**Multiple Choice [30 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 called when you try 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., looks 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 the average of those results does not equal 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 HAD. Reject HA E. None of these

1. Which is the probability of rejecting a true Ho?

A.  B.  C. p-value D. power E. r2

1. What level of confidence should be used if =0.01 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 =0.01 and HA: <o?

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

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

1. **[5 pts]** Completely describe all differences between a population and a sampling distribution.
2. **[5 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?

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

1. **[19 pts]** Male Magnificent Frigatebirds (*Fregata magnificens*) display an enlarged red throat pouch to attract females. However, females may not always be able to see the size of the pouch. Madsen *et al.* (2004) hypothesized that females may judge the size of the pouch from the frequency of the drumming sound produced by the pouch. To determine if pouch size could be predicted from the drumming frequency, Madsen *et al.* recorded the pouch size and frequency of the drumming sound for 18 males (Figure 1).

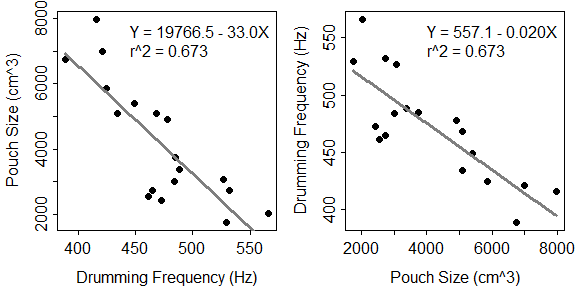


Figure 1. Fitted line plot for pouch size on drumming frequency (Left) and drumming frequency on pouch size (Right).

1. What is the predicted pouch size if the drumming frequency is 500 Hz?
2. In terms of the variables of this problem, interpret the value of the slope? [*use a complete sentence*]
3. What is the residual if the pouch size is 4000 cm3 and the drumming frequency is 450 Hz?

1. How much would one expect the pouch size to change if drumming frequency increased by 100 Hz?
2. What percentage of the variability in pouch size is explained by drumming frequency?
3. **[9 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 |

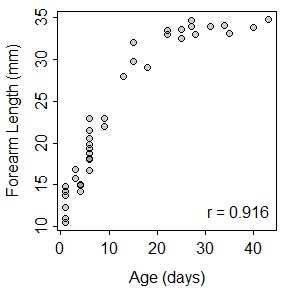
1. What percentage of all dolphin groups were observed in the evening?
2. What percentage of dolphin groups in the evening were observed feeding?
3. What percentage of socializing dolphin groups were observed in the morning?
4. **[10 pts]** Krochmal and Sparks (2007; J. Mammal.) examined the relationship between forearm length and age of 38 Northern Myotis Bats (*Myotis septentrionalis*) captured in Vermillion County, IN. Their results are shown in Figure 3. Use these results to construct a complete bivariate EDA.

Figure 3. Scatterplot for Bat forearm length on age.

**library(NCStats)**

**distrib(#,mean=##,sd=##,lower.tail=FALSE,type=”q”)**

where **#** is replaced with the value of the quantitative variable or the area (i.e., the percentage as a proportion).

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

**sd=##** has ## replaced by the value of the population standard deviation or the SE

**lower.tail=FALSE** is included for a “right-of” calculation

**type=”q”** is included for a reverse calculation

**Use distrib() in RStudio to produce the result(s) needed to answer the next question. Round all answers to *three decimal places* and *include the code* used to produce your result. If you should not answer the question then say so and *be explicit about why you should not*.**

1. **[7 pts]** Renner (1970) examined the content of hydroxymethylfurfurol (HMF) in honey. HMF is an organic compound derived from cellulose without the use of fermentation and is a potential “carbon-neutral” source for fuels. This study found that the distribution of HMF in one kg of honey was strongly right-skewed with a mean of 9.5 g and a standard deviation of 13.5 g. Use this information to answer the questions below.
2. What is the probability that 50 samples of one kg of honey will have an average between 10 and 12 g of HMF?
3. What is the probability that one kg of honey will have less than 10 g of HMG?
4. **[22 pts]** A dean wants to determine if it takes more than 10 minutes, on average, for students to get from one class to another. In an effort to test this hypothesis, she collected a random sample of 100 between-class travel times and found the mean to be 10.12 mins. Assume that it is known from previous studies that the distribution of between-class travel times is symmetric with a standard deviation of 1.60 mins. Test the dean’s hypothesis with α=0.10. Use this information to answer the questions below. Show your work and R code where appropriate.
   1. What are the null and alternative hypotheses? [*Make sure to define the parameter in the hypotheses.*]
   2. Compute the p-value.
   3. What can be concluded about the time it takes for students to move between classes? [*use a complete sentence*]
   4. Construct an appropriate confidence region.
   5. Interpret your confidence region. [*use a complete sentence*]
   6. Define what a Type I and a Type II error would be in this study. [*use complete sentences*]

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

1. **[5 pts XC]** Foresters want to sample enough trees from within a large stand that the mean diameter-at-breast-height (DBH) for all trees in the stand will be estimated to within 10 cm with 99% confidence. Suppose from a pilot study sample of a few trees, that the researchers estimated the standard deviation of DBH among trees to be 60 cm. Use this information to determine how many trees should be sampled to meet the forester’s constraints.
2. **[5 pts XC]** Compute  if HA: >500, =150, n=25, =0.10, and the actual  is 520.