Assignment-10

Ashish Verma

2024-04-23

Section 10.1 Problem #6

Solution

Null Hypothesis: H0:mu1=mu2=m3=m4 Alternate Hypothesis: Ha: At-least one mu is different

```
df_treatment<-4 -1
df_error<-40 -4

t1<-c(20.5,28.1,27.8,27.0,28.0,25.2,25.3,27.1,20.5,31.3)
t2<-c(26.3,24.0,26.2,20.2,23.7,34.0,17.1,26.8,23.7,24.9)
t3<-c(29.5,34.0,27.5,29.4,27.9,26.2,29.9,29.5,30.0,35.6)
t4<-c(36.5,44.2,34.1,30.3,31.4,33.1,34.1,32.9,36.3,25.5)
alpha<-0.01
data <- data.frame(
  value = c(t1, t2, t3, t4),
  treatment = factor(rep(c("t1", "t2", "t3", "t4"), each = 10))
)

result <- aov(value ~ treatment, data = data)

anova_summary <- summary(result)
f_value <- anova_summary[[1]][1,4]
p_value <- anova_summary[[1]][1,5]

cat("F-statistic:", f_value, "\n")
```

```
## F-statistic: 10.84904
```

```
cat("P-value:", p_value, "\n")
```

```
## P-value: 3.199045e-05
```

```
cat('Pvalue',p_value)
```

```
## Pvalue 3.199045e-05
```

```r
if (p_value < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Reject the null hypothesis"
```

Section 10.2 Problem #18

Solution(a)

Null Hypothesis H0:mu1=mu2=mu3=mu4=mu5 Alternate Hypothesis Ha:Not all means are equal.

```r
t1<-c(13,17,7,14)
t2<-c(21,13,20,17)
t3<-c(18,15,20,17)
t4<-c(7,11,18,10)
t5<-c(6,11,15,8)
alpha<-0.05
data <- data.frame(
  value = c(t1, t2, t3, t4, t5),
  treatment = factor(rep(c("t1", "t2", "t3", "t4", "t5"), each=4))
)
result <- aov(value ~ treatment, data = data)
anova_summary <- summary(result)
f_value <- anova_summary[[1]][1,4]
p_value <- anova_summary[[1]][1,5]

cat("F-statistic:", f_value, "\n")
```

```
## F-statistic: 3.485499
```

```r
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.03335772
```

```r
if (p_value < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Reject the null hypothesis"
```

```r
p1<-c(13,21,18,7,6)
p2<-c(17,13,15,11,11)
p3<-c(7,20,20,18,15)
p4<-c(14,17,17,10,8)
```

```r
# Combine the data into a data frame
data <- data.frame(
  values = c(p1, p2, p3, p4),
  group = factor(rep(c("p1", "p2", "p3", "p4"), each=5))
)
result_p <- aov(values ~ group, data = data)

anova_summary <- summary(result_p)
f_value_p <- anova_summary[[1]][1,4]
p_value_p <- anova_summary[[1]][1,5]

cat("F-statistic:", f_value_p, "\n")
```

```
## F-statistic: 0.4117444
```

```r
cat("P-value:", p_value_p, "\n")
```

```
## P-value: 0.7468025
```

```r
if (p_value_p < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Failed to reject the null hypothesis"
```

Solution(b)

```r
tukey_result <- TukeyHSD(result)
tukey_result_p <- TukeyHSD(result_p)
tukey_result
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = value ~ treatment, data = data)
##
## $treatment
##        diff        lwr        upr     p adj
## t2-t1  5.00  -3.276175 13.2761753 0.3754811
## t3-t1  4.75  -3.526175 13.0261753 0.4235109
## t4-t1 -1.25  -9.526175  7.0261753 0.9892929
## t5-t1 -2.75 -11.026175  5.5261753 0.8395387
## t3-t2 -0.25  -8.526175  8.0261753 0.9999807
## t4-t2 -6.25 -14.526175  2.0261753 0.1884779
## t5-t2 -7.75 -16.026175  0.5261753 0.0717704
## t4-t3 -6.00 -14.276175  2.2761753 0.2185546
## t5-t3 -7.50 -15.776175  0.7761753 0.0849318
## t5-t4 -1.50  -9.776175  6.7761753 0.9789688
```

```
tukey_result_p
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = values ~ group, data = data)
##
## $group
##       diff      lwr     upr     p adj
## p2-p1  0.4  -8.4876  9.2876 0.9992038
## p3-p1  3.0  -5.8876 11.8876 0.7703469
## p4-p1  0.2  -8.6876  9.0876 0.9998999
## p3-p2  2.6  -6.2876 11.4876 0.8361704
## p4-p2 -0.2  -9.0876  8.6876 0.9998999
## p4-p3 -2.8 -11.6876  6.0876 0.8042925
```

```
p_values <- tukey_result$p.adj
p_values_p <- tukey_result_p$p.adj
```

Thus, both the procedures do not give the same result. None of the can be said to differ significantly from one another.

