3.3.11

For graphs, please scroll to the bottom.

1. The following table contains information about the robbery rate and percentage of urban population across a set of states:

    1. Sketch the scatterplot and sketch the residual plot (*see bottom*)
    2. Write and interpret the correlation coefficient

        $r = 0.82$. This is a moderate to strong positive correlation that implies that the percentage of urban population has a direct effect on the robberies per 100,000 people.

    3. Write the regression equation and interpret the regression coefficient.

        $y = 3.767x - 117$. There is a positive relationship between the percentage of urban population and the number of robberies and there is an increase in 4 robberies per 100,000 people when the urban population increases by 1%.

    4. Show and interpret the coefficient of determination.

        $r^2 = 0.673$. The coefficient of determination measures how accurate the line of best fit is, meaning that this line of best fit is somewhat accurate for measuring the data.

    5. State whether you think there is a relationship between the two variables, and justify your answer.

        There is a moderate-strong positive relationship between the variables, according to the high positive correlation coefficient and somewhat high coefficient of determination

    6. We know that the percentage of urban population in Idaho is 35%. We also know that the percentage of urban population in Florida is 78%. Predict the robbery rates in each of these states. Are these extrapolations or interpolation, and are they valid predictions?

        Idaho has 15 robberies per 100,000 people, and Florida has 177. Both of these are interpolations since they both are in the range of our data set.

2. Geothermal power is an important source of energy. Since the amount of energy contained in 1 pound of water is a function of its temperature, you might wonder whether water obtained from deeper wells contains more energy per pound.

    1. Do a full regression analysis. (*For graph, see bottom.*)

        $r = 0.33$. In other words, the correlation is weak positive.

        $y = 0.038 + 199$. There is a positive relationship between drill hole depth and average temperature.

        $r^2 = 0.11$. This is bad best fit line to predict the temperature.

        There is a very weak positive relationship between drill hole depth and average temperature, according to the low coefficient of determination, the positive regression equation, and the low correlation coefficient.

    2. Using the regression equation, predict the temperature of the water in a well with an average drill hole depth of 2000m. Is this a reliable prediction?

        According to the equation, the temperature for a 2000 m deep hole would be $276\,C$. However, this value is not very reliable due to the weak correlation and the badness of the line of best fit.

3. Use the potency and temperature data below.

    1. Do a full regression analysis (*For graph, see bottom.*)

        $r = -0.87$. A strong negative correlation can be seen.

        $y = -0.31x + 46$. There is a negative relationship between potency and temperature.

        $r^2 = 0.76$. The line of best fit is very accurate.

        There is a strong negative relationship between potency and temperature, according to the high negative correlation coefficient and the negative regression equation.

When the temperature increases by 1 degree, there is a 0.3 unit decrease in potency. Potency is the response variable.

4. Below is some hypothetical data that shows the Rating and Average Cost for a 30-second advertisement for a number of game shows:

    1. Do a full regression analysis (*For graph, see bottom.*)

        $r = 0.94$. There is a very strong positive correlation.

        $y = 3175x - 4236$. There is a positive relationship.

        $r^2 = 0.88$. There is a very accurate line of best fit.

        There is a strong positive relationship between the two variables, as seen by the high positive correlation coefficient and the positive regression equation.

        When the ratings increase by 0.1, there is a $3175 increase in average cost for a 30-second ad.

5. You have the following data on Per Capita Income in 1987 and the % Births that are "Low Birth Weight" across a sample of states.

    1. Do a full regression analysis (*For graph, see bottom.*)

        $r = 0.11$. This is a weak positive correlation.

