HW 1

Adon Rosen

Date: 2020-02-13

Question 1

A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 4 women from each 10-year age group, beginning with age 40 and ending with age 79, resulting in a total sample size of n=16. Some results follow, where X is the age, and Y is a measure of the muscle mass. Assume a simple linear regression model is appropriate.
For these data:

Mean(X) = 61.69 SD(X) = 14.67
Mean(Y) = 52.21 SD(Y) = 24.16

y-hat = 142.68 – 1.47 *X

SE(Slope) = 0.200 Root MSE = 11.38

a.

Parameters: 1. β_0 = 142.68 = The predicted muscle mass in a women when her X value, age is equal to 0 1. β_1 = -1.47 = The change in muscle mass for every one unit change in age

b.

$\hat{Y}_{63} = 142.68 - 1.47 * 63 = 50.07$

c.

In order to calculate the corellation from the slope first multiply the slope of the regression line by the standard deviation of X and then divide by the standard deviation of Y.

Slope: -1.47

SDx: 14.67

SDy: 24.16

r = (Slope * SDx) / SDy = -0.8925869

d.

State our hypotheses:

$H_0 : r = 0$

$H_a : r \neq 0$

$\alpha = 0.05$

We would like to calculate the t value of the correlation. The t value's formula is: $t = \frac{r\sqrt(n-2)}{\sqrt{1-r^2}}$

t = -7.4072668

p = 3.3156464×10^{-6}

Given this p value, we reject the null hypothesis and conclude there is a non-zero relationship between age and muscle mass.

e.

The 95% C.I. for β_1 = -1.47 +/- 0.429; [-1.899,-1.041]

I will interpret this later

f.

The 95% C.I. for a participant with an age of 63 is equal to [23.043,77.097].

I will interpret this later

g.

The residual is equal to $Y - \hat{Y}$ = 45 - 54.48 = -9.48

h.

PRE = 0.7967114

$\beta_{standardized} = \beta\frac{SD_x}{SD_y}$ = -0.8925869

Here we see that we explain roughly 0.7967 percent of the total variance in the outcome. Furthermore, it appears that the that the standardized coefficient between age and muscle mass is: -0.8925869 which suggests a very strong relationship between these two variables.

i.

If two women differ in age by 10 years the predicted difference in muscle mass will be -14.7 units.

j.

It is difficult to tell if the linear assumption holds with these data. However the provided model fits the data very well.

Question 2

The following SAS output contains an analysis in which 6-year graduation rates were collected over a 6-year period and analyzed for trends. The response variable Y is graduation rate (measured as a percentage). The explanatory variable X is year (beginning with year 1)

a.

$\hat{Y} = 61.52667 + 0.95429 * year$

b.

$\hat{Y}_6 = 61.52667 + 0.95429 * 6 = 67.25241$

c.

The residual is equal to $Y - \hat{Y}$ = 67.8 - 66.29812 = 1.50188

Here the residual is 1.5, meaning the value was underpredicted by 1.5%.

d.

PRE = 0.7541 This PRE suggests an excellent model fit.

e.

The 95% C.I. for β_1 = 0.95429 +/- 0.7565; [0.19779,1.71079]

I will interpret this later

f.

State our hypotheses:

$H_0 : \beta_{year} = 0$

$H_a : \beta_{year} \neq 0$

$\alpha = 0.05$

We reject the null hypothesis that $\beta_{year} = 0$ given the strength of the coefficient $\beta_{year} = 0.95429 (t(4) = 3.50, p = 0.0248)$.