**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.**

\_\_ **D** \_\_\_

\_\_ **D** \_\_\_

\_\_ **D** \_\_\_

\_\_ **A** \_\_\_

\_\_ **C** \_\_\_

\_\_ **E** \_\_\_

\_\_ **E** \_\_\_

\_\_ **D** \_\_\_

\_\_ **D** \_\_\_

\_\_ **C** \_\_\_

\_\_ **C** \_\_\_

\_\_ **B** \_\_\_

\_\_ **B** \_\_\_

\_\_ **C** \_\_\_

\_\_ **D** \_\_\_

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.

Popn … individuals, uses SD, natural variability

Sampling … statistics, uses SE, sampling variability

1. **[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?

Increase n (this is better) or reduce level of confidence

**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
2. In terms of the variables of this problem, interpret the value of the slope? [*use a complete sentence*]

As the predation risk increases by 1% the mean eggs laid decreases by 0.0123, on average

1. What is the predicted mean number of eggs laid if the predation risk is 50%?

1.8911-0.0132\*50 = 1.2311 eggs

1. What is the residual if the predation risk is 75% and the mean number of eggs laid is 0.8?

1.8911-0.0132\*75 = 0.9011 … 0.8-0.9011 = -0.1011

1. How much would one expect the mean number of eggs laid to change if predation risk increased by 20%?

-0.0132\*20 = -0.264 … decrease by 0.264 eggs

1. What percentage of the variability in mean number of eggs laid is explained by predation risk percentage?

38.3%

1. **[12 pts]** Researchers examined the “broad” types of animals (“Mammal”, “Bird”, and “Amph(ibian)/Rep(tile)”) Table 1. Frequency of animals by broad type and zoo location.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Broad Type of Animal** | | |  |
| **Zoo** | **Amph/Rep** | **Bird** | **Mammal** |  |
| **Chicago** | 27 | 66 | 70 | 163 |
| **Minnesota** | 4 | 13 | 52 | 69 |
| **San Antonio** | 168 | 218 | 69 | 455 |
| **San Diego** | 27 | 40 | 109 | 176 |
|  | 226 | 337 | 300 | 863 |

1. What percentage of all animals are in the Minnesota zoo?

=69/863 = 0.0800 🡪 8.0%

1. What percentage of animals in the Chicago zoo are birds?

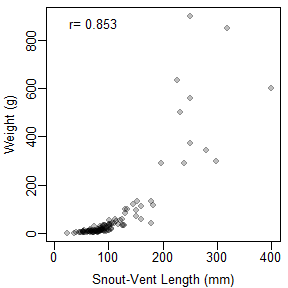
= 66/163 = 0.4049 🡪 40.5%

1. What percentage of animals in the San Diego zoo are mammals?

= 109/176 = 0.6193 🡪 61.9%

1. What percentage of all animals are mammals in the San Diego zoo?

= 109/863 = 0.1263 🡪 12.6%

**[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.

Nonlinear

Positive

Moderately strong

Outliers?

Don’t use r

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?

Cannot answer … sampling distribution is not normal because n<30 but popn is strongly skewed.

1. What is the probability that the mean farm size from a sample of 60 farms will be between 50 and 100 ha?

distrib(100,mean=65.13,sd=107.71/sqrt(60)) = 0.9939

distrib(50,mean=65.13,sd=107.71/sqrt(60)) = 0.1383 🡪 0.9939-0.1383 = 0.8556

1. **[22 pts]** An employee in the Human Resources department of a large firm wants to test whether the mean monthly
   1. What are the null and alternative hypotheses?

Ho: mu>500 vs. Ha: mu>500

* 1. Compute the p-value.

distrib(560,mean=500,sd=150/sqrt(40),lower.tail=FALSE) 🡪 0.0057

* 1. What can be concluded about the company’s current policy for monthly expense claims? [*use a complete sentence*]

It appears that the current monthly expense claims exceed the policy of $500.

* 1. Construct an appropriate confidence region.

distrib(0.99,type=”q”,lower.tail=FALSE) 🡪 z\*=

560-2.326\*150/sqrt(40) = 504.83

* 1. Interpret your confidence region. [*use a complete sentence*]

I am 90% confident that the mean monthly expense claim is greater than $504.83.

* 1. Define what a Type I and a Type II error would be in this study. [*use complete sentences*]

Type I – The mean claim is $500 and we conclude that it is greater than $500.

Type II – The mean claim is not $500 and we conclude that it is $500.

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

**[5 pts XC]** Foresters want to sample enough trees from within a large stand that the mean diameter-at-breast-height

n=(2.576\*60/10)^2 = 238.89 🡪 239 trees

1. **[6 pts XC]** Compute  if HA: >500, =150, n=25, =0.10, and the actual  is 520.

Crit value 🡪 distrib(0.10,mean=500,sd=150/sqrt(25),type=”q”) 🡪 461.553

Beta 🡪 distrib(461.553,mean-520,sd=150/sqrt(25),lower.tail=FALSE) 🡪 0.9743