Assignment 2 - ANOVA - Blocking

Rushabh Barbhaya

2021-04-10

## Questions

1. What assumption must we test to include a variable as a blocking factor?
   * Kurtosis, Skewness, Normality, Independence of Observation, Variance test, and Additivity
   * Blocking technique should also help us reduce the Error
2. Recognize the IV, DV, block and create a table for the following research statement.
   * Independent Variable: Motor Skill Test (condition)
   * Dependent Variable: Performance Score (Performance\_score)
   * Blocking Variable: Age Group (Age)

Specification Table

|  |  |
| --- | --- |
| Variable | #Levels |
| Motor Skill Test | 3 |
| Age(Block) | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Test1 | Test2 | Test3 |
| Age\_60-69 |  |  |  |
| Age\_70\_79 |  |  |  |
| Age\_80+ |  |  |  |

## ANOVA

#### Hypothesis

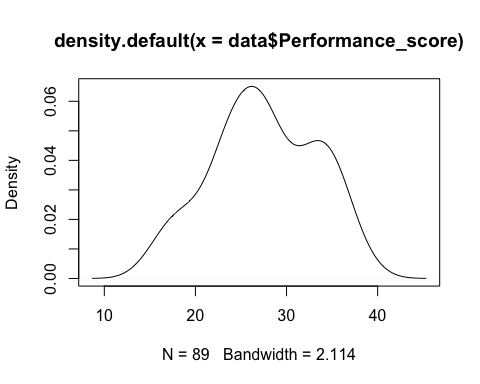
Performance Scores of all elderly people are equal  
 Performance Scores of all elderly people may not be equal

#### Assumptions

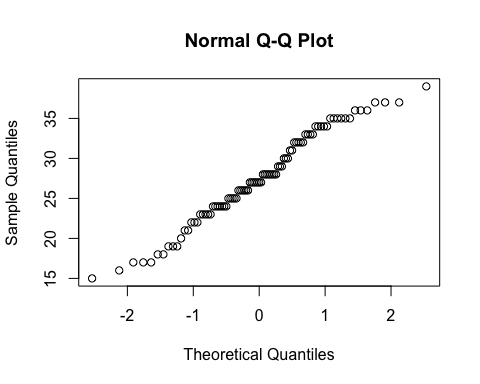
* From the density plot, there seems to have 2 spikes which may affect our results
* We fail to reject the null hypothesis on kurtosis test
* We fail to reject the null hypothesis from skewness test
* We fail to reject the null hypothesis in normality test
* We fail to reject the null hypothesis for variance test
* The largest to smallest ratio of variance is less than 3. Which is not enough evidence to reject the null hypothesis
* Model1 –> Condition\*Age, is not a significant factor
* Model2 –> Condition, is a significant factor
* Model3 –> Age, is a significant factor
* Model4 –> Condition+Age, is cannot be a blocker, it’s a significant factor
* Condition1 –> n = 28, Mean = 32.0714286, SD = 4.2854497
* Condition2 –> n = 30, Mean = 27.8666667, SD = 4.5768031
* Condition3 –> n = 31, Mean = 23.0645161, SD = 4.5529147

#### Summary

library(readxl)  
library(moments)  
library(pgirmess)  
library(pastecs)  
library(compute.es)  
  
data <- read\_excel("Lab3.xlsx")  
summary(data)  
  
# Denisty Plot  
plot(density(data$Performance\_score))



qqnorm(data$Performance\_score)



# Kurtosis Test  
anscombe.test(data$Performance\_score)  
  
# Skewness Test  
agostino.test(data$Performance\_score)  
  
# Normality Test  
shapiro.test(data$Performance\_score)  
  
#Residual Plot  
perf.lm = lm(Performance\_score ~ Condition, data = data)  
perf.res = resid(perf.lm)  
plot(data$Performance\_score, perf.res, ylab = "Residual", xlab = "Condition", main = "Independence of Observation")  
abline(0, 0)



# Variance Test  
bartlett.test(data$Performance\_score, data$Condition)  
tapply(data$Performance\_score, data$Condition, var)  
  
