

Penicillin-.R

Kgomotso Xaba

Two-Way ANOVA of the effects of Penicillin production processes

- The dataset represents the results of 4 different treatments applied during the production of penicillin (treat), with 5 different blends of a particular substance used in the production process(blend) with results recorded from each experimental run(yield). Blend and treat are the explanatory variables and yield is the response variable.

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

library(lme4)

## Loading required package: Matrix

library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

library(tidyr)

##
## Attaching package: 'tidyr'

## The following objects are masked from 'package:Matrix':
##
##      expand, pack, unpack
```

```

library("faraway")
data("penicillin")
force(penicillin)

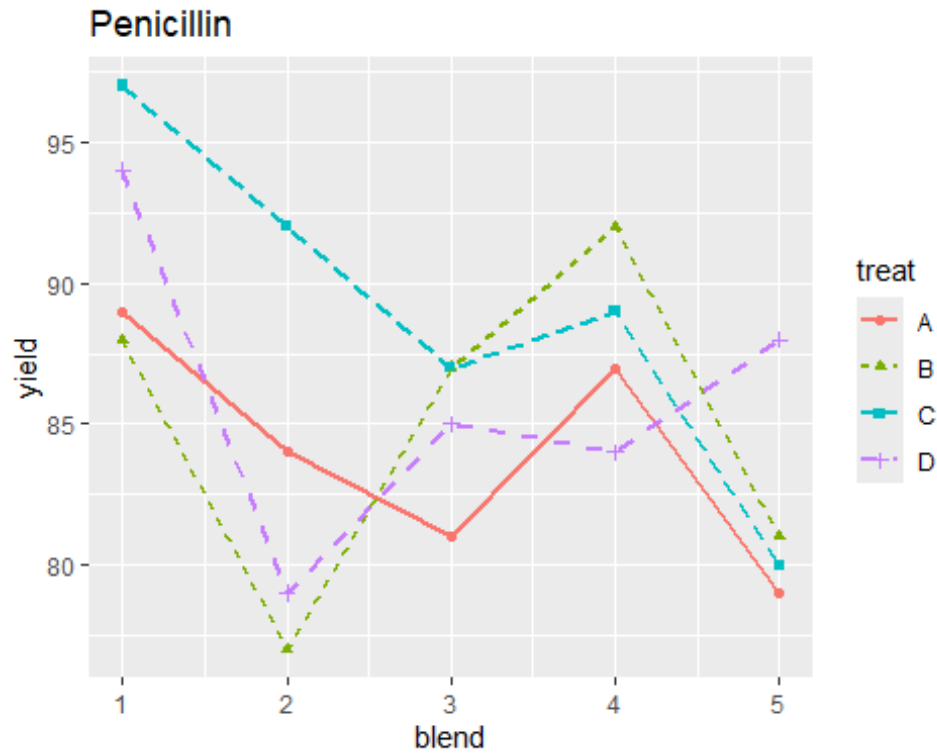
##      treat  blend yield
## 1      A Blend1    89
## 2      B Blend1    88
## 3      C Blend1    97
## 4      D Blend1    94
## 5      A Blend2    84
## 6      B Blend2    77
## 7      C Blend2    92
## 8      D Blend2    79
## 9      A Blend3    81
## 10     B Blend3    87
## 11     C Blend3    87
## 12     D Blend3    85
## 13     A Blend4    87
## 14     B Blend4    92
## 15     C Blend4    89
## 16     D Blend4    84
## 17     A Blend5    79
## 18     B Blend5    81
## 19     C Blend5    80
## 20     D Blend5    88

#changing the data type of the blend entries from character to numeric
#to allow us the plot the graph
penicillin$blend = as.numeric(penicillin$blend)
library(ggplot2)

# Line graph of the 4 treatments on the same axis
ggplot(penicillin,aes(blend,yield,colour = treat)) +
  geom_point(aes(shape = treat)) +
  geom_line(aes(linetype = treat), size = 1) + labs(title = "Penicillin")

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```



- Based on the interaction plot of fitted yield means above we see that it suggests interaction between blend and treat on yield based on the absence of parallel lines and the lines intersecting. Results could possibly have been influenced by the small sample size. We can also deduce that Treat C has the highest mean yield across different blends, followed by treat B, then treat D and treat A having the lowest mean yield recording.