Section 10.3 Problem #26

Solution(a)

```
ip<-c(14.1,13.6,14.4,14.3)
pk<-c(12.8,12.5,13.4,13.0,12.3)
bb<-c(13.5,13.4,14.1,14.3)
cf<-c(13.2,12.7,12.6,13.9)
mz<-c(16.8,17.2,16.4,17.3,18.0)
fl<-c(18.1,17.2,18.7,18.4)

library(dplyr)
```

```
##
## 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
```

```
# Creating a dataframe with the given data
data <- data.frame(
  Value = c(14.1, 13.6, 14.4, 14.3, 12.8, 12.5, 13.4, 13.0, 12.3,
```

```r
            13.5, 13.4, 14.1, 14.3, 13.2, 12.7, 12.6, 13.9, 16.8,
            17.2, 16.4, 17.3, 18.0, 18.1, 17.2, 18.7, 18.4),
  Group = rep(c("ip", "pk", "bb", "cf", "mz", "fl"),
              times = c(4, 5, 4, 4, 5, 4))
)

anova_result <- aov(Value ~ Group, data = data)
anova_summary <- summary(anova_result)

F_value <- anova_summary[[1]]$F[1]
cat("F statistics results",F_value)
```

```
## F statistics results 79.26375
```

```r
p_value <- anova_summary[[1]]$Pr[1]
cat("Pvalue",p_value)
```

```
## Pvalue 1.736909e-12
```

```r
if (p_value < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Reject the null hypothesis"
```

Solution(b)

```r
turkey_result <- TukeyHSD(anova_result)
print(turkey_result)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Value ~ Group, data = data)
##
## $Group
##         diff        lwr        upr     p adj
## cf-bb -0.725 -1.8862516  0.4362516 0.3963073
## fl-bb  4.275  3.1137484  5.4362516 0.0000000
## ip-bb  0.275 -0.8862516  1.4362516 0.9736311
## mz-bb  3.315  2.2133400  4.4166600 0.0000001
## pk-bb -1.025 -2.1266600  0.0766600 0.0775326
## fl-cf  5.000  3.8387484  6.1612516 0.0000000
## ip-cf  1.000 -0.1612516  2.1612516 0.1176619
## mz-cf  4.040  2.9383400  5.1416600 0.0000000
## pk-cf -0.300 -1.4016600  0.8016600 0.9526463
## ip-fl -4.000 -5.1612516 -2.8387484 0.0000000
```

```
## mz-fl -0.960 -2.0616600  0.1416600 0.1107597
## pk-fl -5.300 -6.4016600 -4.1983400 0.0000000
## mz-ip  3.040  1.9383400  4.1416600 0.0000004
## pk-ip -1.300 -2.4016600 -0.1983400 0.0150616
## pk-mz -4.340 -5.3786550 -3.3013450 0.0000000
```

Solution(c)

