**Multiple Choice [15 pts] -- choose the ONE BEST answer for each question by writing the corresponding letter in the blank to the left of the question.**

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

1. 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

1. Which term is defined as “the change in value of the response variable for a unit change in the explanatory variable?”

A. Correlation B. Y-Intercept C. RSS D. Slope E. r2

1. 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

1. Which of the following is NOT a possible value for the coefficient of determination?

A. -0.6 B. 0.34789 C. 0.1 D. 0 E. 1

1. What is it when you are asked to make a prediction outside the domain of the explanatory variable?

A. Residual B. Response C. Extrapolation D. Homoscedastic E. Heteroscedastic

1. 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 (i.e., look like a funnel)?

A. Linear B. Non-linear C. RSS D. Homoscedastic E. Heteroscedastic

-

1. Which of the following statistics is biased?

A. x B. s C. IQR D. All are biased E. None is biased

1. What is the measure of dispersion among statistics from all possible samples?

A. Response B. Mean C. Replicates D. Standard Error E. Standard Deviation

1. What is the measure of sampling variability?

A. Response B. Mean C. Replicates D. Standard Error E. Standard Deviation

1. Which descriptors best describe the situation where multiple samples produce very similar results but those results do not average out to the true value of the parameter?

A. Accurate, Precise B. Accurate, Imprecise C. Inaccurate, Precise D. Inaccurate, Imprecise

-

1. Which term is defined as “the probability of the observed statistic or a value more extreme assuming that the null hypothesis is true?”

A.  B.  C. p-value D. power E. r2

1. What decision should be made if the p-value < ?

**A.** DNR Ho **B.** Reject Ho **C.** DNR HA **D.** Reject HA **E.** None of these

1. Which is the probability of not rejecting a false Ho?

A.  B.  C. p-value D. power E. r2

1. What level of confidence should be used if =0.1 and HA: >o?

A. 0.10 B. 0.50 C. 0.90 D. 0.95 E. 0.99

1. What type of confidence region is constructed if =0.1 and HA: >o?

A. Interval B. Point estimate C. Upper bound D. Lower bound E. Rejection region

**Answer the following questions in the space provided. Show your work where appropriate.**

1. **[9 pts]** Vasey *et al.* (2012; J Anxiety Disord) examined subjects that had a self-described fear of spiders. Each subject was shown a spider and asked to estimate its size. The subject’s estimate of the size of the spider was divided by the actual size of the spider to compute the “spider size discrepancy” (i.e., values <1 mean that the subject underestimated the size of the spider). Additionally, the researchers recorded the subject’s level of distress upon seeing the spider on a scale from 0-100 (higher values mean more distress). The researchers were interested in determining if the subject’s distress level could explain their spider size discrepancy value. Use their results from either Figure 1 or 2 (as appropriate) to answer the questions below the figures.

**Figure 1.** Linear regression of a subject’s distress on the **Figure 2.** Linear regression of a subject’s discrepancy

discrepancy in the estimated size of the spider. in size of the spider on their distress level.

1. In terms of the variables of this problem, interpret the value of the slope?
2. What is the predicted spider size discrepancy value if the subject’s distress level is 80?
3. What is the residual if the subject’s distress level is 30 and the spider size discrepancy value is 0.7?

1. How much would one expect a subject’s spider size discrepancy value to change if the subject could reduce his or her distress level by 20 points?
2. What percentage of the variability in spider size discrepancy value is explained by the subject’s distress level?
3. **[8 pts]** Researchers observed groups of dolphins off the coast of Iceland in 1998. The researchers recorded the time of the day (Morning, Noon, Afternoon, and Evening) and the main activity of the group (travelling, feeding, or socializing). The number of dolphin groups observed by each time of day and activity is shown in Table 1. Use these results to answer the questions below the table. Round all answers to one decimal place.

Table 1. Frequency of Dolphin groups by time of day and type of activity.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Type of Activity** | | |
| **Time of Day** | **Traveling** | **Feeding** | **Socializing** |
| **Morning** | 6 | 28 | 38 |
| **Noon** | 6 | 4 | 5 |
| **Afternoon** | 14 | 0 | 9 |
| **Evening** | 13 | 56 | 10 |

1. What percentage of all dolphin groups were observed in the evening?
2. What percentage of the dolphin groups that were observed in the evening were feeding?
3. What percentage of the dolphin groups that were socializing were observed in the morning?
4. What percentage of the dolphin groups that were observed in the morning were feeding?

