巨量工具期末報告

S05490061\_林姿岑\_資管四

library(readr)

abalone<- read\_csv("abalone.data", col\_names = FALSE)

View(abalone)

# Sex / nominal / -- / M, F, and I (infant) 性別/名義/-/ M，F和I（嬰兒）

# Length / continuous / mm / Longest shell measurement長度/連續/毫米/最長外殼尺寸

# Diameter / continuous / mm / perpendicular to length直徑/連續/毫米/垂直於長度

# Height / continuous / mm / with meat in shell高度/連續/毫米/帶外殼

# Whole weight / continuous / grams / whole abalone整重/連續/克/整隻鮑魚

# Shucked weight / continuous / grams / weight of meat去皮重量/連續重量/克/肉重Viscera weight / continuous / grams / gut weight (after bleeding) 內臟重量/連續重量/克/腸重量（出血後）

# Shell weight / continuous / grams / after being dried殼重/連續/克/乾後

# Rings / integer / -- / +1.5 gives the age in years環/整數/-/ +1.5給出年齡

abaloneBackUp <- abalone

names(abalone)

# [1] "X1" "X2" "X3" "X4" "X5" "X6" "X7" "X8" "X9"

names(abalone) <-c("Sex","Length","Diameter","Height","WholeWeight", "ShuckedWeight", "VisceraWeight", "ShellWeight", "Rings")

names(abalone)

# [1] "Sex" "Length" "Diameter" "Height" "WholeWeight" "ShuckedWeight"

# [7] "VisceraWeight" "ShellWeight" "Rings"

install.packages("dplyr")

library(dplyr)

#使用select()函式選擇要分析的欄位

abaloneNumType <- select(abalone,Length:ShellWeight)#選出Length到ShellWeight之間的所有變數

names(abaloneNumType)

# [1] "Length" "Diameter" "Height" "WholeWeight" "ShuckedWeight" "VisceraWeight"

# [7] "ShellWeight"

#使用filter()函式選要分析的觀察值

(abaloneFilter01 = filter(abalone, Sex == "I"))

#從性別的變數取I出來

# Sex Length Diameter Height WholeWeight ShuckedWeight VisceraWeight ShellWeight Rings

# <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>

# 1 I 0.33 0.255 0.08 0.205 0.0895 0.0395 0.055 7

# 2 I 0.425 0.3 0.095 0.352 0.141 0.0775 0.12 8

# 3 I 0.355 0.28 0.085 0.290 0.095 0.0395 0.115 7

# 4 I 0.38 0.275 0.1 0.226 0.08 0.049 0.085 10

# 5 I 0.24 0.175 0.045 0.07 0.0315 0.0235 0.02 5

# 6 I 0.205 0.15 0.055 0.042 0.0255 0.015 0.012 5

(abaloneFilter02 = filter(abalone, WholeWeight >= mean(WholeWeight) ))

#找出總重量大於等於平均的資料(所有鮑魚的總重量)

# Sex Length Diameter Height WholeWeight ShuckedWeight VisceraWeight ShellWeight Rings

# <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>

# 1 F 0.55 0.44 0.15 0.894 0.314 0.151 0.32 19

# 2 F 0.565 0.44 0.155 0.940 0.428 0.214 0.27 12

# 3 F 0.615 0.48 0.165 1.16 0.513 0.301 0.305 10

# 4 F 0.56 0.44 0.14 0.928 0.382 0.188 0.3 11

#使用mutate()增加新變數欄位

abaloneMutate <- mutate(abalone, LengthVsRings = Length / Rings)

#計算增加新的變數abaloneMutate=總長度除以環數

names(abaloneMutate)

# [1] "Sex" "Length" "Diameter" "Height" "WholeWeight"

# [6] "ShuckedWeight" "VisceraWeight" "ShellWeight" "Rings" "LengthVsRings"

#使用group\_by()函數的功能為設定分組依據，並與summarise()函式合併使用

abaloneMutate %>% group\_by(Sex) %>% summarise(count = n(), WholeLengthMean = mean(Length))

#把用summarise()函式產生出來的資料集依據性別丟給group\_by排序,接著做出個數及平均值出來

# A tibble: 3 x 3

# Sex count WholeLengthMean

# <chr> <int> <dbl>

# 1 F 1307 0.579

# 2 I 1342 0.428

# 3 M 1528 0.561

#4.用ggplot2畫圖

install.packages("ggplot2")

library(ggplot2)

qplot(Diameter,Height,data=abalone,color=Sex)

