

```

library(readxl)
finalexam <- read_excel("Waterfleas.xlsx", sheet = "data", na = "NA")

head(finalexam)
str(finalexam)
colnames(finalexam)

#install packages

install.packages("car")
install.packages("psych")
install.packages("carData")
install.packages("mvtnorm")

library(car)
library(carData)
library(psych)
library(lattice)
library(data.table)
library(plyr)
library(doBy)
library(afex)
library(multcomp)
library(lsmeans)
library(effects)
library(MASS)
library(plot3D)
library(rgl)
library(gmodels)

#Log transformations of the dependent variables

finalexam$LOG2_Fecundity=log2(finalexam$Fecundity)
finalexam

# model 1, ANOVA

fit=aov(LOG2_Fecundity~ Hdepth, data=finalexam)
plot(allEffects(mod=fit))
summary(fit)

#one way ANOVA analysis confirms that Hdepth is statistically significant here.

#Model 1, linear regression
riazmodell1=lm(LOG2_Fecundity~ Hdepth, data = finalexam)
summary(riazmodell1)

#based on linear regression, Hdepth is statistically significant here

#Model 2, ANOVA

```

```

fit1=aov(LOG2_Fecundity~ Timeshift, data=finalexam)
plot(allEffects(mod=fit1))
summary(fit1)

#one way ANOVA analysis confirms that Timeshift is not statistically
significant here.

#Model 2 linear regression

riazmodel2=lm(LOG2_Fecundity~ Timeshift, data = finalexam)
summary(riazmodel2)

#Based on linear regression result Timeshift is statistically insignificant

#Bwplot Fecundity vs Hdepth and Fecundity vs Timeshift

bwplot(LOG2_Fecundity~Hdepth, data=finalexam, fill="blue")
bwplot(LOG2_Fecundity~Timeshift, data=finalexam, fill="blue")
bwplot(LOG2_Fecundity~Timeshift|Hdepth, data=finalexam, fill="blue")

#Model-3 main effect ANOVA, plot and Multiple linear regression
#Main effect ANOVA

fit2=aov(LOG2_Fecundity~ Hdepth+ Timeshift, data=finalexam)
summary(fit2)

plot(allEffects(mod=fit2))
plot(effect(mod=fit2, term = "Timeshift"))
plot(effect(mod=fit2, term = "Hdepth"))

summary.lm(fit2)

#Model-4 full factorial ANOVA. plot and multiple linear regression
#Full Factorial ANOVA

fit3=aov(LOG2_Fecundity~ Hdepth+Timeshift+Hdepth:Timeshift, data=finalexam)
summary(fit3)
plot(allEffects(mod=fit3))
summary.lm(fit3)

# AIC test result

AIC(fit)
AIC(fit1)
AIC(fit2)
AIC(fit3)

#based on AIC, AIC is lowest in model-1 (fit), rest of the model contains
higher AIC,

```

```

#So, I can conclude that model-1 is the best fitted model

#Final MODEL *

riazmodell=lm(LOG2_Fecundity~ Hdepth, data = finalexam)
summary(riazmodell)


# Model diagnosis, outliers and influential observations for "final"
#mODEL Diagnosis
# variance are homogenous

leveneTest(finalexam$LOG2_Fecundity, finalexam$Hdepth) # Levevtest for
homogeneity

max(by(finalexam$LOG2_Fecundity, finalexam$Hdepth,sd))^2/
min(by(finalexam$LOG2_Fecundity, finalexam$Hdepth,sd))^2


residualPlots(riazmodell)
#linearity cannot be assumed

spreadLevelPlot(riazmodell)
# the plot suggests that variance increases

ncvTest(riazmodell)
# Based on ncvtest (Chisquare = 9.23212, Df = 1, p = 0.0023781).
#Formal test dont rejects constant variance,based on p value model is
statistically significant


studres.riazmodell=rstudent(riazmodell) # studentized residuals
hist(studres.riazmodell,
     probability=T,
     col="lightgrey",
     xlim=c(-6,6),
     breaks=12,
     main="Distribution of Studentized Residuals",
     xlab="Studentized residuals")
xfit=seq(-6,6,length=100)
yfit=dnorm(xfit) # normal fit
lines(xfit, yfit, col="red",lwd=2)


# The distribution looks very good
#The Shapiro Wilk's W test using linear regression model
shapiro.test(residuals(riazmodell))
# Formal test confirming normality since W (0.95055) > 0.9
#W = 0.95055, p-value = 0.01219

