**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. 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 name of the variable that you are interested in predicting or explaining?

A. Continuous B. Discrete C. Explanatory D. Response E. Sampling

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. 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 of the following statistics is biased?

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

1. Which four values would be considered “precise and inaccurate” if =30?

A. 28,29,31,32 B. 5,18,42,55 C. 30,31,32,33 D. 10,20,30,40

-

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 false Ho?

A.  B.  C. p-value D. power E. r2

1. What level of confidence should be used if =0.10 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 =0.10 and HA: >o?

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

**Short (Paragraph) Answers -- Answer questions 16 and 17 with complete sentences.**

1. **[6 pts]** Completely describe three major differences between a population distribution and a sampling distribution.
2. **[4 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 questions 18-24 in the space provided. Show your work where appropriate.**

A Sustainable Community Development student examined the relationship between female life expectancy (by country) in 1960 and 2015, with the intent of determining if life expectancy in 1960 helped explain (or could be used to predict) life expectancy in 2015. Her analysis of life expectancy data from the World Bank is in Figure 1.

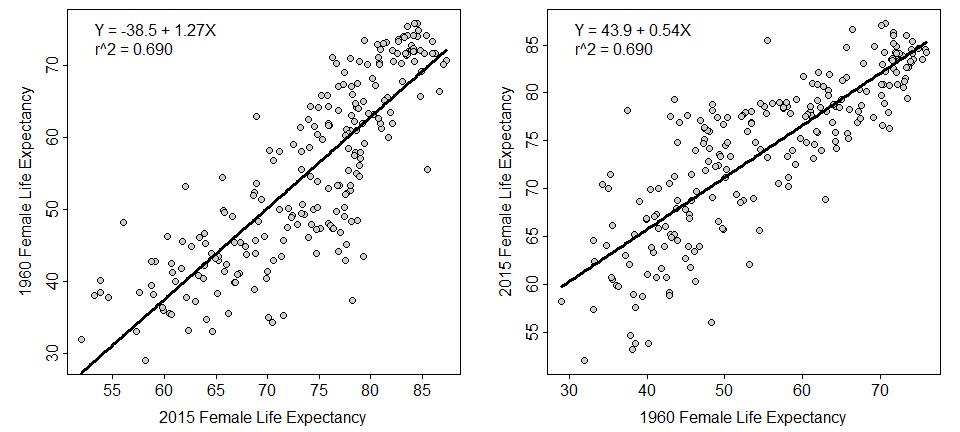
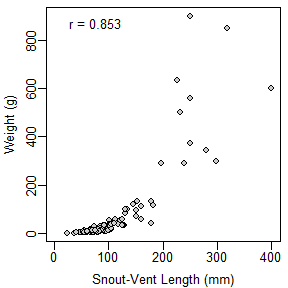


Figure 1. Fitted line plot for 1960 on 2015 (Left) and 2015 on 1960 (Right) female life expectancy by country.

1. **[4 pts]** What is the predicted 2015 life expectancy if the 1960 life expectancy is 55 years?
2. **[4 pts]** In terms of the variables of this problem, interpret the value of the slope? [*use a complete sentence*]
3. **[4 pts]** Compute the residual for Azerbaijan which had a life expectancy of 64 years in 1960 and 72 years in 2015?
4. **[4 pts]** How much would the predicted life expectancy in 2015 differ for two countries with a 20 year difference in life expectancy in 1960?
5. **[3 pts]** What percentage of the variability in 2015 life expectancy is explained by knowing 1960 life expectancy?
6. **[3 pts]** Are you concerned about any aspects of the regression analysis in Figure 1 on the previous page?



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. Use his results in Figure 2 to construct a complete bivariate EDA.

Figure 2. Scatterplot for Iquania weight on snout-to-vent length.

**library(NCStats)**

**distrib(val,mean=meanval,sd=sdval,lower.tail=FALSE,type=”q”)**

where

* **val** is a value of the quant. variable or area (i.e., percentage as a proportion)
* **meanval** is population mean ()
* **sdval** is standard deviation () or error (SE)
* **lower.tail=FALSE** is included for “right-of” calculations
* **type=”q”** is included for reverse calculations

**Round your answers for questions 25-28 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 VERY explicit about why you should not*.**

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.

1. **[3 or 4 pts]** What is the probability that the mean size from a sample of 20 farms will exceed 100 ha?
2. **[3 or 4 pts]** What is the probability that the mean size from a sample of 60 farms will be between 50 and 100 ha?
3. **[3 or 4 pts]** What is the probability that the size of a randomly selected farm will be between 50 and 100 ha?
4. **[3 or 4 pts]** What is the mean size in a sample of 50 farms such that 10% of farms will be larger?

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. **[4 pts]** What are the null and alternative hypotheses?
  2. **[4 pts]** Compute the p-value.
  3. **[3 pts]** What can be concluded about the company’s current policy for monthly expense claims? [*use a complete sentence*]
  4. **[4 pts]** Construct an appropriate confidence region.
  5. **[3 pts]** Interpret your confidence region. [*use a complete sentence*]
  6. **[4 pts]** Define what a Type I and a Type II error would be in this study. [*use complete sentences*]

**The lowest score on questions 35 and 36 will be considered as extra credit. Please show all of your work.**

1. **[5 pts]** 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]** Compute  if HA: <520, =150, n=25, =0.10, and the actual  is 500.