

# Homework 3

Igor Kuivjogi Fernandes

2023-02-09

1 The following output was obtained from a computer program that performed a two-factor ANOVA on a factorial experiment.

Source	DF	SS	MS	F	P
A	1	-	0.0002	-	-
B	-	180.378	-	-	-
Interaction	3	8.479	-	-	0.932
Error	8	158.797	-		
Total	15	347.653			

a) Fill in the blanks in the ANOVA table.

```
# calculating missing p-values
```

```
# P(F > 0.00001), df1 from the A factor, df2 from the Error  
pf(q = 0.00001, df1 = 1, df2 = 8, lower.tail = F)
```

```
## [1] 0.9975543
```

```
# P(F > 3.02917), df1 from the B factor, df2 from the Error  
pf(q = 3.02917, df1 = 3, df2 = 8, lower.tail = F)
```

```
## [1] 0.09334106
```

Source	DF	SS	MS	F	P
A	1	0.0002	0.0002	0.00001	0.9975543
B	3	180.378	60.126	3.02917	0.09334106
Interaction	3	8.479	2.826	0.14237	0.932
Error	8	158.797	19.849		
Total	15	347.653			

b) How many levels were used for factor B?

4 levels because DF from B treatment is  $b - 1 = 3$ , so  $b = 4$ .

c) How many replicates of the experiment were performed?

Degrees of freedom from Error is  $ab(r - 1) = 8$ , then  $2 \times 4 \times (r - 1) = 8$ , then  $8r = 16$ , then  $r = 2$  replicates.

2 Brewer's malt is produced from germinating barley, so brewers like to know under what conditions they should germinate their barley. The following is part of an experiment on barley germination. Barley seeds were divided into 30 lots of 100 seeds, and each lot of 100 seeds was germinated under one of ten conditions chosen at random. The conditions are the ten combinations of weeks after harvest (1, 3, 6, 9, or 12 weeks) and the amount of water used in germination (4 ml or 8 ml). The response is the number of seeds germinating. We are interested in whether the timing and/or amount of water affect germination. Analyze these data to determine how the germination rate depends on the treatments.

ml H <sub>2</sub> O	Age of Seeds (weeks)				
	1	3	6	9	12
4	11	7	9	13	20
	9	16	19	35	37
	6	17	35	28	45
8	8	1	5	1	11
	3	7	9	10	15
	3	3	9	9	25

