Introduction to Anova (Assignment - 2)

Pradeep Paladugula

5/12/2020

Q1) Espresso Data

#Variables  
1) Brewing method (1, 2 and 3) - categorical method  
2) amount fo Creme produced  
  
Hypothesis:  
H0: thier is significant diffence in creame production from three different Brewing methods  
H1: Diffrence is seen in production of creme from method to method   
  
Summary:  
the main reson of this assignment is to work and understand Anova where I considered the Espresso Creme production data between three different brewing methods. I have formated data of categorical variable with method levels. Even though the data appears nearly normal on density normality plot, I have noticed a small skewness in the graph. From agostino test (skew = 0.54679, p-value = 0.1842) as the skewness is not equal to zero, so it can proved that the data has skewness and p-value is greater than the 0.05 which clarifys that the data significantly normal and supports the null hypothesis, which the analaysis is not that clear. The normal curve of the cereme shows the data is not normal, even though if we consider the data is normal. So for the further analysis I have conducted the shapiro normality test (W = 0.92201, p-value = 0.04414) where p-value is less than the 0.05 which says the means of the group is not equal so it supports the alernative hypothesis. from the bartlett test (Bartlett's K-squared = 0.96331, df = 2, p-value = 0.6178) the p-value is greater than 0.05, that defines us that their is no much difference in variance between group variables. from Anova test if we consider (F value = 28.41 Pr(>F) = 4.7e-07 \*\*\*) the f value has got the higher values and p-values is less than the 0.05, as most of the tests support that their is a sifficant difference in creme produciton.

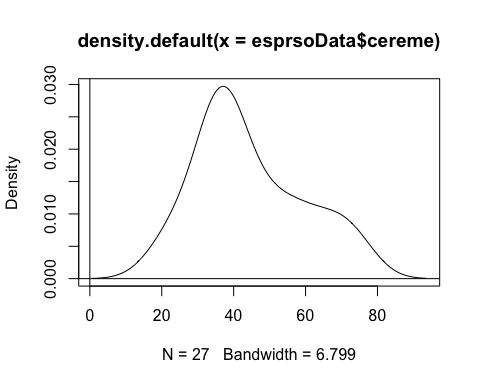
esprsoData <- read.csv('EspressoData.csv')  
  
#Formating the data  
esprsoData$brewmethod <- as.factor(esprsoData$brewmethod)  
str(esprsoData)

## 'data.frame': 27 obs. of 2 variables:  
## $ cereme : num 36.6 39.6 37.7 36 38.5 ...  
## $ brewmethod: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 2 ...

summary(esprsoData)

## cereme brewmethod  
## Min. :21.02 1:9   
## 1st Qu.:35.66 2:9   
## Median :38.52 3:9   
## Mean :44.47   
## 3rd Qu.:55.23   
## Max. :73.19

plot(density(esprsoData$cereme))  
abline(0,0, v = 0)



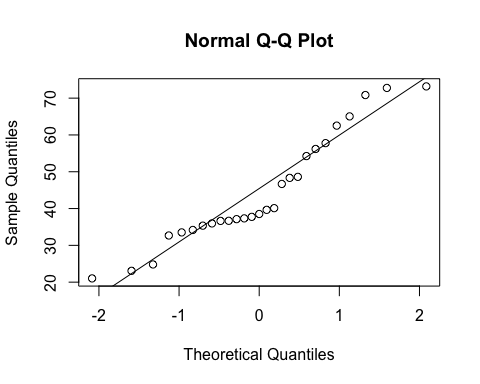
library(moments)  
agostino.test(esprsoData$cereme)

##   
## D'Agostino skewness test  
##   
## data: esprsoData$cereme  
## skew = 0.54679, z = 1.32787, p-value = 0.1842  
## alternative hypothesis: data have a skewness

shapiro.test(esprsoData$cereme)

##   
## Shapiro-Wilk normality test  
##   
## data: esprsoData$cereme  
## W = 0.92201, p-value = 0.04414

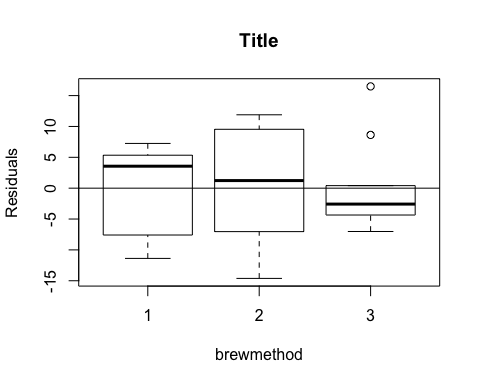
qqnorm(esprsoData$cereme)  
qqline(esprsoData$cereme)



