Week 2\_Workbook

Odelia

3/3/2021

4+5 [1] 9

3^3 [1] 27

sqrt(324) [1] 18

max(iris$Sepal.Length) [1] 7.9

min(iris$Sepal.Length) [1] 4.3

Difference between flower with the longest and shortest Sepal.Length: > 7.9-4.3 [1] 3.6

My dataframe: > This <- 1:100 > is <- c(“Please”) > my <- c(“help”) > plea <- c(“me”) > to <- c(“with”) > you <- c(“this”) > God <- c(“subject”) > df <- data.frame(This,is,my,plea,to,you,God) > df

To rename columns: names(df)[names(df) == ‘old.var.name’] <- ‘new.var.name’ *sub in variable names as necessary*

lm(height ~ weight, data = women) Call: lm(formula = height ~ weight, data = women) Coefficients: (Intercept) weight  
25.7235 0.2872  
Hence, the linear model will be: Height = 0.2872(Weight) + 25.7235

Table for this model: x <- lm(formula = height ~ weight, data = women) sjPlot::tab\_model(x)

Coefficient plot for this model: sjPlot::plot\_model(x)

Prediction plot for this model: p1 <- ggeffects::ggpredict(x) plot(p1)

Explanation for what is being calculated here: sum(womenweight) = 0.4

This equation is calculating the proportion of women whose weights are above 140 against the weights of all women, and expressing this as a fraction of 1. The sum in the equation is meant to sum up all values that are present in the argument as indicated within the brackets. (womenweight) sets the length of the vector to be the weights of all women. Hence, the percentage of the number of women with weights > 140, in relation to the weights of all women, is 0.4 / 40%.

Mean weight of women : mean(women$weight) = 136.7333

Proportion of women who are over the mean weight of women: sum(womenweight))) / length(women$weight) = 0.4666667

What are the advantages and disadvantages of creating more breaks in the Petal.Length indicator? Clarify your answer in a paragraph.

Breaks help to retain alot of information that are lost in histograms with lesser or no breaks at all. Breaks can mitigate the effects of significant outliers that skew coefficients as it is observable in the histogram that the relationship between petal length and frequency is not a smooth one. Breaks can provide the observer with more confidence in the accuracy of the histogram and by extension the predictor variable (petal.length). Breaks also help to identify the petal.length histogram as a multimodal distribution more distinctly. However, breaks also have its disadvantages by causing alot of noise and cluttering the histogram. Breaks also turn a continuous variable into a categorical variable (petal length). Hence, deviant data points are magnified and distort what should have been a smooth relationship between petal length and frequency, which then leads to downward bias in estimation of relationship between the variables, inflation of error variance, and consequently low predictive power of the predictor variable.

Error in this code: mh <- mean(womenweight > mh) / length(womenweight)] where “weight” should have been “height”.

To reorder columns of women dataset for weight to come before height: wdf <- data.frame(womenheight)

To rename columns “w” and “h”: names(wdf)[names(wdf) == ‘women.weight’] <- ‘w’, names(wdf)[names(wdf) == ‘women.height’] <- ‘h’

Read data into R: library(“readr”); testdata<- readr::read\_csv(url(“<https://raw.githubusercontent.com/go-bayes/psych-447/main/data/testdata1.csv>”)); str(testdata)

Save data into data folder: dir.create(“data”); saveRDS(testdata, here::here(“data”, “td.RDS”))

Read data back into R: td <- readRDS(here::here(“data”, “td.RDS”)); str(td)

Write linear model using td dataset: > lm(formula = height ~ weight, data = td) Call: lm(formula = height ~ weight, data = td) Coefficients: (Intercept) weight  
12.74 1.87

The linear model will be: Height = 1.87(weight) + 12.74

Create a coefficient plot: p2 <- lm(formula = height ~ weight, data = td) sjPlot::plot\_model(p2)

Create a prediction plot: p3 <- ggeffects::ggpredict(p2) plot(p3)

How would you interpret the intercept in this model?

The intercept in this model is the mean value of the dependent variable (height) after accounting for all residuals and error variances inherent in the linear regression model in relation to the predictor variable (weight). It is a constant value that is not affected by the value weight, and is meant to account for residuals and error variances in the model.