#直徑跟高度的散佈圖並依據性別來做區分可以發現兩者為正相關r

qplot(Length, data = abalone, fill = Sex)

#長度直方圖依據性別區分

ggplot(abalone, aes(x = abalone$Diameter, y = abalone$WholeWeight)) +geom\_boxplot()

#使用盒狀圖表達直徑跟重量的關系

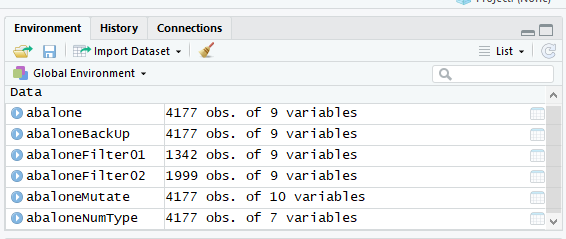
#5 rattle畫5個圖

install.packages("rattle")

install.packages("RGtk2")

library(RGtk2)

library(rattle)

rattle()

巨量工具期末報告書面資料

鮑魚資料集資訊說明：

此資料集內容為通過物理測量預測鮑魚的年齡。鮑魚的年齡是通過將殼切成圓錐形，對其進行染色並通過顯微鏡對其進行計數來確定的，這是一項無聊且耗時的工作。其他更容易獲得的測量值用於預測年齡。要解決該問題，可能需要更多信息，例如天氣模式和位置​​（因此可獲取食物）。

從原始數據中刪除了缺少值的示例（大多數缺少預測值的示例），並且對連續值的範圍進行了縮放以用於ANN（除以200）。

屬性訊息：

資料中給出了屬性名稱，屬性類型，度量單位和簡要說明。環數是要預測的值：可以是連續值，也可以是分類問題。

Name / Data Type / Measurement Unit / Description

-----------------------------

Sex / nominal / -- / M, F, and I (infant) 性別/名義/-/ M，F和I（嬰兒）

Length / continuous / mm / Longest shell measurement長度/連續/毫米/最長外殼尺寸

Diameter / continuous / mm / perpendicular to length直徑/連續/毫米/垂直於長度

Height / continuous / mm / with meat in shell高度/連續/毫米/帶外殼

Whole weight / continuous / grams / whole abalone整重/連續/克/整隻鮑魚

Shucked weight / continuous / grams / weight of meat去皮重量/連續重量/克/肉重Viscera weight / continuous / grams / gut weight (after bleeding) 內臟重量/連續重量/克/腸重量（出血後）

