Homework 4

Igor Kuivjogi Fernandes

2023-02-16

1 The ANOVA from a randomized complete block experiment output is shown below. In this experiment, 30 experimental units were evaluated.

Source	SS	DF	MS	F	P
Treatment	1010.56	4	-	29.84	-
Block	-	-	64.765	-	-
Error	169.33	20	-		
Total	1503.71	-			

a) Fill in the blanks.

```
pf(q = 29.84, df1 = 4, df2 = 20, lower.tail = F)  # for treatment
```

```
## [1] 3.544848e-08
```

```
pf(q = 7.64956, df1 = 5, df2 = 20, lower.tail = F)  # for block
```

```
## [1] 0.0003688504
```

Source	SS	DF	MS	F	P
Treatment	1010.56	4	252.64	29.84	3.544848e-08
Block	323.82	5	64.765	7.64956	0.0003688504
Error	169.33	20	8.4665		
Total	1503.71	29			

b) How many blocks were used in this experiment?
6 blocks

c) What conclusions can you draw?
The treatment effect is significant when at a significance level of $\alpha = 0.05$.
The blocking effect in this case is useful to reduce the error sum of squares.

2 An experiment with 12 hybrids of Brachiaria spp was carried out in a randomized block design with three replications. The variable measured was the leaf protein content (P %).

```r
df <- data.frame(
  hybrid = 1:12,
  b1 = c(6.8, 5.8, 6.8, 5.6, 6.9, 3.9, 6, 4.5, 6.1, 5.3, 5.9, 5.2),
  b2 = c(8.9, 6.4, 8.9, 6.2, 6.1, 4.9, 5.5, 5, 5.3, 6.5, 9, 6.4),
  b3 = c(10, 9, 11, 6.9, 7, 5.2, 7.9, 6.1, 8.5, 9.7, 11.2, 7.6)
)
df_long <- reshape(df, direction = 'long', idvar = 'hybrid', varying = c('b1', 'b2', 'b3'),
                   timevar = 'block', v.names = 'protein')
rownames(df_long) <- 1:nrow(df_long)
df_long$hybrid <- as.factor(df_long$hybrid)
df_long$block <- as.factor(df_long$block)
tibble::glimpse(df_long)
```

```
## Rows: 36
## Columns: 3
## $ hybrid  <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6, 7, 8,~
## $ block   <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ protein <dbl> 6.8, 5.8, 6.8, 5.6, 6.9, 3.9, 6.0, 4.5, 6.1, 5.3, 5.9, 5.2, 8.~
```

a) Formulate the statistical hypotheses H0 and H1 related to the hybrids.
In this example, the hybrid is a treatment, so we can build a hypothesis on this treatment:

H_0 : the mean leaf protein content is equal across all the hybrids
H_1 : at least one mean differs

b) Check the basic assumptions at 5% probability for the purpose of performing the ANAVA (normality of errors: Q-Q Plot; additivity of effects: Tukey test; homoscedasticity: Anscombe and Tukey test (1963)). Interpret the results. Perform the analysis of variance (ANAVA).

First, let's see whether using a blocking effect reduces error variance:

```r
fit <- aov(protein ~ hybrid, data = df_long)
summary(fit)
```

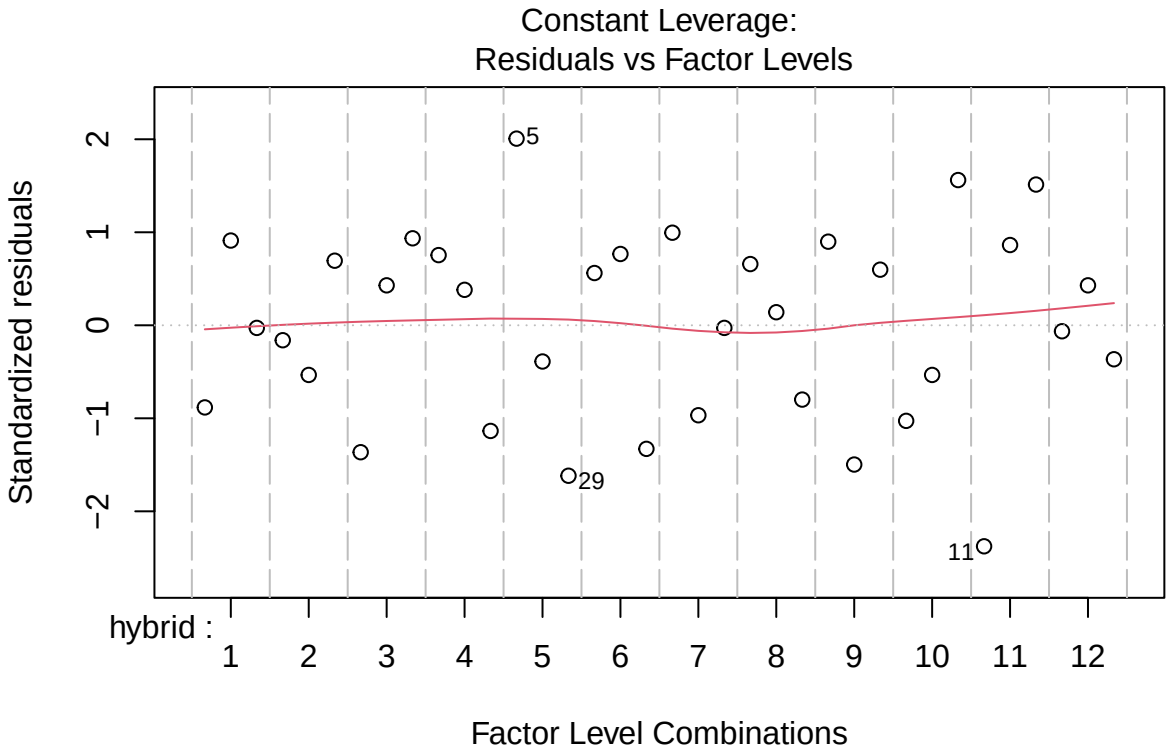
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## hybrid      11  57.00   5.182   2.085 0.0642 .
## Residuals   24  59.65   2.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit_block <- aov(protein ~ hybrid + block, data = df_long)
summary(fit_block)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## hybrid      11  57.00   5.182   6.612 9.00e-05 ***
## block        2  42.41  21.205  27.056 1.18e-06 ***
## Residuals   22  17.24   0.784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding a blocking effect is useful to reduce error variance. In fact, without a blocking effect the treatment effect would not be significant at a significance level of $\alpha = 0.05$ because p-value $= 0.0642 > \alpha$. Let's check the assumptions of ANOVA now:

```
plot(fit_block, which = 5)
```