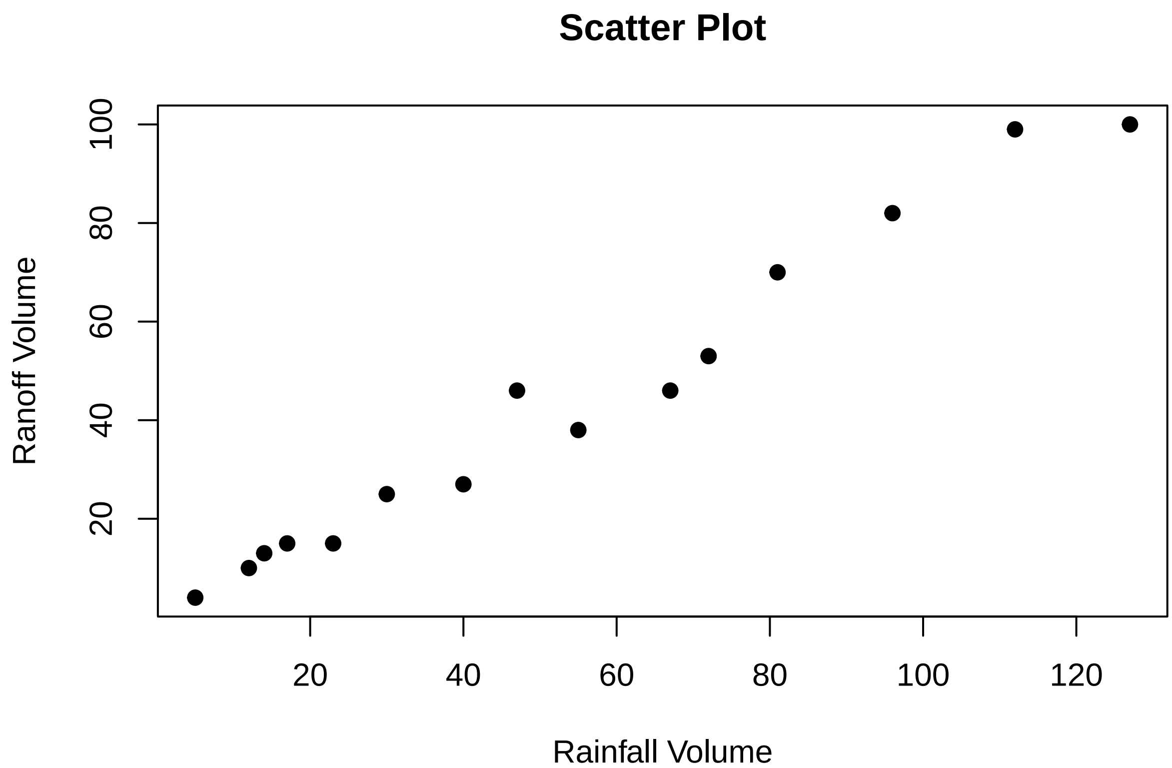
```
thetacap<-(mean(ip)+mean(pk)+mean(bb)+mean(cf))/4 - (mean(mz)+mean(fl))/2
alpha<-0.01
SSC<-((1/16)/4)+((1/16)/5)+((1/16)/4)+((1/16)/4)+((1/4)/5)+((1/4)/4)
tcrit<-qt(1 - alpha/2,20)
margin<-sqrt(tcrit*SSC)
lower_limit<-thetacap - margin
upper_limit<-thetacap + margin

cat("The confidence interval is",lower_limit,upper_limit)
```

```
## The confidence interval is -4.863066 -3.464434
```

Section 12.2 Problem #16

Solution(a)

```
x<-c(5,12,14,17,23,30,40,47,55,67,72,81,96,112,127)
y<-c(4,10,13,15,15,25,27,46,38,46,53,70,82,99,100)
plot(x,y,main="Scatter Plot",pch=19,xlab="Rainfall Volume",ylab="Ranoff Volume")
```



The scatter plot show linear relationship.

Solution(b)

```
model = lm(y~x)
Bo <- coefficients(model)[1]
B1 <- coefficients(model)[2]
cat("The Bo",Bo)
```

```
## The Bo -1.128305
```

```
cat("The B1",B1)
```

```
## The B1 0.8269731
```

Solution(c)

```
ypred<-Bo+B1*50
print(ypred)
```

```
## (Intercept)
##    40.22035
```

Solution(d)

```
A=summary(model)
point_estimate <- A$sigma
cat("The point esitmate of simga",point_estimate)
```

```
## The point esitmate of simga 5.240462
```

Solution(e)

```
R2<-A$r.squared
cat(round(R2,4))
```

```
## 0.9753
```

Section 12.3 Problem #34

Soltuion(a)

The equation of the least squares line is: y = 4.8567 - 0.0747x. The slope of the line, -0.0747, indicates that as air void increases by 1%, the dielectric constant decreases by approximately 0.0747.

Solution(b)

The multiple R-squared value of 0.7797 indicates that approximately 77.97% of the observed variation in the dielectric constant can be explained by the linear relationship between the dielectric constant and air void.

