Practical 5

1940223, Pranav Gopalkrishna

2021-02-13

# **Including necessary packages**

We include "dyplyr" for the filter function...

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.2

##   
## 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

# Population

The chosen population is a set of 150 flowers, whose sepal and petal measurements have been taken. These flowers belong to one of three species, namely Setosa, Versicolor and Virginica.

popl = iris  
head(popl)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

# Strata and variable of interest

Strata are based on the species of the flowers.

stra1 = filter(popl, popl$Species == "virginica")  
stra2 = filter(popl, popl$Species == "setosa")  
stra3 = filter(popl, popl$Species == "versicolor")

The sizes of each stratum is as follows…

N1 = length(stra1$Sepal.Length)  
N2 = length(stra2$Sepal.Length)  
N3 = length(stra3$Sepal.Length)  
c(N1, N2, N3)

## [1] 50 50 50

Variable of interest is sepal length.

# Sampling

Let the sample size be 20. Sampling from each stratum…

set.seed(1)  
sample1 = sample(stra1$Sepal.Length, 20, replace = FALSE)  
  
set.seed(2)  
sample2 = sample(stra2$Sepal.Length, 20, replace = FALSE)  
  
set.seed(3)  
sample3 = sample(stra3$Sepal.Length, 20, replace = FALSE)

# Estimating population's mean sepal length

Sample mean is an unbiased estimator of population mean. So, we will take the means of each sample, to get the estimated mean for each stratum.

sample1\_mean = mean(sample1)  
sample2\_mean = mean(sample2)  
sample3\_mean = mean(sample3)  
c(sample1\_mean, sample2\_mean, sample3\_mean)

## [1] 6.52 5.13 5.84

To estimate the whole population’s mean sepal length, we calculate the combined sample’s mean sepal length…

(sample1\_mean + sample2\_mean + sample3\_mean)/3

## [1] 5.83

# Estimating population's total sepal length

Population size times sample mean is an unbiased estimator of population total. To estimate each stratum’s total sepal length…

stra1\_total = length(stra1$Sepal.Length) \* sample1\_mean  
stra2\_total = length(stra2$Sepal.Length) \* sample2\_mean  
stra3\_total = length(stra3$Sepal.Length) \* sample3\_mean  
c(stra1\_total, stra2\_total, stra3\_total)

## [1] 326.0 256.5 292.0

To estimate the whole population’s total sepal…

stra1\_total + stra2\_total + stra3\_total

## [1] 874.5

# Conclusion

In conclusion, we estimate that the mean sepal length of the whole population to be 5.513333 cm, and the population’s total sepal length to be 874.5cm.