AMS 572 Class Notes

November 19, 2010

Chapter 12 Analysis of Variance (ANOVA)

One-way ANOVA (fixed factors)

$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$

* Goal: compare the means from a (a≥2) different populations.

* It is an extension of the pooled variance t-test.

* Assumptions:

(i) Equal (unknown) population variances $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 = \sigma^2$

(ii) Normal populations

(iii) Independent samples

H_a: these μ_i's are not all equal.

Assumptions: a population, $N(\mu_i, \sigma^2)$, i=1,2,…,a. σ^2 is unknown.

Samples: a independent samples.

Data:

$$
\begin{array}{ccc}
population\,1 & population\,i & population\,a \\
\downarrow & \downarrow & \downarrow \\
sample\,1 & sample\,i & sample\,a \\
n_1 \left\{ \begin{array}{l} X_{11} \\ X_{12} \\ \vdots \\ X_{1n_1} \end{array} \right. &
n_i \left\{ \begin{array}{l} X_{i1} \\ X_{i2} \\ \vdots \\ X_{in_i} \end{array} \right. &
n_a \left\{ \begin{array}{l} X_{a1} \\ X_{a2} \\ \vdots \\ X_{an_a} \end{array} \right.
\end{array}
$$

Balanced design: $n_i = n$

Unbalanced design: otherwise

Derivation of the test

(1) PQ, can be derived

(2) * Union-intersection method. Best method for this type of test as in other regression analysis related tests. Please see AMS 570/571 text book, and also the book by G.A.F. Seber: Linear Regression Model, published by John Wiley for details.

(3) LRT (Likelihood Ratio test)

Test Statistic:

$$F_0 = \frac{MSA}{MSE} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 / (a-1)}{\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 / (N-a)} \overset{H_0}{\sim} F_{a-1,N-a}$$

Total sample size $N = \sum_{i=1}^{a} n_i$.

Sample mean: $\bar{Y}_{1\cdot} \cdots \bar{Y}_{i\cdot} \cdots \bar{Y}_{a\cdot}$ grand mean $\bar{Y}_{\cdot\cdot}$

$$Y_{ij} \overset{iid}{\sim} N(\mu_i, \sigma^2), \quad i = 1, \cdots, a \quad j = 1, \cdots, n_i$$

Balanced design: $n_i = n$

$$Y_{ij} \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\bar{Y}_{1\cdot} \sim N(\mu, \frac{\sigma^2}{n}), \quad \bar{Y}_{i\cdot} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\frac{\sum_{i=1}^{a}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{\sigma^2/n} \sim \chi^2_{a-1}$$

$$\frac{\sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{a(n-1)} = \chi^2_{N-a}$$

$$\frac{\sum_{i=1}^{a}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{\sigma^2/n} \overset{H_0}{\sim} \chi^2_{a-1}$$

Theorem Let $X_i \overset{iid}{\sim} N(\mu, \sigma^2), i = 1, \cdots, n$

(1) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

(2) $\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

(3) $\bar{X}$ and S^2 are independent.

Definition $F = \dfrac{W/k_1}{V/k_2} \sim F_{k_1,k_2}$, where $W \sim \chi^2_{k_1}, V \sim \chi^2_{k_2}$ and they are independent.

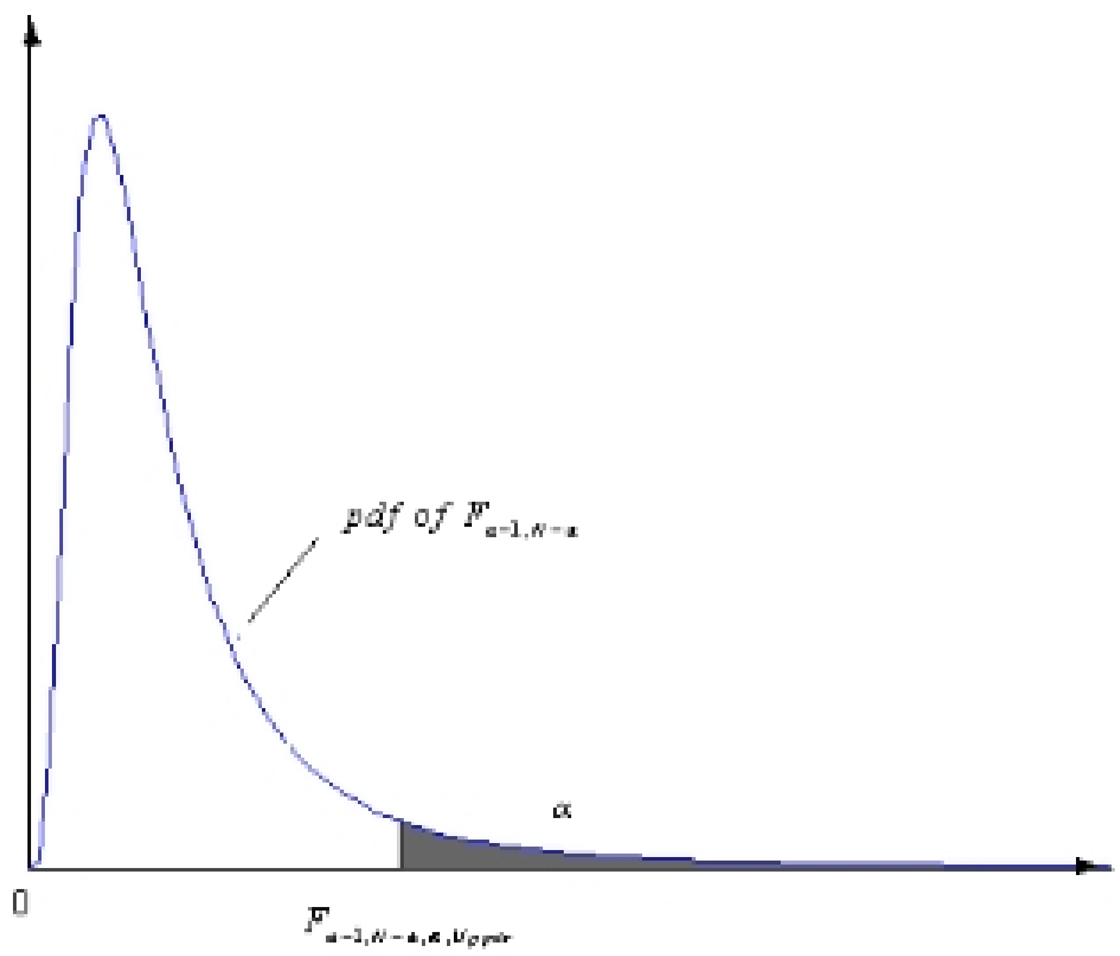
$E(MSA) = \sigma^2 + \dfrac{\sum_{i=1}^{a} n_i(\mu_i - \mu)^2}{a-1}$, $E(MSE) = \sigma^2$.

When H_0 is true: $F_0 \approx 1$, ($\mu_i = \mu$)

H_a is true: $F_0 > 1$.

Intuitively, we reject H_0 in favor of H_a if $F_0 > C$, where C is determined by the significance level as usual:

$\alpha = P(reject\ H_0 \mid H_0) = P(F_0 \geq C \mid H_0)$.



When a=2, $H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 \neq \mu_2$