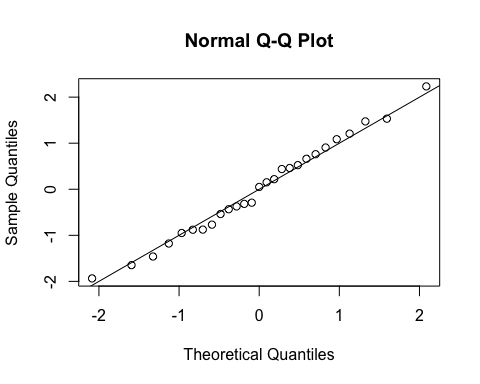
linearRegression = lm(cereme ~ brewmethod, data = esprsoData)   
standardized = rstudent(linearRegression)   
fitted = scale(linearRegression$fitted.values)  
summary(linearRegression, correlation = T)

##   
## Call:  
## lm(formula = cereme ~ brewmethod, data = esprsoData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.62 -6.60 0.41 5.73 16.49   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 32.400 2.819 11.492 3.04e-11 \*\*\*  
## brewmethod2 28.900 3.987 7.248 1.73e-07 \*\*\*  
## brewmethod3 7.300 3.987 1.831 0.0796 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.458 on 24 degrees of freedom  
## Multiple R-squared: 0.7031, Adjusted R-squared: 0.6783   
## F-statistic: 28.41 on 2 and 24 DF, p-value: 4.699e-07  
##   
## Correlation of Coefficients:  
## (Intercept) brewmethod2  
## brewmethod2 -0.71   
## brewmethod3 -0.71 0.50

linearRegressionModal = lm(cereme ~ brewmethod, data=esprsoData)   
residual = resid(linearRegressionModal)  
plot(esprsoData$brewmethod, residual, ylab= "Residuals", xlab = "brewmethod", main = "Title")   
abline(0, 0)



qqnorm(standardized)   
abline(0,1)



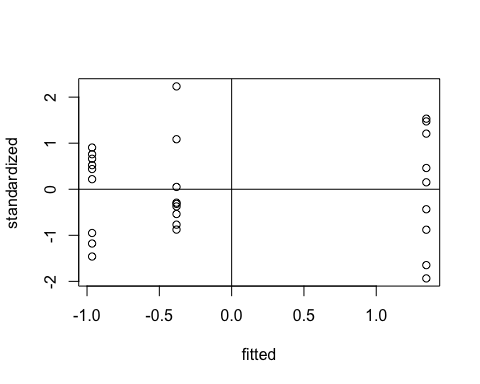
bartlett.test(esprsoData$cereme, esprsoData$brewmethod)

##   
## Bartlett test of homogeneity of variances  
##   
## data: esprsoData$cereme and esprsoData$brewmethod  
## Bartlett's K-squared = 0.96331, df = 2, p-value = 0.6178

tapply(esprsoData$cereme, esprsoData$brewmethod, var)

## 1 2 3   
## 53.29088 102.02220 59.30182

plot(fitted, standardized)   
abline(0,0, v = 0)



summary(aov(cereme ~ brewmethod, data = esprsoData))

## Df Sum Sq Mean Sq F value Pr(>F)   
## brewmethod 2 4065 2032.6 28.41 4.7e-07 \*\*\*  
## Residuals 24 1717 71.5   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

model <- aov(cereme ~ brewmethod, data = esprsoData)  
  
TukeyHSD(model)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = cereme ~ brewmethod, data = esprsoData)  
##   
## $brewmethod  
## diff lwr upr p adj  
## 2-1 28.9 18.942931 38.85707 0.0000005  
## 3-1 7.3 -2.657069 17.25707 0.1811000  
## 3-2 -21.6 -31.557069 -11.64293 0.0000419

library(pastecs)  
library(compute.es)  
  
by(esprsoData$cereme, esprsoData$brewmethod, stat.desc)

## esprsoData$brewmethod: 1  
## nbr.val nbr.null nbr.na min max range   
## 9.0000000 0.0000000 0.0000000 21.0200000 39.6500000 18.6300000   
## sum median mean SE.mean CI.mean.0.95 var   
## 291.6000000 35.9600000 32.4000000 2.4333533 5.6113228 53.2908750   
## std.dev coef.var   
## 7.3000599 0.2253105   
## ------------------------------------------------------------   
## esprsoData$brewmethod: 2  
## nbr.val nbr.null nbr.na min max range   
## 9.0000000 0.0000000 0.0000000 46.6800000 73.1900000 26.5100000   
## sum median mean SE.mean CI.mean.0.95 var   
## 551.7000000 62.5300000 61.3000000 3.3668680 7.7640115 102.0222000   
## std.dev coef.var   
## 10.1006039 0.1647733   
## ------------------------------------------------------------   
## esprsoData$brewmethod: 3  
## nbr.val nbr.null nbr.na min max range   
## 9.000000 0.000000 0.000000 32.680000 56.190000 23.510000   
## sum median mean SE.mean CI.mean.0.95 var   
## 357.300000 37.120000 39.700000 2.566923 5.919334 59.301825   
## std.dev coef.var   
## 7.700768 0.193974

modal12 <- mes(32.4, 61.3, 7.3, 10.1, 9, 9)

