Lesson 11: Adapted from Example 3.2: Pistachios

LTC James K. Starling

Last compiled on 07 February, 2022

Review

1. What did we talk about in the previous lesson? Can you write out a mathematical representation of the model?

We talked about multi-factor experiments.

$y = \beta_0 + \beta_1 \cdot factor_1 + \beta_2 \cdot factor_2.$

2. What is the null and alternative hypothesis with a multi-factor experiment?

The null hypothesis is that factor 1 and factor 2 do not have an effect on the response variable.

$H_0 : \beta_1, \beta_2 = 0$

$H_a : \beta_1 \neq 0, and/or\ \beta_2 \neq 0$

In this chapter we're going to talk about interactions in multi-factor designs.

Background: Pistachios imported from the Middle East are cleaned and bleached with peroxide to turn them white. Some governments have banned bleaching over concerns of health risks; others have explored the impact on the health benefits of pistachios. Researchers (Gazor & Minaei, Dry. Technol., 2005) wanted to investigate the effects of the air velocity of the fan and the drying temperature of the oven on the amount of peroxide which remains (as a percentage) on the pistachios after the bleach process. Two values of air velocity (1.5 and 2.5 mph) and two values of drying temperature (60° and 90°F) were investigated. A full factorial design was conducted such that five batches, each consisting of 24 ounces of nuts, were randomly assigned to each treatment. We would like to determine the optimal settings of temperature and air velocity to minimize the percentage of peroxide remaining.

3. What were the response variables and the explanatory variable(s)? How are these variables classified?

The response variable is the peroxide remaining (percent) with explantory variables of temperature (two levels - 60/90 deg. F.) and air velocity (two levels - 1.5 mph and 2.5 mph).

4. Are any of these explanatory variables blocking variables? Why or why not.

No. The variables can be controlled by the experimenters and are not inherent properties of the pistachios.

Let's have a peek at the sources of variability diagram

Observed variation in: Peroxide remaining (%)	**Sources of explained variation**	**Sources of unexplained variation**
Inclusion criteria • Middle East *Design* • Same oven	• Temperature • Air velocity	• Porousness of the shell • Humidity • Unknown...

5. What are the observed means for each treatment? Create a table similar to the one in Figure 3.2.2. (p. 239).

```
obs.means <-pist.dat %>% group_by(Temperature, AirVelocity) %>%
  summarise(mn=mean(Peroxide)) %>%
  pivot_wider(names_from = Temperature, values_from = mn)
```

```
## `summarise()` has grouped output by 'Temperature'. You can override using the `.groups` argument.
```

```
obs.means
```

AirVelocity	60	90
<fct>	<dbl>	<dbl>
1.5	5.506	1.622
2.5	2.936	2.290

2 rows

6. Conduct a two-variable ANOVA with a two-factor model (don't forget to use effect coding). Calculate the percent of variation explained by each effect. Calculate each F statistic shown on p. 229. Which effects are statistically significant? What is the two-variable statistical model?

```
pist.dat2 <- pist.dat
contrasts(pist.dat2$Temperature) <- contr.sum
contrasts(pist.dat2$AirVelocity) <- contr.sum

pist.lm <- lm(Peroxide~ Temperature + AirVelocity, data=pist.dat2)
anova(pist.lm)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
Temperature	1	25.651125	25.651125	20.343861	0.0003087959
AirVelocity	1	4.522005	4.522005	3.586394	0.0754007680
Residuals	17	21.434925	1.260878	NA	NA

3 rows

```
summary(pist.lm)
```

```
##
## Call:
## lm(formula = Peroxide ~ Temperature + AirVelocity, data = pist.dat2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6555 -0.9645 -0.0340  0.7695  2.1995
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.0885     0.2511  12.301 6.89e-10 ***
## Temperature1   1.1325     0.2511   4.510 0.000309 ***
## AirVelocity1   0.4755     0.2511   1.894 0.075401 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.123 on 17 degrees of freedom
## Multiple R-squared:  0.5847, Adjusted R-squared:  0.5358
## F-statistic: 11.97 on 2 and 17 DF,  p-value: 0.0005707
```

```
SSE <- 21.435
SStemp <- 25.651
SSairvel <- 4.522
SSmodel <- (SStemp + SSairvel)
Fmodel <- (SSmodel / 2) / (SSE / 17); Fmodel
```

