Week Three Homework Solutions
Hwk #70

The data gives the gold medal times for every other Summer Olympics for the women's 100-meter freestyle (swimming).

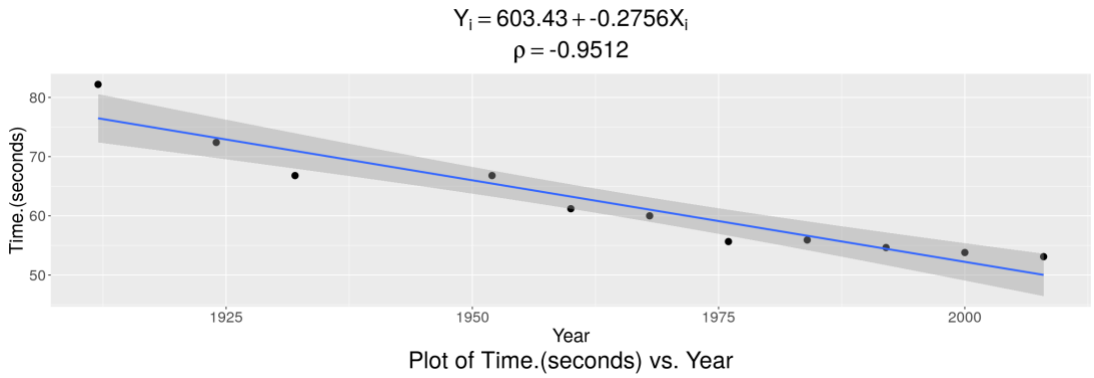
Year	Time (seconds)
1912	82.2
1924	72.4
1932	66.8
1952	66.8
1960	61.2
1968	60.0
1976	55.65
1984	55.92
1992	54.64
2000	53.8
2008	53.1

Table12.22

a. Decide which variable should be the independent variable and which should be the dependent variable.

ANSWER: The independent variable, aka the predictor variable is the year. The dependent variable is the Gold-medal times (seconds) for the Women's 100m.

b. Draw a scatter plot of the data.



c. Does it appear from inspection that there is a relationship between the variables? Why or why not?

ANSWER: Yes, there appears to be a strong, negative linear relationship between the year and the Gold-medal times. Why? Because the data are in a strong line and the correlation value is high (r = -.9512).

d. Calculate the least squares line. Put the equation in the form of: $\hat{y} = a + bx$.

ANSWER: Y = 603.43 – .2756x

e. Find the correlation coefficient. Is the decrease in times significant?

ANSWER: correlation value is high (r = -.9512). Yes, the decrease is significant—it is close to -1.

f. Find the estimated gold medal time for 1932. Find the estimated time for 1984.

ANSWER: Y = 603.43 – .2756(1932) = 70.9708 seconds ←estimated time for 1932

ANSWER: Y = 603.43 – .2756(1984) = 56.6396 seconds ←estimated time for 1984

g. Why are the answers from part f different from the chart values?

ANSWER: The chart values are what were observed. The predicted values are calculated with the equation of the line. There is error involved.

h. Does it appear that a line is the best way to fit the data? Why or why not?

ANSWER: Yes, the coefficient of determination (R-squared) is 90.473%. This means that 90.5% of the variation in Gold-medal times is being explained by the year. This is high.

[R^2=90.473 is found in the output or can be calculated by taking $(-.9512)^2$]

i. Use the least-squares line to estimate the gold medal time for the next Summer Olympics. Do you think that your answer is reasonable? Why or why not?

ANSWER: Y = 603.43 - .2756(2020) = 46.718 seconds ←estimated time for 2020 (if there were Olympics). No this is NOT reasonable since the year 2020 is not in the scope of the data collected. This is extrapolation.

Hwk #71 We are only using # letters and year entered Union.

