midterm1stat631

Jessica Grover

3/15/2022

#Problem 1 #a.Are the lives of these brands of batteries different?

library(emmeans)  
library(lsmeans)

## Warning: package 'lsmeans' was built under R version 4.1.3

## The 'lsmeans' package is now basically a front end for 'emmeans'.  
## Users are encouraged to switch the rest of the way.  
## See help('transition') for more information, including how to  
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.

brands<-c(100,96,92,96,92,76,80,75,84,82,108,100,96,98,100)  
life<-rep(c(1,2,3),each=5)  
data1<-data.frame(brands,life)  
data1$flife<-as.factor(data1$life)  
data1

## brands life flife  
## 1 100 1 1  
## 2 96 1 1  
## 3 92 1 1  
## 4 96 1 1  
## 5 92 1 1  
## 6 76 2 2  
## 7 80 2 2  
## 8 75 2 2  
## 9 84 2 2  
## 10 82 2 2  
## 11 108 3 3  
## 12 100 3 3  
## 13 96 3 3  
## 14 98 3 3  
## 15 100 3 3

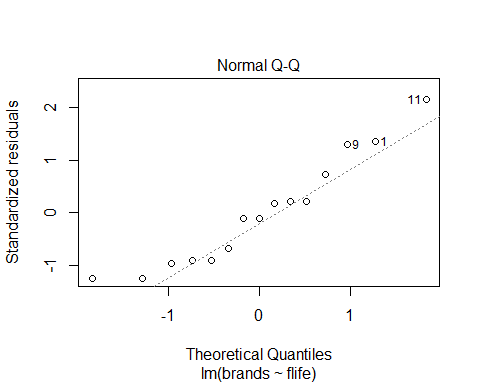
battery<-lm(brands~flife,data=data1)  
anova(battery)

## Analysis of Variance Table  
##   
## Response: brands  
## Df Sum Sq Mean Sq F value Pr(>F)   
## flife 2 1196.1 598.07 38.338 6.141e-06 \*\*\*  
## Residuals 12 187.2 15.60   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

By using Anova test, we got the p-value which is less than 0.5. Therefore, we reject null hypothesis. The means of the brands are different. Hence, the battery life will also be different.

#B. Analyze the normality assumption and state your conclusion.

plot(battery,which=2)

 Here to check the normality we are using qqplot. There is no pattern observed in qqplot. The points are not close to the qqline and hence it is not normal. Normality can also be checked by Shapiro-Wilk’s test

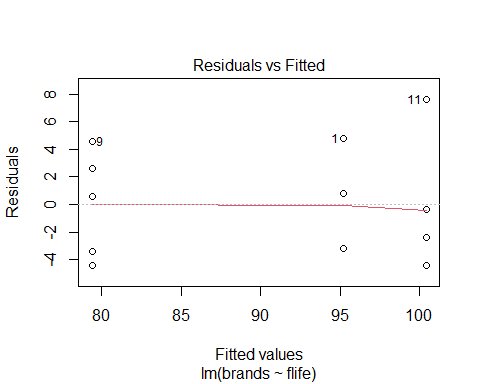
shapiro.test(battery$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: battery$residuals  
## W = 0.92696, p-value = 0.2457

Since p-value=0.2457>0.05, data supports the null hypothesis of normality.

#C.Check the constant variance assumption and state your conclusion.

plot(battery,which = 1)

 Here we are checking constant variance assumptions by residuals vs fitted values plot. There are no clear patterns seen in the above plot. The sample violates the assumptions of constant variance at every fitted value. The levene’s test can be used for homogenety of variance

library(car)

## Loading required package: carData

leveneTest(brands~flife,data=data1)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 0.07 0.9328  
## 12

By performing Levene’s test, we conclude that it is reasonable to consider constant variance as p=9.328>0.05.

#D. Construct 95% simultaneous confidence interval for all pairs of means using Tukey’s method.

lsmResponse<-lsmeans(battery,~flife)  
tk<-summary(contrast(lsmResponse,method="pairwise",adjust = "Tukey",infer=c(T,T),level=0.95))  
tk

## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## 1 - 2 15.8 2.5 12 9.14 22.46 6.325 0.0001  
## 1 - 3 -5.2 2.5 12 -11.86 1.46 -2.082 0.1355  
## 2 - 3 -21.0 2.5 12 -27.66 -14.34 -8.407 <.0001  
##   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 3 estimates   
## P value adjustment: tukey method for comparing a family of 3 estimates

Tukey’s method is used to conduct pairwise mean comparison at certain significant level.Here level is 0.95. Confidence Interval of brand1-brand2 is (9.14,22.46) Confidence Interval of brand1-brand3 is (-11.86,1.46) Confidence Interval of brand1-brand2 is (-27.66,-14.34)

#E. From d, which brand has the shortest life? By considering the means of these three brands, brand 2 has the shortest life.