```
## [1] 11.96503
```

```
Ftemp <- (SStemp / 1) /  (SSE / 17); Ftemp
```

```
## [1] 20.34369
```

```
Fav <- (SSairvel / 1) /  (SSE / 17); Fav
```

```
## [1] 3.586377
```

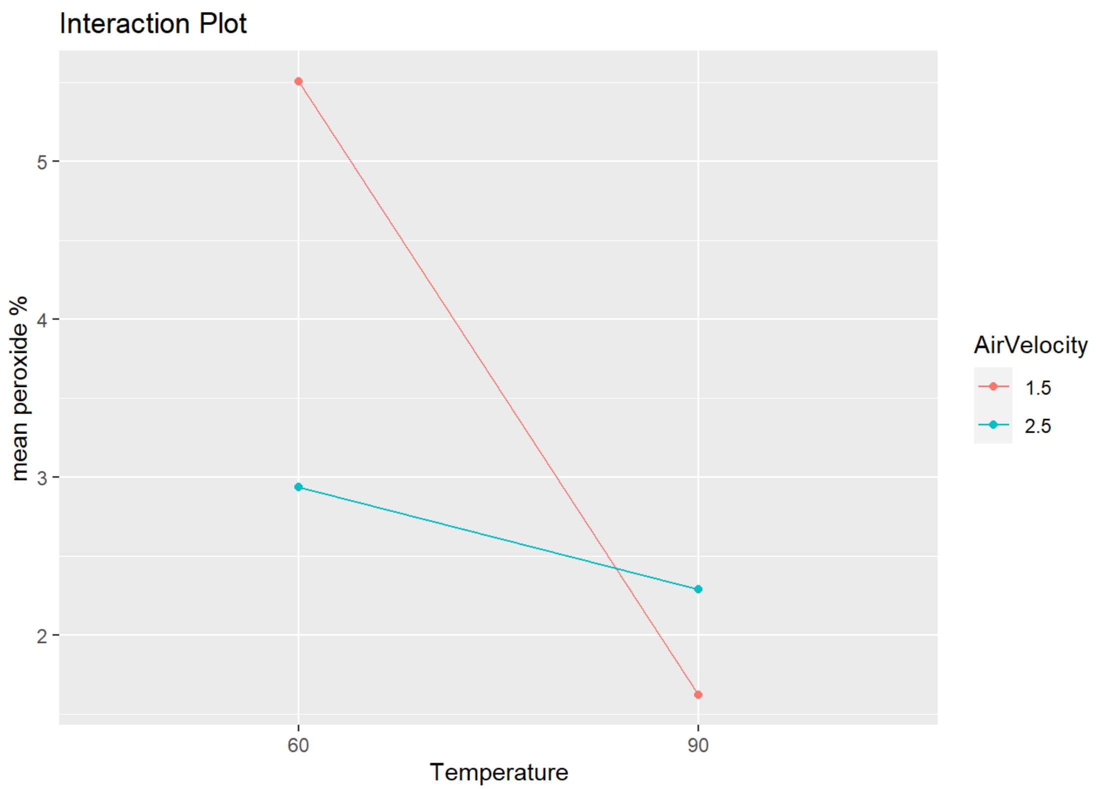
The effect of Temperature appears to be significant (at the alpha = 0.05 level), when accouting for the effects of Airvelcity. The effect of AirVelocity appears to be insignificant (at the alpha = 0.05 level), when accouting for the effects of Temperature.

7. Create an interaction plot. First, group the data by `Temperature` and `AirVelocity` , and then summarize the mean values of Peroxide and assign to `pist.int` . Then use the pipe command and `ggplot()` + `geom_line()` + `geom_point()` + (other stuff that makes sense)

```
pist.int <- pist.dat %>% group_by(Temperature, AirVelocity) %>%
  summarise(mn=mean(Peroxide))
```

