Discussion3

Jing Lyu

1/25/2023

One-Way ANOVA

For this section, we will use PlantGrowth dataset. It contains weights of plants produced under two distinct treatment conditions and a control condition. We will investigate the relationship between conditions and weights.

1. Write down a one-way ANOVA model for this data. Use the factor-effect form.

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j},\ j = 1, \dots, n_i, i = 1, \dots, 3$$

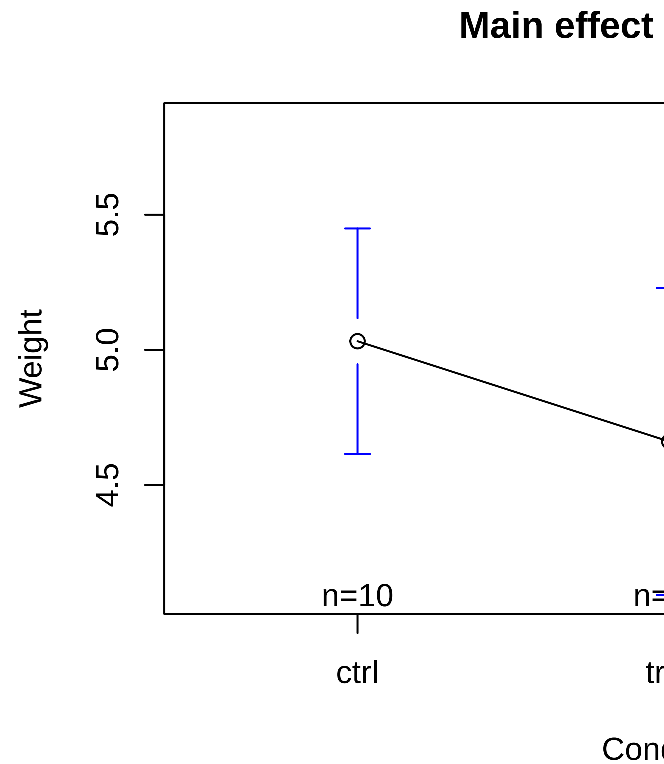
where $\{\alpha_i\}$ satisfies that $\sum_{i=1}^{3} n_i \alpha_i = 0$ and $\{\epsilon_{i,j}\}$ are i.i.d. $N(0, \sigma^2)$.

In this model, α_i represent the effect from the three conditions, which are control ($i = 1$), treatment 1 ($i = 2$) and treatment 2 ($i = 3$). The outcome $Y_{i,j}$ represents the jth subject under ith condition. The mean effect μ represents the mean weight in the population. The errors $\epsilon_{i,j}$ capture any unexplained effects on weights. Values of n_i can be found in the following table.

```
table(PlantGrowth$group)
```

```
##
## ctrl trt1 trt2
##   10   10   10
```

```
library(gplots)
plotmeans(weight ~ group, data = PlantGrowth,
          xlab = "Condition", ylab = "Weight",
          main="Main effect of treatment")
```



2. Obtain the main effects plots. Summarize your findings.

Example observations:

- Apparent differences in weights across condition.
- Largest variability in treatment 1.
- Treatment 1 has the lowest weight.
- Equal sample size under each condition.

```
res.aov <- aov(weight ~ group, data = PlantGrowth)
summary(res.aov)
```

3. Set up the ANOVA table using R for your model. Briefly explain this table. (explain what Df, Sum Sq, Mean Sq, F value, and Pr(>F) mean in this table.)

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## group        2  3.766  1.8832   4.846 0.0159 *
## Residuals   27 10.492  0.3886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Eg: Treatment sum of squares is 3.766. Residual sum of squares is 10.492. F test statistics is 4.846. P-value is 0.0159.

4. Test whether there is any association between conditions and weights. What are the null and alternative hypotheses? P-value is 0.0159 less than 0.05 which indicates significant difference of weights under different conditions.

