ACADGILD

SESSION 8: Exploratory Data Analytics

Assignment 1

Table of Contents

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on Variables & Distributions in R.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

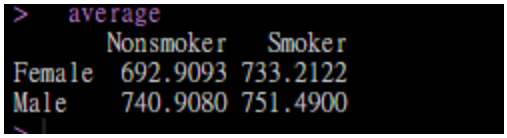
5. Problem Statement

1. **Use the package - RcmdrPlugin.IPSUR.**
   data(RcmdrTestDrive)

   and perform the below operations:

   a. **Calculate the average salary by gender and smoking status.**

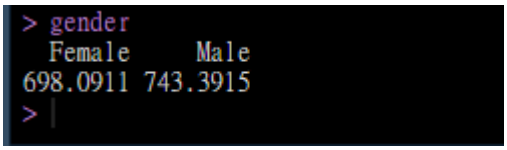
ANS: average=tapply(RcmdrTestDrive$salary, list(RcmdrTestDrive$gender, RcmdrTestDrive$smoking), mean, na.rm =T)



b. Which gender has the highest mean salary?

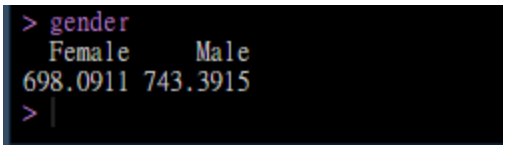
ANS: gender<- tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, mean, na.rm =T)

Gender



c. Report the highest mean salary.

ANS: Male: 743.3915 is the highest salary



d. Compare the spreads for the genders by calculating the standard deviation of salary by gender.

ANS: Sdgender<- tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, sd, na.rm =T)

Sdgender