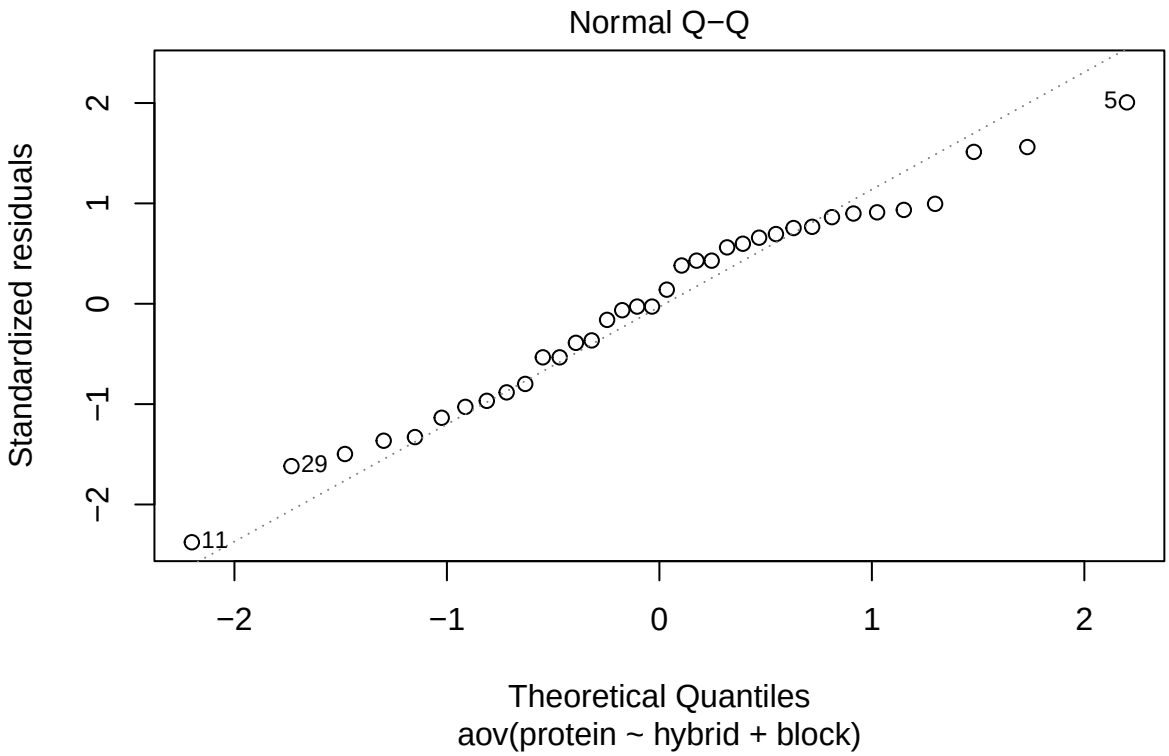
For the Levene's test, the null hypothesis is that the variances are equal across different levels.

```
car::leveneTest(protein ~ hybrid, data = df_long)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group 11  0.5719 0.8324
##       24
```

From the plot, we can see that the variance are homogeneous across the different levels. From the Levene's test, we don't reject the null hypothesis that the variances are equal using a significance level of $\alpha = 0.05$.

```
plot(fit_block, which = 2)
```



For the Shapiro Wilk test, the null hypothesis in this case is that the residuals come from a normal distribution.

```
shapiro.test(fit_block$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit_block$residuals
## W = 0.97646, p-value = 0.6254
```

The Q-Q plot shows that the central points are around the line but there are some points in the tails more far away from the line. From the test, we conclude that the residuals are normally distributed using a significance level of $\alpha = 0.05$, i.e. we don't reject the null hypothesis that the residuals are normally distributed.

