**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 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

**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]** Poysa (2003; Behav. Ecol. Sociobio.) examined the effect of predation risk on the number of eggs laid by Goldeneye ducks (*Bucephala clangula*). Specifically she identified groups of nests that were near each other and then measured relative predation risk (a combination of several other measurements that was then rescaled to be from 0 to 100, with 100 representing the highest predation risk) and mean number of eggs laid for each group of nests. Use her results from either Figure 1 or 2 (as appropriate) to answer the questions below the figures.

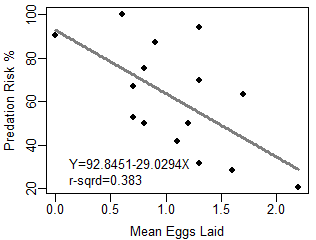
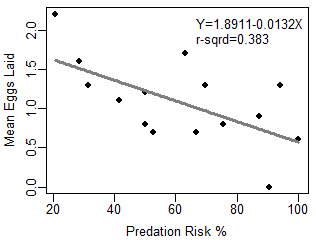
 

Figure 1. Linear regression of predation risk percentage Figure 2. Linear regression of mean eggs laid on

on mean eggs laid. predation risk percentage.

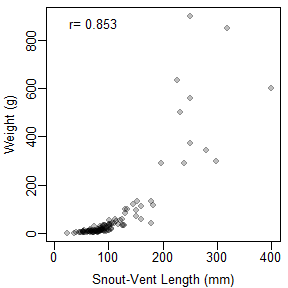
1. What is the predicted mean number of eggs laid if the predation risk is 50%?
2. In terms of the variables of this problem, interpret the value of the slope? [*use a complete sentence*]
3. What is the residual if the predation risk is 75% and the mean number of eggs laid is 0.8?

1. How much would one expect the mean number of eggs laid to change if predation risk increased by 20%?
2. What percentage of the variability in mean number of eggs laid is explained by predation risk percentage?
3. **[12 pts]** Researchers examined the “broad” types of animals (“Mammal”, “Bird”, and “Amph(ibian)/Rep(tile)”) among several zoos. The number of animal types in each zoo is shown in Table 1. Use these results to answer the questions below Table 1. Round all answers to one decimal place.

Table 1. Frequency of animals by broad type and zoo location.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Broad Type of Animal** | | |
| **Zoo** | **Amph/Rep** | **Bird** | **Mammal** |
| **Chicago** | 27 | 66 | 70 |
| **Minnesota** | 4 | 13 | 52 |
| **San Antonio** | 168 | 218 | 69 |
| **San Diego** | 27 | 40 | 109 |

1. What percentage of all animals are in the Minnesota zoo?
2. What percentage of animals in the Chicago zoo are birds?
3. What percentage of animals in the San Diego zoo are mammals?
4. What percentage of all animals are mammals in the San Diego zoo?



1. **[10 pts]** Meiri (2010; J. Zool.) examined the relationship between the weight (g) and snout-to-vent length (mm) of 211 species of Iquania lizards. His results are shown in Figure 3. Use these results to construct a complete bivariate EDA.

Figure 3. Scatterplot for Iquania weight on snout-to-vent length.

**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 four 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]** Alanson (1992; J. Ag. Econ.) examined the size of farms in England in 1939 and 1989. He found that the distribution of farms sizes in 1989 was strongly right-skewed with a mean of 65.13 hectares (ha) and a standard deviation of 107.71 ha. Use this information to answer the questions below.
2. What is the probability that the mean farm size from a sample of 20 farms will exceed 100 ha?
3. What is the probability that the mean farm size from a sample of 60 farms will be between 50 and 100 ha?
4. **[22 pts]** An employee in the Human Resources department of a large firm wants to test whether the mean monthly expense claim by employees is greater than $500 (the amount currently allowed by company policy). Suppose that a random sample of 40 employees is taken and it is found that the mean expense claim for the most recent month was $560. Assume that the standard deviation is $150 and the employee’s boss wants to perform this test at the 1% level. Use this information to answer the questions below. Show your work and R code where appropriate.
   1. What are the null and alternative hypotheses?
   2. Compute the p-value.
   3. What can be concluded about the company’s current policy for monthly expense claims? [*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.