

Review :

1. ANOVA for multi-sample comparison

(a) Assumption

- i. Observations within a treatment and across all treatments are independent
- ii. Normal assumption: $X_{ij} \sim N(\mu_i, \sigma_i^2) \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$
- iii. Equal variance assumption: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$

(b) Hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad H_A$:not all μ_i are equal

(c) Notations

$$\begin{aligned} SSTot &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 = \sum_{\text{all obs}} (x_{ij})^2 - \frac{(x_{..})^2}{N} \\ SSTrt &= \sum_{i=1}^k n_i (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^k \frac{1}{n_i} (x_{i.})^2 - \frac{(x_{..})^2}{N} \\ SSErr &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 = \sum_{i=1}^k (n_i - 1) s_i^2 \end{aligned}$$

(d) Model : $X_{ij} = \mu_i + e_{ij}$ or $X_{ij} = \mu + \alpha_i + e_{ij}$, where $e_{ij} \sim N(0, \sigma^2)$

(e) ANOVA table

| Source | df | SS | MS |
|-----------|-----|-------|-------------------|
| Treatment | k-1 | SSTrt | MSTrt=SSTrt/(k-1) |
| Error | N-k | SSErr | MSErr=SSErr/(N-k) |
| Total | N-1 | SSTot | |

(f) Test statistics : $F = \frac{MSTrt}{MSErr}$ (g) P-value : $P(F \geq F_{k-1, N-k})$ 2. Contrasts : Suppose there are k treatments with treatment means $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$. Let $\lambda_i, i = 1, 2, \dots, k$ be coefficients so that $\sum_{i=1}^k \lambda_i = 0$. Then $\sum_{i=1}^k \lambda_i \bar{X}_i$ is a contrast.3. Testing for Contrast : $H_0 : \sum_{i=1}^k \lambda_i \mu_i = a$

$$T = \frac{\sum_{i=1}^k \lambda_i \bar{X}_i - a}{\sqrt{MSErr} \sqrt{\sum_{i=1}^k \frac{\lambda_i^2}{n_i}}}$$

where $MSErr = S_p^2 = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) s_i^2$ and T has a t-distribution with $d.f. = N - k$ under H_0 .

4. Confidence Interval for $\sum_{i=1}^k \lambda_i \mu_i$

$$\sum_{i=1}^k \lambda_i \bar{X}_i \pm t_{(N-k, \frac{\alpha}{2})} \sqrt{MSErr} \sqrt{\sum_{i=1}^k \frac{\lambda_i^2}{n_i}}$$

5. Testing for $H_0 : \sigma_1^2 = \dots = \sigma_k^2$ --- Use Levene's Test on page 273(course notes)

6. Confidence Interval for σ^2 when $H_o : \sigma_1^2 = \dots = \sigma_k^2$ is true

$$\frac{(N-k)\text{MSErr}}{\chi_{(N-k, \frac{\alpha}{2})}^2} \leq \sigma^2 \leq \frac{(N-k)\text{MSErr}}{\chi_{(N-k, 1-\frac{\alpha}{2})}^2}$$

7. Error Rates for Multiple Comparisons

- (a) Comparison-wise error rate : $P(\text{Reject } H_o | H_o \text{ is true})$ for a single contrast.
 (b) Experiment-wise error rate : Overall error rate for all contrasts.

Practice Problem

1. A researcher is studying the effectiveness of three methods of reducing smoking. He wants to determine whether the mean reduction in the number of cigarettes smoked daily differs from one method to another among men patients. Each smoked 60 cigarettes per day before treatment. Four randomly chosen members of the group pursue method I; four pursue method II ; and so on. The reductions in the number of cigarettes smoked daily are as follows:

| Method | | |
|--------|----|-----|
| I | II | III |
| 50 | 41 | 49 |
| 51 | 40 | 47 |
| 51 | 39 | 45 |
| 52 | 40 | 47 |

Use a one-way analysis of variance to test whether the mean reduction in the number of cigarettes smoked daily is equal for the three methods.

