Exam 1- ECON 3402

- **Directions**: There are 16 multiple-choice questions (worth 1.5 points each) and 3 short-answer/problem-solving exercises worth 21 points for a total of 45 points, which I will then scale up to 60 points. Additionally, there is an *experimental bonus section*, worth five points, which you will tear off, take with you, and turn back into me by October 1, 2013. Please do not collaborate on the experimental section; if you do, you will receive a 0 on that section. You have an hour and fifteen minutes to complete the in-class portion of the exam. All the best.

Here is a list of formulas:

- $\sigma_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x)^2}$
- $Var(x) = \sigma_x^2$
- $r(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} \left[ \left( \frac{x_i - \mu_x}{\sigma_x} \right) \left( \frac{y_i - \mu_y}{\sigma_y} \right) \right]$
- $\hat{y} = \hat{\alpha} + \hat{\beta}x$, where $\alpha = \mu_y - \hat{\beta} \mu_x$ and $\beta = r \frac{\sigma_y}{\sigma_x}$
The next two questions deal with the following hypothetical data set. Suppose I have a sample of five blueberry fritters from Gallery Row coffee. Their weights (in grams) are 150, 150, 145, 147.5, and 145.

1. I sample five more fritters and each weigh 147 grams. What happens to the mean?
   (a) It goes up.
   (b) It goes down.
   (c) It does not change.

2. What happens to the median?
   (a) It goes up.
   (b) It goes down.
   (c) It does not change.

Refer to the graph for the next three questions. In a Business Statistics class with 136, the professor records how much money each student has in their possession during the first class of the semester. The histogram below is of the data collected:

3. The percentage of students with under $10.00 in their possession is closest to:
   (a) 35%
   (b) 50%
   (c) 60%
   (d) 70%

4. The histogram:
   (a) is skewed right
   (b) has an outlier
   (c) is asymmetric
   (d) all of the above

5. The number of students with over $30 in their possession is:
   (a) under 5
   (b) about 10
   (c) about 30
   (d) over 50
6. If we add a large number of data points very close to the mean of our data set, we should expect:

(a) nothing to happen
(b) the standard deviation to increase
(c) the standard deviation to stay the same
(d) the standard deviation to fall

Refer to the following scatterplot for the next two questions.

7. Based on the scatterplot, we would expect the correlation to be:

(a) strong, negative
(b) weak, negative
(c) strong, positive
(d) weak, positive
(e) 0

8. Based on the graph, it is reasonable to think that:

(a) $r < 0$.
(b) The least squares regression line would be upward-sloping, i.e., $\beta > 0$.
(c) Both of these are true.
(d) Neither of these are true.

You are given the following regression equation: $\hat{y} = 8x$, where $x$ is the number of weeks and $y$ is the number of tweets you send out on Twitter.

9. What is the intercept?

(a) 0
(b) 8
(c) none of these

10. You observe some Twitter accounts of your classmates. One of your classmates has been on Twitter for 30 weeks and has sent out 247 tweets. What is the residual?

(a) 247
(b) 240
(c) 7
(d) -7
11. TRUE or FALSE: It is desirable to use a regression line to make predictions outside of the domain of data we used to make the regression line.
   (a) TRUE
   (b) FALSE

12. The box in a box-and-whisker plot shows:
   (a) mean, median, mode
   (b) min, max, mean
   (c) $Q_1$, median, $Q_3$
   (d) $Q_1$, mean, $Q_3$

13. TRUE or FALSE: All outcomes for a normally-distributed variable are equally likely.
   (a) TRUE
   (b) FALSE

14. Responses to a survey are recorded as variables of interest to the researchers. Below are some of the variables from a survey conducted by the U.S. Postal Service. Which of the variables is categorical?
   (a) State of residence
   (b) Number of people, both adults and children, living in the household
   (c) Total household income, before taxes, in 2007
   (d) Age of the respondent

15. Which of the following ideas are most similar to each other?
   (a) stratified random sampling, block design
   (b) simple random sampling, block design
   (c) voluntary response sampling, block design

16. In an experiment, what does it mean for a result to be statistically significant?
   (a) Its magnitude is quantitatively large.
   (b) It is unlikely to have occurred randomly.
   (c) Its meaning has practical importance to the context applied in.
   (d) All of these are true.

17. Free Point? YES/NO.
1. The following data are samples of the length of time it takes to complete an Atlanta Braves game in minutes: 159, 174, 182, 135, 204, 227, 198.

   (a) (1 point) What is the mean of this data?

   (b) (1.5 points) Write down, but do not evaluate the terms associated with the formula for the standard deviation.

   (c) (3 points) Write down the five number summary for this data set.

   (d) (3 points) The actual value of the standard deviation in part (b) is 30.451. Assume the data are normally distributed and use the sample standard deviation as the best guess for the population. Using your answer for part (a), find the following probabilities:

   i. The probability of a game lasting between 180-200 minutes

   ii. The probability of a game lasting more than 214 minutes

   iii. The range of data for which 99.7% of data are modeled as being between, i.e., \( \mu \pm 3\sigma \)

   (e) (1.5 points) You are now told that the data are divided into two categories: the first four observations (159, 174, 182, 135) are games televised on SportSouth, while the next three observations (204, 217, 198) are televised on ESPN. Draw an appropriate graph that includes each observation and captures the two categories of data.
2. (2.5 points) Consider the following two-way table for the type of entree ordered at a wedding reception.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish</td>
<td>113</td>
<td>125</td>
<td>0</td>
</tr>
<tr>
<td>Steak</td>
<td>130</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Chicken</td>
<td>144</td>
<td>159</td>
<td>0</td>
</tr>
<tr>
<td>Vegetarian</td>
<td>4</td>
<td>23</td>
<td>0</td>
</tr>
</tbody>
</table>

(a) What is the likelihood that someone who orders fish is male? Is female?

(b) What percent of people order chicken?

(c) Give the conditional distribution of entree orders given that the person is female.

3. Consider the following daily data on Starbucks sales of coffee mugs and one-pound bags of coffee:

<table>
<thead>
<tr>
<th>Day</th>
<th>Mugs</th>
<th>Pounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>39</td>
</tr>
</tbody>
</table>

(a) (1 point) Calculate the mean number of mugs purchased and the mean number of pounds purchased.

(b) (2 points) The standard deviation of number of mugs purchased is 17.24, while the standard deviation of the number of pounds purchased is 12.19. Write down, but do not evaluate the terms of the correlation coefficient.

(c) (2.5 points) The correlation coefficient is -.925. Using your answer for (a) and the information provided in (b), write down the least squares equation.

(d) (2 points) Interpret the correlation coefficient, the slope of the regression line, and the intercept of the regression line. Write down the r-squared, and answer how much of the variation the least squares equation provides.