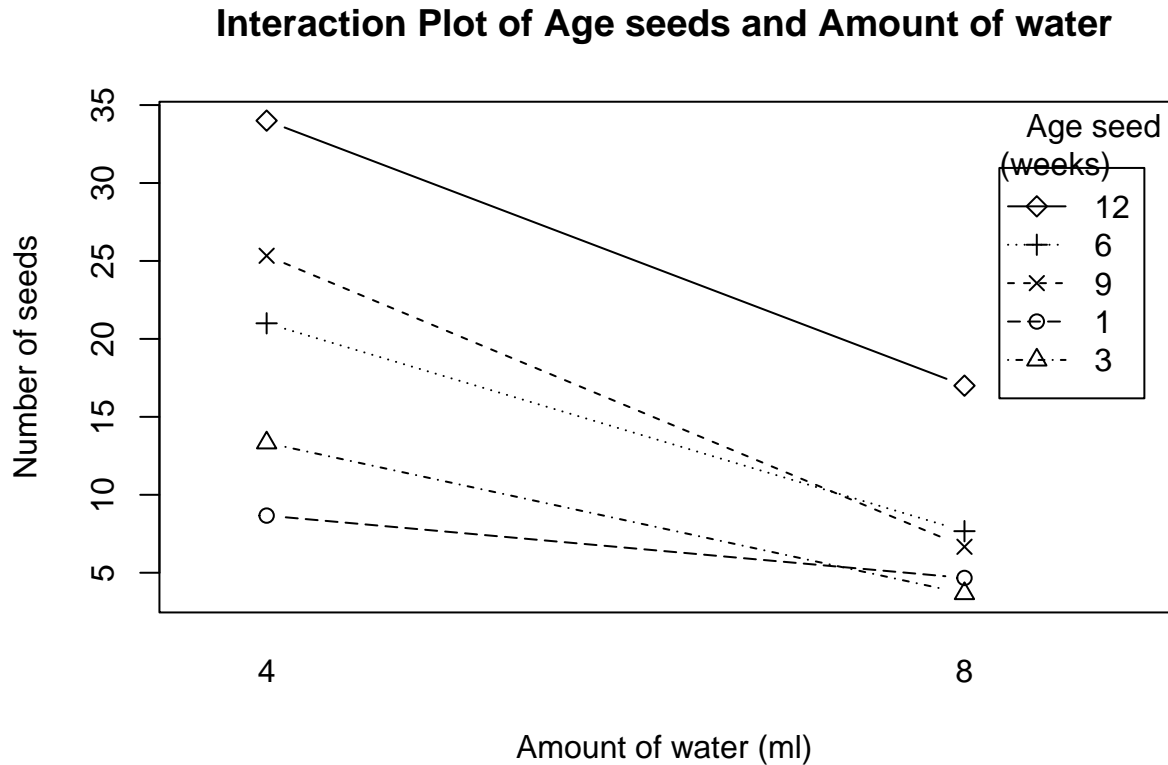
```
df <- expand.grid(h20 = c(4, 8), age_seeds = c(1, 3, 6, 9, 12))
df <- rbind(df, df, df) # 3 reps
df <- df[order(df$h20), ]
rownames(df) <- 1:nrow(df) # fix row numbers
df$h20 <- as.factor(df$h20)
df$age_seeds <- as.factor(df$age_seeds)
```

```
# assign response
df$seeds <- c(
  11, 7, 9, 13, 20,
  9, 16, 19, 35, 37,
  6, 17, 35, 28, 45,
  8, 1, 5, 1, 11,
  3, 7, 9, 10, 15,
  3, 3, 9, 9, 25
)
```

```
tibble::glimpse(df)
```

```
## Rows: 30
## Columns: 3
## $ h20      <fct> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 8, 8, 8, 8, 8, ~
## $ age_seeds <fct> 1, 3, 6, 9, 12, 1, 3, 6, 9, 12, 1, 3, 6, 9, 12, 1, 3, 6, 9, ~
## $ seeds    <dbl> 11, 7, 9, 13, 20, 9, 16, 19, 35, 37, 6, 17, 35, 28, 45, 8, 1~
```

```
with(df, {interaction.plot(h20, age_seeds, seeds, type = 'b',
                           pch = c(1, 2, 3, 4, 5), leg.bty = 'o',
                           main = 'Interaction Plot of Age seeds and Amount of water',
                           xlab = 'Amount of water (ml)', ylab = 'Number of seeds',
                           trace.label = 'Age seed\n(weeks)')})
```



In all different harvesting weeks (1, 3, 6, 9, 12) we observe a decreasing of count of seeds when we increase the water amount of water (in ml) from 4 to 8, however, the count of seeds decreases a lot (from ~35 to ~20) when we change the amount of water from 4 to 8 ml and harvest after 12 weeks (losango symbol), whereas a smaller decrease is seen when we change the amount of water from 4 to 8 ml but harvest after only 1 week (circle symbol).

When looking only to the Amount of water, using 4 mls rather than 8 always produced a larger number of seeds, despite the harvesting weeks.

