

One-way ANOVA: First parametrisation

AVIJNANDAN ROY
Roll-BS2023

Let's now try to fit the model. Now, we have the data set and we want to fit or model this way-

R-code

```
dat=read.csv('data.csv')
with(dat,boxplot(yield~variety:tilt))
with(dat,interaction.plot(tilt,variety,yield))
fit=lm(yield~variety+tilt+variety:tilt,dat=dat)
fit=lm(yield~variety*tilt,dat=dat)
```

We had our model-

$$y_{ijk} = +\mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad i\text{- variety, } j\text{- slope, } k\text{- number of the variety in that slope}$$

Here in this model we have written the command for **fit** where we have presented the code in two ways. These two codes provides same result. Now, for the first code we have used **variety** for α_i , **slope** for β_j and **variety:tilt** for γ_{ij} and our ϵ_{ijk} is automatically supplied by software by default. Now, this combination is such a commonly occurring thing that 'r' provides a simple abbreviation for this which we have presented in the second code for fit and most of the statisticians/data scientists generally use the second code. For the second code we are just using the modified model-

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

.So after running the code for fit we get our output-

```
> fit=lm(yield~variety*tilt,dat=dat)
> fit
```

Call:

```
lm(formula = yield ~ variety * tilt, data = dat)
```

Coefficients:

| | |
|-----------------------|---------------------------|
| (Intercept) | varietyJaya |
| 250.162 | 10.368 |
| varietyTaichung | tiltslope |
| 30.670 | -9.586 |
| varietyJaya:tiltslope | varietyTaichung:tiltslope |
| 24.562 | -18.168 |

Here we get the *Intercept* term for $\hat{\mu}$. For *variety* we have three varieties and we have already guessed that one of varieties will be dropped due to lack of identifiability. So, they have not mentioned IR8. So, our $\hat{\alpha}_2$ is varietyJaya(10.368), $\hat{\alpha}_3$ is varietyTaichung(30.670) and similarly $\hat{\beta}_2$ is tiltslope(-9.586) and now again for the γ_{ij} 's there are lot of things that are unidentifiable and they have dropped them. So what they have left with is $\hat{\gamma}_{22}$ (24.562) and $\hat{\gamma}_{32}$ (-18.168). Notice as the rank of the matrix is 6 they are always going to give 6 estimates.

Now, it will be a good instructive exercise to write down the design matrix and try to figure out the rank and figure out which of the things that 'r' actually drops. So, if we have i going from 1 to I , j goes from 1 to J , then exactly what is the rank try to come up with a simple formula and the formula is pretty simple that is rank of the design matrix is IJ . A little hint can be given that is out the by thinking what of the γ_{ij} 's will be dropped.

We will learn later when the γ terms will be zero or non-zero. For the time being we shall see the summary of the fit.

```
> summary(fit)

Call:
lm(formula = yield ~ variety * tilt, data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-0.3600 -0.1685  0.0360  0.1095  0.4580

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      250.16200    0.09586  2609.68  <2e-16 ***
varietyJaya       10.36800    0.13557   76.48  <2e-16 ***
varietyTaichung   30.67000    0.13557  226.24  <2e-16 ***
tiltslope        -9.58600    0.13557  -70.71  <2e-16 ***
varietyJaya:tiltslope  24.56200    0.19172  128.12  <2e-16 ***
varietyTaichung:tiltslope -18.16800    0.19172  -94.76  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2143 on 24 degrees of freedom
Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
F-statistic: 2.604e+04 on 5 and 24 DF,  p-value: < 2.2e-16
```

The first part gives you the test of hypothesis and we can check rank of the fit that is-

```
> fit$rank
[1] 6 $
```