University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

ANOVA - F test

In the one-way ANOVA problem if we let

$$\bar{\mu} = \frac{\sum_{i=1}^{n} n_i \mu_i}{N},$$

it can be shown (see proof below) that:

a. $$E(BSS) = (k-1)\sigma^2 + \sum_{i=1}^{n} n_i(\mu_i - \bar{\mu})^2.$$

b. $$E(WSS) = (N-k)\sigma^2$$

Proof:

a.

$$E(BSS) = E\left(\sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2\right) = E\left(\sum_{i=1}^{k} n_i\bar{y}_i^2 - 2\bar{y}\sum_{i=1}^{k} n_i\bar{y}_i + \bar{y}^2\sum_{i=1}^{k} n_i\right).$$

In the previous expression we can substitute

$$\sum_{i=1}^{k} n_i = N, \text{ and } \sum_{i=1}^{k} n_i\bar{y}_i = N\bar{y}$$

to get

$$E(BSS) = E\left(\sum_{i=1}^{k} n_i\bar{y}_i^2 - 2\bar{y}N\bar{y} + N\bar{y}^2\right) = E\left(\sum_{i=1}^{k} n_i\bar{y}_i^2 - N\bar{y}^2\right) =$$

$$\sum_{i=1}^{k} n_i E(\bar{y}_i^2) - NE(\bar{y}^2) =$$

$$\sum_{i=1}^{k} n_i\left(\text{var}(\bar{y}_i) + (E(\bar{y}_i))^2\right) - N\left(\text{var}(\bar{y}) + (E(\bar{y}))^2\right) =$$

$$\sum_{i=1}^{k} n_i\left(\frac{\sigma^2}{n_i} + \mu_i^2\right) - N(\frac{\sigma^2}{N} + \bar{\mu}^2).$$

We get the last term in the previous expession because:

$$E(\bar{y}) = E\left(\frac{n_1\bar{y}_1 + n_2\bar{y}_2 + \cdots + n_k\bar{y}_k}{N}\right) = \frac{\sum_{i=1}^{k} n_i\mu_i}{N} = \bar{\mu}.$$

Therefore

$$E(BSS) = \sum_{i=1}^{k} n_i\left(\frac{\sigma^2}{n_i} + \mu_i^2\right) - N(\frac{\sigma^2}{N} + \bar{\mu}^2) = k\sigma^2 + \sum_{i=1}^{k} n_i\mu_i^2 - \sigma^2 - N\bar{\mu}^2 \quad \Rightarrow$$

$$E(BSS) = (k-1)\sigma^2 + \sum_{i=1}^{k} n_i(\mu_i - \bar{\mu})^2.$$

b.

$$E(WSS) = E\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2\right) =$$

$$E\left(\sum_{j=1}^{n_1}(y_{1j}-\bar{y}_1)^2 + \sum_{j=1}^{n_2}(y_{2j}-\bar{y}_2)^2 + \cdots + \sum_{j=1}^{n_k}(y_{kj}-\bar{y}_k)^2\right) =$$

$$E\left((n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots (n_k-1)s_k^2\right),$$

where s_i^2 is the sample variance for each group. We also know that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. Therefore, since $E(\chi^2_{n-1}) = n-1$, the expected value of each term is:

$$E\left((n_i-1)s_i^2\right) = (n_i-1)\sigma^2.$$

Finally:

$$E(WSS) = (n_1-1)\sigma^2 + (n_2-1)\sigma^2 + \cdots + (n_k-1)\sigma^2 = (N-k)\sigma^2.$$

Conclusion:
Using the results of part (a) and part (b) we see that if H_0 is true then $E(\frac{BSS}{k-1}) = \sigma^2$ and $E(\frac{WSS}{N-K}) = \sigma^2$. But when H_0 is not true we have

$$E(\frac{BSS}{k-1}) = \sigma^2 + \sum_{i=1}^{k} n_i \frac{(\mu_i-\bar{\mu})^2}{k-1} > \sigma^2.$$

Therefore small values of the ratio

$$F = \frac{\frac{BSS}{k-1}}{\frac{WSS}{N-k}}$$

favor H_0, while large values favor H_a. We reject H_0 if $F > F_{1-\alpha;k-1,N-k}$.