Soltuion(c)

Null Hypothesis H0:beta=0

Alternate Hypothesis Ha:beta!=0

```
alpha<-0.01
data <- data.frame(y = c(4.55, 4.49, 4.50, 4.47, 4.47, 4.45, 4.40, 4.34, 4.43, 4.43, 4.42, 4.40, 4.33, 4
model <- lm(y ~ x, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.058429 -0.027528 -0.003202  0.021798  0.057253
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  4.858691   0.059768  81.293  < 2e-16 ***
## x           -0.074676   0.009923  -7.526 1.21e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03551 on 16 degrees of freedom
## Multiple R-squared:  0.7797, Adjusted R-squared:  0.766
## F-statistic: 56.63 on 1 and 16 DF,  p-value: 1.214e-06
```

```r
coef_summary <- summary(model)$coefficients
coef_x <- coef_summary["x", ]
t_value <- coef_x["t value"]
p_value <- coef_x["Pr(>|t|)"]

cat("T testing results",t_value)
```

```
## T testing results -7.525603
```

```r
cat("Pvalue results",p_value)
```

```
## Pvalue results 1.214344e-06
```

```r
if (p_value < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Reject the null hypothesis"
```

Soltuion(d)

Null Hypothesis:H0:beta=-0.064 Alternate Hypothesis:Ha:beta<-0.064

```r
beta<--0.064
betanotcap<-0.074676   #from table
SSE <- coef_x["Std. Error"]
ttesting <- (-betanotcap+beta)/SSE
p_value <- pt(ttesting,16)
cat("P_value results",p_value)
```

```
## P_value results 1.098479e-10
```

```r
if (p_value < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Reject the null hypothesis"
```

Section 12.4 Problem #46

Soluion(a)

```
x <- c(0.718, 0.808, 0.924, 1.000, 0.667, 0.529, 0.514, 0.559,
       0.766, 0.470, 0.726, 0.762, 0.666, 0.562, 0.378, 0.779,
       0.674, 0.858, 0.406, 0.927, 0.311, 0.319, 0.518, 0.687,
       0.907, 0.638, 0.234, 0.781, 0.326, 0.433, 0.319, 0.238)
y <- c(0.428, 0.480, 0.493, 0.978, 0.318, 0.298, 2.224, 0.198,
       0.326, 2.336, 0.765, 0.190, 0.066, 2.221, 2.898, 0.836,
       0.126, 0.305, 2.577, 0.779, 2.707, 2.610, 2.648, 2.145,
       1.007, 2.090, 21.132, 0.538, 21.098, 2.581, 2.862, 2.551)
SXX<-1.48193150
SYY<-11.82637622
SXY<-3.83071088
sigma_xi<-19.404
sigma_yi<--.594
sigma_y2<-11.835795
sigma_xy<-3.497811

betacap<-SXY/SXX
betanotcap<-(sigma_yi/32) - (betacap*sigma_xi/32)
cat("The fitted simple regression is",betacap,betanotcap)
```

```
## The fitted simple regression is 2.584945 -1.586008
```

```
SST<-sigma_y2 - (sigma_yi^2)/32
SSE<-sigma_y2 -(betanotcap*sigma_yi) - (betacap*sigma_xy)
R2<- 1 - (SSE/SST)
cat("The proportion of observed variation is given by",R2)
```

```
## The proportion of observed variation is given by 0.8433747
```

Soluion(b)

```
alpha<-0.05
SSE<-sigma_y2 -(betanotcap*sigma_yi) - (betacap*sigma_xy)
SE<- sqrt(SSE/(32-2))
sb1 = SE/sqrt(SXX)
cat("SB1",sb1)
```

```
## SB1 0.2041045
```

```
df<-30
tcrit<- qt(1 - alpha/2,df)
cat("Tcrit",tcrit)
```

```
## Tcrit 2.042272
```

```r
lower_limit<-betacap - (tcrit*sb1)
upper_limit<-betacap + (tcrit*sb1)
cat("The 95% confidence interval are",lower_limit,upper_limit)
```

```
## The 95% confidence interval are 2.168108 3.001782
```

Soluion(c)

```r
predicted_x<-0.6
yprime<-betanotcap+(betacap*predicted_x)
cat("Yprime",yprime)
```

```
## Yprime -0.03504152
```

```r
degree_of_freedom<-30
alpha<-0.05
tcrit<- qt(1 - alpha/2,df)
sqrt_result<-sqrt(0.03125+((predicted_x - (sigma_xi/32))^2)/SXX)
margin<-tcrit*SE*sqrt_result
lower_limit<- yprime - margin
upper_limit<- yprime + margin
cat("The 95% confidence interval are",lower_limit,upper_limit)
```

```
## The 95% confidence interval are -0.1247835 0.0547005
```

