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:

$$H_0 : \text{the mean leaf protein content is equal across all the hybrids}$$
$$H_1 : \text{at least one mean differs}$$

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).

Firstly, 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:

```
par(mfrow = c(2, 2))
plot(fit, which = 5)
plot(fit, which = 1)
plot(fit, which = 2)
plot(residuals(fit_block) ~ rownames(df_long), main = 'Residuals vs Exp. Unit',
     font.main = 1, data = df_long, xlab = 'Experimental unit', ylab = 'Residuals')
abline(h = 0, lty = 2)
```
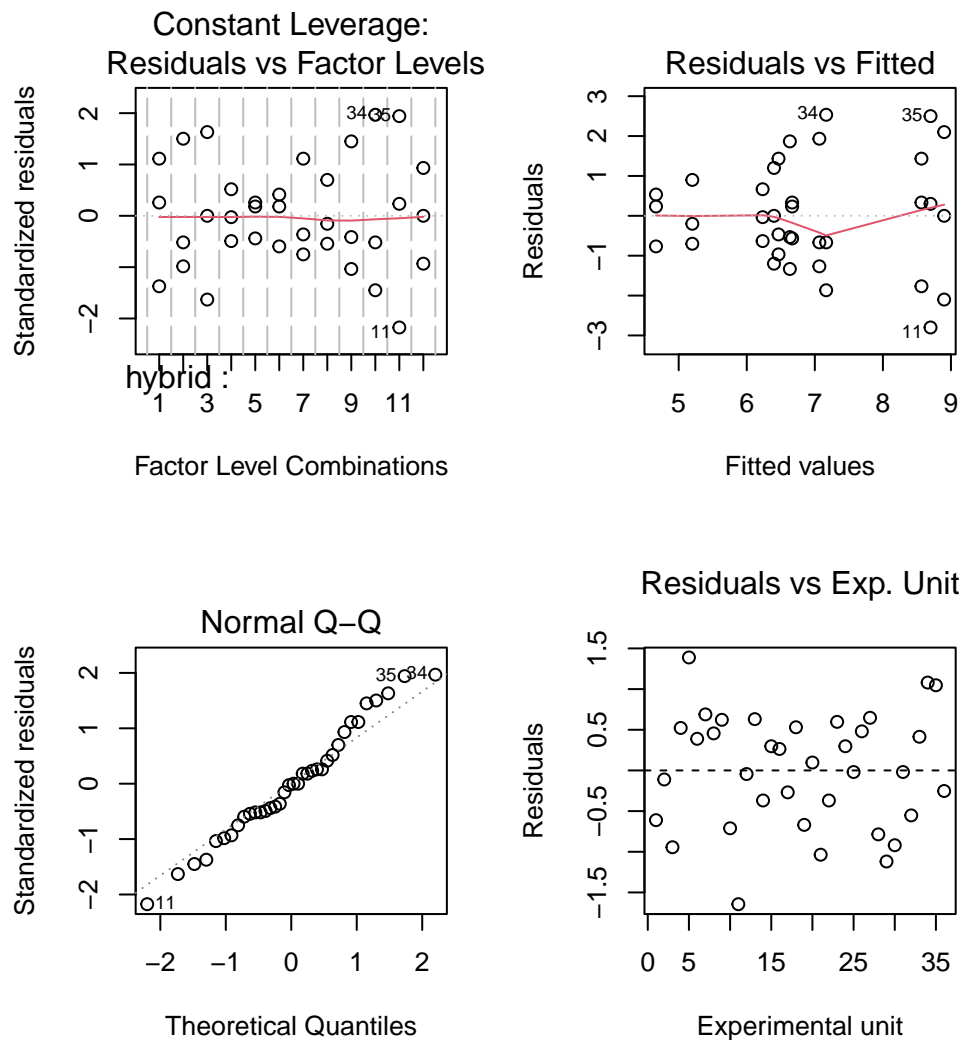

Some hybrids have a lower variance than others.

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.

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

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

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.

The Residual VS experimental Units plots shows an horizontal pattern across the zero, which is good.

c) Which hybrid performed best?

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