Assignment 1

Rushabh Barbhaya

4/2/2021

## Analysis Of Variance

Exploring the data

library(moments)  
data = read.csv("EspressoData.csv")  
head(data)

## cereme brewmethod  
## 1 36.64 1  
## 2 39.65 1  
## 3 37.74 1  
## 4 35.96 1  
## 5 38.52 1  
## 6 21.02 1

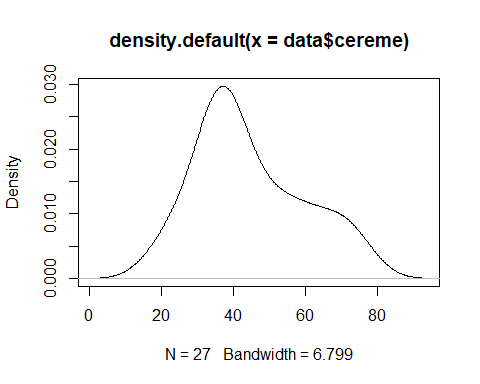
str(data)

## 'data.frame': 27 obs. of 2 variables:  
## $ cereme : num 36.6 39.6 37.7 36 38.5 ...  
## $ brewmethod: int 1 1 1 1 1 1 1 1 1 2 ...

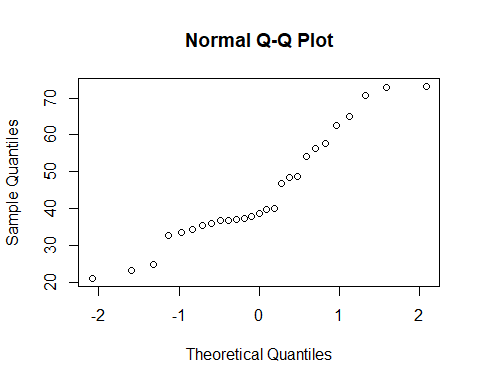
summary(data)

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

plot(density(data$cereme))



qqnorm(data$cereme)

 ## Skewness Test

There is no skewness  
 There is skewness in the data  
The p-value is greater than 0.1. we cannot accept the null hypothesis

agostino.test(data$cereme)

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

## Normality Test

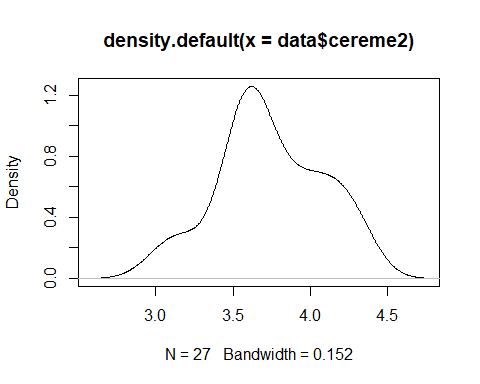
The data is normally distributed  
 The data is not normally distributed  
p-value is less than 0.05. we cannot reject the null hypothesis

shapiro.test(data$cereme)

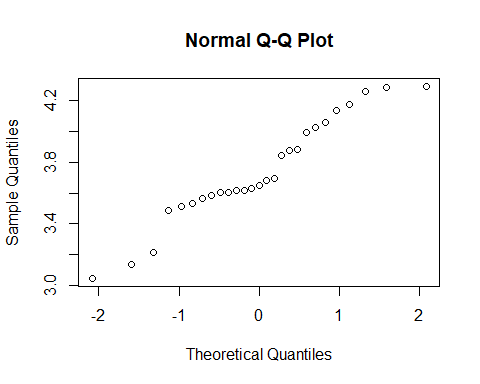
##   
## Shapiro-Wilk normality test  
##   
## data: data$cereme  
## W = 0.92201, p-value = 0.04414

performing log correction plotting density and residual plots

data$cereme2 = log(data$cereme)  
plot(density(data$cereme2))



qqnorm(data$cereme2)



agostino.test(data$cereme2)

##   
## D'Agostino skewness test  
##   
## data: data$cereme2  
## skew = -0.068827, z = -0.172938, p-value = 0.8627  
## alternative hypothesis: data have a skewness

shapiro.test(data$cereme2)

##   
## Shapiro-Wilk normality test  
##   
## data: data$cereme2  
## W = 0.95017, p-value = 0.2164

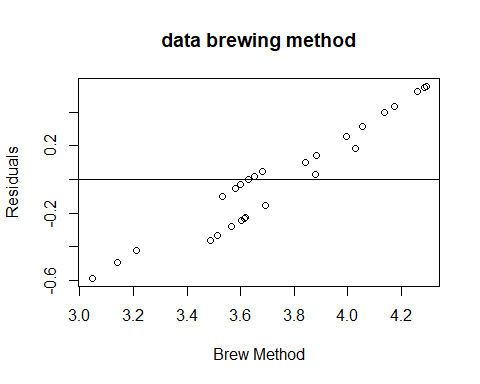
eruption.lm = lm(cereme2 ~ brewmethod, data = data)  
summary(eruption.lm)

##   
## Call:  
## lm(formula = cereme2 ~ brewmethod, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.58830 -0.23968 -0.00305 0.21702 0.55230   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.52679 0.16878 20.896 <2e-16 \*\*\*  
## brewmethod 0.10698 0.07813 1.369 0.183   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3315 on 25 degrees of freedom  
## Multiple R-squared: 0.06977, Adjusted R-squared: 0.03256   
## F-statistic: 1.875 on 1 and 25 DF, p-value: 0.1831

eruption.res = resid(eruption.lm)  
summary(eruption.res)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.588300 -0.239680 -0.003054 0.000000 0.217016 0.552303

plot(data$cereme2, eruption.res, ylab = "Residuals", xlab = "Brew Method", main = "data brewing method")  
abline(0,0)