2. Samples of 6 different brands of margarine were analyzed to determine the level of physiologically active polyunsaturated fatty acids (PAPFUA, on percentages). The data is shown below:

| A | B | C | D | E | F |
|------|------|------|------|------|------|
| 15.4 | 12.8 | 13.5 | 14.1 | 18.8 | 18.1 |
| 14.8 | 12.5 | 13.4 | 13.6 | 17.2 | 18.7 |
| 15.6 | 13.4 | 14.1 | 14.4 | 18.4 | 18.4 |
| 14.4 | 13.0 | 14.3 | 14.3 | 17.3 | |
| 16.0 | 12.3 | | | 18.0 | |
| 15.2 | | | | | |

- (a) Compute an ANOVA table for the data. Is there evidence of differences among the true PAPFUA percentages for the different brands?
 (b) A is an “old” brand while the other 5 are “new” brands. Find a 95% C.I. for the mean difference between the PAPFUA % for A and the average of PAPFUA % for the other 5 new brands.
 (c) Brands B, C and D are soybean-based while brands E and F are corn-based. Conduct a test for the following:

$$H_0 : \frac{\mu_B + \mu_C + \mu_D}{3} = \frac{\mu_E + \mu_F}{2} \quad \text{vs} \quad H_a : \frac{\mu_B + \mu_C + \mu_D}{3} < \frac{\mu_E + \mu_F}{2}$$

- (d) Carry out a test for equality of variance for the 6 brands.
 (e) Compute 95% C.I. for the underlying error variance.

Solution:

1.

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad VS \quad H_A : \text{not all } \mu \text{ are equal}$$

We get $F=108.07$, $F \sim F(2, 9)$, $K=6$, $N=12$

$$P(F \geq 108.07) = 0$$

```
>y=c(data1[,1],data[,2],data[,3])
>x=c(rep(1,4), rep(2,4), rep(3,4))
>x=factor(x)
fm1=lm(y~x)
>anova(fm)
then, we can get F=108.07.
```

2. (a)

```
>y=c(data2[,1],data2[,2],data2[,3],data2[,4],data2[,5],data2[,6])
```

```
>x=c(rep(1,6),rep(2,5),rep(3,4),rep(4,4),rep(5,5),rep(6,3))
```

```
>x=factor(x)
```

```
>fm2a=lm(y~x)
```

```
>anova(fm2a)
```

```
>confint.lm(fm2a)
```

```
>y=c(data3[,1],data3[,2],data3[,3],data3[,4],data3[,5],data3[,6])
```

```
>x=c(rep(1,6),rep(2,4),rep(3,4),rep(4,4),rep(5,4),rep(6,2))
```

```
>x=factor(x)
```

```
>fm3a=lm(y~x)
```

```
>anova(fm3a)
```

then, we can get $F=84.55$, $MSE=0.259$ in fm2a

we can get $P=0.213$ in fm3a.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_6 \quad H_A : \text{not all } \mu \text{ are equal}$$

We get $F=84.55$, $F \sim F(5, 21)$, $K=6$, $N=27$

$$P(F \geq 84.55) = 0$$

(b)

$$\sum_{i=1}^6 \lambda_i \bar{x}_i, \quad \lambda_1 = 1, \quad \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = -\frac{1}{5}$$

So, the 95% C.I. is

$$\sum_{i=1}^6 \lambda_i \bar{X}_i \pm t_{(21,0.025)} \sqrt{MSE_{err}} \sqrt{\sum_{i=1}^6 \frac{\lambda_i^2}{n_i}}$$

$$=(-0.801, 0.337)$$

(c)

$$\sum_{i=1}^6 \lambda_i \bar{x}_i, \quad \lambda_1 = 0, \quad \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{3}, \quad \lambda_5 = \lambda_6 = -\frac{1}{2}$$

$$H_0 : \sum_{i=1}^6 \lambda_i \bar{x}_i = 0 \quad VS \quad H_1 : \sum_{i=1}^6 \lambda_i \bar{x}_i > 0$$

We can get

$$T = \frac{\sum_{i=1}^6 \lambda_i \bar{x}_i}{\sqrt{\text{MSErr}} \sqrt{\sum_{i=1}^6 \frac{\lambda_i^2}{n_i}}} = \frac{-4.575}{0.5089 \times 0.4595} = -19.58. \quad df = N - K = 21$$

So,

$$P\text{-value} = P(T_{21} < -19.56), \text{ We can't reject } H_0$$

(d)

$$\text{We have } H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2, \quad \text{And } P = 0.213$$

(e) From (d), we can't reject H_0 . So, using 6 on the handout the 95% C.I. for σ^2 is:

$$\frac{(N-k)\text{MSErr}}{\chi_{(N-k, \frac{\alpha}{2})}^2} \leq \sigma^2 \leq \frac{(N-k)\text{MSErr}}{\chi_{(N-k, 1-\frac{\alpha}{2})}^2}$$

The C.I. is (0.1533, 0.5291)