# ANOVA with Blocking  
model1 = aov(Performance\_score ~ factor(Condition)\*factor(Age), data = data)  
summary(model1)  
model2 = aov(Performance\_score ~ factor(Condition), data = data)  
summary(model2)  
model3 = aov(Performance\_score ~ factor(Age), data = data)  
summary(model3)  
model4 = aov(Performance\_score ~ factor(Condition)+factor(Age), data = data)  
summary(model4)  
  
anova(model1, model2)  
  
#Pairwise t test  
pairwise.t.test(data$Performance\_score, data$Condition, paired = FALSE, p.adjust.method = "bonferroni")  
  
#Tukey's test  
TukeyHSD(model1)  
  
by(data$Performance\_score, data$Condition, stat.desc)  
mes(27.8666667, 32.0714286, 4.5768031, 4.2854497, 30, 28)  
mes(23.0645161, 27.8666667, 4.5529147, 4.5768031, 31, 30)  
mes(23.0645161, 32.0714286, 4.5529147, 4.2854497, 31, 28)

## Age Performance\_score Condition   
## Min. :1 Min. :15.00 Min. :1.000   
## 1st Qu.:1 1st Qu.:24.00 1st Qu.:1.000   
## Median :2 Median :27.00 Median :2.000   
## Mean :2 Mean :27.52 Mean :2.034   
## 3rd Qu.:3 3rd Qu.:32.00 3rd Qu.:3.000   
## Max. :3 Max. :39.00 Max. :3.000   
##   
## Anscombe-Glynn kurtosis test  
##   
## data: data$Performance\_score  
## kurt = 2.2365, z = -2.0554, p-value = 0.03984  
## alternative hypothesis: kurtosis is not equal to 3  
##   
##   
## D'Agostino skewness test  
##   
## data: data$Performance\_score  
## skew = -0.11171, z = -0.45976, p-value = 0.6457  
## alternative hypothesis: data have a skewness  
##   
##   
## Shapiro-Wilk normality test  
##   
## data: data$Performance\_score  
## W = 0.9755, p-value = 0.09018  
##   
##   
## Bartlett test of homogeneity of variances  
##   
## data: data$Performance\_score and data$Condition  
## Bartlett's K-squared = 0.14381, df = 2, p-value = 0.9306  
##   
## 1 2 3   
## 18.36508 20.94713 20.72903   
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(Condition) 2 1199.0 599.5 313.667 <2e-16 \*\*\*  
## factor(Age) 2 1549.6 774.8 405.389 <2e-16 \*\*\*  
## factor(Condition):factor(Age) 4 22.6 5.7 2.961 0.0246 \*   
## Residuals 80 152.9 1.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(Condition) 2 1199 599.5 29.89 1.4e-10 \*\*\*  
## Residuals 86 1725 20.1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(Age) 2 1550 774.9 48.48 7.97e-15 \*\*\*  
## Residuals 86 1374 16.0   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(Condition) 2 1199.0 599.5 286.9 <2e-16 \*\*\*  
## factor(Age) 2 1549.6 774.8 370.8 <2e-16 \*\*\*  
## Residuals 84 175.5 2.1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## Analysis of Variance Table  
##   
## Model 1: Performance\_score ~ factor(Condition) \* factor(Age)  
## Model 2: Performance\_score ~ factor(Condition)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 80 152.91   
## 2 86 1725.19 -6 -1572.3 137.1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Pairwise comparisons using t tests with pooled SD   
##   
## data: data$Performance\_score and data$Condition   
##   
## 1 2   
## 2 0.0017 -   
## 3 6e-11 0.0002  
##   
## P value adjustment method: bonferroni   
## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Performance\_score ~ factor(Condition) \* factor(Age), data = data)  
