ASSIGNMENT 3

Salma Elhassa

2/25/2022

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed R codes and figures.

## Read the compensation data set

library(readr)  
setwd("~/Desktop/BIO 47120/datasets-master")  
compensation <- read\_csv("compensation.csv")

## Rows: 40 Columns: 3  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Grazing  
## dbl (2): Root, Fruit  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

dim(compensation)

## [1] 40 3

str(compensation, give.attr = FALSE)

## spec\_tbl\_df [40 × 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Root : num [1:40] 6.22 6.49 4.92 5.13 5.42 ...  
## $ Fruit : num [1:40] 59.8 61 14.7 19.3 34.2 ...  
## $ Grazing: chr [1:40] "Ungrazed" "Ungrazed" "Ungrazed" "Ungrazed" ...

##There are 3 variables and 40 observations. Root and Fruit variables are numeric type (floats) and the Grazing variable is a categorical type (string).

head(compensation) # returns the first 6 rows (observations) of the compensation data frame

## # A tibble: 6 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 6.22 59.8 Ungrazed  
## 2 6.49 61.0 Ungrazed  
## 3 4.92 14.7 Ungrazed  
## 4 5.13 19.3 Ungrazed  
## 5 5.42 34.2 Ungrazed  
## 6 5.36 35.5 Ungrazed

tail(compensation) # returns the last 6 rows (observations) of the compensation data frame

## # A tibble: 6 × 3  
## Root Fruit Grazing  
## <dbl> <dbl> <chr>   
## 1 9.35 98.5 Grazed   
## 2 7.07 40.2 Grazed   
## 3 8.16 52.3 Grazed   
## 4 7.38 46.6 Grazed   
## 5 8.52 71.0 Grazed   
## 6 8.53 83.0 Grazed

dim(compensation) # returns the dimensions of the data frame, rows x columns

## [1] 40 3

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

glimpse(compensation) # returns a transposed version of the data frame, up to some display limit. Also shows number of rows and columns and variable types.

## Rows: 40  
## Columns: 3  
## $ Root <dbl> 6.225, 6.487, 4.919, 5.130, 5.417, 5.359, 7.614, 6.352, 4.975,…  
## $ Fruit <dbl> 59.77, 60.98, 14.73, 19.28, 34.25, 35.53, 87.73, 63.21, 24.25,…  
## $ Grazing <chr> "Ungrazed", "Ungrazed", "Ungrazed", "Ungrazed", "Ungrazed", "U…

summary(compensation) # returns summary statistics for the compensation data frame. In this case it returns the quartiles, median, mean, min and max

## Root Fruit Grazing   
## Min. : 4.426 Min. : 14.73 Length:40   
## 1st Qu.: 6.083 1st Qu.: 41.15 Class :character   
## Median : 7.123 Median : 60.88 Mode :character   
## Mean : 7.181 Mean : 59.41   
## 3rd Qu.: 8.510 3rd Qu.: 76.19   
## Max. :10.253 Max. :116.05

##choose colums

select(compensation, Fruit)

## # A tibble: 40 × 1  
## Fruit  
## <dbl>  
## 1 59.8  
## 2 61.0  
## 3 14.7  
## 4 19.3  
## 5 34.2  
## 6 35.5  
## 7 87.7  
## 8 63.2  
## 9 24.2  
## 10 64.3  
## # … with 30 more rows

##choose rows

slice(compensation, 2:10)

## # A tibble: 9 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 6.49 61.0 Ungrazed  
## 2 4.92 14.7 Ungrazed  
## 3 5.13 19.3 Ungrazed  
## 4 5.42 34.2 Ungrazed  
## 5 5.36 35.5 Ungrazed  
## 6 7.61 87.7 Ungrazed  
## 7 6.35 63.2 Ungrazed  
## 8 4.97 24.2 Ungrazed  
## 9 6.93 64.3 Ungrazed

slice(compensation, c(2, 3, 10))

## # A tibble: 3 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 6.49 61.0 Ungrazed  
## 2 4.92 14.7 Ungrazed  
## 3 6.93 64.3 Ungrazed

##find the rows where it is true that Fruit is >80 return ##them as a data frame

filter(compensation, Fruit > 80)

## # A tibble: 9 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 7.61 87.7 Ungrazed  
## 2 7.00 80.6 Ungrazed  
## 3 10.3 116. Grazed   
## 4 9.04 84.4 Grazed   
## 5 8.99 80.3 Grazed   
## 6 8.98 82.4 Grazed   
## 7 9.84 105. Grazed   
## 8 9.35 98.5 Grazed   
## 9 8.53 83.0 Grazed

#Set the categories for row filter to "lo\_hi\_fruit"  
#conditional row filtering  
lo\_hi\_fruit <- filter(compensation, Fruit > 80 | Fruit < 20)  
lo\_hi\_fruit

