P8130 Recitation 4: Oct 9th/11th

Yutao Liu

October 5, 2017

Key Words: t Test for Linear Contrasts in One-Way ANOVA

1. Basic Setting

- There are k treatment groups
- Subjects are classified by only one factor, treatment
- The ith group has n_i subjects, so the total number of subjects $N = \sum_{i=1}^{k} n_i$

2. One-way ANOVA Model:

- $Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$

a. μ is the overall mean response in the population
b. α_i is the difference between mean respone of the ith group and the overall mean response in the population
c. $\mu_i \overset{def}{=} \mu + \alpha_i$ is the mean response of the ith group in the population

3. Linear Contrast

- A linear combination of the estimated group means $\widehat{\mu_L} = \sum_{i=1}^{k} c_i \widehat{\mu}_i = \sum_{i=1}^{k} c_i \bar{Y}_i$ is a linear contrast if $\sum_{i=1}^{k} c_i = 0$, where $\boldsymbol{c} = (c_1, \ldots, c_k)'$ is called a contrast coefficients vector
- $\mu_L = E[\widehat{\mu_L}] = \sum_{i=1}^{k} c_i E[\bar{Y}_i] = \sum_{i=1}^{k} c_i \mu_i = \sum_{i=1}^{k} c_i(\mu + \alpha_i) = \sum_{i=1}^{k} c_i \mu + \sum_{i=1}^{k} c_i \alpha_i = \sum_{i=1}^{k} c_i \alpha_i$
- Examples of various contrast vectors:

a. Comparison between two individual groups, e.g., $\boldsymbol{c} = (c_1, c_2, c_3, \ldots, c_k)' = (1, 0, 0, \ldots, -1)'$
b. Comparison of one group with others, e.g., $\boldsymbol{c} = (c_1, c_2, c_3 \ldots, c_k)' = (1/2, 1/2, 0, \ldots, -1)'$

- Distribution of linear contrasts: $E[\widehat{\mu_L}] = \mu_L$, $Var(\widehat{\mu_L}) = \sigma^2 \sum_{i=1}^{k} c_i^2/n_i$

$$\frac{\widehat{\mu_L} - \mu_L}{\sigma\sqrt{\sum_{i=1}^{k} c_i^2/n_i}} \sim N(0,1)$$

Estimating σ^2 with $\widehat{\sigma^2} = s^2$, the Within MS from the one-way ANOVA table,

$$\frac{\widehat{\mu_L} - \mu_L}{s\sqrt{\sum_{i=1}^{k} c_i^2/n_i}} = \frac{\widehat{\mu_L} - \mu_L}{\sqrt{Within\ MS}\sqrt{\sum_{i=1}^{k} c_i^2/n_i}} = \frac{\widehat{\mu_L} - \mu_L}{se(\widehat{\mu_L})} \sim t_{N-k}$$

- Inference of linear contrasts:

a. Test $H_0 : \mu_L = 0$ vs $H_a : \mu_L \neq 0$ at level of significance α. Reject H_0 if

$$|t| = |\frac{\widehat{\mu_L}}{s\sqrt{\sum_{i=1}^{k} c_i^2/n_i}}| = |\frac{\widehat{\mu_L}}{se(\widehat{\mu_L})}| > t_{N-k,1-\alpha/2}$$

b. $(1-\alpha)100\%$ confidence interval of μ_L

$$[\widehat{\mu_L} \pm t_{N-k,1-\alpha/2} \times se(\widehat{\mu_L})]$$

Real Data Example: Rosner, Chapter 12.4, Example 12.9 (p567)

Pulmonary Disease Suppose we want to compare the pulmonary function of the group of smokers who inhale cigarettes with that of the group of nonsmokers. The three groups of inhaling smokers in Table 12.1 could just be combined to form one group of 600 inhaling smokers. However, these three groups were selected so as to be of the same size, whereas in the general population the proportions of light, moderate, and heavy smokers are not likely to be the same. Suppose large population surveys report that 70% of inhaling smokers are moderate smokers, 20% are heavy smokers, and 10% are light smokers. How can inhaling smokers as a group be compared with nonsmokers?

The estimation and testing of hypotheses for linear contrasts is used for this type of question.

Summary statistics in each group

```
cbind(means, stddev, smplsize)
```

```
##      means stddev smplsize
## [1,]  3.78   0.79      200
## [2,]  3.30   0.77      200
## [3,]  3.32   0.86       50
## [4,]  3.23   0.78      200
## [5,]  2.73   0.81      200
## [6,]  2.59   0.82      200
```

Number of groups k

```
k <- length(means); k
```

```
## [1] 6
```

Within MS s^2

```
stddev^2
```

```
## [1] 0.6241 0.5929 0.7396 0.6084 0.6561 0.6724
```

```
smplsize - 1
```

```
## [1] 199 199  49 199 199 199
```

```
stddev^2 * (smplsize - 1)
```

```
## [1] 124.1959 117.9871  36.2404 121.0716 130.5639 133.8076
```

```
WithinSS <- sum(stddev^2 * (smplsize - 1) ); WithinSS
```

```
## [1] 663.8665
```

```
WithinMS <- WithinSS / sum( (smplsize - 1) ); WithinMS
```

```
## [1] 0.6358875
```

Contrast vector $\boldsymbol{c} = (1, 0, 0, -.1, -.7, -.2)'$

```
contrstvec <- c(1, 0, 0, -.1, -.7, -.2); round(sum(contrstvec), 9)
```

```
## [1] 0
```

Linear contrast $\widehat{\mu_L}$

```
means * contrstvec
```

```
## [1]  3.780  0.000  0.000 -0.323 -1.911 -0.518
```

```r
L <- sum(means * contrstvec); L
```

```
## [1] 1.028
```

standard error of linear contrast

```r
contrstvec^2/smplsize
```

```
## [1] 0.00500 0.00000 0.00000 0.00005 0.00245 0.00020
```

```r
sum(contrstvec^2/smplsize)
```

```
## [1] 0.0077
```

```r
se <- (WithinMS * sum(contrstvec^2/smplsize) ) %>% sqrt(.); se
```

```
## [1] 0.0699738
```

t statistics

```r
L/se
```

```
## [1] 14.69121
```

Degrees of freedom

```r
sum(smplsize) - k
```

```
## [1] 1044
```

Conlousion: this linear contrast is very highly significant ($p < .001$), and the inhaling smokers as a group have strikingly worse pulmonary function than the nonsmokers.