Act06ANOVA1w Stat 463/563

Dr. Islam

Due Date: Oct 27, 2022

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Data below refers to measurements of a response due to four treatment groups:

x1: 36 48 41 50 42 47 48 49 40 42  
x2: 39 45 37 47 54 48 42 47 41 46  
x3: 52 48 50 46 53 45 46 45 43 58  
x4: 45 38 44 36 43

Apply ANOVA to answer the following questions:

1. Perform the overall significance of ANOVA (i.e., the significance of the ANOVA F-test for where refers to mean of the population the sample comes from)
2. Draw a boxplot by days and comment on the equality of means of various days.
3. Perform the test of HOV via Bartlett’s test, and make a conclusion about the acceptance or rejection of the homogeneity of the variance at 5% level of significance.
4. Perform test of normality via Shapiro and Wilk Test (shapiro.test) and Kolmogorov-Smirnov Test (ks.test)
5. Assess the normality normality via Q-Q plot
6. Compute 90% CI for the treatment effects.

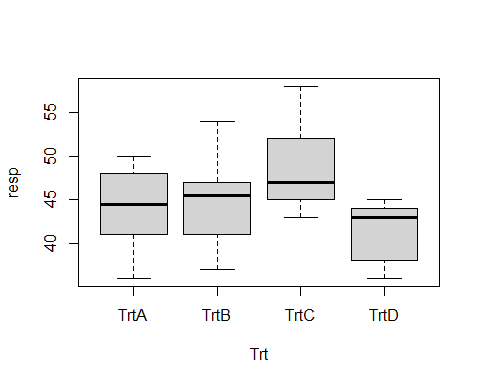
x1=scan(text="36 48 41 50 42 47 48 49 40 42")  
x2=scan(text="39 45 37 47 54 48 42 47 41 46")  
x3=scan(text="52 48 50 46 53 45 46 45 43 58")  
x4=scan(text="45 38 44 36 43")  
n1=length(x1)  
n2=length(x2)  
n3=length(x3)  
n4=length(x4)  
Trt=rep(c("TrtA","TrtB","TrtC","TrtD"),c(n1,n2,n3,n4))  
resp=c(x1,x2,x3,x4)  
#aov(resp~Trt), works.  
dat=data.frame(resp,Trt)  
# (a)  
res=aov(resp~Trt, data=dat)  
summary(res)

Df Sum Sq Mean Sq F value Pr(>F)   
Trt 3 207.3 69.09 3.17 0.038 \*  
Residuals 31 675.7 21.80   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(res)

Analysis of Variance Table  
  
Response: resp  
 Df Sum Sq Mean Sq F value Pr(>F)   
Trt 3 207.27 69.090 3.1698 0.03803 \*  
Residuals 31 675.70 21.797   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (b)  
boxplot(resp ~ Trt, data = dat)



# The boxplot reveals that equality of means could be of suspect.  
  
# (c) Bartlett test of HOV  
bartlett.test(resp~Trt, data=dat)

Bartlett test of homogeneity of variances  
  
data: resp by Trt  
Bartlett's K-squared = 0.25405, df = 3, p-value = 0.9684

# (d)   
# Extract the residuals from the aov() result object  
resid<- residuals(res)  
  
#Shapiro-Wilk test of normality  
shapiro.test(resid )

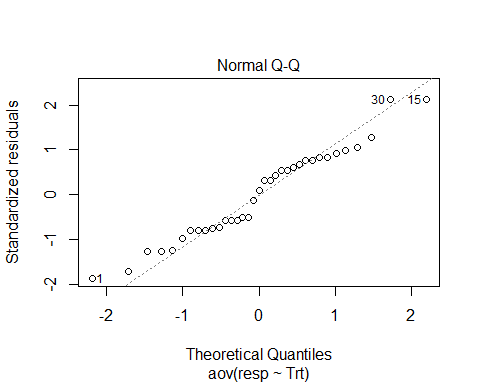
Shapiro-Wilk normality test  
  
data: resid  
W = 0.96131, p-value = 0.2502

# Kolmogorov-Smirnov test of normality  
ks.test(resid,"pnorm", mean=mean(resid), sd=sd(resid))

Warning in ks.test.default(resid, "pnorm", mean = mean(resid), sd = sd(resid)):  
ties should not be present for the Kolmogorov-Smirnov test

Asymptotic one-sample Kolmogorov-Smirnov test  
  
data: resid  
D = 0.15419, p-value = 0.3761  
alternative hypothesis: two-sided

# (e) Requests 2nd plot with q-q plot  
plot(res, 2)



# (f) Confidence interval of treatment effects  
# note that intercept refers to overall mean (mu)  
# other lines refers to Tau\_i; recall mu\_i=mu+Tau\_i;  
confint(res, level=0.90)