#F. For the brand selected in (e), the manufacturer decides to replace without charge any battery that fails in less than 80 weeks, what percentage of batteries of this brand would the company expect to replace?

pnorm((80-79.4)/sqrt(15.6))

## [1] 0.5603714

The company is expected to replace 56.03% of brand 2 batteries.

#Problem 2 #a.

library(emmeans)  
data1<-c(97,96,92,95,83,87,78,81,85,84,78,79,64,72,63,74,52,56,44,50,48,58,49,53)  
res<-rep(c(1,2,3,4,5,6),each=4)  
level1<-data.frame(data1,res)  
level1$fres<-as.factor(level1$res)  
level1

## data1 res fres  
## 1 97 1 1  
## 2 96 1 1  
## 3 92 1 1  
## 4 95 1 1  
## 5 83 2 2  
## 6 87 2 2  
## 7 78 2 2  
## 8 81 2 2  
## 9 85 3 3  
## 10 84 3 3  
## 11 78 3 3  
## 12 79 3 3  
## 13 64 4 4  
## 14 72 4 4  
## 15 63 4 4  
## 16 74 4 4  
## 17 52 5 5  
## 18 56 5 5  
## 19 44 5 5  
## 20 50 5 5  
## 21 48 6 6  
## 22 58 6 6  
## 23 49 6 6  
## 24 53 6 6

power1=powerTransform(data1~fres,data = level1)  
summary(power1)

## bcPower Transformation to Normality   
## Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd  
## Y1 1.8985 1 0.5304 3.2666  
##   
## Likelihood ratio test that transformation parameter is equal to 0  
## (log transformation)  
## LRT df pval  
## LR test, lambda = (0) 7.775487 1 0.005296  
##   
## Likelihood ratio test that no transformation is needed  
## LRT df pval  
## LR test, lambda = (1) 1.726905 1 0.18881

The transformation is not required as the lambda value is 1. #b. We are rejecting null hypothesis as the value of p is significant. The treatments are independent and at least one of them is not equivalent.

nit<-lm(data1~fres,data=level1)  
anova(nit)

## Analysis of Variance Table  
##   
## Response: data1  
## Df Sum Sq Mean Sq F value Pr(>F)   
## fres 5 6398.3 1279.67 71.203 3.197e-11 \*\*\*  
## Residuals 18 323.5 17.97   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

lsmType<-lsmeans(nit,~fres)  
summary(contrast(lsmType,list("Nitrogen Vs Non nitrogen"=c(1,0,-1,-1,1,0))),infer = c(T,T),side="two sided")

## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## Nitrogen Vs Non nitrogen -4.25 4.24 18 -13.2 4.66 -1.003 0.3294  
##   
## Confidence level used: 0.95

#c. No, there is no significant quadratic effects of nitrogen under non irrigated condition. The value of p=0.3294>0.05, it is not significant at 95% of confidence level.

#d.

summary(contrast(lsmType,list("Irrigation effect"=c(1,-1,0,0,1,-1))),  
infer=c(T,T),side = "two-sided")

## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## Irrigation effect 11.2 4.24 18 2.34 20.2 2.654 0.0162  
##   
## Confidence level used: 0.95

There is a significant effect of irrigation. The interval is positive between irrigation and non irrigation. This indicates that non irrigation is better than irrigation with 95% confidence level.

#e.

tk<- summary(contrast(lsmType,method="pairwise",adjust=("Tukey"),infer=c(T,T),level = 0.95))  
tk

## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## 1 - 2 12.75 3 18 3.22 22.28 4.253 0.0054  
## 1 - 3 13.50 3 18 3.97 23.03 4.503 0.0032  
## 1 - 4 26.75 3 18 17.22 36.28 8.924 <.0001  
## 1 - 5 44.50 3 18 34.97 54.03 14.845 <.0001  
## 1 - 6 43.00 3 18 33.47 52.53 14.344 <.0001  
## 2 - 3 0.75 3 18 -8.78 10.28 0.250 0.9998  
## 2 - 4 14.00 3 18 4.47 23.53 4.670 0.0022  
## 2 - 5 31.75 3 18 22.22 41.28 10.592 <.0001  
## 2 - 6 30.25 3 18 20.72 39.78 10.091 <.0001  
## 3 - 4 13.25 3 18 3.72 22.78 4.420 0.0038  
## 3 - 5 31.00 3 18 21.47 40.53 10.341 <.0001  
## 3 - 6 29.50 3 18 19.97 39.03 9.841 <.0001  
## 4 - 5 17.75 3 18 8.22 27.28 5.921 0.0002  
## 4 - 6 16.25 3 18 6.72 25.78 5.421 0.0005  
## 5 - 6 -1.50 3 18 -11.03 8.03 -0.500 0.9955  
##   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 6 estimates   
## P value adjustment: tukey method for comparing a family of 6 estimates

There is no 0 in any intervals. This indicates that the mean response of the 1st group is significantly larger than those of the other groups.There is a strong evidence that it gives us the best result of big bluestem.