3 - regression



11.1: The time it takes to transmit a file always depends on the file size. Suppose you transmitted 30 files, with the average size of 126 Kbytes and the standard deviation of 35 Kbytes. The average transmittance time was 0.04 seconds with the standard deviation of 0.01 seconds. The correlation coefficient between the time and the size was 0.86.

Based on this data, fit a linear regression model and predict the time it will take to transmit a 400 Kbyte file.

 $b_1 = (0.86) \left(\frac{0.01}{35} \right)$ G(x)= b1 x +b0 6(409 = 0.107326 by = 0.00024571428 bo = 0.04 - (b1)(126) 60 = 0.00904



112. The following statistics were obtained from a sample of size n = 75:

– the predictor variable X has mean 32.2, variance 6.4;

164) (28)

– the response variable Y has mean 8.4, variance 2.8; and the sample covariance between X and Y is 3.6.

= 0.85 (a) Estimate the linear regression equation predicting Y based on X.

b1=0.85 (28)

b, = 0.56 bo = (8.4) - (0.56)(32.2)

(6(x)= 0.56x-9.7

At a gas station, 180 drivers were asked to record the mileage of their cars and the number of miles per gallon. The results are summarized in the table.

| i innes per ganon | . The results ar | e summarized in the |
|-------------------|------------------|---------------------|
| | Sample mean | Standard deviation |
| Mileage | 24,598 | 14,634 × |
| Miles per gallon | 23.8 | 3.4 4 |

The sample correlation coefficient is r = -0.17.

b1 = - 0.17 (3.4) 6(x) = bo+b1.x bo = 23 8 - b1. (24,598)

(a) Compute the least squares regression line which describes how the number of miles bo = 24.775per gallon depends on the mileage. What do the obtained slope and intercept mean in this situation?

11.8. Anton wants to know if there is a relation between the number of hours he spends preparing for his weekly quiz and the grade he receives on it. He keeps records for 10 weeks

It turns out that on the average, he spends 3.6 hours a week preparing for the quiz, with the standard deviation of 0.5 hours. His average grade is \mathfrak{S}_{2} (out of 100), with the standard deviation of \mathfrak{A}_{2} . The correlation between the two variables is r=0.62.

(a) Find the equation of the regression line predicting the quiz grade based on the time spent on preparation.

| $b_1 = 0.62 \left(\frac{19}{0.5} \right) = 17.36$ | | | | |
|--|--|--|--|--|
| bo=82-b1(36)=19.504 | | | | |
| 6(x)=19.504+(17.36)x | | | | |

| Year | Population mln. people | Year | Population mln. people | Year | Population mln. people |
|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|---------------------------|
| 1950 1955 1960 1965 1970 | 2558 2782 3043 3350 3712 | 1975 1980 1985 1990 1995 | 4089 4451 4855 5287 5700 | 2000 2005 2010 2015 2020 | 6090 6474 6864 ? |

Example 11.3 (WORLD POPULATION). In Example 11.1, x_i is the year, and y_i is the world population during that year. To estimate the regression line in Figure 11.1, we compute

$$\bar{x} = 1980; \quad \bar{y} = 4558.1;$$

$$S_{xx} = (1950 - \bar{x})^2 + \dots + (2010 - \bar{x})^2 = 4550;$$

 $S_{xy} = (1950 - \bar{x})(2558 - \bar{y}) + \dots + (2010 - \bar{x})(6864 - \bar{y}) = 337250, 300$

Then

$$b_1 = S_{xy}/S_{xx} = 74.1$$

$$b_0 = \bar{y} - b_1\bar{x} = -142201.$$

The estimated regression line is

$$\widehat{G}(x) = b_0 + b_1 x = -142201 + 74.1x.$$

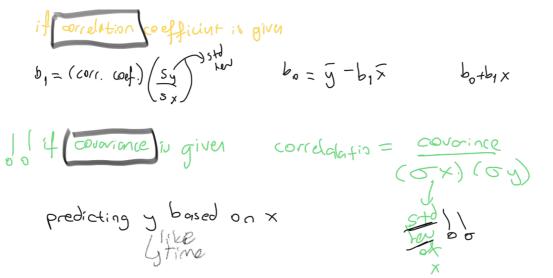
We conclude that the world population grows at the average rate of 74.1 million every year.

We can use the obtained equation to predict the future growth of the world population. Regression predictions for years 2015 and 2020 are

$$\widehat{G}(2015) = b_0 + 2015 \, b_1 = 7152 \text{ million people}$$

$$\widehat{G}(2020) = b_0 + 2020 \, b_1 = 7523 \text{ million people}$$

 \Diamond



The time it takes to copy a file from one disk to another always depends on the file size. Assume that you have copied 50 files, with the average size of 10 MB and the standard deviation of 100 ms. The correlation coefficient between the time and the size was 0.8.

Based on this data, fit a linear regression model and predict the time it will take to transmit a 90 MB file (in ms).

$$b_1 = 0.8 \left(\frac{100}{5} \right) = 16$$
 $b_0 = 200 - |6.10| = 200$

Answer: 1480

690)=1480