5 % 95 %  
(Intercept) 41.7967817 46.803218  
TrtTrtB -3.2400852 3.840085  
TrtTrtC 0.7599148 7.840085  
TrtTrtD -7.4357012 1.235701

1. A bank management is struggling in finding the best way to staff a branch office on different days. The management knows that it will need more tellers on Fridays than on other days, but the management is trying to find if it will have constant tellers across the rest of the week. It collects data for the number of transactions each day for two months. Here are the collected data:

Mondays: 278, 325, 300, 258, 279, 311, 314, 266, 312  
Tuesdays: 245, 281, 303, 287, 275, 245, 230, 301, 257  
Wednesdays: 290, 199, 311, 268, 244, 295, 257, 275  
Thursdays: 253, 281, 242, 230, 280, 288, 258, 265

The management wishes to test:

against

for at least two groups,

1. What test should you recommend to the bank management?
2. What assumptions make the test recommended in (a) valid?
3. Report SST (SS Total), SSB (SS between Treatment), SSE (SS Residual).
4. Perform the test recommended in (a) at 5% level of significance.
5. What decision are you conveying to the bank management?
6. Perform multiple comparision at significance leve by **pairwise.t.test()** or **TukeyHSD()**.

Notes:  
TukeyHSD() suppports conf.level=option, whereas confint() supports level= option to control % CI sought for.

trt1<-c(278, 325, 300, 258, 279, 311, 314, 266, 312);  
trt2<-c(245, 281, 303, 287, 275, 245, 230, 301, 257);  
trt3<-c(290, 199, 311, 268, 244, 295, 257, 275);  
trt4<-c(253, 281, 242, 230, 280, 288, 258, 265);  
n1=length(trt1)  
n2=length(trt2)  
n3=length(trt3)  
n4=length(trt4)  
resp<-c(trt1,trt2, trt3,trt4)  
days<-rep(c("Mon","Tue","Wed","Thu"),c(n1,n2,n3,n4))  
dat=data.frame(resp,days)  
#(a) Use an ANOVA test  
#(b)Normality of samples, homogeneity of group variances and independence of groups are required for the validity of ANOVA test.  
#(c)+(d)+)(e) can be obtained using summary() or anova() to aov objects  
res<-aov(resp~days, data=dat)  
summary(res)

Df Sum Sq Mean Sq F value Pr(>F)   
days 3 5151 1717.1 2.393 0.0881 .  
Residuals 30 21527 717.6   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(res)

Analysis of Variance Table  
  
Response: resp  
 Df Sum Sq Mean Sq F value Pr(>F)   
days 3 5151.3 1717.09 2.393 0.08809 .  
Residuals 30 21526.7 717.56   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#(f)  
TukeyHSD(res, conf.level = 0.90)

Tukey multiple comparisons of means  
 90% family-wise confidence level  
  
Fit: aov(formula = resp ~ days, data = dat)  
  
$days  
 diff lwr upr p adj  
Thu-Mon -31.541667 -62.70216 -0.3811695 0.0942008  
Tue-Mon -24.333333 -54.56346 5.8967892 0.2385288  
Wed-Mon -26.291667 -57.45216 4.8688305 0.2033480  
Tue-Thu 7.208333 -23.95216 38.3688305 0.9447898  
Wed-Thu 5.250000 -26.81389 37.3138870 0.9791822  
Wed-Tue -1.958333 -33.11883 29.2021638 0.9987581

pairwise.t.test(resp,days)

Pairwise comparisons using t tests with pooled SD   
  
data: resp and days   
  
 Mon Thu Tue   
Thu 0.13 - -   
Tue 0.26 1.00 -   
Wed 0.26 1.00 1.00  
  
P value adjustment method: holm

pairwise.t.test(resp,days, p.adjust="bonferroni", alt="t")

Pairwise comparisons using t tests with pooled SD   
  
data: resp and days   
  
 Mon Thu Tue   
Thu 0.13 - -   
Tue 0.38 1.00 -   
Wed 0.31 1.00 1.00  
  
P value adjustment method: bonferroni

pairwise.t.test(resp,days, p.adjust="bonferroni", alt="l")

Pairwise comparisons using t tests with pooled SD   
  
data: resp and days   
  
 Mon Thu Tue   
Thu 0.065 - -   
Tue 0.190 1.000 -   
Wed 0.157 1.000 1.000  
  
P value adjustment method: bonferroni

#boxplot(resp ~ days, data = dat)