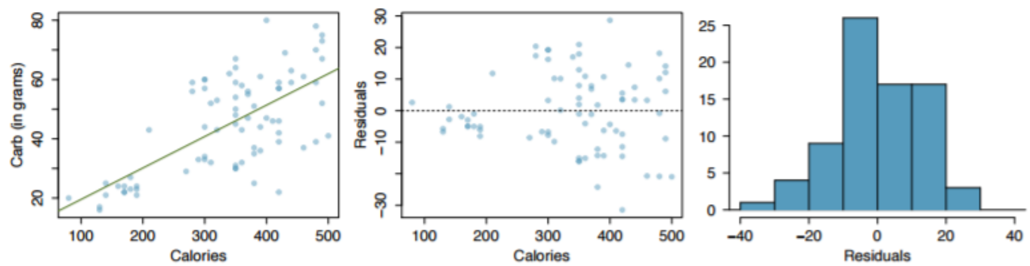
Leland Randles DATA606

Homework, Chapter 7

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.[21] 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.

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

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

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

(a) Fairly linear. More variability on the right of the plot than the left. Not many outliers.

(b) Calories is the explanatory variable and carbohydrates is the response variable.

(c) To predict the amount of carbohydrates in a menu item based on the calories (since only calories are listed by Starbucks).

(d) The linearity condition, the residuals are nearly normal, there is nearly constant variability, and the observations appear to be independent. So, yes.

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.

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

(c) Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

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

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

(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?

(a) You can estimate the slope of the least squares line by taking the standard deviation of shoulder girth divided by the standard deviation of the height and multiplying it by the correlation: (10.37 / 9.41) * 0.67 = 0.738

Then the point estimates can be used to build the equation:

y – 171.14 = 0.738 * (x – 107.2)

y = 0.738x + 92.0264

(b) The y-intercept suggests that someone who is 92.0264 centimeters tall would have a shoulder girth of 0, and that each additional centimeter of shoulder girth would add 0.738 centimeters of height.

(c) $R^2 = 0.67^2 = 0.4489$. This means 44.89% of the variability in height can be explained by shoulder girth.

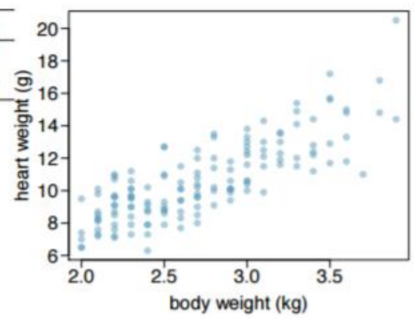
(d) Height = (0.738 * 100) + 92.0264 = 165.8264 centimeters

(e) The residual is 160 – 165.8264, or -5.8264 centimeters. The residual helps determine how well the linear model fits the data set.

(f) No, one should not extrapolate outside the model.

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.

(b) Interpret the intercept.

(c) Interpret the slope.

(d) Interpret R^2.

(e) Calculate the correlation coefficient.

(a) Heart weight = -0.357 + 4.034 * body weight

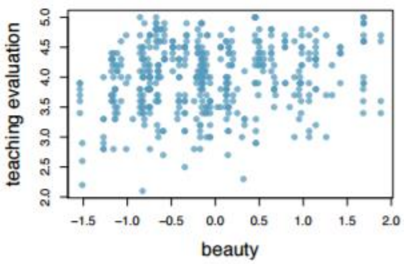
(b) It says that if someone had a body weight of zero kilograms, their heart weight would be -0.357 grams (of course neither is possible).

(c) The slope is telling us that for every one-kilogram increment up or down in body weight, the heart weight increases (or decreases) by 4.034.

(d) The R squared of 64.66% tells us that 64.66% of the variability in heart weight can be explained by body weight.

(e) The correlation coefficient is the sqaure root of R^2, which is 0.804.

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 effectiveness 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.[24] The scatterplot below shows the relationship between these variables, and also provided is a regression output for predicting teaching evaluation score from beauty score.

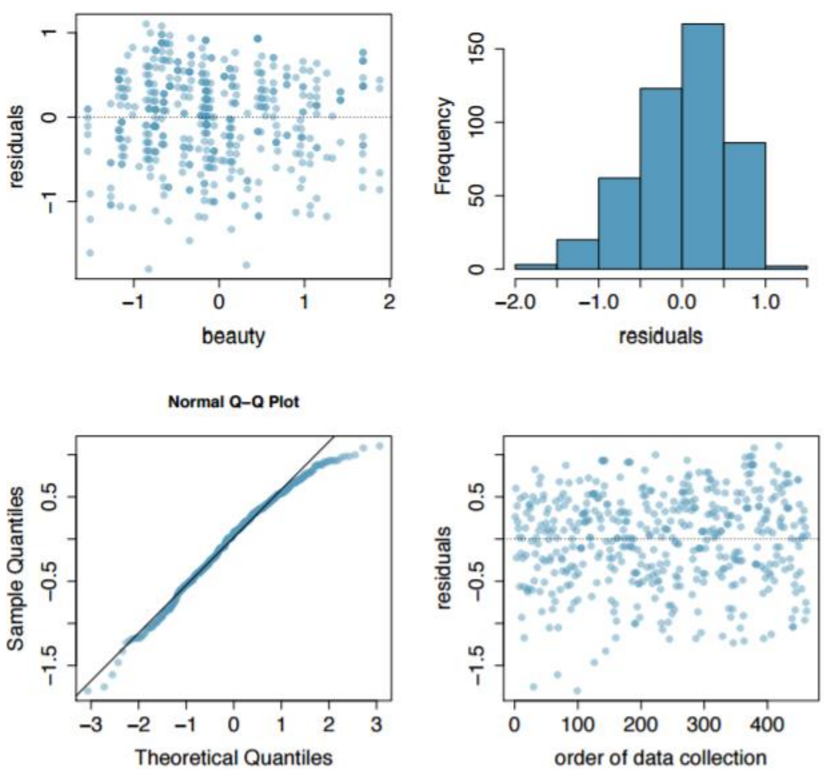


	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.010	0.0255	157.21	0.0000
beauty		0.0322	4.13	0.0000

(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.

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

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



(a) y – (- 0.0883) = ((0.0322 / 0.0255) * *R*) * (x – 3.9983). We don't have the correlation *R*, so not sure how to compute.

(b) The p-values are very low, which suggests beauty DOES have a strong relationship to teacher ratings.

(c) Conditions:

    a. Linearity – Q-Q plot suggests linearity
    b. Nearly normal residuals – some left skew, but pretty close
    c. Constant variability – Looks pretty good on the residual plot
    d. Independent observations – there is no reason to believe the observations were not independent