#here we are checking the significance of interaction between treatments and blends

```
anova_interactions = aov(yield ~ treat*blend, penicillin)
summary(anova_interactions)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## treat         3   70.0   23.33   0.877 0.4801
## blend         1   90.0   90.00   3.383 0.0907 .
## treat:blend    3   80.8   26.93   1.013 0.4211
## Residuals    12  319.2   26.60
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova= aov(yield ~ treat+blend, penicillin)
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## treat         3    70    23.33   0.875 0.4759
## blend         1    90    90.00   3.375 0.0861 .
## Residuals    15   400    26.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#variance components and proportional allocation

```
model= lmer(yield ~ treat+blend+(1|blend), penicillin)
var_components = as.data.frame(VarCorr(model))
var_components$proportion = var_components$vcov/sum(var_components$vcov)
var_components
```

```
##      grp      var1 var2      vcov      sdcor proportion
## 1  blend (Intercept) <NA> 9.791673 3.129165 0.3420671
## 2 Residual      <NA> <NA> 18.833331 4.339739 0.6579329
```

#examines fitment, normality, outliers, indepedece of observations
plot(anova)

- H_0 : The interaction between blend and treatment is not statistically significant

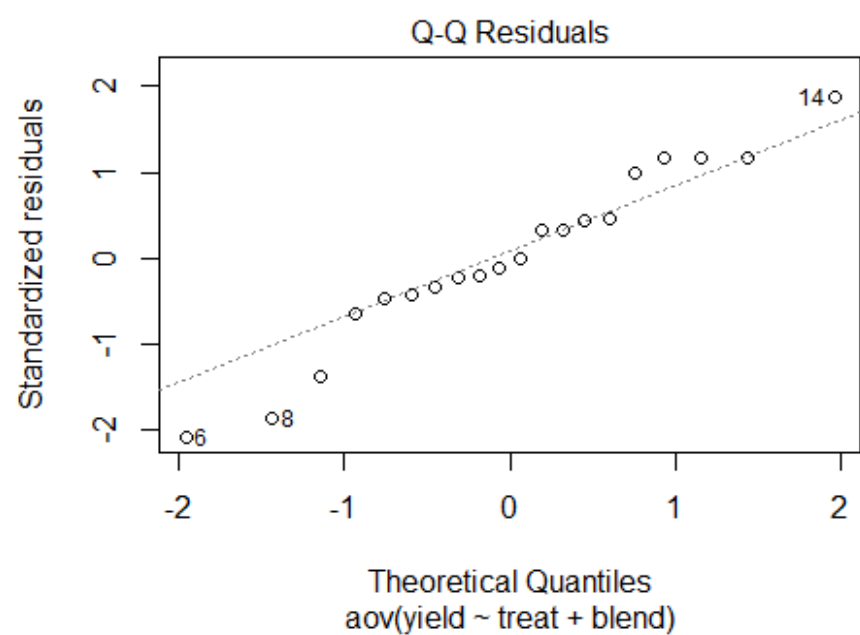
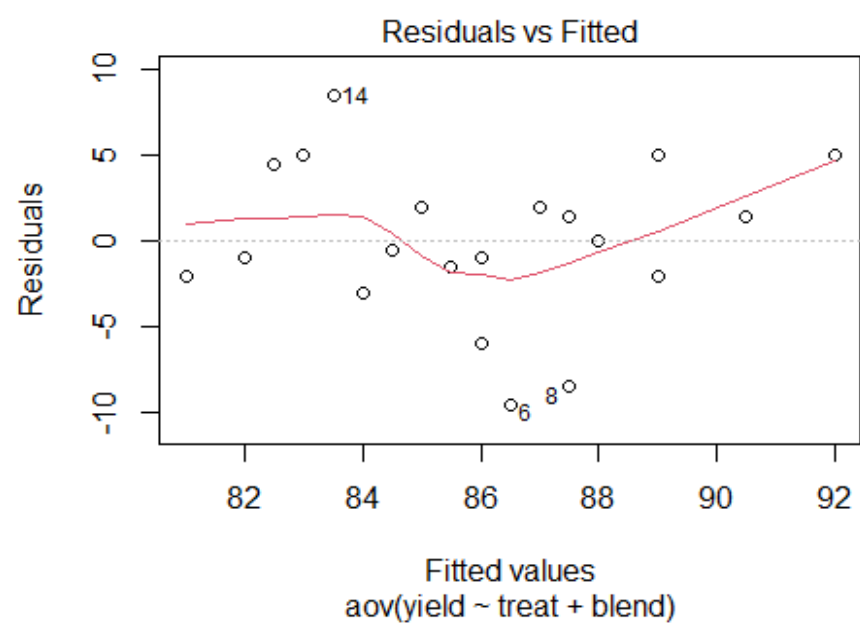
H_1 : There is a significant interaction effect between treatment factor and blend

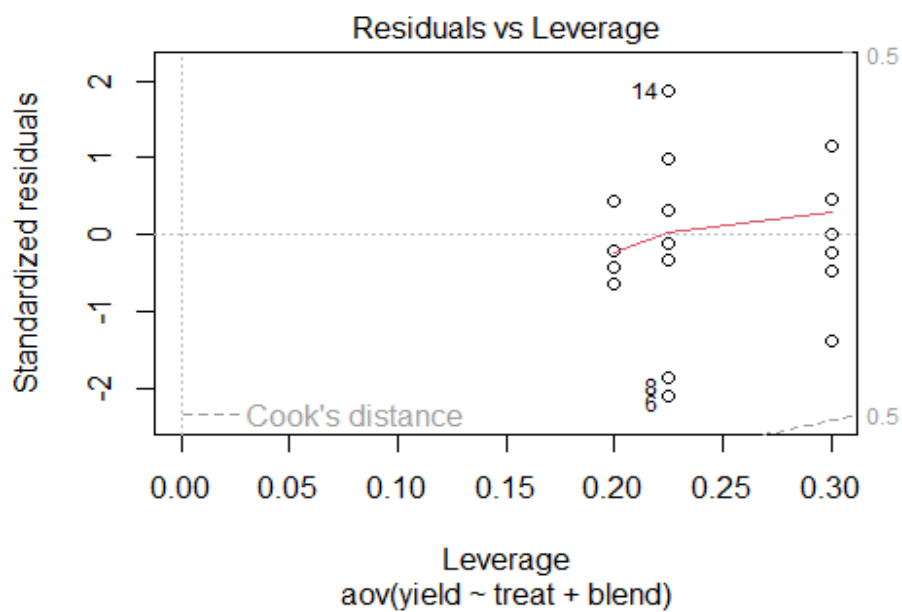