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \ \text{ v.s. } \ H_A : \text{not all } \alpha_i \text{ are the zero.}$$

Contrasts

In this section, we will use the salaries dataset. It contains data on the salaries of different professors. We will investigate the relationship between ranks of professors and salaries.

```
library(car)
df=Salaries
head(df)
```

```
##        rank discipline yrs.since.phd yrs.service  sex salary
## 1      Prof          B            19          18 Male 139750
## 2      Prof          B            20          16 Male 173200
## 3  AsstProf          B             4           3 Male  79750
## 4      Prof          B            45          39 Male 115000
## 5      Prof          B            40          41 Male 141500
## 6 AssocProf          B             6           6 Male  97000
```

```
levels(df$rank)
```

```
## [1] "AsstProf"  "AssocProf" "Prof"
```

```
table(df$rank)
```

```
##
##  AsstProf AssocProf      Prof
##        67        64       266
```

One-way ANOVA table:

```
aov1 = aov(salary ~ rank, df)
summary.lm(aov1)
```

```
##
## Call:
## aov(formula = salary ~ rank, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -68972 -16376  -1580  11755 104773
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      80776       2887  27.976  < 2e-16 ***
## rankAssocProf    13100       4131   3.171  0.00164 **
## rankProf         45996       3230  14.238  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23630 on 394 degrees of freedom
## Multiple R-squared:  0.3943, Adjusted R-squared:  0.3912
## F-statistic: 128.2 on 2 and 394 DF,  p-value: < 2.2e-16
```

We access the information in the form of regression output using the *summary.lm* command.

Global test shows the difference exists among the means.

Questions:

(1) Do non-tenured position (AsstProf) and tenured position (AssocProf and Prof) have different salary?

(2) Is there a difference of salary within tenured position (AssocProf vs Prof)?

Denote mean salary of each group as μ_1:AsstProf, μ_2:AssocProf, μ_3:Prof.

Gloal test:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad vs. \quad H_A : \text{they are not all equal.}$$

Contrast 1: In the first contrast, we group AssocProf and Prof into the treatment condition. Then the test becomes

$$H_0 : \mu_1 = \frac{\mu_2 + \mu_3}{2} \quad vs. \quad H_A : \mu_1 \neq \frac{\mu_2 + \mu_3}{2}$$

The contrast we are interested in is $\mu_1 - (\mu_2 + \mu_3)/2$ or $2\mu_1 - (\mu_2 + \mu_3)$ with $c_1 = 2, c_2 = c_3 = -1$.

Constrast 2:

$$H_0 : \mu_2 = \mu_3 \quad vs. \quad H_A : \mu_2 \neq \mu_3$$

The constrast we are interested in is $\mu_2 - \mu_3$ with $c_1 = 1, c_2 = -1$.

Assign the contrasts to the variable `rank` in the dataset df.

```
contrast1 = c(2,-1,-1)
contrast2 = c(0,1,-1)
contrasts(df$rank) = cbind(contrast1, contrast2)
contrasts(df$rank) # check
```

```
##           contrast1 contrast2
## AsstProf          2         0
## AssocProf        -1         1
## Prof             -1        -1
```

We now analyze our contrasts by rerunning the same ANOVA command that we ran before. However, because now R has more information on the structure of the variable `rank` in the form of contrasts, the output will be different.

```
aov2 = aov(salary ~ rank, df)
summary.lm(aov2)
```

```
##
## Call:
## aov(formula = salary ~ rank, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -68972 -16376  -1580  11755 104773
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     100475       1459  68.855   <2e-16 ***
## rankcontrast1    -9849       1108  -8.892   <2e-16 ***
## rankcontrast2   -16448       1645  -9.997   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23630 on 394 degrees of freedom
## Multiple R-squared:  0.3943, Adjusted R-squared:  0.3912
## F-statistic: 128.2 on 2 and 394 DF,  p-value: < 2.2e-16
```

Both contrasts are significant, meaning that becoming tenured affects professors' salaries and so does moving up among tenured positions.