Introducing linear regression

Helen Burn

6/12/2020

Showing relationships for discussion

Open up the linear regression Little App. (See footnote[1]). In this activity we'll explore three different data sets available through the app. For each data set, the variable indicated below plays the role of the response variable. Some explanatory variables are listed as sub-points.

1. In the Data Tab in the top toll bar, set Source Package to `Little Apps`. For Data set choose `NHANES2`. Response variable: BMI. It's important for students to know what BMI is. Explanation from the CDC & BMI calculator for students.
   - age (r = 0.5 reasonable scatterplot to assume linearity)
   - income (r = -0.07) shows a very diffuse scatter plot but also helps demo the app to students.
   - pulse: weak relationship
   - systolic: weak-to-moderate relationship
   - diastolic: has outliers
   - sleep_hour: weak-to-moderate. But has a negative relationship
2. Set Source Package to `mosaic`, and then Data set choose `CPS85`. Response variable: wage
   - age
   - education
3. Set Source Package to `Little Apps`. For Data set `Natality_2014` Response variable: `mager`, mother's age
   - `fagecomb`, father's age. Moderate size correlation. Ask what real-world phenomenon accounts for the correlation.

Open-ended exploring

Systolic blood pressure from the `NHANES` data.

Background: Explain to students what is the difference between the systolic and diastolic blood pressure. Each time the heart beats, the blood pressure in the arteries goes up. It quickly rises to a maximum and then decays until the next beat. Systolic is the maximum blood pressure each beat, diastolic the minimum. The "pulse pressure" is the difference between the two. See this site on blood pressure.

Tasks

1. Determinine three explanatory variables that are predictive of systolic blood pressure.

Write down the names of the explanatory variables here . . .

2. For each of the three variables, in the Graph tab in the upper tool bar, check the box to show the aux plot, then using and estimate from the graph, list the strength of the relationship both as a fraction of the variation explained and as the change in systolic blood pressure per unit change of the explanatory variable. Note: the T graph labelled as Model is the amount of variation explained by the model. The

[1] https://maa-statprep.shinyapps.io/Little_App_Regression/

T graph labelled raw is the amount of variation for the data. The fraction of he variation explained is the ratio of the model T to the raw T. The change in systolic blood pressure per unit change of the explanatory variable is the slope of the line.

Fill in the table with your answers. . . .

variable name	fraction of variation	change of response per unit change in explanatory

3. Then check whether those three explanatory variables explain diastolic blood pressure as well.

Which of systolic or diastolic blood pressure is better explained by the explanatory variables? . . .

Housing Prices

Set Source Package to `mosaic`, and then Data set choose `SaratogaHouses`. Response variable: `price`

a Determinine three explanatory variables that are predictive of house price.

Write down the names of the explanatory variables here . . .

b For each of the three variables , list the strength of the relationship both as a fraction of the variation explained and as the change in price of the explanatory variable.

Fill in the table with your answers. . . .

variable name	fraction of variation	change of response per unit change in explanatory

Version 0.3, 2020-08-13