##   
## $`factor(Condition)`  
## diff lwr upr p adj  
## 2-1 -4.204762 -5.072310 -3.337214 0  
## 3-1 -9.006912 -9.867679 -8.146146 0  
## 3-2 -4.802151 -5.647707 -3.956594 0  
##   
## $`factor(Age)`  
## diff lwr upr p adj  
## 2-1 -4.516166 -5.369099 -3.663233 0  
## 3-1 -10.310345 -11.177377 -9.443313 0  
## 3-2 -5.794179 -6.647112 -4.941246 0  
##   
## $`factor(Condition):factor(Age)`  
## diff lwr upr p adj  
## 2:1-1:1 -2.922222e+00 -4.947474 -0.8969700 0.0005097  
## 3:1-1:1 -8.022222e+00 -10.047474 -5.9969700 0.0000000  
## 1:2-1:1 -2.922222e+00 -4.947474 -0.8969700 0.0005097  
## 2:2-1:1 -8.722222e+00 -10.747474 -6.6969700 0.0000000  
## 3:2-1:1 -1.276768e+01 -14.748843 -10.7865103 0.0000000  
## 1:3-1:1 -9.666667e+00 -11.744532 -7.5888017 0.0000000  
## 2:3-1:1 -1.342222e+01 -15.447474 -11.3969700 0.0000000  
## 3:3-1:1 -1.872222e+01 -20.747474 -16.6969700 0.0000000  
## 3:1-2:1 -5.100000e+00 -7.071236 -3.1287642 0.0000000  
## 1:2-2:1 1.421085e-14 -1.971236 1.9712358 1.0000000  
## 2:2-2:1 -5.800000e+00 -7.771236 -3.8287642 0.0000000  
## 3:2-2:1 -9.845455e+00 -11.771369 -7.9195406 0.0000000  
## 1:3-2:1 -6.744444e+00 -8.769697 -4.7191922 0.0000000  
## 2:3-2:1 -1.050000e+01 -12.471236 -8.5287642 0.0000000  
## 3:3-2:1 -1.580000e+01 -17.771236 -13.8287642 0.0000000  
## 1:2-3:1 5.100000e+00 3.128764 7.0712358 0.0000000  
## 2:2-3:1 -7.000000e-01 -2.671236 1.2712358 0.9674422  
## 3:2-3:1 -4.745455e+00 -6.671369 -2.8195406 0.0000000  
## 1:3-3:1 -1.644444e+00 -3.669697 0.3808078 0.2078478  
## 2:3-3:1 -5.400000e+00 -7.371236 -3.4287642 0.0000000  
## 3:3-3:1 -1.070000e+01 -12.671236 -8.7287642 0.0000000  
## 2:2-1:2 -5.800000e+00 -7.771236 -3.8287642 0.0000000  
## 3:2-1:2 -9.845455e+00 -11.771369 -7.9195406 0.0000000  
## 1:3-1:2 -6.744444e+00 -8.769697 -4.7191922 0.0000000  
## 2:3-1:2 -1.050000e+01 -12.471236 -8.5287642 0.0000000  
## 3:3-1:2 -1.580000e+01 -17.771236 -13.8287642 0.0000000  
## 3:2-2:2 -4.045455e+00 -5.971369 -2.1195406 0.0000001  
## 1:3-2:2 -9.444444e-01 -2.969697 1.0808078 0.8583631  
## 2:3-2:2 -4.700000e+00 -6.671236 -2.7287642 0.0000000  
## 3:3-2:2 -1.000000e+01 -11.971236 -8.0287642 0.0000000  
## 1:3-3:2 3.101010e+00 1.119844 5.0821766 0.0001161  
## 2:3-3:2 -6.545455e-01 -2.580459 1.2713685 0.9750171  
## 3:3-3:2 -5.954545e+00 -7.880459 -4.0286315 0.0000000  
## 2:3-1:3 -3.755556e+00 -5.780808 -1.7303033 0.0000028  
## 3:3-1:3 -9.055556e+00 -11.080808 -7.0303033 0.0000000  
## 3:3-2:3 -5.300000e+00 -7.271236 -3.3287642 0.0000000  
##   
## data$Condition: 1  
## nbr.val nbr.null nbr.na min max range   
## 28.0000000 0.0000000 0.0000000 25.0000000 39.0000000 14.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 898.0000000 33.0000000 32.0714286 0.8098739 1.6617239 18.3650794   
## std.dev coef.var   
## 4.2854497 0.1336220   
## ------------------------------------------------------------   
## data$Condition: 2  
## nbr.val nbr.null nbr.na min max range   
## 30.0000000 0.0000000 0.0000000 20.0000000 35.0000000 15.