Univariate Statistical Analysis

Lecture 11

Analysis of Variance
Chapter 15

Single-Factor ANOVA and the *F* Test

A **single-factor analysis of variance (ANOVA)** problem involves a comparison of k population or treatment means $\mu_1, \mu_2, \ldots, \mu_k$. The objective is to test

H_0: $\mu_1 = \mu_2 = \mu_3$

against

H_a: At least two of the μ's are different from each other, $\mu_1 \neq \mu_2 \neq \mu_3$

2

Single-Factor ANOVA and the *F* Test

Watch video:

https://www.youtube.com/watch?v=WcmzS3nEUqo

Understanding Analysis of Variance (ANOVA) including Excel - Statistics Help

3

Refer to Text Example 15.2 (p.764) Heart Attack (Data source p.762 Table 15.1).

$\overline{x_1} = 10.89,\ \overline{x_2} = 11.25,\ \overline{x_3} = 11.37,\ \overline{x_4} = 11.75$

$\bar{\bar{x}} = 11.315$ is the average of all samples.

$s_1 = 0.69,\ s_2 = 0.74,\ s_3 = 0.91,\ s_4 = 1.07$

$n_1 = n_2 = n_3 = n_4 = 35$

$k = 3$ is the number of groups

$N = n_1 + n_2 + n_3 + n_4 = 35 + 35 + 35 + 35 = 140$ is the total samples

4

A measure of **differences among the sample means** is the **treatment sum of squares**, denoted by **SSTr** and given by

$$\text{SSTr} = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \cdots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

$= 35(10.89 - 11.315)^2 + 35(11.25 - 11.315)^2 + 35(11.37 - 11.315)^2 + 35(11.75 - 11.315)^2$

$= 6.322 + 0.148 + 0.106 + 6.623$

=13.199

5

A measure of **variability within the *k* samples**, called **error sum of squares** and denoted by **SSE**, is

$$\text{SSE} = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2$$

$= (35-1)(0.69)^2 + (35-1)(0.74)^2 + (35-1)(0.91)^2 + (35-1)(1.07)^2$

$= 101.888$

6

A **single-factor analysis of variance (ANOVA)** problem involves a comparison of k population or treatment means $\mu_1, \mu_2, ..., \mu_k$. The objective is to test

H_0: $\mu_1 = \mu_2 = \mu_3$

against

H_a: $\mu_1 \neq \mu_2 \neq \mu_3$

Each sum of squares has an associated df:

treatment $df_1 = k - 1 = 4 - 1 = 3$

error $df_2 = N - k = 140 - 4 = 136$

A **mean square** is a sum of squares divided by its df. In particular,

mean square for treatments = MStr = $\frac{\text{SSTr}}{k-1} = \frac{13.199}{4-1} = 4.4$

mean square for error = MSe = $\frac{\text{SSE}}{N-k} = \frac{101.888}{140-4} = 0.749$

All F tests in this text are upper-tailed, so P-values are areas under the F curve to the right of the calculated values of F.

Test value: $F = \frac{\text{MSTr}}{\text{MSE}} = \frac{4.4}{0.749} = 5.87$

Critical value, given

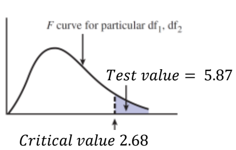
$df_1 = 3$

$df_2 = 136$ -> 120 (closest value in the table)

Level of significant = 0.05

Refer to Appendix TABLE 6 (p.795 – 798)

The critical value is 2.68



Conclusion:
Reject H_o. There is **enough evidence** to conclude that at least one population mean is different that the other means.