## # A tibble: 13 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 4.92 14.7 Ungrazed  
## 2 5.13 19.3 Ungrazed  
## 3 7.61 87.7 Ungrazed  
## 4 7.00 80.6 Ungrazed  
## 5 4.43 18.9 Ungrazed  
## 6 10.3 116. Grazed   
## 7 9.04 84.4 Grazed   
## 8 6.11 15.0 Grazed   
## 9 8.99 80.3 Grazed   
## 10 8.98 82.4 Grazed   
## 11 9.84 105. Grazed   
## 12 9.35 98.5 Grazed   
## 13 8.53 83.0 Grazed

filter(compensation, Grazing == "Ungrazed")

## # A tibble: 20 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 6.22 59.8 Ungrazed  
## 2 6.49 61.0 Ungrazed  
## 3 4.92 14.7 Ungrazed  
## 4 5.13 19.3 Ungrazed  
## 5 5.42 34.2 Ungrazed  
## 6 5.36 35.5 Ungrazed  
## 7 7.61 87.7 Ungrazed  
## 8 6.35 63.2 Ungrazed  
## 9 4.97 24.2 Ungrazed  
## 10 6.93 64.3 Ungrazed  
## 11 6.25 52.9 Ungrazed  
## 12 5.45 32.4 Ungrazed  
## 13 6.01 53.6 Ungrazed  
## 14 5.93 54.9 Ungrazed  
## 15 6.26 64.8 Ungrazed  
## 16 7.18 73.2 Ungrazed  
## 17 7.00 80.6 Ungrazed  
## 18 4.43 18.9 Ungrazed  
## 19 7.30 75.5 Ungrazed  
## 20 5.84 46.7 Ungrazed

### returns the first 6 rows (observations) of the compensation data frame

head(compensation)

## # A tibble: 6 × 3  
## Root Fruit Grazing   
## <dbl> <dbl> <chr>   
## 1 6.22 59.8 Ungrazed  
## 2 6.49 61.0 Ungrazed  
## 3 4.92 14.7 Ungrazed  
## 4 5.13 19.3 Ungrazed  
## 5 5.42 34.2 Ungrazed  
## 6 5.36 35.5 Ungrazed

compensation <- mutate(compensation, logFruit = log(Fruit))  
head(compensation)

## # A tibble: 6 × 4  
## Root Fruit Grazing logFruit  
## <dbl> <dbl> <chr> <dbl>  
## 1 6.22 59.8 Ungrazed 4.09  
## 2 6.49 61.0 Ungrazed 4.11  
## 3 4.92 14.7 Ungrazed 2.69  
## 4 5.13 19.3 Ungrazed 2.96  
## 5 5.42 34.2 Ungrazed 3.53  
## 6 5.36 35.5 Ungrazed 3.57

##sort rows

arrange(compensation, Fruit)

## # A tibble: 40 × 4  
## Root Fruit Grazing logFruit  
## <dbl> <dbl> <chr> <dbl>  
## 1 4.92 14.7 Ungrazed 2.69  
## 2 6.11 15.0 Grazed 2.70  
## 3 4.43 18.9 Ungrazed 2.94  
## 4 5.13 19.3 Ungrazed 2.96  
## 5 4.97 24.2 Ungrazed 3.19  
## 6 5.45 32.4 Ungrazed 3.48  
## 7 5.42 34.2 Ungrazed 3.53  
## 8 5.36 35.5 Ungrazed 3.57  
## 9 6.96 38.9 Grazed 3.66  
## 10 7.07 40.2 Grazed 3.69  
## # … with 30 more rows

##chaining with %>% ##provides a mechanism for chaining commands with a new forward-pipe operator

compensation %>%  
 filter(Fruit > 80) %>%  
 select(Root)

## # A tibble: 9 × 1  
## Root  
## <dbl>  
## 1 7.61  
## 2 7.00  
## 3 10.3   
## 4 9.04  
## 5 8.99  
## 6 8.98  
## 7 9.84  
## 8 9.35  
## 9 8.53

##summarize by groups

compensation %>%  
 group\_by(Grazing) %>%  
 summarise(meanFruit = mean(Fruit))

## # A tibble: 2 × 2  
## Grazing meanFruit  
## <chr> <dbl>  
## 1 Grazed 67.9  
## 2 Ungrazed 50.9

compensation %>%  
 group\_by(Grazing) %>%  
 summarise(  
 meanFruit = mean(Fruit),  
 sdFruit = sd(Fruit))

## # A tibble: 2 × 3  
## Grazing meanFruit sdFruit  
## <chr> <dbl> <dbl>  
## 1 Grazed 67.9 25.0  
## 2 Ungrazed 50.9 21.8

##transform by groups

compensation\_mean\_centred <- compensation %>%  
 group\_by(Grazing) %>%  
 mutate(Fruit\_minus\_mean = Fruit - mean(Fruit))

##regression by groups

library(broom)  
compensation\_lms <- compensation %>%  
 group\_by(Grazing) %>%  
 do(tidy(lm(Fruit ~ Root, data=.)))