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 836.0000000 27.5000000 27.8666667 0.8356061 1.7090064 20.9471264   
## std.dev coef.var   
## 4.5768031 0.1642393   
## ------------------------------------------------------------   
## data$Condition: 3  
## nbr.val nbr.null nbr.na min max range   
## 31.0000000 0.0000000 0.0000000 15.0000000 30.0000000 15.0000000   
## sum median mean SE.mean CI.mean.0.95 var   
## 715.0000000 24.0000000 23.0645161 0.8177276 1.6700226 20.7290323   
## std.dev coef.var   
## 4.5529147 0.1973991   
## Mean Differences ES:   
##   
## d [ 95 %CI] = -0.95 [ -1.49 , -0.4 ]   
## var(d) = 0.08   
## p-value(d) = 0   
## U3(d) = 17.17 %   
## CLES(d) = 25.15 %   
## Cliff's Delta = -0.5   
##   
## g [ 95 %CI] = -0.93 [ -1.47 , -0.4 ]   
## var(g) = 0.07   
## p-value(g) = 0   
## U3(g) = 17.5 %   
## CLES(g) = 25.44 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.43 [ -0.62 , -0.2 ]   
## var(r) = 0.01   
## p-value(r) = 0   
##   
## z [ 95 %CI] = -0.46 [ -0.73 , -0.2 ]   
## var(z) = 0.02   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0.18 [ 0.07 , 0.48 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = -1.72 [ -2.7 , -0.73 ]   
## var(lOR) = 0.25   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = -6.13   
## Total N = 58Mean Differences ES:   
##   
## d [ 95 %CI] = -1.05 [ -1.59 , -0.52 ]   
## var(d) = 0.07   
## p-value(d) = 0   
## U3(d) = 14.64 %   
## CLES(d) = 22.85 %   
## Cliff's Delta = -0.54   
##   
## g [ 95 %CI] = -1.04 [ -1.57 , -0.51 ]   
## var(g) = 0.07   
## p-value(g) = 0   
## U3(g) = 14.95 %   
## CLES(g) = 23.14 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.47 [ -0.65 , -0.25 ]   
## var(r) = 0.01   
## p-value(r) = 0   
##   
## z [ 95 %CI] = -0.51 [ -0.77 , -0.25 ]   
## var(z) = 0.02   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0.15 [ 0.06 , 0.39 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = -1.91 [ -2.88 , -0.94 ]   
## var(lOR) = 0.25   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = -5.85   
## Total N = 61Mean Differences ES:   
##   
## d [ 95 %CI] = -2.03 [ -2.66 , -1.4 ]   
## var(d) = 0.1   
## p-value(d) = 0   
## U3(d) = 2.1 %   
## CLES(d) = 7.52 %   
## Cliff's Delta = -0.85   
##   
## g [ 95 %CI] = -2.01 [ -2.63 , -1.39 ]   
## var(g) = 0.1   
## p-value(g) = 0   
## U3(g) = 2.24 %   
## CLES(g) = 7.79 %   
##   
## Correlation ES:   
##   
## r [ 95 %CI] = -0.72 [ -0.82 , -0.57 ]   
## var(r) = 0   
## p-value(r) = 0   
##   
## z [ 95 %CI] = -0.9 [ -1.17 , -0.64 ]   
## var(z) = 0.02   
## p-value(z) = 0   
##   
## Odds Ratio ES:   
##   
## OR [ 95 %CI] = 0.02 [ 0.01 , 0.08 ]   
## p-value(OR) = 0   
##   
## Log OR [ 95 %CI] = -3.69 [ -4.83 , -2.55 ]   
## var(lOR) = 0.34   
## p-value(Log OR) = 0   
##   
## Other:   
##   
## NNT = -5.05   
## Total N = 59