Let's now check the additivity of effects using the tukey test for additivite effects:

```
daewr::Tukey1df(data.frame(df_long$protein, df_long$hybrid, df_long$block))
```

```
## Registered S3 method overwritten by 'DoE.base':
##   method           from
##   factorize.factor conf.design

## Source           df     SS        MS        F     Pr>F
## A                11   57.0022    5.182
```

```
## B                2   42.4106    21.2053
## Error           22   17.2428    3.1351
## NonAdditivity    1   6.1902    6.1902    11.76    0.0025
## Residual        21   11.0526    0.5263
```

We reject the null hypothesis that the effects are additive, i.e. we can see that there's interaction between the treatment and the block using a significance level of $\alpha = 0.05$.

c) Which hybrid performed best?

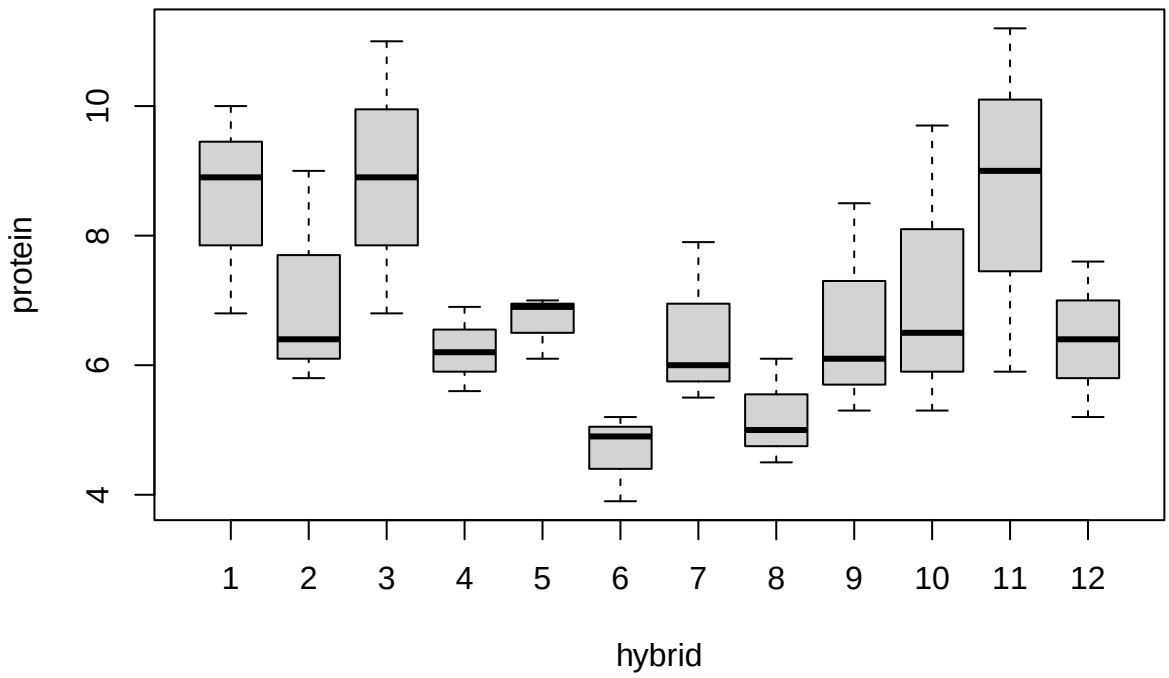
```
lsmeans::lsmeans(fit, ~hybrid)
```

```
##  hybrid lsmean   SE df lower.CL upper.CL
##  1        8.57 0.91 24     6.69    10.45
##  2        7.07 0.91 24     5.19     8.95
##  3        8.90 0.91 24     7.02    10.78
##  4        6.23 0.91 24     4.35     8.11
##  5        6.67 0.91 24     4.79     8.55
##  6        4.67 0.91 24     2.79     6.55
##  7        6.47 0.91 24     4.59     8.35
##  8        5.20 0.91 24     3.32     7.08
##  9        6.63 0.91 24     4.75     8.51
##  10       7.17 0.91 24     5.29     9.05
##  11       8.70 0.91 24     6.82    10.58
##  12       6.40 0.91 24     4.52     8.28
##
## Confidence level used: 0.95
```

The hybrid with the highest protein mean (i.e. performed the best) was the 3rd one.

d) Create a graph that shows the performance of different hybrids.

```
boxplot(protein ~ hybrid, data = df_long)
```


protein
4
6
8
10
1
2
3
4
5
6
7
8
9
10
11
12
hybrid