

Multiple Choice [10 pts]-- choose the ONE BEST answer for each question by writing the corresponding letter in the blank to the left of the question.

- _____ 1. Which graph would be used to explore the relationship between amount of water vapor in a cloud mass (ppm) and snowfall amount (inches)?
A. Barplot B. Histogram C. Dot Chart D. Scatterplot E. Pie Chart
- _____ 2. Which association occurs when most of the individuals are above average for one variable but below average for the other variable?
A. Negative B. Neutral C. Positive D. Strong E. Weak
- _____ 3. Which of the following is not a possible value for the correlation coefficient?
A. -1 B. 0 C. 0.1 D. 0.34789 E. 1.2
- _____ 4. Which of the following correlation coefficients represents the weakest relationship?
A. -1 B. 0 C. 0.1 D. 0.34789 E. 1.2
- _____ 5. What is the name of the variable that we are interested in predicting or explaining?
A. Continuous B. Discrete C. Explanatory D. Response E. Sampling
- _____ 6. What is the vertical difference between an observed and predicted value of the response variable?
A. Explanatory B. Frequency C. Predicted D. Residual E. Response
- _____ 7. Which of the following coefficients of determination represents the most precise predictions?
A. -1 B. 0 C. 0.1 D. 0.34789 E. 1.2
- _____ 8. Which word best describes the situation where the best-fit line goes through the middle of the points on a scatterplot but the points are unequally scattered around that line?
A. linear B. non-linear C. RSS D. homoscedastic E. heteroscedastic
- _____ 9. What type of study is it if the researcher sends a survey to the entire population, but only some of the individuals return it?
A. Convenience B. Inference C. Simple Random D. Regression E. Voluntary Response
- _____ 10. What type of study is it if every individual has the same chance of being selected for the sample?
A. Convenience B. Inference C. Simple Random D. Regression E. Voluntary Response

Answer the following questions, showing your work as necessary (including writing R code) on a separate sheet.

11. [6 pts] Assume that a wholesale distributor of fertilizer products knows that the weekly demand for 5-5-6 N-P-K fertilizer is normally distributed with a mean of 12 tons and a standard deviation of 3.6 tons. Use this information to answer the questions below. [Note: R suggestions are on the last page.]
- a) What is the probability that the demand for one week will exceed 14 tons?
 - b) What is the probability that the demand for one week is between 10 and 15 tons?
 - c) What weekly demand can the distributor expect for the top 25% of weeks?

Short (Paragraph) Answers -- Answer FOUR of the following questions with complete sentences on a separate sheet of paper. Circle the questions below that you have chosen to answer and make sure to clearly label your answers on the separate sheet. Each question is worth 3 points.

12. What are the two major goals of regression? Give a specific example to illustrate each goal.
13. What are the two major assumptions of regression? Draw plots that illustrate situations where these assumptions are (a) both met, (b) one is violated, and (c) the other is violated.
14. Describe the major differences between an observational and experimental study.
15. What are the major principles of experimental design and why is each important?
16. Describe several situations (be specific) where observational studies are valuable.
17. Completely describe all differences between a population and a sampling distribution.

Answer the following three questions on a separate sheet. Show your work where appropriate.

18. [6 pts] A group of marine biologists from California studied the foraging ecology of northern elephant seals (*Mirounga angustirostris*) off the California coast (Le Boeuf *et al.*, 2000). Part of their analysis required that they record, for each observed seal, the month that it was observed and the sex of the seal. The raw frequency table is shown below. Present your answers below to one decimal place.

Sex	Month			Total
	Jun	Jul	August	
Male	8	7	12	27
Female	5	2	13	20
Total	13	9	25	47

- What percent of all elephant seals were male and captured in June?
 - What percent of the elephant seals captured in July were male?
 - What percent of the female elephant seals were captured in June?
19. [7 pts] Students designed a simple experiment to determine if the duration of a D cell battery differed significantly among four different types (Duracell, Rayovac, Energizer, and Eveready) used at three different temperatures (5, 15, 25°C). Each battery was placed into a high-drain (i.e., higher energy usage) machine held in a chamber where the temperature could be strictly controlled. The time (in seconds) it took until the battery was “dead” was recorded. Twelve batteries of each type were available to the students. Use this information to answer the following questions.
- What is the response variable? _____
 - What is/are the factor(s)? _____
 - What is/are the number of levels? _____
 - What is the number of treatments? _____
 - What is the number of replicates per treatment? _____
20. [5 pts] Government researchers examined the relationship between the median price of a house and the percent of the population that was of a “lower status” (as defined by household income, percent unemployed, etc.) for a large sample of communities across the United States. Their results are shown in Figure 1. Also note that $r = -0.587$. In the space below, construct an appropriate bivariate EDA for this information.