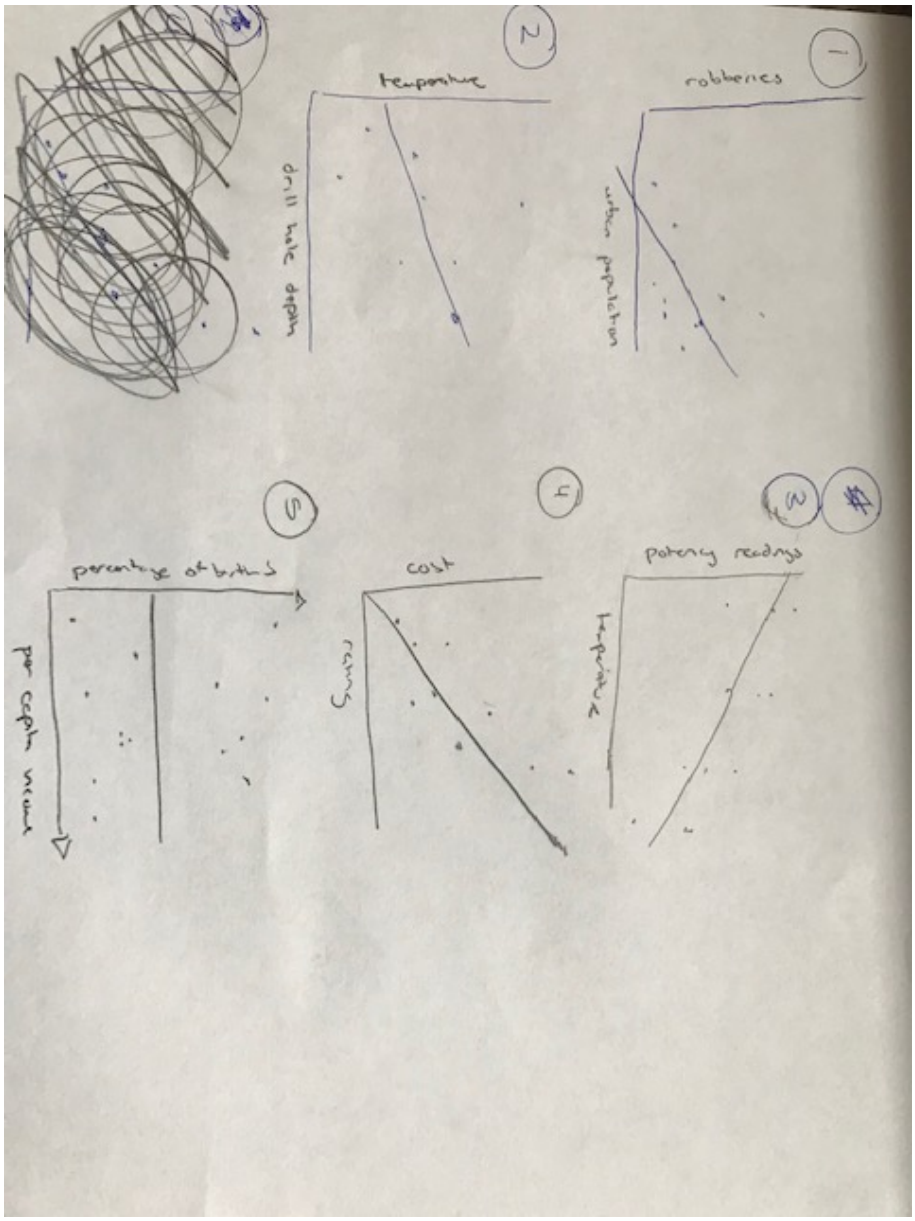
        $y = 0.000112x + 4.99$. There is a positive relationship, but it's practically constant throughout.

        $r^2 = 0.012$. There is very low accuracy in the line of best fit.

        There is a very weak positive relationship between increase in per capita income and decrease in percentage of low birth weights, shown by the low correlation coefficient and the practically constant regression equation.

        An increase in per capita income is not associated with a decrease in the percentage of low birth weights; however, we don't really know due to the extremely low correlation coefficient.

Graphs


1
robberies
urban population
2
temperature
drill hole depth
3
potency readings
temperature
4
cost
rating
5
percentage of births
per capita income