .903}
i. .946
j. Cannot be determined from the given data.
8. Suppose you perform a sequence of Bernoulli trials each of which has probability of success \( p = .3 \). What is the probability that you will have exactly 1 or 2 success in 4 trials?

a. .79
b. .77
c. .73
d. .70
e. **.68**
f. .65
g. .62
h. .59
i. .55
j. .52

9. The heights of American men and of Norwegian men are well described by Normal Models. For American men the mean is 69 inches and the standard deviation is 3 inches. For Norwegian men the mean is 71 and the standard deviation is again 3. Suppose an American man and a Norwegian man are selected at random. What is the probability that the American man is taller?

a. .198
b. .215
c. .231
d. .245
e. .267
f. .296
g. **.319**
h. .334
i. .355
j. .378
10. A Gallup Poll in March of this year asked 500 people "Generally speaking, do you think nuclear power plants are safe or not safe?" 57% said "Safe", the remainder answered "Not Safe" or "Unsure". What is the 90% confidence interval for estimating the fraction of the population who believe nuclear power plants are safe.

a. (.565, .575)
b. (.561, .579)
c. (.558, .582)
d. (.552, .588)
e. (.548, .592)
f. (.543, .597)
g. (.540, .600)
h. (.539, .601)
i. (.537, .603)
j. (.534, .606)

11. You want to test the hypothesis that the average daily high temperature in Kansas city is 65°. The data you have is that the high temperature in Kansas City for the past four Mondays were 70°, 75°, 66° and 68°. What is value of the t-statistic you would use? How many degrees of freedom does it have? Do you reject the hypothesis against a two sided alternative.

a. 2.3146, df = 3. Do not reject the null hypothesis.
b. 2.3146, df = 4. Do not reject the null hypothesis.
c. 2.3146, df = 5 Do not reject the null hypothesis.
d. 2.3146, df = 3, Reject the null hypothesis.
e. 2.3146, df = 4, Reject the null hypothesis.
f. 3.2134, df = 5, Reject the null hypothesis.
g. 3.2134, df = 3, Do not reject the null hypothesis.
h. 3.2134, df = 4, Do not reject the null hypothesis.
i. 3.2134, df = 5, Do not reject the null hypothesis.
j. 3.2134, df = 3, Do not reject the null hypothesis.
12. You want to test the null hypothesis that a six sided die is fair using a Chi-square test with significance level $\alpha = .05$. You roll the die 96 times and obtain these results:

<table>
<thead>
<tr>
<th>Face</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of times</td>
<td>20</td>
<td>10</td>
<td>15</td>
<td>25</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

A Chi-square test compares the observed data with the expected values. What is the expected number of times the face 4 shows up? The computed value of the test statistic $\chi^2 = 9.625$. What is the number of degrees of freedom (DF) associated with this value. Do you reject the null hypothesis?

a. 15, 5, Reject
b. 15.5, 5, Do not Reject.
c. 16, 5, Reject.
d. 16.5, 5, Reject
e. 15, 5, Do not reject.
f. 16, 6, Reject.
g. 16.6, Do not reject.
h. 15.5, 6, Reject.
i. 17, 5, Reject.
j. 16, 7, Reject.

$\chi^2 = 9.625$ was distributed.

A table of critical values for $\chi^2$ was distributed.

16, 5, do not reject. Answer c on the correction sheet.
13. The book contains an example of regression of \(\%\text{Body Fat}(Y)\) on \(\text{Waist}(X)\). The computer output is

Dependent variable is: \(\%\text{BF}\)
R-squared = 67.8%
s = 4.713 with 250 - 2 = 248 degrees of freedom

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef</th>
<th>SE(Coeff)</th>
<th>t-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-42.734</td>
<td>2.717</td>
<td>-15.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Waist</td>
<td>1.70</td>
<td>0.0743</td>
<td>22.9</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Which (if any) of the following statements are correct?

(1) In the data set that was used the regression model accounts for roughly 68% of the variability in \(\%\text{Body Fat}\).

(2) The predicted level of \(\%\text{Body Fat}\) for a person with a 35 inch waist is

\[
\%\text{Body Fat} = -42.73 + 1.70(35) = 16.77
\]

(3) The very small value of \(p\) associated with the variable \(\text{Waist}\) as well as the fact that the standard error for the coefficient of \(\text{Waist}\) is much smaller than the coefficient make us very confident that the association we see between the variables is not due to chance.