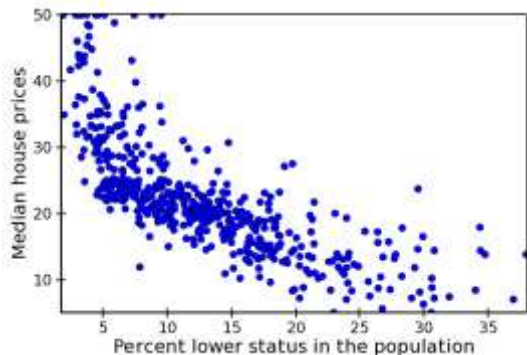


Figure 1. Scatterplot of median house prices versus the percentage of the population classified as “lower status.”

21. [13 pts] Lowe (2012) investigated the effect of climate change on the abundance of the *Gyrinophilus porphyriticus* salamander in a stream in New Hampshire. One part of his study was to examine the abundance (number of individuals) of the salamander over a 12-year period to see if the variability in abundance could be explained by year. The results for this study are shown in Figure 2 and 3. Use these results to answer the questions below.
- What is the response variable?
 - In terms of the variables of this problem, what is the equation of the best-fit line?
 - In terms of the variables of this problem, interpret the value of the slope?
 - What is the predicted abundance for 2008?
 - What is the residual for 2005 if the abundance is 28?
 - What percentage of the variability in abundance is explained by year?
 - How much would one expect abundance to change in five years?

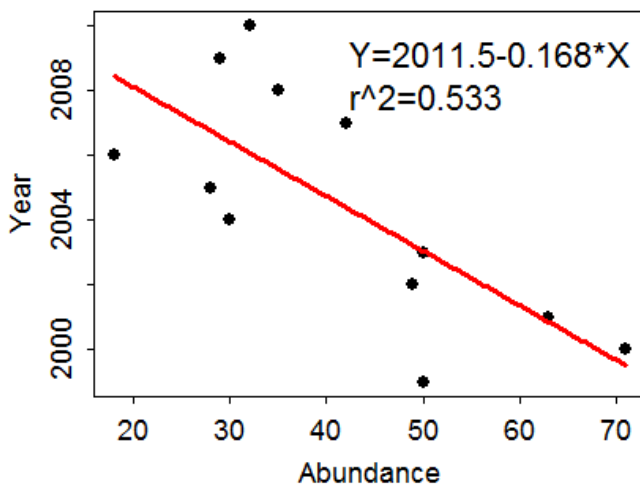


Figure 2. Linear regression of year on abundance of salamanders.

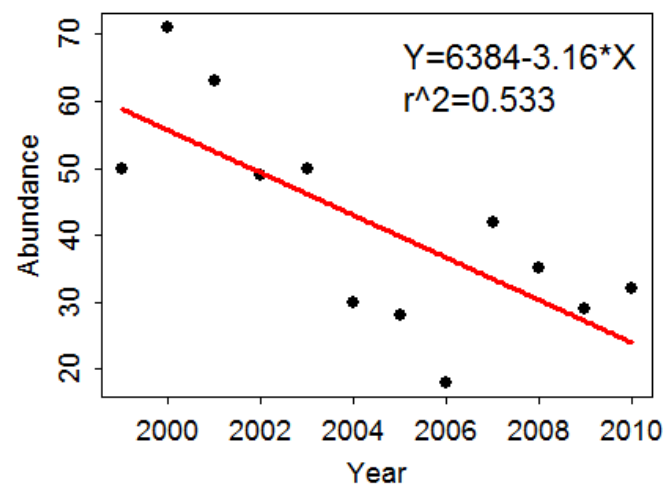


Figure 3. Linear regression of salamander abundance on year.

```
library(NCStats)
distrib(x,mean=##,sd=##,lower.tail=XXXXX,type="X")
```

where **x** is replaced with the value of the quantitative variable or the area

mean=## has ## replaced by the value of the mean

sd=## has ## replaced by the value of the standard deviation

lower.tail=XXXXX has XXXXX replaced with TRUE (the default) for a "left-of" and FALSE for a "right-of" calculation

type="X" has X replaced with p (the default) for a forward and q for a reverse question