Let's check whether these factors and the interaction are significant.

```
fit <- aov(seeds ~ h20 * age_seeds, data = df)
summary(fit)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## h20         1 1178.1   1178.1   19.723 0.000251 ***
## age_seeds    4 1321.1    330.3    5.529 0.003645 **
## h20:age_seeds 4  208.9     52.2    0.874 0.496726
## Residuals   20 1194.7     59.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using a significance level of  $\alpha = 0.05$ , we see that both the amount of water and harvesting weeks affect germination (i.e. are significant), however, the interaction between both treatments is not significant to the germination because  $p\text{-value} = 0.4906 > \alpha$ .

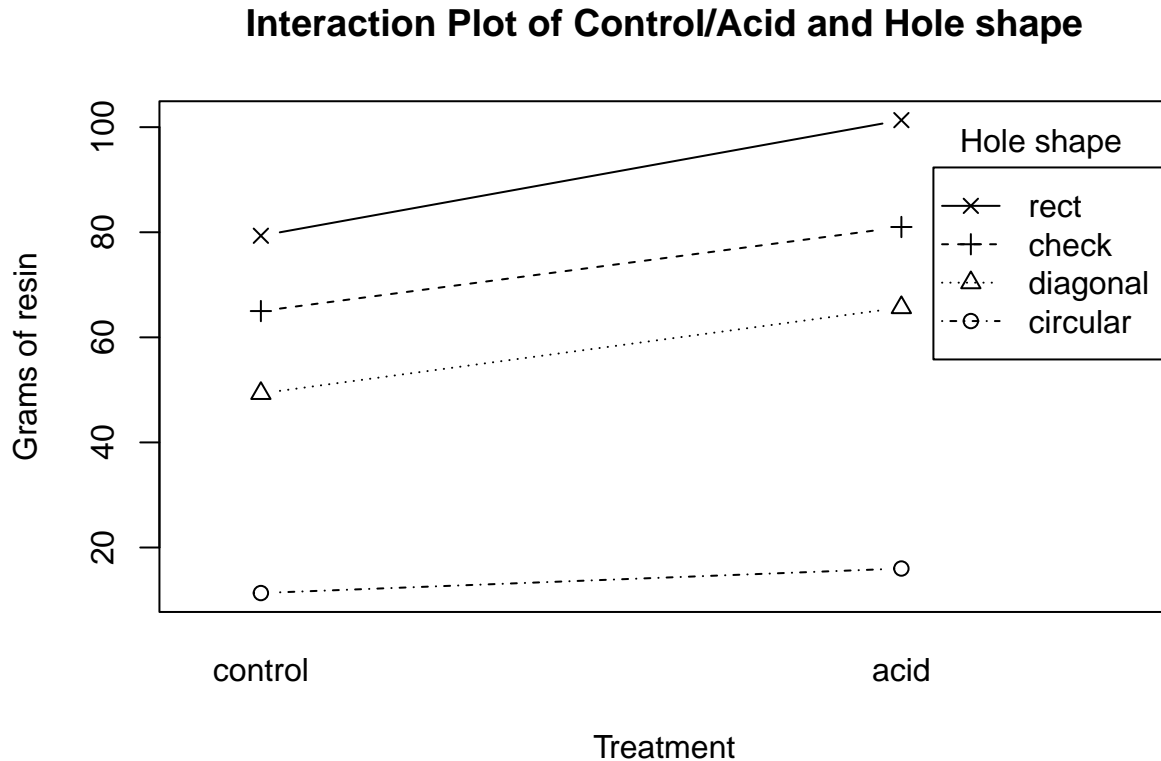
**3** Pine oleoresin is obtained by tapping the trunks of pine trees. Tapping is done by cutting a hole in the bark and collecting the resin that oozes out. This experiment compares four shapes for the holes and the efficacy of acid treating the holes. Twenty-four pine trees are randomly selected from a plantation, and the 24 are assigned randomly to the eight combinations of whole shape (circular, diagonal slash, check, rectangular) and acid treatment (yes or no). The response is the total grams of resin collected from the hole (data from Low and Bin Mohd. Ali 1985). Analyze these data to determine how the treatments affect resin yield. Include the Tukey HSD test in your analysis.

	Circular	Diagonal	Check	Rect.
Control	9	43	60	77
	13	48	65	70
	12	57	70	91
Acid	15	66	75	97
	13	58	78	108
	20	73	90	99

