Assignment 4

Terence Lau

11/18/2018

# Assignment 4

## Question One

1. The data file **Real-Estate.txt** contains information on the homes sold in the Denver area during the year 2003. The varialbes in this data file are as follow:

|  |  |
| --- | --- |
| Price | Sell price in $1000 |
| Bedrooms | Number of bedrooms |
| Size | Size of the home in square feet |
| Pool | Swimming Pool (1=yes, 0=no) |
| Distance | Distatnce from the home to the center of the city |
| Township | Township No. |
| Garage | Garage attached (1=yes, 0=no) |
| Baths | Number of bathrooms |

Use R to:

1. Develop a 95 percent confidence interval for the mean selling price of the homes.
2. Develop a 98 percent confidence interval for the mean distance the home is from the center of the city.
3. Develop a 90 percent confidence interval for the proportion of homes with an attached garage.
4. Conduct a test of hypothesis to determine if the mean selling price of homes with an attached garage and swimming pool is greater than $200,0000. Use the 0.05 significance level. Specify your conclusion.

### Answer

#### Part A

estate <- read.table("Real-Estate.txt", header = T)  
  
t.test(estate$Price) # we can get the conf interval

##   
## One Sample t-test  
##   
## data: estate$Price  
## t = 48.097, df = 104, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 211.9868 230.2189  
## sample estimates:  
## mean of x   
## 221.1029

For the 95 percent confidence interval is

#### Part B

# we can change the conf.int of the t.test   
t.test(estate$Price, conf.level = 0.98) # we can get the conf interval

##   
## One Sample t-test  
##   
## data: estate$Price  
## t = 48.097, df = 104, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 98 percent confidence interval:  
## 210.2413 231.9644  
## sample estimates:  
## mean of x   
## 221.1029

For the 98 percent confidence interval is

#### Part C

prop.test(71, 105, conf <- 0.90)

##   
## 1-sample proportions test with continuity correction  
##   
## data: 71 out of 105, null probability conf <- 0.9  
## X-squared = 55.979, df = 1, p-value = 7.326e-14  
## alternative hypothesis: true p is not equal to 0.9  
## 95 percent confidence interval:  
## 0.5769309 0.7623647  
## sample estimates:  
## p   
## 0.6761905

The 90 percent confidence interval is

#### Part D

PPG <- estate$Price[estate$Pool == 1 & estate$Garage == 1]  
t.test(PPG, mu = 200)

##   
## One Sample t-test  
##   
## data: PPG  
## t = 1.7683, df = 22, p-value = 0.09088  
## alternative hypothesis: true mean is not equal to 200  
## 95 percent confidence interval:  
## 197.6843 229.1157  
## sample estimates:  
## mean of x   
## 213.4

#### Part E

home\_with\_pool <- estate$Price[estate$Pool == 1]  
home\_without\_pool <- estate$Price[estate$Pool == 0]  
home\_with\_pool\_less <- home\_with\_pool[home\_with\_pool < home\_without\_pool]

## Warning in home\_with\_pool < home\_without\_pool: longer object length is not  
## a multiple of shorter object length

home\_with\_pool\_less <- na.omit(home\_with\_pool\_less)  
binom.test(length(home\_with\_pool\_less),length(home\_with\_pool))

##   
## Exact binomial test  
##   
## data: length(home\_with\_pool\_less) and length(home\_with\_pool)  
## number of successes = 29, number of trials = 38, p-value =  
## 0.001658  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
## 0.5975876 0.8855583  
## sample estimates:  
## probability of success   
## 0.7631579

alpha <- .02  
z.half.alpha = qnorm(1 - alpha/2)   
c(-z.half.alpha, z.half.alpha)

## [1] -2.326348 2.326348

#### Part G

home\_less\_median <- estate$Price[estate$Price < median(estate$Price)]  
home\_large\_median <- estate$Price[estate$Price >= median(estate$Price)]  
var.test(home\_less\_median, home\_large\_median)$p.value

## [1] 0.0004382044

The p-value is 0.00004382044

#### Part H

## Question Two

The following is sample information. Using R to test the hypothesis at the 0.05 significance level that the treatment means are equal and answer the following questions.

|  |  |  |
| --- | --- | --- |
| Treatment1 | Treatment2 | Treatment3 |
| 9 | 13 | 10 |
| 7 | 20 | 9 |
| 11 | 14 | 15 |
| 9 | 13 | 14 |
| 12 | - | 15 |
| 10 | - | - |

1. State the null hypothesis and the alternate hypothesis.
2. Give the value of SST, SSW and SSG.
3. Generate an ANOVA table.
4. State your decision regarding the null hypothesis.
5. If the null hypothesis is rejected, can we conclude that treatment 2 and treatment 3 differ? Use the 0.05 significance level.