```

```

outlierTest(riazmodell1) # based on Bonferroni p= 0.02, observation 2 is an
outlier
#No Studentized residuals with Bonferroni p < 0.05
#Largest |rstudent|:
#rstudent unadjusted p-value Bonferroni p
#2 -3.779943      0.00035901      0.022977

influenceIndexPlot(riazmodell1,vars=c("Studentized","Bonf"))
#observation 2 has a negative residual, meaning that it is lower than the
predicted value of the model, and
#significantly falls out outside the overall distribution of the residuals.

influenceIndexPlot(riazmodell1,vars="Cook")
# there are two influential observations, 1 and 2, We could therefore decide
to refit our model after leaving out
#these last outlier influential observations, nr 1 and 2

# Conclusion

# The model "final" meets most of the assumptions, so we keep this model.

Anova(riazmodell1, type = 3)
summary(riazmodell1)

#####

#Without log transformation
#####

# model 1, ANOVA

fitr=aov(Fecundity~ Hdepth, data=finalexam)
plot(allEffects(mod=fitr))
summary(fitr)

#one way ANOVA analysis confirms that Hdepth is statistically significant here.

#Model 1, linear regression
riazmodell11=lm(Fecundity~ Hdepth, data = finalexam)
summary(riazmodell11)

#based on linear regression, Hdepth is statistically significant here

#Model 2, ANOVA

fitr1=aov(Fecundity~ Timeshift, data=finalexam)
plot(allEffects(mod=fitr1))

```

```

summary(fitr1)

#one way ANOVA analysis confirms that Timeshift is not statistically
significant here.

#Model 2 linear regression

riazmodel21=lm(Fecundity~ Timeshift, data = finalexam)
summary(riazmodel21)

#Based on linear regression result Timeshift is statistically insignificant

#Bwplot Fecundity vs Hdepth and Fecundity vs Timeshift

bwplot(Fecundity~Hdepth, data=finalexam, fill="blue")
bwplot(Fecundity~Timeshift, data=finalexam, fill="blue")
bwplot(Fecundity~Timeshift|Hdepth, data=finalexam, fill="blue")

#Model-3 main effect ANOVA, plot and Multiple linear regression
#Main effect ANOVA

fitr2=aov(Fecundity~ Hdepth+ Timeshift, data=finalexam)
summary(fitr2)

plot(allEffects(mod=fitr2))
plot(effect(mod=fitr2, term = "Timeshift"))
plot(effect(mod=fitr2, term = "Hdepth"))

summary.lm(fitr2)

#Model-4 full factorial ANOVA. plot and multiple linear regression
#Full Factorial ANOVA

fitr3=aov(Fecundity~ Hdepth+Timeshift+Hdepth:Timeshift, data=finalexam)
summary(fitr3)
plot(allEffects(mod=fitr3))
summary.lm(fitr3)

# AIC test result

AIC(fitr)
AIC(fitr1)
AIC(fitr2)
AIC(fitr3)

#based on AIC, AIC is lowest in model-1 (fit), rest of the model contains
higher AIC,
#So, I can conclude that model-1 is the best fitted model

```

```

vif(fitr2)

#Final MODEL *

riazmodell1=lm(Fecundity~ Hdepth, data = finalexam)
summary(riazmodell1)


# Model diagnosis, outliers and influential observations for "final"
#mODEL Diagnosis
# variance are homogenous

leveneTest(finalexam$LOG2_Fecundity, finalexam$Hdepth) # Levevtest for
homogeneity

max(by(finalexam$Fecundity, finalexam$Hdepth,sd))^2/
min(by(finalexam$Fecundity, finalexam$Hdepth,sd))^2


residualPlots(riazmodell1)
#linearity cannot be assumed

spreadLevelPlot(riazmodell1)
# the plot suggests that variance increases

ncvTest(riazmodell1)
# Based on ncvtest
#Variance formula: ~ fitted.values
#Chisquare = 0.594455, Df = 1, p = 0.4407
#Formal test rejects constant variance,based on p value model is statistically
insignificant


studres.riazmodell1=rstudent(riazmodell1) # studentized residuals
hist(studres.riazmodell1,
     probability=T,
     col="lightgrey",
     xlim=c(-6,6),
     breaks=12,
     main="Distribution of Studentized Residuals",
     xlab="Studentized residuals")
xfit=seq(-6,6,length=100)
yfit=dnorm(xfit) # normal fit
lines(xfit, yfit, col="red",lwd=2)


# The distribution looks very good
#The Shapiro Wilk's W test using linear regression model
shapiro.test(residuals(riazmodell1))
#W = 0.96333, p-value = 0.05433
# Formal test confirming normality since W (0.96) > 0.9

```

```

outlierTest(riazmodell1) # observation 38 is an outlier
#No Studentized residuals with Bonferroni p < 0.05
#Largest |rstudent|:
#  rstudent unadjusted p-value Bonferroni p
#38 2.008432          0.04903          NA

influenceIndexPlot(riazmodell1,vars=c("Studentized","Bonf"))

influenceIndexPlot(riazmodell1,vars="Cook") # but there are no influential
observations, so no problems there

# Conclusion

# The model "final" meets most of the assumptions, so we keep this model.

Anova(riazmodell, type = 3)
summary(riazmodell1)

```