## Variance Test

The variance in each group is the same  
 The variance in each group is not the same  
p-value is greater than 0.1. we cannot accept the hypothesis

bartlett.test(data$cereme2, data$brewmethod)

##   
## Bartlett test of homogeneity of variances  
##   
## data: data$cereme2 and data$brewmethod  
## Bartlett's K-squared = 1.3633, df = 2, p-value = 0.5058

tapply(data$cereme2, data$brewmethod, var)

## 1 2 3   
## 0.06130530 0.02883379 0.03150783

## ANOVA

summary(aov(cereme2 ~ factor(brewmethod), data = data))

## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(brewmethod) 2 1.9797 0.9898 24.41 1.64e-06 \*\*\*  
## Residuals 24 0.9732 0.0405   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

model2 <- aov(cereme2 ~ factor(brewmethod), data = data)  
summary(model2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(brewmethod) 2 1.9797 0.9898 24.41 1.64e-06 \*\*\*  
## Residuals 24 0.9732 0.0405   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Post hoc test

For pairwise test, brew method 2-3 is different from other  
Brew method 2 has a lot of variance as compared to others  
The effect for method 2 is too large

pairwise.t.test(data$cereme2, data$brewmethod, paired = FALSE, p.adjust.method = "bonferroni")

##   
## Pairwise comparisons using t tests with pooled SD   
##   
## data: data$cereme2 and data$brewmethod   
##   
## 1 2   
## 2 1.3e-06 -   
## 3 0.10080 0.00034  
##   
## P value adjustment method: bonferroni

library(pgirmess)  
kruskalmc(data$cereme2, factor(data$brewmethod), data = data)

## Multiple comparison test after Kruskal-Wallis   
## p.value: 0.05   
## Comparisons  
## obs.dif critical.dif difference  
## 1-2 14.666667 8.957452 TRUE  
## 1-3 3.666667 8.957452 FALSE  
## 2-3 11.000000 8.957452 TRUE

TukeyHSD(model2)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = cereme2 ~ factor(brewmethod), data = data)  
##   
## $`factor(brewmethod)`  
## diff lwr upr p adj  
## 2-1 0.6506828 0.41362616 0.8877395 0.0000013  
## 3-1 0.2139631 -0.02309354 0.4510198 0.0823041  
## 3-2 -0.4367197 -0.67377635 -0.1996630 0.0003265

library(compute.es)  
library(pastecs)  
by(data$cereme2, data$brewmethod, stat.desc)

## data$brewmethod: 1  
## nbr.val nbr.null nbr.na min max range   
## 9.00000000 0.00000000 0.00000000 3.04547437 3.68009095 0.63461658   
## sum median mean SE.mean CI.mean.0.95 var   
## 31.07286511 3.58240721 3.45254057 0.08253302 0.19032149 0.06130530   
## std.dev coef.var   
## 0.24759907 0.07171503   
## ------------------------------------------------------------   
## data$brewmethod: 2  
## nbr.val nbr.null nbr.na min max range   
## 9.00000000 0.00000000 0.00000000 3.84331581 4.29305880 0.44974299   
## sum median mean SE.mean CI.mean.0.95 var   
## 36.92901038 4.13564644 4.10322338 0.05660171 0.13052379 0.02883379   
## std.dev coef.var   
## 0.16980514 0.04138335   
## ------------------------------------------------------------   
## data$brewmethod: 3  
## nbr.val nbr.null nbr.na min max range   
## 9.00000000 0.00000000 0.00000000 3.48676327 4.02873881 0.54197554   
## sum median mean SE.mean CI.mean.0.95 var   
## 32.99853309 3.61415591 3.66650368 0.05916815 0.13644200 0.03150783   
## std.dev coef.var   
## 0.17750446 0.04841246

mes(3.45254057, 4.10322338, 0.24759907, 0.16980514, 9, 9)

## Mean Differences ES:   
##   
## d [ 95 %CI] = -3.06 [ -4.43 , -1.7 ]   
## var(d) = 0.48   
## p-value(d) = 0   
## U3(d) = 0.11 %   
## CLES(d) = 1.51 %   
## Cliff's Delta = -0.97   
##   
## g [ 95 %CI] = -2.92 [ -4.22 , -1.62 ]   
## var(g) = 0.44   
## p-value(g) = 0   
## U3(g) = 0.18 %   
## CLES(g) = 1.95 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.85 [ -0.94 , -0.64 ]   
## var(r) = 0   
## p-value(r) = 0   
##   
## z [ 95 %CI] = -1.26 [ -1.77 , -0.76 ]   
## var(z) = 0.07   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0 [ 0 , 0.05 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = -5.56 [ -8.03 , -3.09 ]   
## var(lOR) = 1.59   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = -5   
## Total N = 18

mes(3.66650368, 4.10322338, 0.17750446, 0.16980514, 9, 9)

## Mean Differences ES:   
##   
## d [ 95 %CI] = -2.51 [ -3.75 , -1.28 ]   
## var(d) = 0.4   
## p-value(d) = 0   
## U3(d) = 0.6 %   
## CLES(d) = 3.77 %   
## Cliff's Delta = -0.92   
##   
## g [ 95 %CI] = -2.39 [ -3.57 , -1.22 ]   
## var(g) = 0.36   
## p-value(g) = 0   
## U3(g) = 0.83 %   
## CLES(g) = 4.52 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.8 [ -0.92 , -0.53 ]   
## var(r) = 0.01   
## p-value(r) = 0   
##   
## z [ 95 %CI] = -1.1 [ -1.6 , -0.59 ]   
## var(z) = 0.07   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0.01 [ 0 , 0.1 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = -4.56 [ -6.8 , -2.32 ]   
## var(lOR) = 1.31   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = -5.01   
## Total N = 18