# 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 | 10 | 14 | 24 | 10 | 18 | 100 |
| Score | | | | | | | | |

1. (2 points each)

i. Suppose you are running an experiment where you attempt to have a cat perform a trick. You wish to continue with attempts until the cat performs a trick 4 times. What distribution would you use to model this experiment?

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 is not an example of a continuous random variable?

a) The weight gain in pounds per month for a calf
b) The price for cheesecake of New York Style cheesecake
c) The time it takes you to finish this statistics test
d) The number of mistakes made by a typist on a randomly chosen page of a physics thesis

v. Which of the following statements is (are) true about a Poisson distribution with parameter $\lambda$?

a) The mean of the distribution is $\lambda$
b) The variance of the distribution is $\lambda$
c) The parameter $\lambda$ must be greater than zero
d) All of the above statements are true
e) None of the above statements are true

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 sample of 3 chips is selected for inspection from a shipment containing 14 memory chips. Note that 5 of the 14 chips are defective.

   a. (3pts) *How many* ways are there to select 3 chips from the shipment of 14?

   b. (4pts) What is the probability that *all 3* of the chips selected are defective?

   c. (5pts) What is the probability that *at least* 1 of the 3 chips selected is defective?

3. Ice cream cartons from a certain manufacturer have an average weight of μ=12 ounces, with a standard deviation of σ=1 ounce. You take a random sample of 100 cartons.

   a. (4 pts) Find the expected value and standard deviation of the *average weight* ( $\overline{X}$ ) in the sample of 100 cartons.

   b. (4 pts) Find the (approximate) probability that the *average weight* among these 100 cartons is between 11.8 and 12.15 ounces.

   c. (2 pts) What theorem are you using to calculate the probability in b? What assumption is necessary in order for this theorem to hold?

4. 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 530 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.

5. An instructor collects data from their class describing the average number of television shows students watch during the week. In a class size of 25, it was found that the sample mean was 3.4 shows per week and a standard deviation of 1.2. Assume the population is normally distributed.

   a. (10 pts) Conduct a hypothesis test to determine if the population mean of television shows watched per week is less than 3.5. Use significance level $\alpha=0.05$. 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% confidence interval for the true mean shows watched per week.

5. (continued)

c. (5 pts) Calculate a 99% prediction interval for the amount of shows watched per week for the next randomly selected student.

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) | 1 | 2 | 3 |
| x | 0 | 0.4 | 0.1 | 0.2 |
| | 1 | 0.1 | 0.2 | 0.0 |

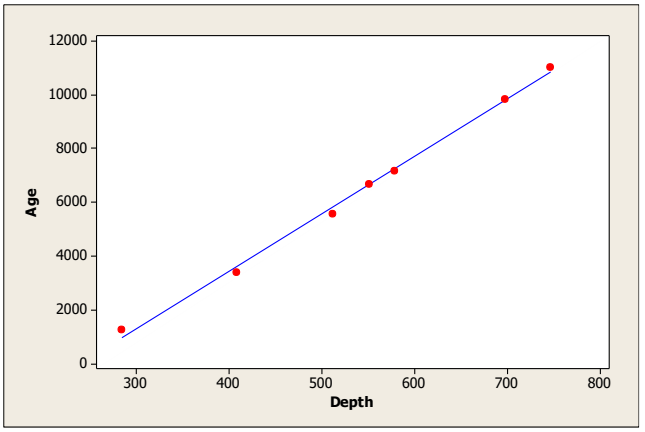
7. The following is a part of a regression analysis output on, ages (in calendar years before 1950) [variable: "Age"] of several sediment samples taken at various depths ranging from 284.5 through746.5cm [variable: "Depth" ] in Lago di Fimon, a lake in Italy.

Model Summary

| S | R-sq | R-sq(adj) |
|---|---|---|
| 220.344 | 99.65% | 99.58% |

Coefficients

| Term | Coef | SE Coef | t Stat | P-value |
|---|---|---|---|---|
| Intercept | -5129.03 | 315.6212 | -16.2506 | 0.7892 |
| Depth | 21.38951 | 0.564217 | 37.91009 | 2.41E-07 |




a. (3 pts) Does the Age and the Depth have a significant linear relationship at α=0.05? Provide reasoning to support your answer.

b.

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

ii. (3 pts) Find the correlation coefficient (r) between Depth and Age, and comment on the strength and direction.

7. (continued)

c. (4 pts) Suppose a new sediment sample is planned to be taken at a depth of 473.55cm. What would be the predicted age (round to the nearest whole number)?

d. (2 pts) Why shouldn't one rely on the interpretation of the intercept term here?

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