```
## `summarise()` has grouped output by 'Temperature'. You can override using the `.groups` argument.
```

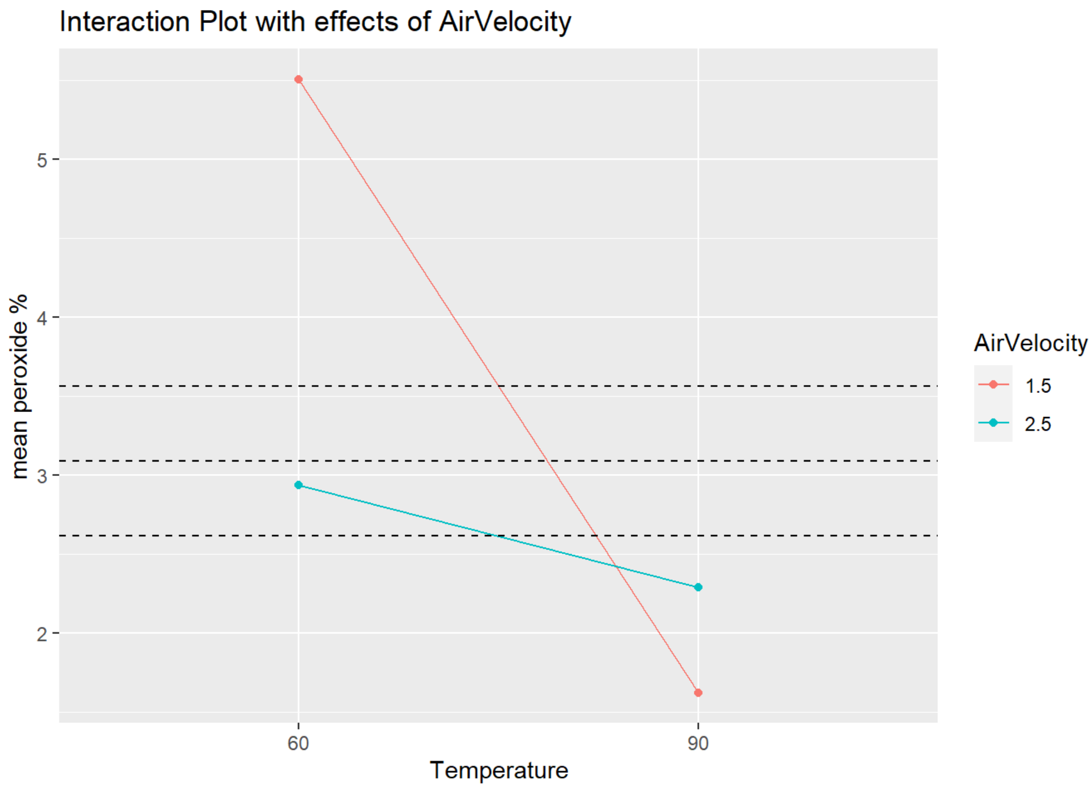
```
p <- pist.int %>% ggplot(aes(x=Temperature, y=mn, group= AirVelocity, color=AirVelocity)) +
  geom_line() +
  geom_point() +
  labs(title="Interaction Plot", y="mean peroxide %")
p
```



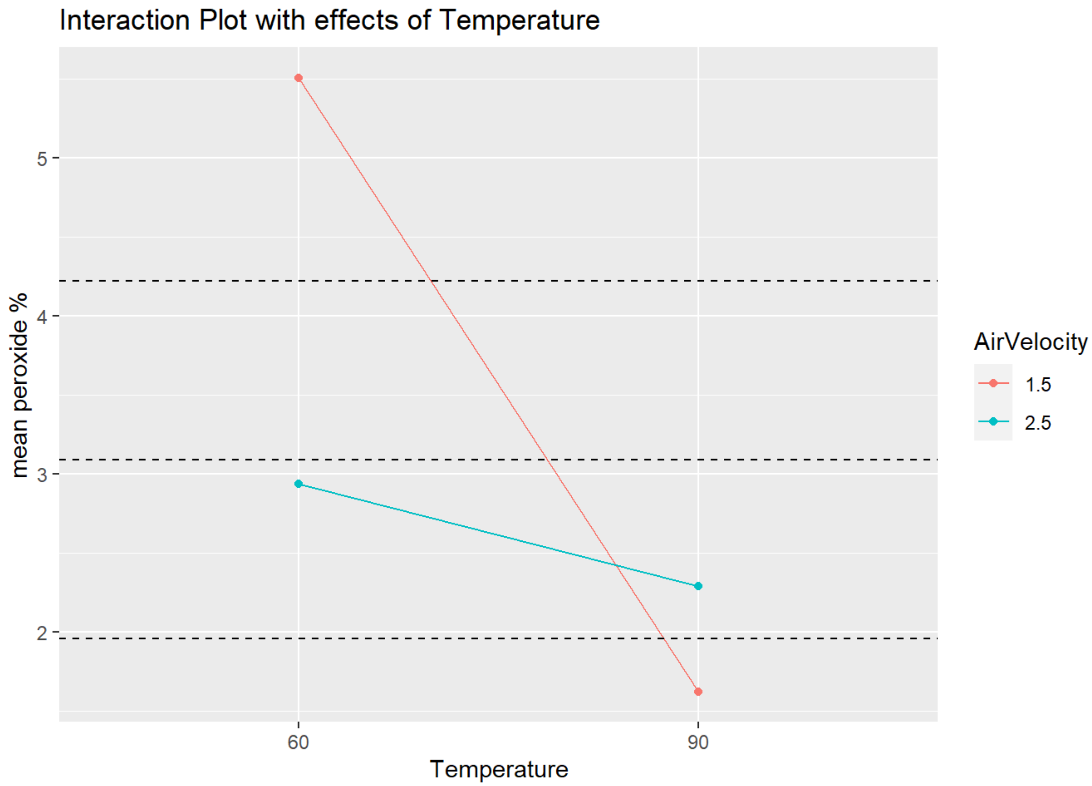
8. Using the previous figure, add vertical lines accounting for the effect of the air velocity and (separately) the effects of temperature. Use `geom_hline` to plot.

