Shenduo Zhang

One-way ANOVA fixed effect model

We are interested in the an crop yield experiment. There are three type of fertilizer been used to the crop. We are interested if they have influence on the mean of yield of crop.

1.1 Data

There are total of 96 samples with a balanced design of 32 samples in each group.

```
> head(crop.data)
  density block fertilizer    yield
1       1     1          1 177.2287
2       2     2          1 177.5500
3       1     3          1 176.4085
4       2     4          1 177.7036
5       1     1          1 177.1255
6       2     2          1 176.7783
```

And the data has passed the homogenetiy of variance test, the Barlett's test. The data is stored in crop.data in the code folder.

1.2 ANOVA model

Denote the total mean across all sample as μ, and the difference between each group as α_i. We have the following ANOVA model

$$y_{ij} = \mu + \alpha_i + e, \quad \forall j = 1, 2, \ldots, 32, \forall i = 1, 2, 3. \tag{1}$$

where e is normal distribution around zero, and $\sum_i \alpha_i = 0$.

We are interested in testing the hypothesis $H_0 : \alpha_i = 0, \forall i = 1, 2, 3$ against $H_1 : \alpha_i \neq 0$ for some i.

1.3 Computation

To test the hypothesis, we need to use the F-test. Constuct the F-statistics

$$F = \frac{(\sum_i n_i \bar{y}_i^2 - \frac{y_{..}^2}{n})/(3-1)}{(\sum_i (n_i - 1)s_i^2)/(n-3)} \tag{2}$$

where $y_{..}$ is the sum of all observations across all groups, n is the total number of observations over all groups, s_i^2 is the sample variance for the ith group. This is also known as the $BetweenMS/WithinMS$.

We shall reject the null hypothesis with large value of F. More specificly, under the null hypothesis, F follows a F-distribution $F(2, n-3)$, hence the p-value can be given by computing the upper tail at the observation value.

```
> #自编程序
> one_wayANOVA(d1,d2,d3,0.05)
$`The value of Statistics`
[1] 7.862752

$`reject null hypothesis?`
[1] TRUE

$`p-value`
[1] 0.0006999158
```

1.4 Numerical result

We first compute the value of F-statistics, and determine if we shall reject the null hypothesis by coparing the p-value and requiring siginificance level.

We can see that we shall reject the null hypothesis at siginificance level 0.05.

The package in R gives the following identical results.

```
> #R里自带程序结果
> model <- aov(yield~fertilizer,data = crop.data)
> summary(model)
            Df Sum Sq Mean Sq F value Pr(>F)
fertilizer   2   6.07  3.0340   7.863  7e-04 ***
Residuals   93  35.89  0.3859
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

All of the result indicates that there is a siginificant difference in the mean of the crop yield across different groups of crop.