### Answer

* Null hypothesis: The means are not equal.
* Althernative hypothesis: The means are equal.

treatment1 <- c(9, 7,11,9,12,10)  
treatment2 <- c(13,20,14,13)  
treatment3 <- c(10,9,15,14,15)  
  
mean\_total <- (sum(treatment1) + sum(treatment2) + sum(treatment3)) / (length(treatment1) + length(treatment2) + length(treatment3))  
  
SST <- sum(sum((treatment1 - mean\_total)^2) + sum((treatment2 - mean\_total)^2) + sum((treatment3 - mean\_total)^2))  
  
SSW\_treatment1 <- sum((treatment1 - mean(treatment1))^2)  
  
SSW\_treatment2 <- sum((treatment2 - mean(treatment2))^2)  
  
SSW\_treatment3 <- sum((treatment3 - mean(treatment3))^2)  
  
SSW\_total <- SSW\_treatment1 + SSW\_treatment2 + SSW\_treatment3  
  
SSG <- length(treatment1) \* (mean(treatment1) - mean\_total)^2 + length(treatment2) \* (mean(treatment2) - mean\_total)^2 + length(treatment3) \* (mean(treatment3) - mean\_total)^2  
  
paste("SSW = ", SSW\_total, "SSG = ", SSG, "SST = ", SST)

## [1] "SSW = 82.5333333333333 SSG = 70.4 SST = 152.933333333333"

df\_total <- 14  
df\_within <- 12  
df\_between <- 2   
  
MSG <- SSG / df\_between   
MSW <- SSW\_total / df\_within  
  
F\_value <- MSG / MSW   
  
  
paste("df of SSG = ", df\_between, "df of SSW", df\_within, "df of total", df\_total, "MSG = ", MSG, " MSW = ", MSW, "F = ", F\_value )

## [1] "df of SSG = 2 df of SSW 12 df of total 14 MSG = 35.2 MSW = 6.87777777777778 F = 5.11793214862682"

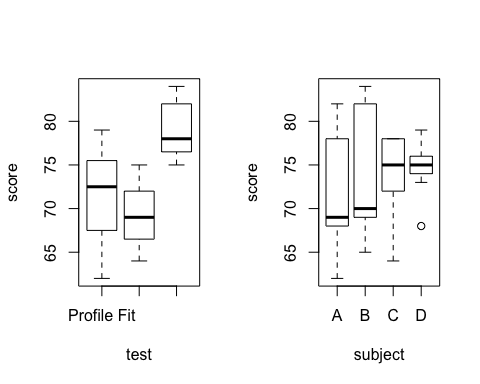
treatment <- c(9, 7, 11, 9, 12, 10, 13, 20, 14, 13, 10, 9, 15, 14, 15)  
treatment1 <- c(9, 7, 11, 9, 12, 10)  
treatment2 <- c(13, 20, 14, 13)  
treatment3 <- c(10, 9, 15, 14, 15)  
  
na <- length(treatment1)  
nb <- length(treatment2)  
nc <- length(treatment3)  
type.treatment <- c(rep("1",na),rep("2",nb),rep("3",nc))  
treatment.data <- data.frame(treatment, type.treatment)  
oneway.test(treatment~type.treatment, data = treatment.data, var.equal = T)

##   
## One-way analysis of means  
##   
## data: treatment and type.treatment  
## F = 5.1179, num df = 2, denom df = 12, p-value = 0.0247

## Question Three

A psychologist is working with three types of aptitude tests that may be given to prospective management trainees. In deciding how to structure the testing process, an important issue is the possibility of interaction between test takers and test type. If there was no interaction, only one type of test would be needed. Three tests of each type are given to members of each of four groups of subject type. These were distinguished by rating of poor, fair, good, and excellent in preliminary interview. The scores obtained are listed in the following table.

psych\_test <- data.frame(score<-c(65,68,62,69,71,67,75,75,78,74,79,76,72,69,69,70, 69, 65,64 ,72, 65,68,73,75,78, 82, 80,83, 82, 84,78, 78, 75,76, 77, 75), test <- gl(3, 12, 36, labels = c("Profile Fit", "Mindbender", "Psych Out")), subject <- gl(4, 3, 36, labels = c("A", "B", "C", "D")))   
  
op <- par(mfrow = c(1, 2))  
plot(score ~ test + subject, data = psych\_test)



bartlett.test(score ~ test, data = psych\_test) ## ## Bartlett test of homogeneity of variances ## ## data: time by toxicant

