

Lsn 14

Clark

Admin

Our text discusses a study in Vitamin C levels using 3 brands and 2 temperatures. There are six glasses of orange juice in their study. Here the data is:

```
OJ1 <- read.table("http://www.isi-stats.com/isi2/data/OJ-n6.txt",header=T)
OJ1
```

```
##   VitC      Brand      Temp
## 1 2.38 GreatValue Refrigerator
## 2 2.32 GreatValue      Room
## 3 2.34      Meijer Refrigerator
## 4 1.86      Meijer      Room
## 5 2.01 MinuteMaid Refrigerator
## 6 2.01 MinuteMaid      Room
```

They look at the data:

```
OJ1 %>% ggplot(aes(x=Brand,y=VitC,color=Temp))+
  geom_line(aes(group=Temp))+geom_point()
```

Since they just had MA376 they decide to fit this model:

```
contrasts(OJ1$Brand)=contr.sum
contrasts(OJ1$Temp)=contr.sum
OJ.lm<-lm(VitC~Brand*Temp,data=OJ1)
coef(OJ.lm)
```

```
## (Intercept)      Brand1      Brand2      Temp1 Brand1:Temp1
## 2.15333333 0.19666667 -0.05333333 0.09000000 -0.06000000
## Brand2:Temp1
## 0.15000000
```

```
levels(OJ1$Brand)
```

```
## [1] "GreatValue" "Meijer"      "MinuteMaid"
```

```
levels(OJ1$Temp)
```

```
## [1] "Refrigerator" "Room"
```

So their fitted model is:

```
anova(OJ.lm)
```

```
## Warning in anova.lm(OJ.lm): ANOVA F-tests on an essentially perfect fit are
## unreliable
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: VitC
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## Brand      2  0.12413  0.062067
```

```
## Temp      1  0.04860  0.048600
```

```
## Brand:Temp 2  0.06840  0.034200
```

```
## Residuals  0  0.00000
```

What's going on?

Remember that any statistical test or confidence interval relies on being able to compute a variance.

```
OJ1 %>% group_by(Brand,Temp)%>%summarize(Mean=mean(VitC),Sd=sd(VitC))
```

```
## # A tibble: 6 x 4
```

```
## # Groups:   Brand [3]
```

```
##   Brand      Temp      Mean      Sd
```

```
##   <fct>      <fct>    <dbl> <dbl>
```

```
## 1 GreatValue Refrigerator  2.38   NaN
```

```
## 2 GreatValue Room          2.32   NaN
```

```
## 3 Meijer     Refrigerator  2.34   NaN
```

```
## 4 Meijer     Room          1.86   NaN
```

```
## 5 MinuteMaid Refrigerator  2.01   NaN
```

```
## 6 MinuteMaid Room          2.01   NaN
```

Without replication we CANNOT compute a variance so we cannot do any statistical tests. Our book calls a blocked design with replication a **generalized block design**. A picture:

A quick way to see if we can fit a model is to do a degrees of freedom calculation.

If we didn't have replication, what models could we fit:

Note how we can continue to expand this as we get more replication. Let's say we are also interested in Water PH. How many treatments do we have if we consider brand as a blocking variable (not a treatment). So, within each block we could run a 2×2 factorial design. Again a picture:

How many glasses of orange juice would we need in our study?

If we have 12 glasses of orange juice what would our model be?

Our ANOVA table would be:

What would happen if we include an interaction between water, brand, and temperature?

Say we want to include all interactions, our model would become:

```
OJ3=read.table("http://www.isi-stats.com/isi2/data/OJ-n24.txt",header=T)

full.lm<-lm(VitC1000~Brand*Temp*Water,data=OJ3)
anova(full.lm)
```

```
## Analysis of Variance Table
##
## Response: VitC1000
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Brand      2  0.70081  0.35040  934.4111 6.745e-14 ***
## Temp       1  0.17682  0.17682  471.5111 5.329e-11 ***
## Water      1  0.05415  0.05415  144.4000 4.762e-08 ***
## Brand:Temp  2  0.23666  0.11833  315.5444 4.221e-11 ***
## Brand:Water 2  0.00167  0.00084    2.2333  0.149780
## Temp:Water  1  0.00482  0.00482   12.8444  0.003755 **
## Brand:Temp:Water 2 0.00016 0.00008    0.2111  0.812628
## Residuals  12  0.00450  0.00037
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3 Way interaction... The interaction between temperature and water depends on brand. (Figure 3.3.10)

```
OJ3 %>% ggplot(aes(x=Temp,y=predict(full.lm),color=Water))+
  geom_line(aes(group=Water))+geom_point() +
  facet_wrap(vars(Brand))
```

What happens if we don't add 3 way

```
OJ3=read.table("http://www.isi-stats.com/isi2/data/OJ-n24.txt",header=T)
mod.lm<-lm(VitC1000~Brand+Temp+Water+Brand:Temp+Water:Brand+Water:Temp,data=OJ3)
anova(mod.lm)
```

```
## Analysis of Variance Table
##
## Response: VitC1000
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Brand      2  0.70081  0.35040 1053.093 5.474e-16 ***
## Temp       1  0.17682  0.17682  531.399 1.556e-12 ***
## Water      1  0.05415  0.05415  162.741 4.249e-09 ***
## Brand:Temp  2  0.23666  0.11833  355.623 9.989e-13 ***
## Brand:Water 2  0.00167  0.00084    2.517  0.116463
## Temp:Water  1  0.00482  0.00482   14.476  0.001933 **
## Residuals  14  0.00466  0.00033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our test is more powerful if we don't consider everything. Couple of thoughts. Some folks believe that we shouldn't interact blocks with our treatments. Let's read through pg. 249