homework_ch7_JimmyNg

Jimmy Ng

November 1, 2018

Question 7.24

(a) There's a positive relationship (or correlation) between number of calories and amount of carbohydrates (in gram).
(b) The explanatory (or predicting) variable is number of calories, whereas the response (or predicted) variable is the amount of carbohydrates (in gram).
(c) We want to fit a regression line to the data because we want to predict amount of carbohydrates (in gram) from the number of calories from Starbucks food menu.
(d) No, the residual plot (the chart in the middle) shows a pattern - meaning the variance is not equal or we are having an issue of heteroscedasticity.

Question 7.26

(a) slope = (9.41 / 10.37) * 0.67 point_slope_equation = slope(x) - slope*(107.2) + 171.14 regression_line = 0.6079749*(x) + 105.9651

(b) The slope is positive, i.e. 0.6079749, meaning every single unit (cm) of increase in girth would have approximately associated with 0.61 increase in height (cm) plus 105.9651 (cm).

(c) R-squared = 0.67^2 = 0.4489. That means, the independent variable (girth) is accounting for approximately 45% of the variance in the dependent variable (height).

(d) 0.6079749*(x) + 105.9651 0.6079749*(100) + 105.9651 = 166.7626 The study's height is approximatey 166.8cm.

(e) residual = actual - predicted 160cm - 166.8cm = -6.76cm The negative residual indicates that the prediction model overestimates the student's height by 6.76cm.

(f) No, because it is out of predicting range. We should not apply the model to data outside of its range (such as 0cm girth which makes no sense at all).

Question 7.30

(a) regression_line = 4.034(x) - 0.357
(b) that's the base value where x is equal to 0 (which is impossible). If x is equal to 1, then the formula would have predicted the outcome is equal to 4.034*(1) - 0.357 = 3.677
(c) every single unit increase in body weight (in kg) would have 4.034 times increase in heart weight (in g).
(d) 64.66% - that's the percent variance in heart weight that is accounted by body weight. Put it this way, 64.66% of the variance found in heart weight can be explained by body weight.
(e) sqrt(0.6466) = 0.8041144. The correlation coefficient, r, is approximately equal to 0.804.

Question 7.40

(a) y = mx + b 3.9983 = m(-0.0883) + 4.010 m = (3.9983 - 4.010) / -0.0883 m = 0.1325028

(b) Yes, the slope is positive - meaning the two variables go in the same positive direction

(c)

residual plot - checking the variance of residuals, in this case, the plot displays a random pattern and homoscedasticity is assured.

histogram of residuals - checking the nearly-normal-residuals, in this case, it meets the requirement.

qq plot - once again, it's checking the assumption of normal distribution and it is satisfied.

residual plot (order of data collection) - yes it checks out too. There's no dependence/consecutive orders in data collected - unlike something we usually find in time series or markov chain.