Lsn 12 - AY23

Clark

Admin

Pistachios imported form 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 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 on the pistachios after the bleach process. Two values of air velocity and two values of drying temperature 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.

How many treatments do we have?

How would the randomization be conducted?

Sources of variation diagram:

Statistical model:

```
pistachio = read.table("http://www.isi-stats.com/isi2/data/pistachioStudySubset.txt",header=T)
```

Here we're going to treat temperature and air velocity as factors not as continuous random variables. What's the downside of this?

```r
pistachio.fac<-pistachio %>% mutate(Temperaturef=as.factor(Temperature),AirVelocityf=as.factor(AirVeloci
contrasts(pistachio.fac$Temperaturef)=contr.sum
contrasts(pistachio.fac$AirVelocityf)=contr.sum
full.lm<-lm(Peroxide ~ Temperaturef+AirVelocityf,data=pistachio.fac)
coef(full.lm)
```

```
##   (Intercept) Temperaturef1 AirVelocityf1
##        3.0885        1.1325        0.4755
```

The fitted model is:

So according to the model, the best combination is:

```r
gr.means=pistachio.fac%>%group_by(Temperaturef,AirVelocityf)%>%summarize(mean.perox=mean(Peroxide))
```

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

```r
gr.means
```

```
## # A tibble: 4 x 3
## # Groups:   Temperaturef [2]
##   Temperaturef AirVelocityf mean.perox
##   <fct>        <fct>             <dbl>
## 1 60           1.5                5.51
## 2 60           2.5                2.94
## 3 90           1.5                1.62
## 4 90           2.5                2.29
```

Hmmm. . . .

```r
gr.means %>% ggplot(aes(x=Temperaturef,y=mean.perox,color=AirVelocityf))+
  geom_line(aes(group=AirVelocityf))+geom_point()
```

This is evidence of a **Statistical Interaction**. An interaction means that the effect is modified by the presence of another variable. In our experiment, if the temperature is 60, then it'd be more advantageous to have an air velocity of 2.5, but if the temperature is 90, then it'd be more advantageous to have an air velocity of 1.5.

Note that we are **not** saying that Air Velocity and Temperature are confounders. If we know the air velocity do we gain any knowledge of the temperature?

Our sources of variation diagram could now be:

Our statistical model is:

Our null and alternative for testing interaction is:

An appropriate statistic might be the difference of the differences, here we get:

```
(gr.means$mean.perox[1]-gr.means$mean.perox[2])-(gr.means$mean.perox[3]-gr.means$mean.perox[4])
```

```
## [1] 3.238
```

To see if this statistic is significant, as before, we need to know what the distribution under H_0 is, so we could do:

```
M=1000
stats.df=data.frame(rep=seq(1,M),stat=NA)
for(i in 1:M){
  pistachio.shuf=pistachio.fac
  pistachio.shuf$Tempshuf=sample(pistachio.shuf$Temperaturef)
  pistachio.shuf$Airshuf=sample(pistachio.shuf$AirVelocityf)
  gr.shuf=pistachio.shuf%>%group_by(Tempshuf,Airshuf)%>%
    summarize(mean.perox=mean(Peroxide),.groups = 'keep')
  stats.df[i,]$stat=(gr.shuf$mean.perox[1]-gr.shuf$mean.perox[2])-
    (gr.shuf$mean.perox[3]-gr.shuf$mean.perox[4])
}
```

So under H_0 how rare would it be to observe a 3.24?

```
stats.df %>% filter(abs(stat)>3.24)%>%summarise(pval=n()/M)
```

```
##    pval
## 1 0.034
```

Estimating interaciton effects might seem a bit tricky. In R we fit:

```
interact.lm<-lm(Peroxide ~ Temperaturef+AirVelocityf+Temperaturef:AirVelocityf,data=pistachio.fac)
coef(interact.lm)
```

```
##                 (Intercept)               Temperaturef1
##                      3.0885                      1.1325
##               AirVelocityf1 Temperaturef1:AirVelocityf1
##                      0.4755                      0.8095
```

Figuring out which combination gets a positive .8095 and which gets a -.8095 can be straight forward if we recall that $\alpha_1 = -\alpha_2$ and $\beta_1 = -\beta_2$.

Our ANOVA goes from

```
anova(full.lm)
```

```
## Analysis of Variance Table
##
## Response: Peroxide
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Temperaturef  1 25.651 25.6511 20.3439 0.0003088 ***
## AirVelocityf  1  4.522  4.5220  3.5864 0.0754008 .
## Residuals    17 21.435  1.2609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

to

```
anova(interact.lm)
```

```
## Analysis of Variance Table
##
## Response: Peroxide
##                           Df  Sum Sq Mean Sq F value    Pr(>F)
## Temperaturef               1 25.6511 25.6511 49.2751 2.895e-06 ***
## AirVelocityf               1  4.5220  4.5220  8.6866 0.0094633 **
## Temperaturef:AirVelocityf  1 13.1058 13.1058 25.1759 0.0001263 ***
## Residuals                 16  8.3291  0.5206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

See what happens to the Residuals.

Post Hoc tests can them proceed.

```
TukeyHSD(aov(Peroxide ~ Temperaturef*AirVelocityf,data=pistachio.fac))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Peroxide ~ Temperaturef * AirVelocityf, data = pistachio.fac)
##
## $Temperaturef
##         diff       lwr       upr   p adj
```

```
## 90-60 -2.265 -2.949023 -1.580977 2.9e-06
##
## $AirVelocityf
##           diff       lwr        upr     p adj
## 2.5-1.5 -0.951 -1.635023 -0.2669765 0.0094633
##
## $`Temperaturef:AirVelocityf`
##                 diff          lwr        upr     p adj
## 90:1.5-60:1.5 -3.884 -5.189540746 -2.5784593 0.0000014
## 60:2.5-60:1.5 -2.570 -3.875540746 -1.2644593 0.0001988
## 90:2.5-60:1.5 -3.216 -4.521540746 -1.9104593 0.0000150
## 60:2.5-90:1.5  1.314  0.008459254  2.6195407 0.0482522
## 90:2.5-90:1.5  0.668 -0.637540746  1.9735407 0.4805448
## 90:2.5-60:2.5 -0.646 -1.951540746  0.6595407 0.5081100
```

```
#Note * gives all main effects and interactions
```

For different combinations, our fitted models are: