***Choose the one best answer for each question below by writing the corresponding letter in the blank to the left of the question. Each question is worth 1 point.***

**1.** What symbol is used to represent the population standard deviation?

\_\_\_\_\_ **A.**  **B.**  **C.** Q3 **D.** x **E.** s

**2.** What symbol is used to represent the population mean?

\_\_\_\_\_ **A.**  **B.**  **C.** Q3 **D.** x **E.** s

**3.** 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

**4.** What is the vertical difference between an observed and predicted value of the response variable?

\_\_\_\_\_ **A.** Explanatory **B.** Frequency **C.** Predicted **D.** Response **E.** Residual

**5.** Which of the following is NOT described in a bivariate EDA for quantitative variables?

\_\_\_\_\_ **A.** Association **B.** Form **C.** Outliers **D.** Shape **E.** Strength

***Answer questions 6-12 in the provided space.***

1. **[4 pts]** What are the two major goals of linear regression?
2. **[3 pts]** Define RSS and explain how it is used to find the “best-fit line.”

The 2018 Environmental Performance Index (EPI) ranks 180 countries on 24 performance indicators covering environmental health and ecosystem vitality. These metrics provide a gauge for how close countries are to established environmental policy goals and, thus, offers a scorecard that highlights leaders in environmental performance and provides guidance for countries that aspire to be leaders in sustainability. Data related to the EPI are used in the following two questions.

1. **[10 pts]** The researchers examined the relationship between an environmental health index (related to air quality, water quality, and lead exposure) and the overall EPI for 180 nations. Construct a complete bivariate EDA from their results in Figure 1.

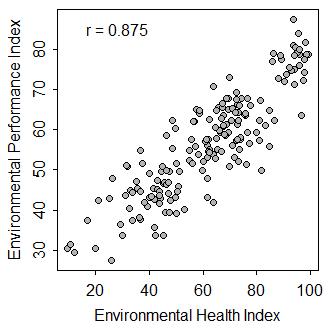


Figure 1. Scatterplot for overall environmental performance and environmental health indices.

1. **[10 pts]** The researchers also examined the relationship between the overall EPI and gross domestic product (GDP) per capita for the 180 countries. Construct a complete bivariate EDA from their results in Figure 2.

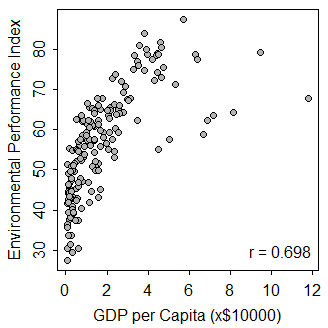


Figure 2. Scatterplot of environmental performance index and gross domestic product per capita.

1. **[16 pts]** The face colors of endangered Gouldian Finch (Erythrura gouldiae) living in the savannahs of northern Australia have either red, yellow, or black faces. Their bills are also either red, yellow, or black, but bill color is not always the same as face color. All birds in a sample were classified according to their face and bill color as shown in Table 1. Use these results to answer the questions below the table *to one decimal place and show your work*.

Table 1. Frequency of finches by bill color and face color.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Face Color** | | |
| **Bill Color** | **Black** | **Red** | **Yellow** |
| **Black** | 16 | 5 | 6 |
| **Red** | 19 | 20 | 6 |
| **Yellow** | 18 | 22 | 22 |

1. What percentage of black-faced finches had a red bill?
2. What percentage of all finches had a yellow face and a yellow bill?
3. What percentage of all finches had a red face?
4. What percentage of red-billed finches had a black face?
5. **[22 pts]** The amount of recyclable paper materials produced per day for a small college campus is normally distributed with a mean of 550 lbs and a standard deviation of 125 lbs. The sustainability coordinator for the campus has defined a “heavy-use” day as having more than 750 lbs and “low-use” day as having less than 500 lbs of recyclable paper material. Use this information to answer the questions below. *Please write the R code that you used along with your final numerical answer to one decimal place.*
6. What percentage of days would be categorized as “heavy-use”?
7. What percentage of days would be categorized as neither “heavy-use” nor “low-use”?
8. What should the weight of recyclable paper materials produced be changed to so that 15% of days would be considered “heavy-use”?
9. What is the IQR for weight of recyclable paper materials produced per day?
10. What is the median weight of recyclable paper materials produced per day?

**library(NCStats)**

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

where

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

1. **[30 pts]** Department of Health researchers examined the relationship between average number of decayed teeth per person and average sugar consumption per person for a variety of countries with the hope of determining if amount of tooth decay could be explained by amount of sugar consumed. The results from their study are shown in Figure 3. Use this information to answer the questions below. *Show your work as necessary*.

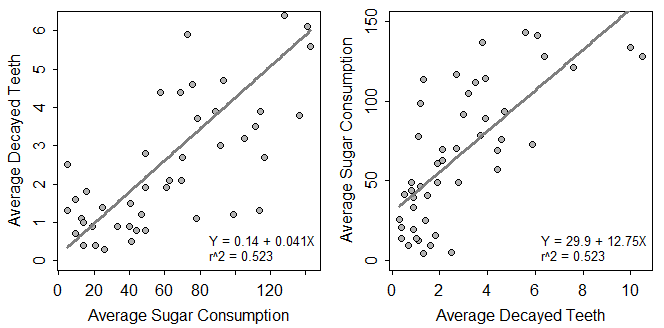


Figure 3. Fitted line plot for average tooth decay per person on average sugar consumption (Left) and average sugar consumption on average tooth decay per person (Right) for 47 countries. Sugar consumption is in grams of sugar consumed per person per day.

1. In terms of the variables of this problem, interpret the slope value? [*use a complete sentence*]
2. In terms of the variables of this problem, interpret the intercept value? [*use a complete sentence*]
3. How much would one expect average tooth decay to change if average sugar consumption increased by 20 g?
4. What is the predicted average tooth decay if average sugar consumption is 170 g?
5. What is the predicted average tooth decay if average sugar consumption is 30 g?
6. What is the residual if average tooth decay is 2.20 and average sugar consumption is 80 g?
7. What percentage of variability in average tooth decay is explained by average sugar consumption?
8. What is the correlation coefficient between average sugar consumption and average tooth decay?
9. Do you have concerns about this regression? [*thoroughly explain your answer, whether you have any concerns or not.*]