ANOVA One-Way

Defn (One-Way ANOVA) is a hypothesis test where

$H_0: \mu_0 = \mu_1 = \ldots = \mu_k$

H_a: not H_0

∄ Z-stat in ANOVA;

Z-stat equivalent is F in ANOVA

$F \sim$ F-distn (SSB df, SSW df)

$$F\text{-stat} = \frac{mSSB}{mSSW},$$

where mSSB = mean of sum of sq's btwn (SSB)

mSSW = mean of sum of sq's wtn (SSW)

Assumptns

1. ∃ random samples collectd fr k popns
2. k popns are normally distributd w/ variances $\sigma_1^2 = \ldots = \sigma_k^2$ (ie homoskedastic)

   mean $\mu_1, \ldots, \mu_k$

   $[N \sim (\sigma_i^2, \mu_i)]$;

   ie k popns are iid & $N \sim (\sigma_i^2, \mu_i)$

   note: moderate departures fr these assumptns will not seriously affect appropriateness of test
3. For generality, let sample sizes be unequal and let n_i, for $i = 1, \ldots k$, be the # of observns in the sample drawn fr. the i^{th} popn.

- mSSB & mSSW have a Chi-distn

Eg of when to use ANOVA

I used ANOVA when checking if a feature was useful when analyzing Ames Housing Data. Sale prices by month sold had very close means ie mean sale price in Jan was near mean sale price in Feb. I used ANOVA to test if the mean sale prices of all months were equal & thus month sold was possibly not a useful feature in predicting sale price.

$$SSB = \sum_{\text{all k}} \sum_{\substack{\text{all samples} \\ \text{in grp k}}} (X_{ij} - \bar{X}_m)^2$$

$$SSW = \sum_{\text{all k}} n_i (\bar{X}_i - \bar{X}_{total})$$

grp	A	B	C
samples	X_{1A}	X_{1B}	X_{1C}
	⋮	⋮	⋮
	$X_{nA,A}$	$X_{nB,A}$	$X_{nC,C}$

SSW is looking @ diffs w/in a grp

SSB is looks @ differences btwn a grp

$$mSSB = \frac{SSB}{df_B} = \frac{SSB}{N-1}$$

df = degrees of freedom

$$mSSW = \frac{SSW}{df_W} = \frac{SSW}{N-k}$$