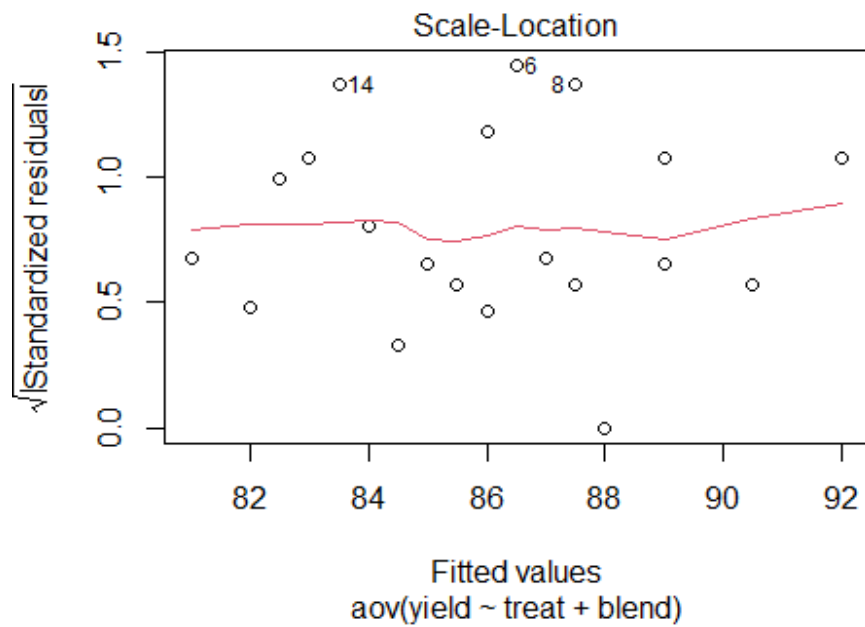
Since the p-value = 0.4211 > 0.05, we fail to reject the null hypothesis i.e. There is no significant interaction between treatments and blends on the yield

- H_0 : The type of blend or treatment do not have a significant effect on yield.

H_1 : The type of blend or treatment has a significant effect on yield.

Since the p-value=0.0861 > 0.05 & Treat p-value=0.4759 > 0.05, we fail to reject the null hypothesis, they each do not have a significant effect on yield.





#examine whether the distribution of the residuals is Normal or Not
`shapiro.test(residuals(anova))`

```
##
## Shapiro-Wilk normality test
##
## data: residuals(anova)
## W = 0.96232, p-value = 0.5911
```

#examine whether the variance of the residuals is constant or Not
bptest(anova)

```
##  
## studentized Breusch-Pagan test  
##  
## data: anova  
## BP = 2.788, df = 4, p-value = 0.5939
```

- Based on the Shapiro test $p\text{-value}=0.5911>0.05$, we fail to reject the null hypothesis i. e. residuals follow a Normal distribution. This can also be confirmed by the QQ-plot given the outcomes are grouped around the normality line with a few outliers also indicated
- Based on the BP test $p\text{-value}=0.5939>0.05$, we fail to reject the null hypothesis i. e. homoscedasticity exists. i.e. variance of residuals are constant across observations. The residuals & fitted plot shows linearity since values are equally spread on either side of horizontal line and it shows there are outliers.