```
# similar to the previous question
df <- expand.grid(acid = c('control', 'acid'), shape = c('circular', 'diagonal', 'check', 'rect'))
df <- rbind(df, df, df)
df$acid <- factor(df$acid, c('control', 'acid')) # control level first
df$shape <- as.factor(df$shape)
df <- df[order(df$acid), ]
rownames(df) <- 1:nrow(df)
df$resin <- c(
  9, 43, 60, 77,
  13, 48, 65, 70,
  12, 57, 70, 91,
  15, 66, 75, 97,
  13, 58, 78, 108,
  20, 73, 90, 99
)
tibble::glimpse(df)
```

```
## Rows: 24
## Columns: 3
## $ acid <fct> control, control, control, control, control, control, control, c~
## $ shape <fct> circular, diagonal, check, rect, circular, diagonal, check, rect~
## $ resin <dbl> 9, 43, 60, 77, 13, 48, 65, 70, 12, 57, 70, 91, 15, 66, 75, 97, 1~
```

```
with(df, {interaction.plot(acid, shape, resin, type = 'b',
                           pch = c(1, 2, 3, 4), leg.bty = 'o',
                           main = 'Interaction Plot of Control/Acid and Hole shape',
                           xlab = 'Treatment', ylab = 'Grams of resin',
                           trace.label = 'Hole shape'))})
```



When looking only to the first treatment (control or acid), we can see that despite the hole shape the grams of resin are always bigger when using the control rather the acid treatment.

In addition, the rate of grams of resin seems bigger when changing from control to acid treatment when using the rectangular shape hole rather the circular one, for example.

The lines are quite parallel, which means that we would not expect an significant interaction.

We can fit an two-way ANOVA first and check the results:

```
fit <- aov(resin ~ acid * shape, data = df)
summary(fit)
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## acid         1   1305      1305  28.955 6.12e-05 ***
## shape        3  19407      6469 143.493 8.93e-12 ***
## acid:shape    3    237         79   1.756   0.196
## Residuals   16    721         45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this result, using a significance level of  $\alpha = 0.05$ , we can see that both control/acid treatment and hole

shape treatments affect the quantity of grams of resin, i.e, they are significant. However, the interaction is not significant because the p-value > 0.05.

```
fit_tukey <- TukeyHSD(fit, ordered = F)
fit_tukey
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = resin ~ acid * shape, data = df)
##
## $acid
##          diff      lwr      upr    p adj
## acid-control 14.75 8.939027 20.56097 6.12e-05
##
## $shape
##          diff      lwr      upr    p adj
## diagonal-circular 43.83333 32.742396 54.92427 0.0000000
## check-circular    59.33333 48.242396 70.42427 0.0000000
## rect-circular     76.66667 65.575729 87.75760 0.0000000
## check-diagonal    15.50000  4.409063 26.59094 0.0051399
## rect-diagonal     32.83333 21.742396 43.92427 0.0000014
## rect-check        17.33333  6.242396 28.42427 0.0019665
##
## $'acid:shape'
##          diff      lwr      upr    p adj
## acid:circular-control:circular 4.6666667 -14.313864 23.64720 0.9866494
## control:diagonal-control:circular 38.0000000 19.019469 56.98053 0.0000733
## acid:diagonal-control:circular 54.3333333 35.352803 73.31386 0.0000007
## control:check-control:circular 53.6666667 34.686136 72.64720 0.0000008
## acid:check-control:circular 69.6666667 50.686136 88.64720 0.0000000
## control:rect-control:circular 68.0000000 49.019469 86.98053 0.0000000
## acid:rect-control:circular 90.0000000 71.019469 108.98053 0.0000000
## control:diagonal-acid:circular 33.3333333 14.352803 52.31386 0.0003337
## acid:diagonal-acid:circular 49.6666667 30.686136 68.64720 0.0000024
## control:check-acid:circular 49.0000000 30.019469 67.98053 0.0000029
## acid:check-acid:circular 65.0000000 46.019469 83.98053 0.0000001
## control:rect-acid:circular 63.3333333 44.352803 82.31386 0.0000001
## acid:rect-acid:circular 85.3333333 66.352803 104.31386 0.0000000
## acid:diagonal-control:diagonal 16.3333333 -2.647197 35.31386 0.1199600
## control:check-control:diagonal 15.6666667 -3.313864 34.64720 0.1477801
## acid:check-control:diagonal 31.6666667 12.686136 50.64720 0.0005852
## control:rect-control:diagonal 30.0000000 11.019469 48.98053 0.0010360
## acid:rect-control:diagonal 52.0000000 33.019469 70.98053 0.0000013
## control:check-acid:diagonal -0.6666667 -19.647197 18.31386 1.0000000
## acid:check-acid:diagonal 15.3333333 -3.647197 34.31386 0.1636475
## control:rect-acid:diagonal 13.6666667 -5.313864 32.64720 0.2651715
## acid:rect-acid:diagonal 35.6666667 16.686136 54.64720 0.0001547
## acid:check-control:check 16.0000000 -2.980531 34.98053 0.1332427
## control:rect-control:check 14.3333333 -4.647197 33.31386 0.2199118
## acid:rect-control:check 36.3333333 17.352803 55.31386 0.0001247
## control:rect-acid:check -1.6666667 -20.647197 17.31386 0.9999817
## acid:rect-acid:check 20.3333333 1.352803 39.31386 0.0313384
## acid:rect-control:rect 22.0000000 3.019469 40.98053 0.0174419
```

