In‌ ‌The‌ ‌Name‌ ‌of‌ ‌God‌ ‌

Statistical‌ ‌Inference‌ ‌HW#6 ‌

Amin‌ ‌Asadi‌ Sarijalou‌

810196410‌ ‌

Spring 1400

Contents

[Problem 1 3](#_Toc74601608)

[1-1 3](#_Toc74601609)

[1-2 3](#_Toc74601610)

[a) 3](#_Toc74601611)

[b) 4](#_Toc74601612)

[1-3 4](#_Toc74601613)

[a) 4](#_Toc74601614)

[b) 4](#_Toc74601615)

[Problem 2 4](#_Toc74601616)

[Problem 3 5](#_Toc74601617)

[3-1 5](#_Toc74601618)

[3-2 5](#_Toc74601619)

[Problem 4 5](#_Toc74601620)

[a) 5](#_Toc74601621)

[b) 6](#_Toc74601622)

[Problem 5 6](#_Toc74601623)

[a. 6](#_Toc74601624)

[b. 7](#_Toc74601625)

[c. 7](#_Toc74601626)

[d. 7](#_Toc74601627)

[e. 7](#_Toc74601628)

[f. 7](#_Toc74601629)

[Problem 6 8](#_Toc74601630)

[a) 8](#_Toc74601631)

[b) 9](#_Toc74601632)

[c) 10](#_Toc74601633)

[d) 11](#_Toc74601634)

[e) 12](#_Toc74601635)

[f) 13](#_Toc74601636)

# Problem 1

## 1-1

## 1-2

### a)

* Linearity
  + relationship between the explanatory and response variable should be linear
* Nearly Normal Residuals:
  + residuals should be nearly normally distributed
* Constant variability
  + variability of points around the least square line should be roughly constant

### b)

## 1-3

### a)

True

### b)

False, it is a measure of (the strength of linear) association between two numerical variables.

# Problem 2

The outlier is point: (5, 160),

because it falls away from the cloud of points.

Effects:

* The correlation coefficient (R) will increase and get closer to 1.
* (y-intercept) of the regression line will decrease and slope will increase.
* The standard deviation of the residuals will decrease.

# Problem 3

### 3-1

No, because p-value for slope is 0.660 which is (much) greater than significance level (0.01) so we fail to reject and hence there isn’t enough evidence to claim that there is a linear relationship between these variables for all players in this population.

### 3-2

a is the correct choice

Since so result of the p-value method will be equivalent to the result of 95% confidence interval method.

Therefore, because the Null Value = 0 falls into the confidence interval, we fail to reject and hence Nader can’t conclude a linear relationship between how much toothpaste he uses and how long he brushes.

