Advanced Bioinformatics 2023 assessment

M2108433

2023-04-05

Task 3.1.

sum(5:55)

## [1] 1530

Task 3.2.

sumfun <- function(n) {  
 sum(5:n)  
}  
sumfun(10)

## [1] 45

sumfun(20)

## [1] 200

sumfun(100)

## [1] 5040

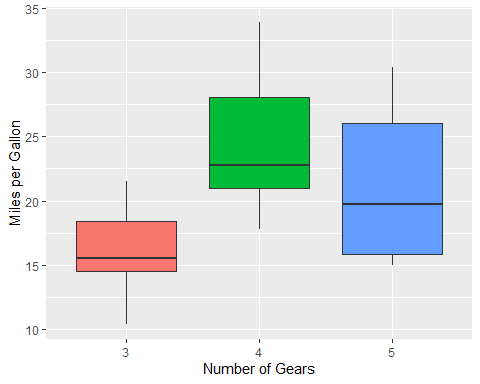
Task 3.3.

fibonacci <- numeric(12)  
fibonacci[1] <- 1  
fibonacci[2] <- 1  
for (i in 3:12) {  
 fibonacci[i] <- fibonacci[i-1] + fibonacci[i-2]  
}  
  
print(fibonacci)

## [1] 1 1 2 3 5 8 13 21 34 55 89 144

Task 3.4.

library(ggplot2)  
data(mtcars)  
ggplot(data=mtcars, aes(x=factor(gear), y= mpg, fill=factor(gear))) +  
geom\_boxplot() +  
labs(x = "Number of Gears", y = "Miles per Gallon") +  
theme(legend.position = "none")



Task 3.5.

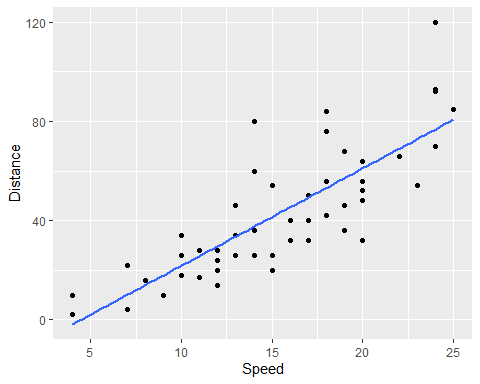
data(cars)  
cars\_regression <- lm(data = cars, dist ~ speed)  
  
summary(cars\_regression)

##   
## Call:  
## lm(formula = dist ~ speed, data = cars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -29.069 -9.525 -2.272 9.215 43.201   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -17.5791 6.7584 -2.601 0.0123 \*   
## speed 3.9324 0.4155 9.464 1.49e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.38 on 48 degrees of freedom  
## Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438   
## F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

Task 3.6.

library(ggplot2)  
data(cars)  
ggplot(data=cars, aes(x=speed, y=dist)) +  
geom\_point() +  
geom\_smooth(method='lm', se = FALSE) +  
labs(x = "Speed", y = "Distance")

## `geom\_smooth()` using formula = 'y ~ x'



Task 3.7.

breaking\_distance <- cars$speed^2  
  
reaction\_time\_regression <- lm(data = cars, dist ~ breaking\_distance)  
  
reaction\_time\_regression

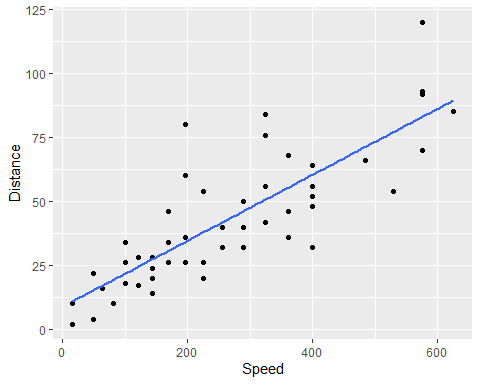
##   
## Call:  
## lm(formula = dist ~ breaking\_distance, data = cars)  
##   
## Coefficients:  
## (Intercept) breaking\_distance   
## 8.860 0.129

reaction\_time <- 0.129\*2   
coverted\_reaction\_time <- reaction\_time/3600  
  
coverted\_reaction\_time

## [1] 7.166667e-05

library(ggplot2)  
data("cars")  
ggplot(data=cars, aes(x=breaking\_distance, y=dist)) +  
geom\_point() +  
geom\_smooth(method='lm', se = FALSE) +  
labs(x = "Speed", y = "Distance")

## `geom\_smooth()` using formula = 'y ~ x'



RNA-seq assessment

Task 3.8.

counts <- read.csv(file = "C:/Users/diogo/OneDrive/Ambiente de Trabalho/RNAseq/exercise1\_counts.csv", header = T, row.names = 1)  
head(counts)

## control\_FFa1.bam control\_FFa2.bam control\_FFa3.bam mutant\_KOa1.bam  
## 497097 16 16 0 21  
## 100503874 20 0 0 64  
## 100038431 0 0 2 0  
## 19888 11 0 10 113  
## 20671 14 16 0 40  
## 27395 465 193 596 436  
## mutant\_KOa2.bam mutant\_KOa3.bam mutant\_KOb1.bam mutant\_KOb2.bam  
## 497097 16 27 20 0  
## 100503874 0 4 5 0  
## 100038431 0 8 0 0  
## 19888 0 26 14 6  
## 20671 8 33 33 12  
## 27395 686 572 1378 1901  
## mutant\_KOb3.bam  
## 497097 2  
## 100503874 2  
## 100038431 0  
## 19888 16  
## 20671 24  
## 27395 1553

sam\_des <- read.table("C:/Users/diogo/OneDrive/Ambiente de Trabalho/RNAseq/exercise1\_sample\_description.info", sep = "\t", header = TRUE)  
head(sam\_des)

## filename sample condition batch  
## 1 control\_FFa1.bam FFa1 FFa a  
## 2 control\_FFa2.bam FFa2 FFa b  
## 3 control\_FFa3.bam FFa3 FFa c  
## 4 TMC\_mutant\_KOa1.bam KOa1 KOa a  
## 5 TMC\_mutant\_KOa2.bam KOa2 KOa b  
## 6 TMC\_mutant\_KOa3.bam KOa3 KOa c

Task 3.9.

col\_data <- data.frame(sample = sam\_des$sample, condition = sam\_des$condition, batch = sam\_des$batch)  
head(col\_data)

## sample condition batch  
## 1 FFa1 FFa a  
## 2 FFa2 FFa b  
## 3 FFa3 FFa c  
## 4 KOa1 KOa a  
## 5 KOa2 KOa b  
## 6 KOa3 KOa c

all(colnames(counts) == col\_data$name)

## [1] TRUE

task 3.10

suppressPackageStartupMessages(library(DESeq2))  
  
dds <- DESeqDataSetFromMatrix(countData = counts, colData = col\_data, design =~ condition)

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in  
## design formula are characters, converting to factors

dds

## class: DESeqDataSet   
## dim: 26301 9   
## metadata(1): version  
## assays(1): counts  
## rownames(26301): 497097 100503874 ... 100040384 100040400  
## rowData names(0):  
## colnames(9): control\_FFa1.bam control\_FFa2.bam ... mutant\_KOb2.bam  
## mutant\_KOb3.bam  
## colData names(3): sample condition batch

Task 3.11.

rld <- rlog(dds)  
vst <- varianceStabilizingTransformation(dds)

Task 3.12. Task 3.13. Task 3.14. Task 3.15.

ChIP-seq assessment

Task 3.16. Task 3.17. Task 3.18. Task 3.19.