From ANOVA, we saw that both treatments are significant (i.e. at least one treatment level differs from another). Now, with the Tukey HSD test, we can see that all the differences are significant as well (e.g., acid - control regarding the first treatment, and diagonal - circular and check - circular regarding the second treatment). In addition, ANOVA shows us the interaction was not significant, so we don't need to check it using a post hoc test such as Tukey HSD).

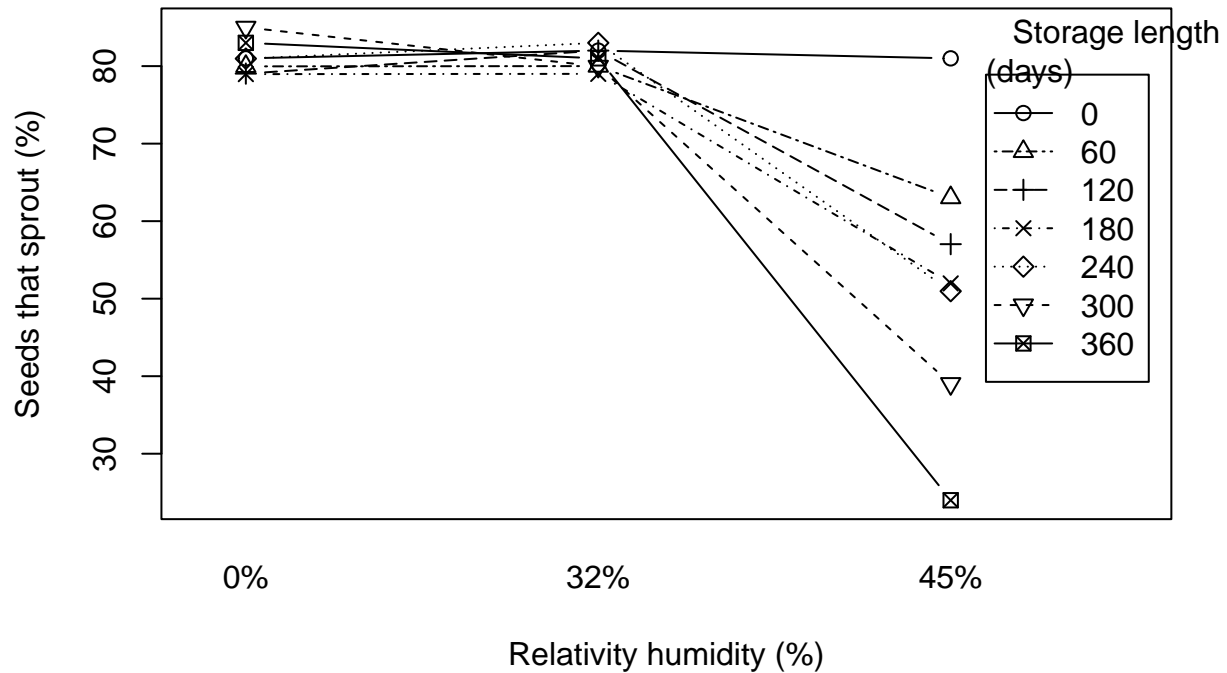
**4 Big sagebrush** is often planted in range restoration projects. An experiment is performed to determine the effects of storage length and relative humidity on the viability of seeds. Sixty-three batches of 300 seeds each are randomly divided into 21 groups of three. These 21 groups each receive a different treatment, namely the combinations of storage length (0, 60, 120, 180, 240, 300, or 360 days) and storage relative humidity (0, 32, or 45%). After the storage time, the seeds are planted, and the response is the percentage of seeds that sprout. Use the data set HW3\_Q4.xlsx and analyze these data for the effects of the factors on viability.

```
df <- expand.grid(rel_hum = c('0%', '32%', '45%'), storage_length = c(0, 60, 120, 180, 240, 300, 360))
df <- rbind(df, df, df) # 3 reps
df$rel_hum <- as.factor(df$rel_hum)
df$storage_length <- as.factor(df$storage_length)
df <- df[order(df$rel_hum), ]
rownames(df) <- 1:nrow(df)
df$seeds_sprout <- c(
  82.1, 78.6, 79.8, 82.3, 81.7, 85, 82.7,
  79, 80.8, 79.1, 75.5, 80.1, 87.9, 84.6,
  81.9, 80.5, 78.2, 79.1, 81.1, 82.1, 81.7,
  83.1, 78.1, 80.4, 77.8, 83.8, 82, 81,
  80.5, 83.6, 81.8, 80.4, 83.7, 77.6, 78.9,
  82.4, 78.3, 83.8, 78.8, 81.5, 80.3, 83.1,
  83.1, 66.5, 52.9, 52.9, 52.2, 38.6, 25.2,
  78.9, 61.4, 58.9, 54.3, 51.9, 37.9, 25.8,
