Stat 101
Summer 2023 Session 1
Dr. Lily Yen

Activity 8-1

Show all your work

Name: _____
Number: _____
Signature: _____
Score: ___/10

State all spreadsheet functions used.

Problem 1: The spreadsheet under Activity 8 in eLearn contains data on Starbucks food menu regarding the number of calories along with fat, carbohydrates, fibre, and protein in grams in each food item.

Make a scatterplot with calories on the x-axis and fat on the y axis.

a. According to the scatterplot, describe the relationship between the number of calories and the amount of fat (in grams) that Starbucks food menu items contain.

The data show a linear trend, the more calories the more fat or protein.

b. In this scenario, identify the explanatory and the response variables.

We used the number of calories as the explanatory variable and fat, carbohydrates, fibre, and protein content as the response variables.

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

The linear model helps us predict fat or protein contents according to the number of calories. It also helps us see whether the linear relation is strong or weak.

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

Linearity: the data show a linear trend. **Nearly normal residuals**: the data fall around the line without obvious outliers. **Constant variability**: the variance is roughly constant. **Independent observations**: these are not time series observations which are highly correlated, thus, not independent. Therefore, least squares regression can be applied to these data.

e. Use the spreadsheet functions to carry out a regression analysis. From the summary report, answer the following.

(1) Write out the linear model.
(2) Interpret the y-intercept.
(3) Interpret the slope.
(4) Interpret R^2.
(5) Calculate the correlation coefficient.

(1) The linear model for calories vs. fat is $y = -3.48985 + 0.051174x$

(2) The y-intercept is the amount of fat in grams when the food contains no (0) calories. Clearly, this does not make sense.

(3) The slope is the gain in grams per increase in one calories.

(4) An interpretation of $R^2 = 0.582558$ is that the number of calories in a food item explains 58% of the variability in the amount of fat (in grams) for that particular food item.

(5) The correlation coefficient is given by *Multiple R* in the Summary Output of the Regression analysis without the sign. Since the regression line clearly has a positive slope, we conclude that the correlation coefficient $R = 0.763255$.

f. Do the data provide strong evidence that an increase in calories is associated with an increase in fat? State the null and alternative hypotheses, report the p-value, and state your conclusion.

$H_0 : \beta_1 = 0$, $H_A : \beta_1 \neq 0$, p-value is 1.1×10^{-15}, with a $T = 10.1622$. This means that the statistic falls in the rejection region. Thus, we reject the null hypothesis in favour of H_A because the data provide strong evidence that the true slope parameter is different from zero.