Looking at Data and Simple Linear Regression -Homework

Rob McCulloch

10/7/2017

Contents

---

1. Looking at the House Price Data

Let's again use the midcity house price data (midcity.csv) that we looked at in class.

We'll focus on 4 variables:

- Nbhd: which of 3 neighborhoods the house is in.
- SqFt: size of the house in square feet.
- Brick: is the house brick, yes/no.
- Price: price of the house in dollars.

1.1

To look at the variable SqFt, make a histogram.
Get the mean and standard deviation.

1.2

To look at the variable Brick make a table telling us how many of our house are made of brick and how many are not.

In R:
`hd=read.csv("midcity.csv")`.
`table(hd$Brick)`.
In excel: easy to do with as a pivot table.

1.3

Is SqFT related to Price ? Plot SqFt vs. Price.
Is Brick related to Price ? Compute the averager price for brick and nonbrick houses.
Is Brick and SqFT related to Price ? plot SqFT *and* brick vs. Price.
In R: `plot(hd$SqFt,hd$Price,col=hd$Brick,xlab="size",ylab="price",pch=16)`.

1.4

Is Brick related to neighborhood?
Make a two-way table giving the number of observations (houses) in each of the 2*3 subsets of observations from the 2 categories for Brick and the 3 categories for Nbhd.

In R:
`table(hd`$Nbhd$, hd`Brick)`

In Excel: pivot table. Put Nbhd in the Row Labels, Brick in the column labels, and any variable in the Values Pane. Click the drop-down menu beside the variable name in the "Values" pane, and then click "Value Field Settings". You will get a list of summary functions to choose from. (for example with the midcity data, we chose Average). Use the counts summary.

What percent of houses are made of brick in each of the three neighborhoods?

1.5

What is the correlation between SqFt and Price?

Make a new variable which is size = SqFt/1000.

What is the correlation between size and Price?

Make a new variable which is price = Price/1000.

What is the correlation between size and price?

How do the three correlations compare? Why does this make sense?

---

2. The Shock Absorber Data

The data comes from a company which supplies a major automobile manufacturer with shock absorbers. An important characteristic is the "force transferred through the shock absorber when the shank is forced out of the cylinder". If you don't know what that really means, don't worry, neither do I.

What we do need to understand is that the manufacturer only considers the shock to be an acceptable part if the force measurement is between 485 and 585.

The shock manufacturer and the auto manufacturer are arguing over the following issue. Before the shock is finally shipped, it is filled with gas. After it is filled with gas, it becomes very difficult to measure the force characteristic we are interested in. The shock manufacturers would like to make the measurement before the shock is filled with gas. The auto maker is concerned that there may be a difference in the force before and after the shock is filled with gas and so would like to make the measurement after it is filled.

The shock maker claims that there is little difference between the before and after measurement so that the before measurement can be used.

To investigate this we have the before (column 1, reboundb) and the after (column 2, rebounda) measurements on 35 shocks (in shock.csv).

Get the shock data (shock.csv) from the webpage.

2.1

Plot reboundb vs. rebounda.

Is there a linear relationship?

2.2

Give an interval which should contain roughly 95% of the rebounda numbers.

2.3

What is the correlation between reboundb and rebounda?

2.4

Regress rebounda on reboundb.

What is the equation of the line.

2.5

Suppose reboundb is 600, what is your prediction for rebounda?