**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 (default) for a “left-of” and FALSE for a “right-of” calculation

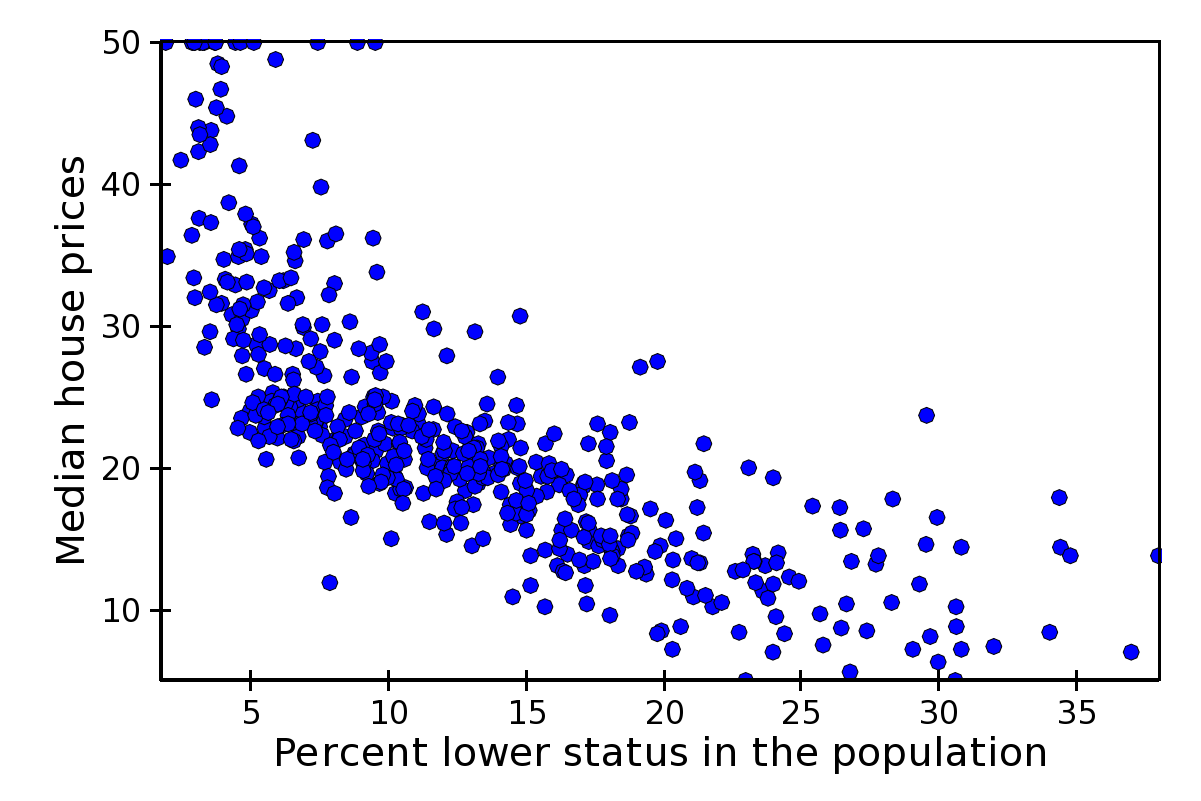
**type=”X”** has X replaced with p (default) for a forward and q for a reverse question

**Use distrib() in RStudio to produce the result(s) needed to answer the next question. Round all answers to four decimal places and include the code used to produce your result. Hints about distrib() are on the previous page.**

1. **[6 pts]** Assume that a wholesale distributor of fertilizer products knows that the weekly demand for 5-Nitrogen, 5-Phosphorous, 6-Potassium fertilizer is only slightly right-skewed with a mean of 12 tons and a standard deviation of 3.6 tons. Use this information to answer the questions below.
2. What is the probability that the demand for one week will exceed 14 tons?
3. What is the probability that the mean demand for a sample of 20 weeks will exceed 14 tons?
4. What is the probability that the mean demand for a sample of 40 weeks will be between 11.5 and 13 tons?

**Complete a thorough bivariate EDA appropriate to the type of variable for the following question.**

1. **[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 3. Also note that r = -0.587.



**Figure 3**. Scatterplot of median house prices versus the percentage of the population classified as “lower status.”

1. **[12 pts]** The golden rectangle is a rectangle with a width-to-length ratio of 0.618 (i.e., if the width is divided by the length of the golden rectangle then the resultant value is 0.618). Evidence for the golden rectangle can be found in several works by ancient Greeks and Egyptians. Recently, anthropologists measured the width-to-length ratios of beaded rectangles used by the Shoshoni Indians of America to decorate their leather goods. The results from their sample are shown in Table 2.

Table 2. Summary statistics for the width-to-length ratio of a sample of Shoshoni beaded rectangles.

n mean sd min Q1 median Q3 max

20 0.6605 0.0925 0.5530 0.6060 0.6410 0.6765 0.9330

Further assume that the distribution of ratios is approximately symmetric, that the standard deviation is 0.1, the anthropologists were interested in determining if the golden rectangle was or was not evident in the beadwork of the Shoshoni Indians, and they tested this at the 10% level. Use this information to answer the questions below.

* 1. What are the null and alternative hypotheses?
  2. Compute the p-value.
  3. What can you conclude about the beadwork of the Shoshoni and the golden rectangle?
  4. Construct an appropriate confidence region.
  5. Interpret your confidence region.
  6. Define what a Type I and a Type II error would be in this study.

**Short (Paragraph) Answers -- Answer the following questions with complete sentences.**

1. **[3 pts]** Completely describe all differences between a population and a sampling distribution.
2. **[3 pts]** Describe choices that you, as a researcher, can make to reduce the margin-of-error. Which is the best choice to make and why?

**The following questions are worth extra credit. Show all of your work.**

1. **[3 pts XC]** Biologists often visually estimate the number of geese in a flock to determine relative abundance. Some researchers believe, however, that biologists poorly estimate the true size of the flock with their visual estimates. To test this hypothesis, researchers will show a slide with a certain number of geese on it to a random sample of biologists and ask each biologist to estimate the number of geese on the slide. Assume that the standard deviation for the estimates is 40, the researchers want the final estimate to be within 5 birds, and that they will use a 10% significance level. Compute what sample size they should use.
2. **[4 pts XC]** Compute power if HA: >50, =15, n=25, =0.10, and the actual  is 52.