```
mod.mean <- 3.0885;
av.adj <- 0.4755;
temp.adj <- 1.1325;

p + geom_hline(yintercept = mod.mean, linetype = "dashed") +
  geom_hline(yintercept = mod.mean - av.adj, linetype = "dashed") +
  geom_hline(yintercept = mod.mean + av.adj, linetype = "dashed") +
  labs(title="Interaction Plot with effects of AirVelocity")
```




```
p + geom_hline(yintercept = mod.mean, linetype = "dashed") +
  geom_hline(yintercept = mod.mean - temp.adj, linetype = "dashed") +
  geom_hline(yintercept = mod.mean + temp.adj, linetype = "dashed") +
  labs(title="Interaction Plot with effects of Temperature")
```



9. How do we calculate the main effects for `Temperature` and `AirVelocity` ? If there were no effects, what would we expect the lines to look like?

```
overallmean <- mean(pist.dat$Peroxide)
obs.means
```

AirVelocity <fct>	60 <dbl>	90 <dbl>
1.5	5.506	1.622
2.5	2.936	2.290

2 rows

```
# Temperature effects:
avg15vel <- (5.51 + 1.62)/2
avg25vel <- (2.94+2.29) /2
c(overallmean - avg15vel, overallmean - avg25vel)
```

```
## [1] -0.4765  0.4735
```

```
# AirVelocity effects:
avg60deg <- (5.51+2.94)/2
avg90deg <- (1.62+2.29)/2
c(overallmean - avg60deg, overallmean - avg90deg)
```

```
## [1] -1.1365  1.1335
```

We see that the effects of the Temperature are +/- 1.1335, and the effects of the AirVelocity are +/- 0.4735. The lines would be parallel if there were no effects.

10. Calculate the `difference in the differences`. How do you interpret this answer?

```
obs.means
```

AirVelocity <fct>	**60** <dbl>	**90** <dbl>
1.5	5.506	1.622
2.5	2.936	2.290

2 rows

```
#diff when 60 deg:
5.51 - 2.94
```

```
## [1] 2.57
```

```
#diff when 90 deg:
1.62-2.29
```

```
## [1] -0.67
```

```
# diff of diffs
2.57-(-0.67)
```

```
## [1] 3.24
```

```
# diff when 1.5 mph
5.51-1.62
```

```
## [1] 3.89
```

```
# diff when 2.5 mph
2.94-2.29
```

```
## [1] 0.65
```

```
# diff of diffs
3.89-0.65
```

```
## [1] 3.24
```

11. Calculate the interaction effects by subtracting the predicted means from the observed means.

```
# copy obs means
pred.means <- obs.means
#replace entries
pred.means[1,'60'] <- predict(pist.lm, data.frame(Temperature='60',AirVelocity='1.5'))
pred.means[1,'90'] <- predict(pist.lm, data.frame(Temperature='90',AirVelocity='1.5'))
pred.means[2,'60'] <- predict(pist.lm, data.frame(Temperature='60',AirVelocity='2.5'))
pred.means[2,'90'] <- predict(pist.lm, data.frame(Temperature='90',AirVelocity='2.5'))
# subtract
obs.means - pred.means
```

```
## Warning in Ops.factor(left, right): '-' not meaningful for factors
```

AirVelocity <lgl>	60 <dbl>	90 <dbl>
NA	0.8095	-0.8095
NA	-0.8095	0.8095

2 rows

12. Verify the numbers of the interaction effects by using a linear model.

```
pist.int.lm <- lm(Peroxide~Temperature*AirVelocity, data=pist.dat2)
anova(pist.int.lm)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
Temperature	1	25.651125	25.651125	49.275073	2.894649e-06
AirVelocity	1	4.522005	4.522005	8.686642	9.463326e-03
Temperature:AirVelocity	1	13.105805	13.105805	25.175875	1.263170e-04
Residuals	16	8.329120	0.520570	NA	NA

4 rows

```
summary(pist.int.lm)
```

```
##
## Call:
## lm(formula = Peroxide ~ Temperature * AirVelocity, data = pist.dat2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9000 -0.5255 -0.1130  0.4715  1.3900
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3.0885     0.1613  19.144 1.88e-12 ***
## Temperature1                1.1325     0.1613   7.020 2.89e-06 ***
## AirVelocity1                0.4755     0.1613   2.947 0.009463 **
## Temperature1:AirVelocity1   0.8095     0.1613   5.018 0.000126 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7215 on 16 degrees of freedom
## Multiple R-squared:  0.8386, Adjusted R-squared:  0.8083
## F-statistic: 27.71 on 3 and 16 DF,  p-value: 1.422e-06
```

References

Nathan Tintle et al.(2019). Intermediate Statistical Investigations for U.S. Military Academy at West Point.