Shell weight / continuous / grams / after being dried殼重/連續/克/乾後

Rings / integer / -- / +1.5 gives the age in years環/整數/-/ +1.5給出年齡

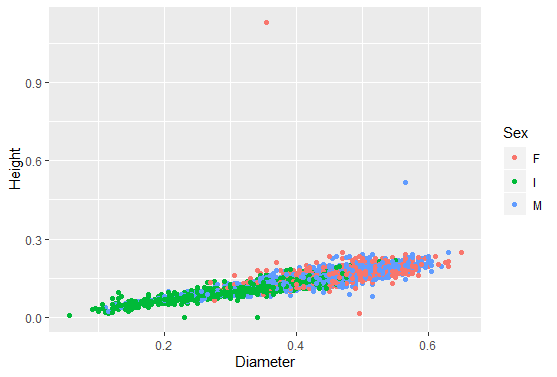


Figure 1：#直徑跟高度的散佈圖並依據性別來做區分可以發現兩者為正相關

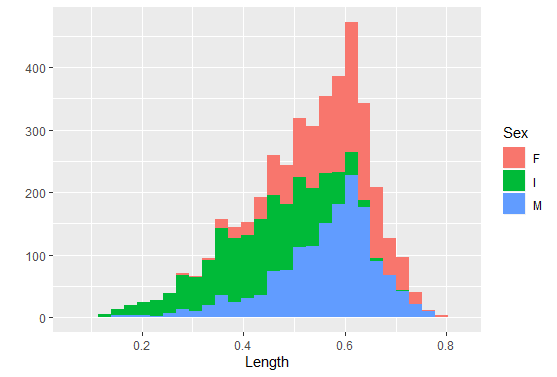


Figure 2：長度直方圖依據性別區分，不同性別鮑魚之長度統計

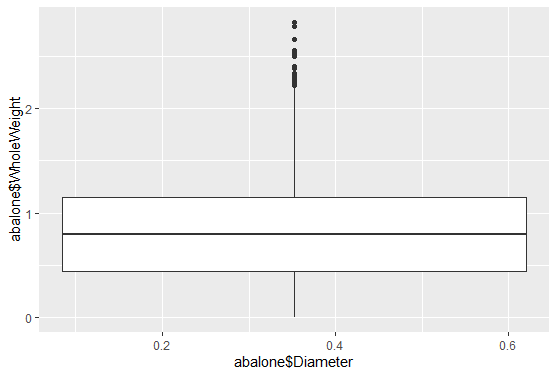


Figure ：#使用盒狀圖表達直徑跟重量的關係

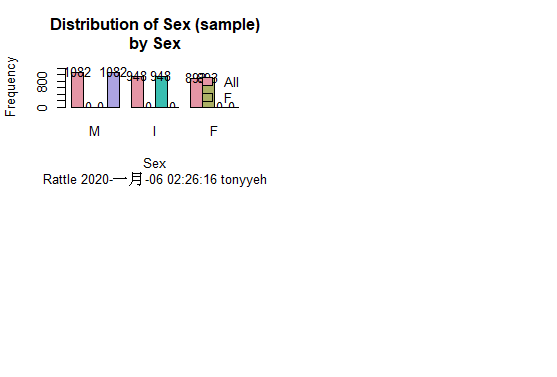


Figure 4：鮑魚性別累計分佈圖

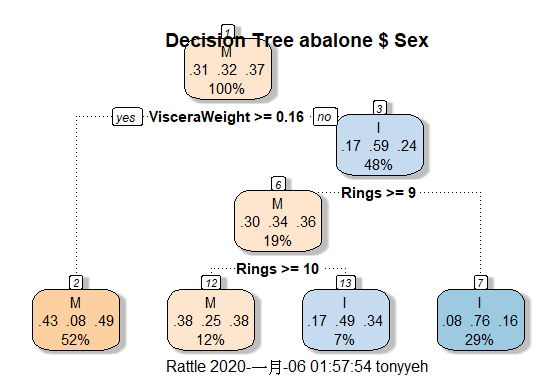


Figure 5：使用鮑魚資料集進行決策樹分析，可以看出以性別為區分變數時，52%性別為M的鮑魚內臟重量>=0.16；性別為I的鮑魚內臟重量皆小於性別為M之鮑魚。

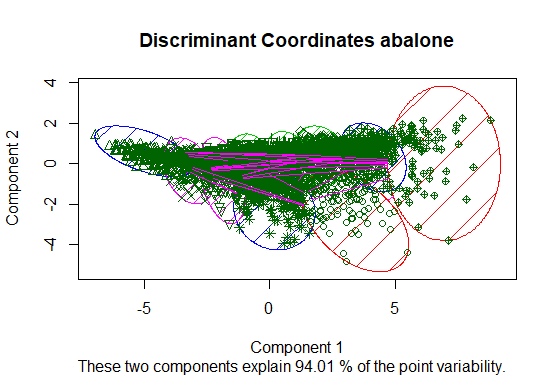


Figure 6：判別鮑魚個體座標圖，依據圖表可看出鮑魚基本上依據其特徵可區分為三種，基本上這兩個座標說明了變異點的94.01％

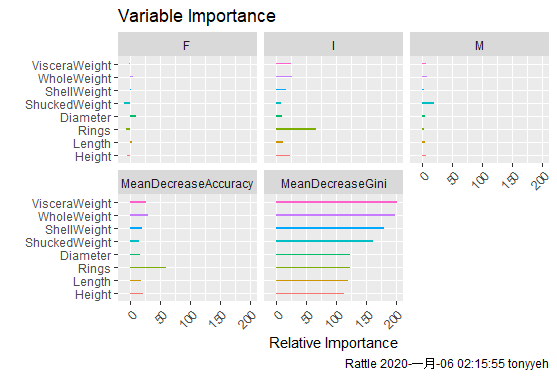


Figure 7：變數重要性分析圖表，可看出每個變數的相對重要性

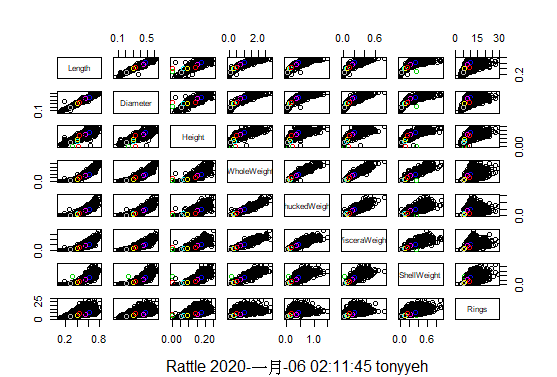


Figure ：各項資料分析多層級叢圖