Soluion(d)

```r
predicted_x<-0.6
yprime<-betanotcap+(betacap*predicted_x)
cat("Yprime",yprime)
```

```
## Yprime -0.03504152
```

```r
degree_of_freedom<-30
alpha<-0.05
tcrit<- qt(1 - alpha/2,df)
sqrt_result<-sqrt(1.03125+((predicted_x - (sigma_xi/32))^2)/SXX)
margin<-tcrit*SE*sqrt_result
lower_limit<- yprime - margin
upper_limit<- yprime + margin
cat("The 95% confidence interval are",lower_limit,upper_limit)
```

```
## The 95% confidence interval are -0.550351 0.480268
```

Soluion(e)

```
predicted_x<-0.7
yprime<-betanotcap+(betacap*predicted_x)
cat("Yprime",yprime)
```

```
## Yprime 0.2234529
```

```
degree_of_freedom<-30
alpha<-0.05
tcrit<- qt(1 - alpha/2,df)
sqrt_result<-sqrt(0.03125+((predicted_x - (sigma_xi/32))^2)/SXX)
margin<-tcrit*SE*sqrt_result
lower_limit<- yprime - margin
upper_limit<- yprime + margin
cat("The 95% confidence interval are",lower_limit,upper_limit)
```

```
## The 95% confidence interval are 0.1256285 0.3212774
```

Since 0 is not contained in interval, there is enough evidence to conclude that true average astringency for a tannin concentration of 0.75 is something other than 0.

Section 12.5 Problem #58

Solution(a)

```
x<-c(4200,3600,3750,3675,4050,2770,4870,4500,3450,2700,3750,3300)
y<-c(370,340,375,310,350,200,400,375,285,225,345,285)
n<-length(x)
xbar<-mean(x)
ybar<-mean(y)

Sxx <- sum(x^2)-(sum(x)^2)/n
Syy <- sum(y^2)-(sum(y)^2)/n
Sxy <- sum(x*y) - (sum(x)*sum(y))/n

corr_r<-Sxy/sqrt(Sxx*Syy)
cat('Correlation coefficient',corr_r)
```

```
## Correlation coefficient 0.9231564
```

Solution(b)

Same output.The correlation coefficient is independent of origin and scale. It won't change by adding, subtraction, multiplying and dividing.

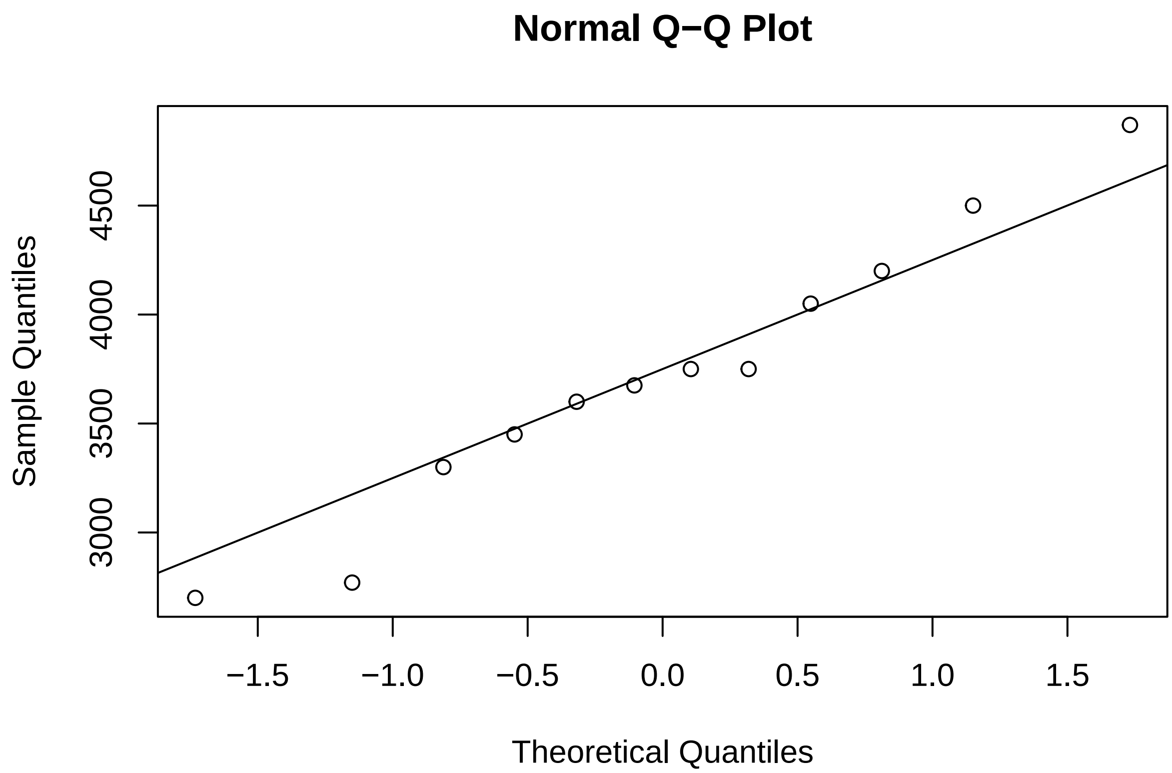
Solution(c)

