Illapani_Homework_HW7

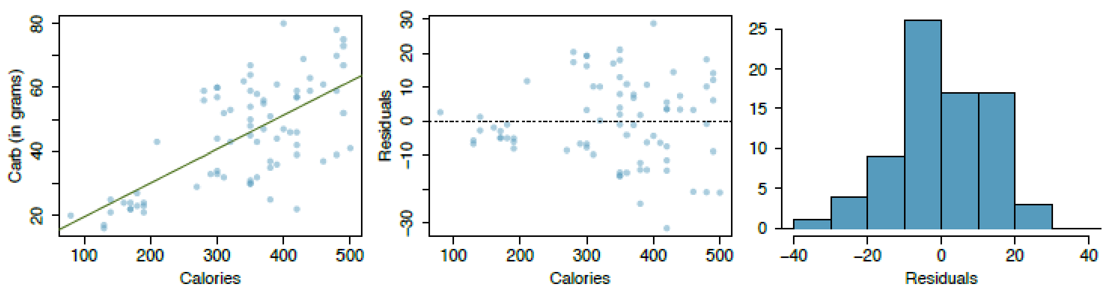
Srini Illapani

November 8, 2015

Introduction to Linear Regresion

7.24 Nutrition at Starbucks, Part I.

The scatterplot below shows the relationship between the number of calories and amount of carbohydrates (in grams) Starbucks food menu items contain. Since Starbucks only lists the number of calories on the display items, we are interested in predicting the amount of carbs a menu item has based on its calorie content.



(a) Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.

The relationship between the number of calories and the amount of carbohydrates is linear and positive. An increase in the number of calories, relates to an increase in the amount of carbohydrates

(b) In this scenario, what are the explanatory and response variables?

Explanatory variable is Calories on the x axis. Response variable is Carbohydrates on the y axis.

(c) Why might we want to fit a regression line to these data?

To predict the direction and the amount of carbohydrates in Starbucks food items based on the number of calories.

(d) Do these data meet the conditions required for fitting a least squares line?

Yes. The data appears to have linearity although not perfect, the normal residuals appear to be distributed nomoral. There appears to be a constant variability and the observations are independent.

7.26 Body measurements, Part III.

Exercise 7.15 introduces data on shoulder girth and height of a group of individuals. The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

(a) Write the equation of the regression line for predicting height.

```
# pH = b0 + b1 * shoulder girth

b1 <- 0.67 * (9.41 / 10.37)
b0 <- (b1 * (0 - 107.2)) + 171.14

b1

## [1] 0.6079749

b0

## [1] 105.9651
```

The equation for predicting height is, pH = 105.96 + 0.607 * shoulder girth

(b) Interpret the slope and the intercept in this context.

The intercept is the height in centimeters at girth of 0 cm. The slope is the increase in number of centimeters in height for increase in shoulder girth.

(c) Calculate R2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

```
R2 <- 0.67^2
R2

## [1] 0.4489
```

The linear model explains 45% of the variation in height data.

(d) A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.

```
b0 = 105.9651
b1 = 0.6079749
shoulder_girth = 100

h = b0 + (b1 * shoulder_girth)
h

## [1] 166.7626
```

(e) The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.

```
height = 160
e <- height - 166.7626
e

## [1] -6.7626
```

Residual is -6.7626 It is a negative residual and is helpful in evaluating how well a linear model fits a data set.

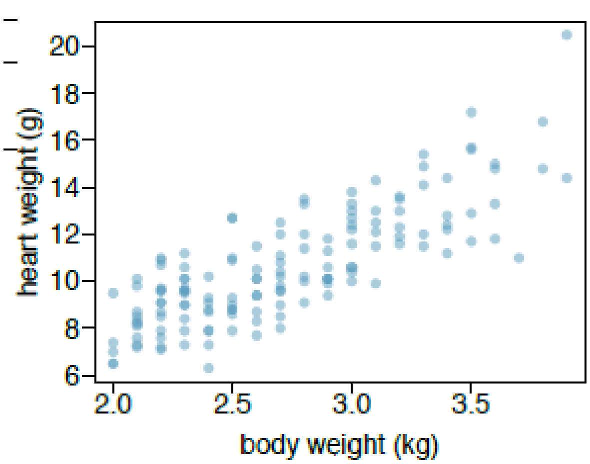
(f) A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?

The scatter plot has shoulder girth starting at 80 cms. It would not be appropriate to use this linear model to predict the height of one year old.

7.30 Cats, Part I.

The following regression output is for predicting the heart weight (in g) of cats from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cats.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.357	0.692	-0.515	0.607
body wt	4.034	0.250	16.119	0.000
$s = 1.452$	$R^2 = 64.66\%$		$R^2_{adj} = 64.41\%$	



(a) Write out the linear model.

cat_heart$_{weight}$grams = -0.357 + (4.034 * cat_body$_{weight}$kilograms)

(b) Interpret the intercept.

The intercept value -0.357 means the model will predict a negative heart weight when the cat's body weight is zero. This is not a meaningful value for the given context.