# Problem 4

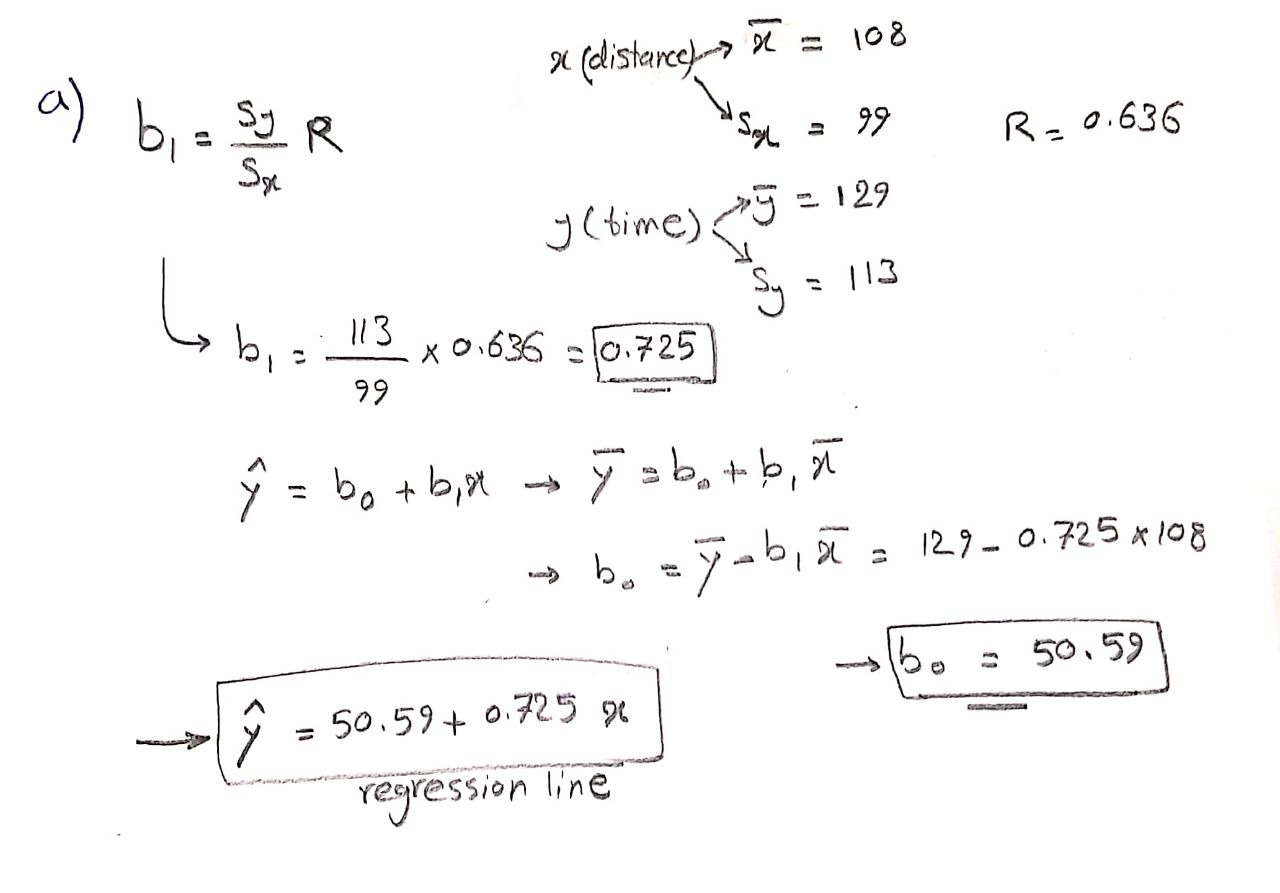
### a)

### b)

Yes. According the given table, for a two-tailed test, the p-value is 0. Here we need a one-tailed test. Since the p-value for a one-tailed test is ½ of that of two-tailed test, and because half of the two-tailed p-value is still 0, so we have sufficient evidence to reject in favor of and conclude that the slope is positive.

# Problem 5

## a.



## b.

Interpretation of Slop: For each 1 mile increase in the distance between two stops, we would expect that the time of travelling between those points, to be more on average by 0.725 minutes.

Interpretation of Intercept: For two stops having zero distance between each other, it takes on average 50.59 minutes to travel between them.

## c.

It tells us that about 40% of the variability in the time of travel between stops is explained by the model (which include distance between stops as the variable).

Also 1 - 0.40 = 60% of the variability in the time of travel between stops is explained by other variables not included in the model.

## d.

## e.

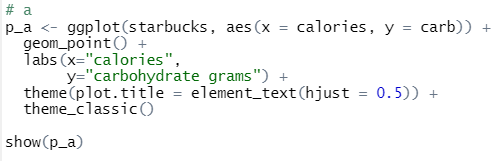
The residual is positive. This means that we have underestimated the response variable (time of travel) by about 42.62 minutes.