  81, 61.2, 59.3, 48.7, 48.8, 40.6, 21
)
tibble::glimpse(df)

## Rows: 63
## Columns: 3
## $ rel_hum      <fct> 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, ~
## $ storage_length <fct> 0, 60, 120, 180, 240, 300, 360, 0, 60, 120, 180, 240, 3~
## $ seeds_sprout  <dbl> 82.1, 78.6, 79.8, 82.3, 81.7, 85.0, 82.7, 79.0, 80.8, 7~

with(df, {interaction.plot(rel_hum, storage_length, seeds_sprout, type = 'b',
  pch = c(1, 2, 3, 4, 5, 6, 7), leg.bty = 'o',
  main = 'Interaction Plot of Relativity humidity and Storage length',
  xlab = 'Relativity humidity (%)', ylab = 'Seeds that sprout (%)',
  trace.label = 'Storage length\n(days)')})
```

## Interaction Plot of Relativity humidity and Storage length



From the interaction plot we can see that when changing the relativity humidity from 32% to 45% the percentage of seeds that sprout varies depending of the days of storage length, which seems to show there is an interaction.

When changing the relativity humidity from 0% to 32% the percentage of seeds that sprout seems not to change that much when using different storage lengths. With very different slope lines across the last two levels of relativity humidity, we could suspect that there are interactions.

```
fit <- aov(seeds_sprout ~ rel_hum * storage_length, data = df)
summary(fit)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## rel_hum        2  11476    5738  1178.93 <2e-16 ***
## storage_length  6   1789     298    61.25 <2e-16 ***
## rel_hum:storage_length 12  4154     346    71.12 <2e-16 ***
## Residuals      42     204        5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using a significance level of  $\alpha = 0.05$ , we conclude that both treatments are significant. The interaction interaction is also significant, as we suspected from the interaction plot.