(c) Interpret the slope.

The slope of 4.035 indicates the change of y value over the x. A 1 kilo change in the body weight would predict a 4.035 gram change in weight of heart.

(d) Interpret R2.

R2 = 64.66%
The linear model describes 64.66% of the variation in the heart weight.

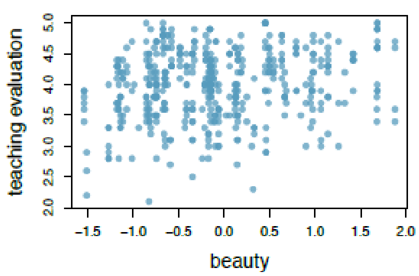
(e) Calculate the correlation coefficient.

```
R2 = 0.6466
r <- sqrt(0.6466)
r

## [1] 0.8041144
```

7.40 Rate my professor.

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching e???ectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. Researchers at University of Texas, Austin collected data on teaching evaluation score (higher score means better) and standardized beauty score (a score of 0 means average, negative score means below average, and a positive score means above average) for a sample of 463 professors. The scatterplot below shows the relationship between these variables, and also provided is a regression output for predicting teaching evaluation score from beauty score.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.010	0.0255	157.21	0.0000
beauty	$\boxed{}$	0.0322	4.13	0.0000

(a) Given that the average standardized beauty score is -0.0883 and average teaching evaluation score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the information provided in the model summary table.

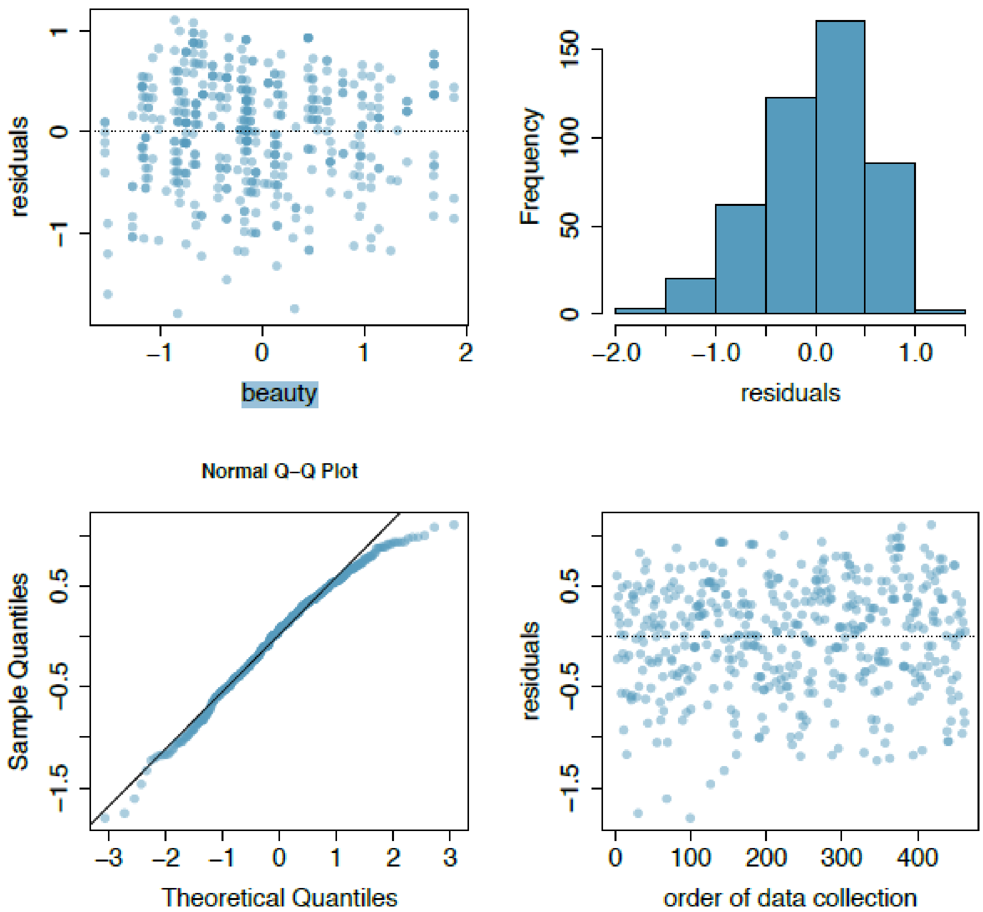
```
b1 <- (3.9983 - 4.010)/-0.0883
b1

## [1] 0.1325028
```

(b) Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.

The data does provide convincing evidence that the slope of the teaching valuation is positive as the p value is 0.000

(c) List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots. beauty



Linearity - There is a no clear trend in the scatterplot. The linearity condition is satisfied without other data supporting the same.

Nearly normal residuals - As shown in the residuals distribution and plot diagram, they are nearly normal.

Constant variability - The variability is roughly constant using the residuals bar chart.

Independent observations - It would be acceptable to assume the observations are independent using the residuals order of data colletion chart.