Same output.The correlation coefficient is independent of origin and scale. It won't change by adding, subtraction, multiplying and dividing

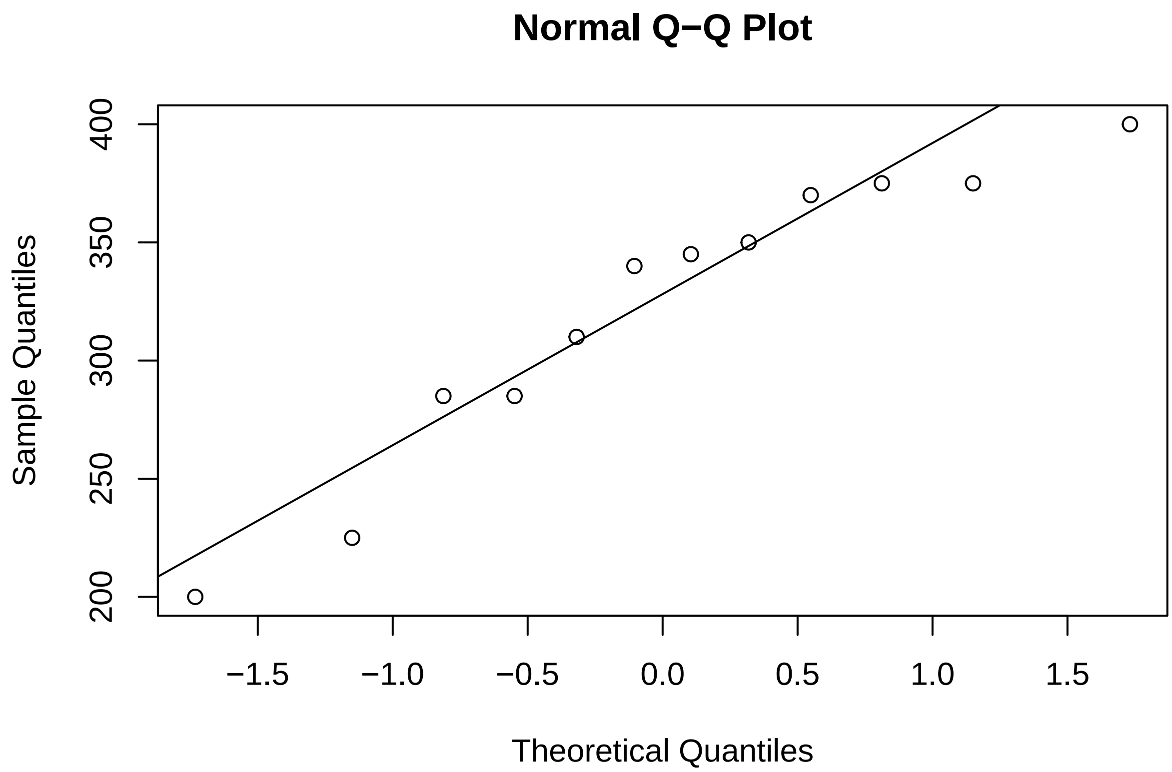
Solution(d)

Null Hypothesis:H0:p=0 Alternate Hypothesis Ha:p!=0

```
qqnorm(x)
qqline(x)
```




```
alpha<-0.05
qqnorm(y)
qqline(y)
```




```r
n<-12
df<- n -2
ttesting<- corr_r * sqrt(df)/sqrt(1-(corr_r^2))
cat("T Statics results",ttesting)
```

```
## T Statics results 7.593888
```

```r
p_value <- 2 * (1 - pt(abs(ttesting), df))
cat("P value",p_value)
```

```
## P value 1.852845e-05
```

```r
if (p_value < alpha) {
  print("Reject the null hypothesis")
} else {
  print("Failed to reject the null hypothesis")
}
```

```
## [1] "Reject the null hypothesis"
```