Analysis of Variance (ANOVA)

ANOVA is used to compare many means to find evidence against them all being equal

We have many samples of some variable from distinct populations or subpopulations, we want to know if there is evidence that these samples come from distributions with different means.

Sample 1	Sample 2	. . . . . .	Sample k
$\bar{X}_1 = \dots$	$\bar{X}_2 \dots$		$\bar{X}_k \dots$
$S_1 = \dots$	$S_2 \dots$		$S_k \dots$
$n_1 = \dots$	$n_2 \dots$		$n_k \dots$

Assume (Confirm...)

All values within and between groups are independent

All samples should be approximately

Assume all variances are equal $(\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2)$

Hypothesis Test

H_0: The population means for each sample are equal.

$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

H_a: at least one is different.

$$\mu_i \neq \mu_j \quad i \neq j$$

H_a: at least one is different

$$\mu_i \neq \mu_j \quad i \neq j$$

What we're going to compare is the variable between means with the variability within samples.

MSE (Mean Squared Error)

$$MSE = \frac{SSE}{df_E} \qquad SST = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$df_E = (n-k) \qquad = \frac{1}{n-k}(SST - SSG)$$

$$= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \dots + (n_k-1)s_k^2}{\underbrace{(n-k)}_{df_E}}$$

MSG (Mean Square Between Groups)

$$MSG = \frac{1}{df_G} \cdot SSG$$

$$(df_G = k-1) \qquad = \frac{1}{(k-1)} \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2$$

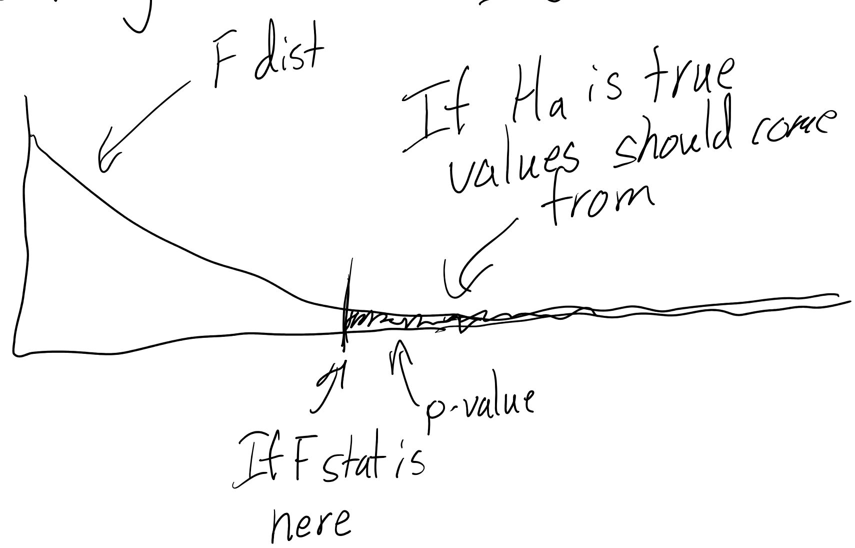
$\bar{x}$ is the grand mean of all the data.

Under H_0: $$F = \frac{MSG}{MSE} \sim F_{dist}(df_1 = k-1, df_2 = n-k)$$

MSG → $df = k-1$

MSE → $df = n-k$

Under H_a: The value we get should be large



If p-value is large we fail to reject.

fail to reject.

here

If p-value is small we can conclude that one of the means is different.

To determine which we can conduct multiple difference of means tests.

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

If we conduct multiple tests to find the different mean, we should conduct them at a significance level of $\alpha^* = \alpha / K$ $K = \frac{k(k-1)}{2}$

IE: If $\alpha = 5\%$ and $k = 3$, we must conduct $\frac{3 \cdot 2}{2} = 3$ tests, and they should only be significant if p-val $< .05/3$

Bonferroni Correction