this time: ANOVA

read: LN pp. L-302 → L-322

AMS 7
31 May 18

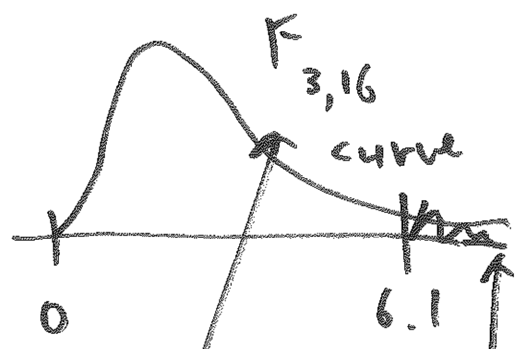
next time: categorical data analysis

today: LN pp. L-269 → ①

$$n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + n_3(\bar{y}_3 - \bar{y})^2 + n_4(\bar{y}_4 - \bar{y})^2$$

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

$$\frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2 + (n_3 - 1)\cdot s_3^2 + (n_4 - 1)\cdot s_4^2}{\underbrace{n_1 + n_2 + n_3 + n_4}_{n} - \underset{I}{4}} = \hat{\sigma}^2$$



long-run hist. of F if null true

$$F = \frac{MS_B}{MS_W}$$

& if $\sigma_1 = \dots = \sigma_I = \sigma$ is okay & if pop. hists in all groups are normal

$$F_{I-1,\, n-I}$$

if $\underline{p}$ is small, reject null

$\underline{p} = 0.0056 = 0.56\%$

if $\underline{p} \le 5\%$, result is statsig.

these differences among the 4 groups are (highly) statsig)

$$\hat{SE}(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{(earlier)}$$

but if $\sigma_1 = \sigma_2 = \sigma$

you would create a "pooled variance estimate" $s^2 = \hat{\sigma}^2$

& your new $\widehat{SE}$ formula would look like $\widehat{SE}_{\substack{\text{pooled} \\ \text{variance}}}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$

$$= s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$