## f.

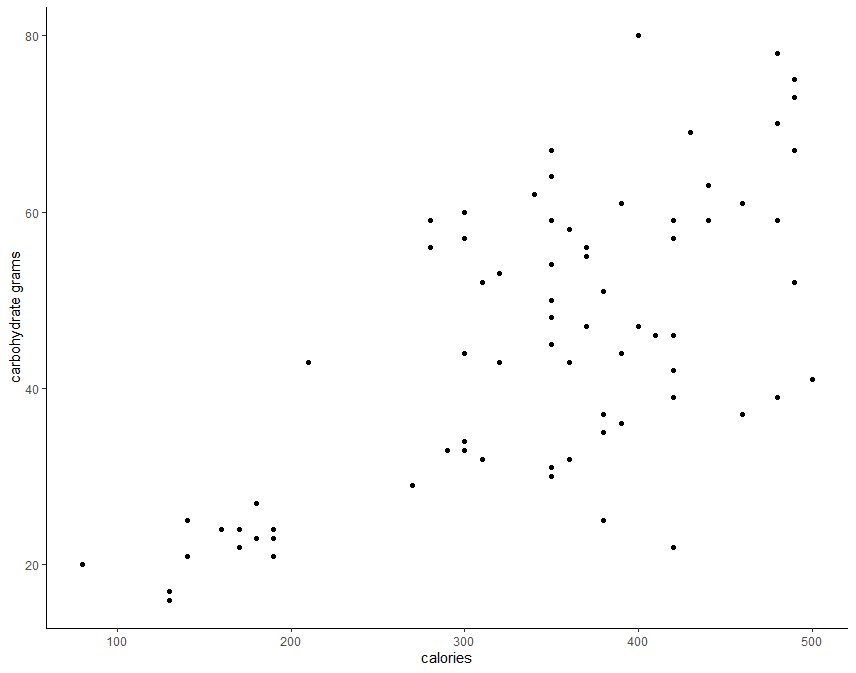
No, we cannot. Because in fitting out model, we don’t consider this much distant points. 500 miles distance, is a very long distance because it is standard deviation away from the mean of the distances which we consider.

# Problem 6

## a)



**The Data Points:**



Scatterplot Description: I see that there seem to be a positive linear association between Carbohydrate Grams and number of Calories contained in each food item.

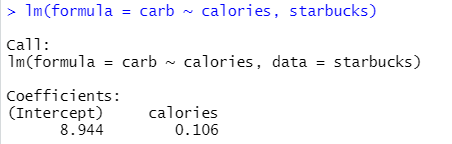
## b)

Explanatory Variable: Calories

Response Variable: Carbohydrate Grams

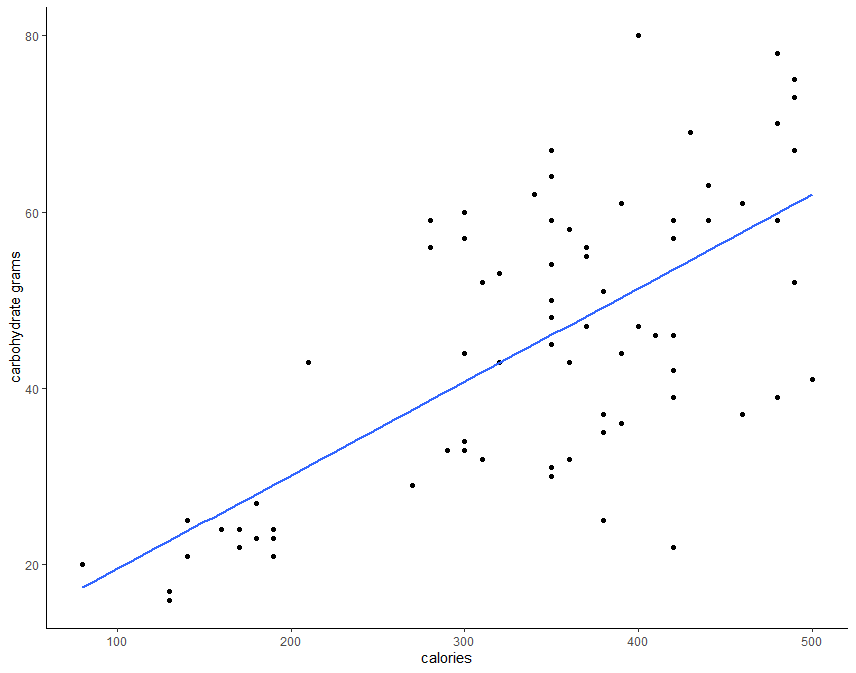
We construct the problem in this way because we want to fit a model in order to predict the amount of Carbohydrate based on number of Calories.

