Statistics 3115 – Spring 2015

Final Exam

Name: ________________________________

Thursday, May 14, 2015
Time: 120 minutes
Instructors: Brittany Cuchta & Iranga Mendis

Instructions:

- *Do not open the exam until I say you may.*
- *Circle your final answers where appropriate.*
- ***Show as much work as possible to obtain partial credit.***
- *Use your time wisely - **move on if you get stuck**.*
- *As always, you are expected to exhibit Academic Integrity during the exam.*

Materials Allowed:

- *You may have a calculator that cannot communicate with other devices.*
- *A single crib sheet (front & back) is allowed for this exam (No worked out examples).*

Problem	1	2	3	4	5	6	7	Total
Points	12	12	14	10	24	10	18	100
Score								

1. (2 points each)

i. You have a very nice collection of marbles, some of which are classic antique glass marbles. Unfortunately, all your marbles (both the antiques and the new synthetic ones) are in one bag. You wish to show some of your antique marbles to one of your friends. If you pick out 5 marbles at a time, what distribution could you use to calculate the probability that you get at least one antique marble?

ii. Suppose that in the binomial probability mass function $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\lambda > 0$. Then the binomial distribution approaches which distribution? Give name of distribution and parameter value(s).

iii. For a given sample space S of an experiment, a ______________________ is any rule that associates a number with each possible outcome in S.

iv. Which of the following statements are not correct?

a. A discrete random variable X can assume only a finite number of possible values.

b. A discrete random variable X is any random variable whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on.

c. A random variable X is said to be continuous if its set of possible values consists of an entire interval on the number line.

d. Number of students in your statistics class is an example of a discrete random variable.

v. When monitoring the status of a computer system over time, with the number of breakdowns constituting the events of interest, the appropriate probability distribution is:

a. Binomial distribution
b. Negative binomial distribution
c. Poisson distribution
d. Hypergeometric distribution

vi. Which of the following would decrease the width of a confidence interval? ***Circle all that apply***.

a. Increase the sample size.
b. Decrease the sample size.
c. Increase the level of confidence.
d. Decrease the level of confidence.

2. a. (4 pts) The call letters of a radio station must have four letters. The first letter must be a K or a W. How many different station call letters can be made if repetition of letters is allowed?

b. Five men and six women belong to a university club. The club has been asked to select four of their members to represent them on a float in the local St. Pat's parade.

   i. (4 pts) What is the probability of selecting all four women to represent the club on the float?

   ii. (4 pts) What is the probability of selecting four members so that exactly two men and two women represent the club on the float?

3. A Gallup poll was conducted in November 2013 to investigate shopping trends during the 2013 holiday season. As part of this study, they found that 522 out of 1000 national adults that were polled said that they are likely to do holiday shopping online that year.

a. (10 pts) Conduct a hypothesis test using significance level $\alpha=0.05$ to determine if the true proportion of national adults likely to do holiday shopping online in 2013 is *greater than* 50%. To receive full credit, please show all of the hypothesis testing steps that were presented in class, including writing your final conclusion in sentence form.

b. (4 pts) Calculate the p-value.

4. It is known that dragon eggs have an average weight of 5 lb. with a population standard deviation of 2 lbs. You take a random sample of 100 dragon eggs.

a. (4 pts) Find the expected value and standard deviation of the average weight in the sample of 100 eggs.

b. (4 pts) Find the (approximate) probability that the average weight among these 100 eggs is between 4.8 and 5.2 lbs.

c. (2 pts) What theorem are you using to calculate the above probability? What assumptions must be met in order to use this theorem?

5. A study was done regarding the mean cooking time for the average Thanksgiving turkey. In a sample of 20 frozen turkeys, it was found that the sample mean was 5.2 hours, with a standard deviation of 30 minutes. Assume the population is normally distributed.

a. (10 pts) Conduct a hypothesis test to determine if the population mean of turkey cook-time is less than 5.5 hours. Use significance level α=0.05. To receive full credit, please show all of the hypothesis testing steps that were presented in class, including writing your final conclusions in sentence form.

b. (5 pts) Calculate a 99% prediction interval for the cooking time of the next randomly selected turkey.

c. (5 pts) Calculate a 99% confidence interval for the true mean turkey cooking time.

d. (4 pts) Describe the type I and type II errors in the context of this problem.

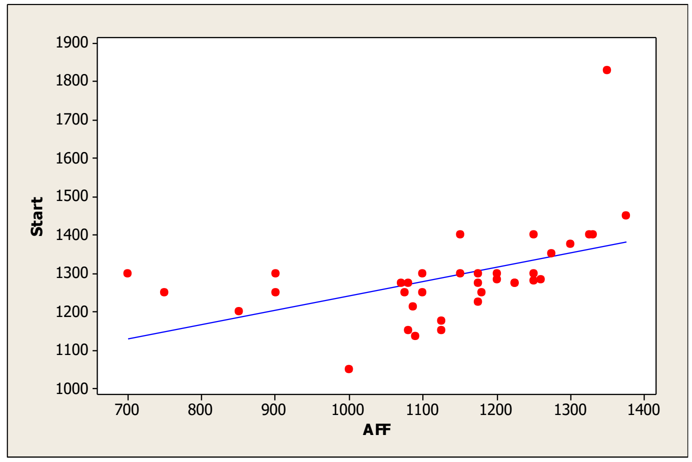
6. (10 pts) A computer program can make subroutine calls either to subroutine A or subroutine B. Let X be the number of calls to subroutine A and Y the number of calls to subroutine B. Find the correlation coefficient of X and Y. Comment on the strength and direction of the relationship. The joint pmf of X and Y is given below.

		y		
	p(x,y)	0	1	2
x	0	0.0	0.2	0.2
	1	0.3	0.1	0.2

7. Given below is a Minitab® output for a simple linear regression on a Pizza Franchise chain, predicting startup cost in thousands of dollars ("Start") via annual franchise fee in thousands of dollars ("AFF"). Reference: *Business Opportunity Handbook*

Predictor	Coef	SE Coef	T	P
Constant	867.6	135.1	6.42	0.000
AFF	0.3732	0.1179	3.16	0.003

S = 110.626 R-Sq = 22.8% R-Sq(adj) = 20.5%



a. (3 pts) Does the Startup cost and the annual franchise fee have a significant linear relationship at $\alpha=0.05$? Provide reasoning to support your answer.

b. (2 pts) Report and interpret the Coefficient of Determination

c. (3 pts) Find the correlation coefficient (r) between annual franchise fee and Startup cost, and comment on the strength and direction.

7. (continued)

d. (4 pts) Suppose someone wants to open a new store with an annual franchise fee of $1100 (in thousands of dollars). What would be the predicted Startup cost (round to the nearest whole number)?

e. (2 pts) What is the (approximate) scope of the model for this data set?

f. (4 pts) List the assumptions of a Simple Linear Regression Model