## Mean Differences ES:   
##   
## d [ 95 %CI] = -3.28 [ -4.69 , -1.86 ]   
## var(d) = 0.52   
## p-value(d) = 0   
## U3(d) = 0.05 %   
## CLES(d) = 1.02 %   
## Cliff's Delta = -0.98   
##   
## g [ 95 %CI] = -3.12 [ -4.47 , -1.78 ]   
## var(g) = 0.47   
## p-value(g) = 0   
## U3(g) = 0.09 %   
## CLES(g) = 1.36 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.87 [ -0.95 , -0.67 ]   
## var(r) = 0   
## p-value(r) = 0   
##   
## z [ 95 %CI] = -1.32 [ -1.83 , -0.81 ]   
## var(z) = 0.07   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0 [ 0 , 0.03 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = -5.95 [ -8.51 , -3.38 ]   
## var(lOR) = 1.71   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = -5   
## Total N = 18

modal12

## N.total n.1 n.2 d var.d l.d u.d U3.d cl.d cliffs.d pval.d g var.g  
## 1 18 9 9 -3.28 0.52 -4.69 -1.86 0.05 1.02 -0.98 0 -3.12 0.47  
## l.g u.g U3.g cl.g pval.g r var.r l.r u.r pval.r fisher.z var.z  
## 1 -4.47 -1.78 0.09 1.36 0 -0.87 0 -0.95 -0.67 0 -1.32 0.07  
## l.z u.z OR l.or u.or pval.or lOR l.lor u.lor pval.lor NNT  
## 1 -1.83 -0.81 0 0 0.03 0 -5.95 -8.51 -3.38 0 -5

modal13 <- mes(32.4, 39.7, 7.3, 7.7, 9, 9)

## Mean Differences ES:   
##   
## d [ 95 %CI] = -0.97 [ -1.95 , 0 ]   
## var(d) = 0.25   
## p-value(d) = 0.07   
## U3(d) = 16.53 %   
## CLES(d) = 24.57 %   
## Cliff's Delta = -0.51   
##   
## g [ 95 %CI] = -0.93 [ -1.86 , 0 ]   
## var(g) = 0.23   
## p-value(g) = 0.07   
## U3(g) = 17.71 %   
## CLES(g) = 25.62 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.46 [ -0.76 , 0.01 ]   
## var(r) = 0.03   
## p-value(r) = 0.07   
##   
## z [ 95 %CI] = -0.5 [ -1 , 0.01 ]   
## var(z) = 0.07   
## p-value(z) = 0.07   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0.17 [ 0.03 , 1.01 ]   
## p-value(OR) = 0.07   
##   
## Log OR [ 95 %CI] = -1.76 [ -3.54 , 0.01 ]   
## var(lOR) = 0.82   
## p-value(Log OR) = 0.07   
##   
## Other:   
##   
## NNT = -6.05   
## Total N = 18

modal13

## N.total n.1 n.2 d var.d l.d u.d U3.d cl.d cliffs.d pval.d g var.g  
## 1 18 9 9 -0.97 0.25 -1.95 0 16.53 24.57 -0.51 0.07 -0.93 0.23  
## l.g u.g U3.g cl.g pval.g r var.r l.r u.r pval.r fisher.z var.z l.z  
## 1 -1.86 0 17.71 25.62 0.07 -0.46 0.03 -0.76 0.01 0.07 -0.5 0.07 -1  
## u.z OR l.or u.or pval.or lOR l.lor u.lor pval.lor NNT  
## 1 0.01 0.17 0.03 1.01 0.07 -1.76 -3.54 0.01 0.07 -6.05

modal23 <- mes(61.3, 39.7, 10.1, 7.7, 9, 9)

## Mean Differences ES:   
##   
## d [ 95 %CI] = 2.41 [ 1.19 , 3.62 ]   
## var(d) = 0.38   
## p-value(d) = 0   
## U3(d) = 99.19 %   
## CLES(d) = 95.55 %   
## Cliff's Delta = 0.91   
##   
## g [ 95 %CI] = 2.29 [ 1.14 , 3.45 ]   
## var(g) = 0.35   
## p-value(g) = 0   
## U3(g) = 98.9 %   
## CLES(g) = 94.74 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = 0.79 [ 0.51 , 0.92 ]   
## var(r) = 0.01   
## p-value(r) = 0   
##   
## z [ 95 %CI] = 1.06 [ 0.56 , 1.57 ]   
## var(z) = 0.07   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 78.46 [ 8.69 , 707.96 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = 4.36 [ 2.16 , 6.56 ]   
## var(lOR) = 1.26   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = 1.35   
## Total N = 18

modal23

## N.total n.1 n.2 d var.d l.d u.d U3.d cl.d cliffs.d pval.d g var.g  
## 1 18 9 9 2.41 0.38 1.19 3.62 99.19 95.55 0.91 0 2.29 0.35  
## l.g u.g U3.g cl.g pval.g r var.r l.r u.r pval.r fisher.z var.z l.z  
## 1 1.14 3.45 98.9 94.74 0 0.79 0.01 0.51 0.92 0 1.06 0.07 0.56  
## u.z OR l.or u.or pval.or lOR l.lor u.lor pval.lor NNT  
## 1 1.57 78.46 8.69 707.96 0 4.36 2.16 6.56 0 1.35