## c)





**The Fitted Regression Line:**

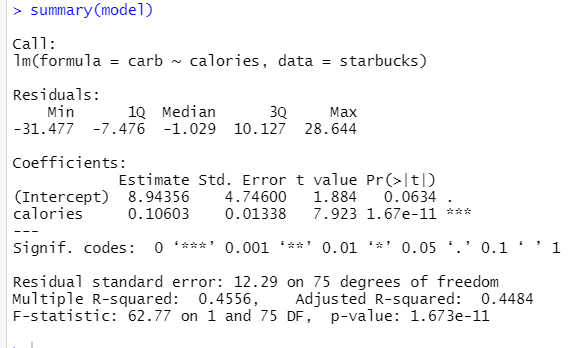


## d)

Interpretation of Slop: With increasing the number of Calories by 1, we would expect that the amount of Carbohydrate increases on average by 0.106 grams.

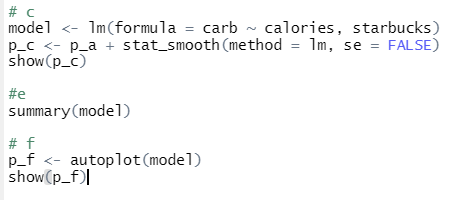
Interpretation of Intercept: If a food has zero Calories, we would expect that it has on average grams of Carbohydrate.

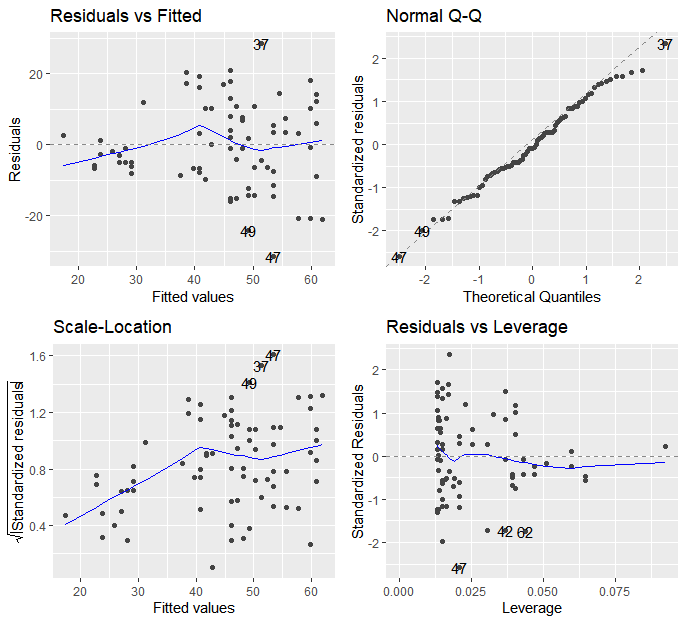
## e)



As we see in the summary of the fitted model, which means that about 45% of the variability of the response variable (grams of Carbohydrate) is explained by the model (which includes number of Calories as the variable).

## f)





As we see in the Residuals Vs Fitted plot, the residuals seem to be different from the normal distribution at the tails. We can check that in the QQ-Plot beside it.

The QQ-Plot tells us that the same thing. The residuals are not normally distributed at tails.

We conclude that the fitted model is a good predictor in the middle parts, but it isn’t a good fit at the tails.