##   
## Bartlett test of homogeneity of variances  
##   
## data: score by test  
## Bartlett's K-squared = 3.864, df = 2, p-value = 0.1449

bartlett.test(score ~ subject, data = psych\_test)

##   
## Bartlett test of homogeneity of variances  
##   
## data: score by subject  
## Bartlett's K-squared = 5.9082, df = 3, p-value = 0.1162

rats.aov <- aov(score ~ test \* subject, data = psych\_test)   
summary(rats.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## test 2 622.1 311.03 50.895 2.33e-09 \*\*\*  
## subject 3 43.7 14.56 2.382 0.0945 .   
## test:subject 6 366.8 61.14 10.005 1.46e-05 \*\*\*  
## Residuals 24 146.7 6.11   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Question Four

Refer to the data Real\_Estate.txt. Use the selling price of the home as the dependent variable and determine the regression equation with number of bedrooms, size of the house, whether there is a pool, whether there is an attached garage, distance from the home to the center of the city, and the number of bathrooms as independent variables.

### Part A

Write out the regression equation. Interpret the coefficient on the distance from the home to the center of the city, the coefficient on size of house, the coefficient on garage and pool.

### Part B

Develop a correlation matrix. Which independent variable has the strongest correlation with the dependent variable?

### Part C

Find and interpret the value of and R\_adj^2 .

real.Estate <- read.table("Real-Estate.txt", header = T)  
names(real.Estate)

## [1] "Price" "Bedrooms" "Size" "Pool" "Distance" "Township"  
## [7] "Garage" "Baths"

attach(real.Estate)  
pri.dis.size.gar <- pri.dis.size.gar <- lm(Price ~ Distance + Size + Garage, data = real.Estate)  
summary(pri.dis.size.gar)

##   
## Call:  
## lm(formula = Price ~ Distance + Size + Garage, data = real.Estate)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -76.727 -24.177 -3.554 26.139 92.896   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 78.70898 36.68671 2.145 0.0343 \*   
## Distance -1.44596 0.79896 -1.810 0.0733 .   
## Size 0.05995 0.01467 4.088 8.75e-05 \*\*\*  
## Garage 44.70816 8.25365 5.417 4.13e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 36.9 on 101 degrees of freedom  
## Multiple R-squared: 0.4042, Adjusted R-squared: 0.3865   
## F-statistic: 22.84 on 3 and 101 DF, p-value: 2.29e-11

attach(real.Estate)

## The following objects are masked from real.Estate (pos = 3):  
##   
## Baths, Bedrooms, Distance, Garage, Pool, Price, Size, Township

part.a <- cbind.data.frame(Price, Distance, Size, Garage)  
cof <- cov(part.a)  
detach(real.Estate)  
print(cof)

## Price Distance Size Garage  
## Price 2218.91913 -79.6739286 4346.08516 11.6557418  
## Distance -79.67393 23.7549451 -142.03297 -0.8233516  
## Size 4346.08516 -142.0329670 61831.50183 9.7069597  
## Garage 11.65574 -0.8233516 9.70696 0.2210623

cor.test(x = real.Estate$Bedrooms, y = real.Estate$Price)

##   
## Pearson's product-moment correlation  
##   
## data: real.Estate$Bedrooms and real.Estate$Price  
## t = 5.3654, df = 103, p-value = 4.996e-07  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3028403 0.6048591  
## sample estimates:  
## cor   
## 0.4673771

library(psych)  
res2 <- corr.test(part.a)  
res2$r

## Price Distance Size Garage  
## Price 1.0000000 -0.3470312 0.37104159 0.52627394  
## Distance -0.3470312 1.0000000 -0.11719450 -0.35929488  
## Size 0.3710416 -0.1171945 1.00000000 0.08302732  
## Garage 0.5262739 -0.3592949 0.08302732 1.00000000

res2$ci.adj

## lower.adj upper.adj  
## 1 -0.5364115 -0.1243777  
## 2 0.1337786 0.5680747  
## 3 0.3128950 0.6890812  
## 4 -0.3271805 0.1038210  
## 5 -0.5534810 -0.1280593  
## 6 -0.1103947 0.2703897