State	# letters in name	Year entered the Union	Rank for entering the Union	Area (square miles)
Alabama	7	1819	22	52,423
Colorado	8	1876	38	104,100
Hawaii	6	1959	50	10,932
Iowa	4	1846	29	56,276
Maryland	8	1788	7	12,407
Missouri	8	1821	24	69,709
New Jersey	9	1787	3	8,722
Ohio	4	1803	17	44,828
South Carolina	13	1788	8	32,008
Utah	4	1896	45	84,904
Wisconsin	9	1848	30	65,499

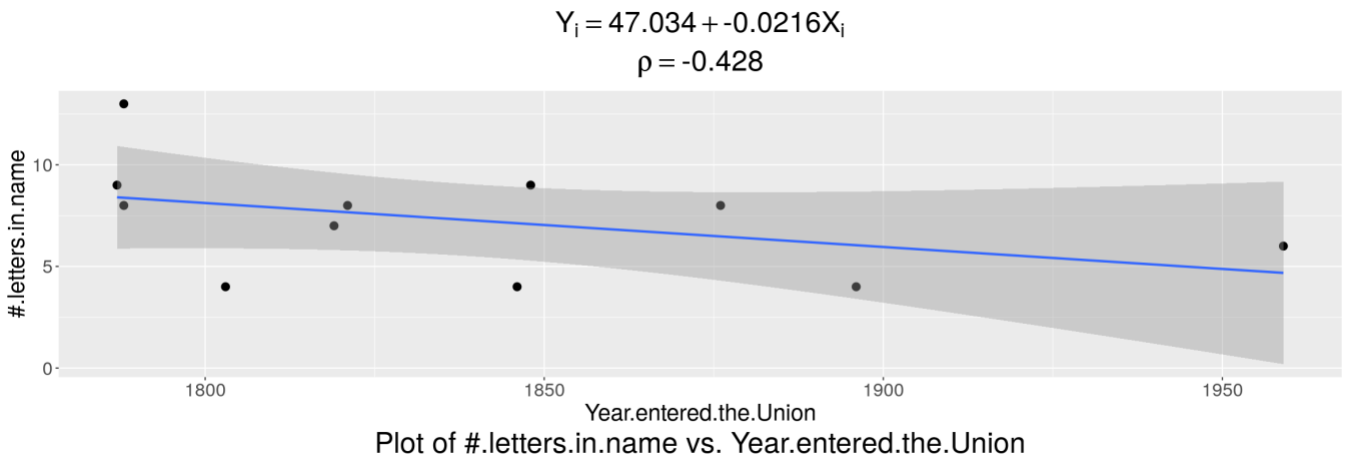
Table12.23

We are interested in whether or not the **number of letters in a state name** depends upon the **year** the state entered the Union.

a. Decide which variable should be the independent variable and which should be the dependent variable.

ANSWER: The independent variable, aka the predictor variable, is the year. The dependent variable is the number of letters in a state name.

b. Draw a scatter plot of the data.



Plot of #.letters.in.name vs. Year.entered.the.Union

c. Does it appear from inspection that there is a relationship between the variables? Why or why not?

ANSWER: No, there does not appears to be a relationship since it appears that there is a weak, negative linear relationship between the year and the number of letters in the name. Why? Because the correlation value is low (r = -.428).

d. Calculate the least-squares line. Put the equation in the form of: $\hat{y} = a + bx$.

ANSWER: Y = 47.034 – .0216x

e. Find the correlation coefficient. What does it imply about the significance of the relationship?

ANSWER: The correlation value is low (r = -.428). The relationship does not appear significant.

f. Find the estimated number of letters (to the nearest integer) a state would have if it entered the Union in 1900. Find the estimated number of letters a state would have if it entered the Union in 1940.

ANSWER: Y = 47.034 – .0216(1900) = 5.994 Rounded to nearest integer is 6 letters.

ANSWER: Y = 47.034 – .0216(1940) = 5.13 Rounded to nearest integer is 5 letters.

g. Does it appear that a line is the best way to fit the data? Why or why not?

ANSWER: No, the coefficient of determination (R-squared = 18.318%) indicates that only 18.3% of the variation in # of letters in the state name is being explained by the regression line that uses year. This is not high and indicates that the line does not fit the data well. (One could also say that the correlation value is low).

h. Use the least-squares line to estimate the number of letters a new state that enters the Union this year would have. Can the least squares line be used to predict it? Why or why not?

ANSWER: Y = 47.034 – .0216(2020) = 3.402 ← rounded to 3 letters.
No this is NOT reasonable since the year 2020 is way out of the scope of the data collected. This is extrapolation.