(4) Based on this analysis we would estimate that on average a person with a one inch larger waist would have a \(\%\text{Body Fat}\) measurement 1.7 percentage points higher.

\[\text{a. All}\]
\[\text{b. None}\]
\[\text{c. 1, 2, and 3}\]
\[\text{d. 1, 2, and 4}\]
\[\text{e. 1, 3 and 4}\]
\[\text{f. 2, 3, and 4}\]
\[\text{g. 1 and 2}\]
\[\text{h. 3 and 4}\]
\[\text{i. 1 and 4}\]
\[\text{j. 2 and 3}\]
14. In the book there is discussion of an experiment to compare the performance of containers designed to keep hot beverages hot. Four brands were tested, 8 containers of each brand. The average loss of temperature for beverages stored in the each of the four brands was:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Average temperature loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUPPS</td>
<td>10.2</td>
</tr>
<tr>
<td>Nissan</td>
<td>2.8</td>
</tr>
<tr>
<td>SIGG</td>
<td>16.0</td>
</tr>
<tr>
<td>Starbucks</td>
<td>10.3</td>
</tr>
</tbody>
</table>

To investigate further a one way analysis of variance was done which produced the following output:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container</td>
<td>3</td>
<td>714</td>
<td>238</td>
<td>10.7</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>28</td>
<td>622</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>1338</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What do you conclude from this analysis?

a. Accept the null hypothesis; conclude there appears to be no significant difference in the means.

b. Accept the null hypothesis; conclude that there does appear to be a significant difference in the means.

c. Accept the null hypothesis; conclude that the brand Nissan performs significantly better than the others.

d. Reject the null hypothesis; conclude there appears to be no significant difference in the means.

e. Reject the null hypothesis; conclude that there does appear to be a significant difference in the means.

f. Reject the null hypothesis; conclude that the brand Nissan performs significantly better than the others.
15. The book contains an example of multiple regression of %Body Fat on Waist and Height. The
computer output is

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
<th>F-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>12216.5</td>
<td>2</td>
<td>6108.28</td>
<td>307</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>4912.25</td>
<td>247</td>
<td>19.8877</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE(Coeff)</th>
<th>t-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.10088</td>
<td>7.888</td>
<td>-0.403</td>
<td>0.6870</td>
</tr>
<tr>
<td>Waist</td>
<td>1.77309</td>
<td>0.0715</td>
<td>24.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Height</td>
<td>-0.60154</td>
<td>0.1089</td>
<td>-5.47</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Which (if any) of the following statements are correct?

(1) In the data set that was used, the regression model accounts for roughly 100\% - 71\% = 29\% of the variability in %Body Fat.

(2) The predicted level of %Body Fat for a person with a 35 inch waist who is 65 inches tall is %Body Fat = -3.10 + 1.77(35) -.60(65) = 19.85

(3) We should reject both of the following hypotheses:
   \( H_1 \) : the coefficient of Waist is zero.
   \( H_2 \) : the coefficient of Height is zero.

(4) The variable Waist explains roughly three times as much of the variability of %Body Fat as is explained by the variable Height.

a. All
b. None
c. 1, 2, and 3
d. 1, 2, and 4
e. 1, 3 and 4
f. 2, 3, and 4
g. 1 and 2
h. 3 and 4
i. 1 and 4
j. 2 and 3
16. You have collected information from men and women about how much they liked various movies. Also you have placed the movies in two categories; drama and comedy. You wish to know if, on average, men’s and women’s views about movies are the same; to know if, on average, people’s views on the two types of movies are the same; and to know if there might be some more complicated relation; for instance women might especially like comedies. The best way to analyze the data would be to do

a. A t-test comparing the sexes and a t-test comparing the movie categories.
b. A paired t-test comparing the sexes and a paired t-test comparing the movie categories.
c. A t-test comparing the sexes and a t-test comparing the movie categories, and a Chi-square test of homogeneity
d. A paired t-test comparing the sexes and a paired t-test comparing the movie categories, and a Chi-square test of homogeneity
e. A one way Analysis of Variance comparing sexes and another one way Analysis of Variance comparing movie categories
f. A one way Analysis of Variance comparing sexes and another comparing movie categories and a Chi-square test for independence.
g. A two way Analysis of Variance and a Chi-square test for independence
h. A two way Analysis of Variance with interaction term
i. A two way Analysis of Variance with interaction term and Chi-square test for independence