this time: ANOVA

next time: categorical data analysis

read: LN pp. L-259 → L-311

AMS 7 20 Jul 2016

this time: LN pp. L-269 ①

please go to myucsc & fill out the anonymous course evaluation before midnight Fri 22 Jul

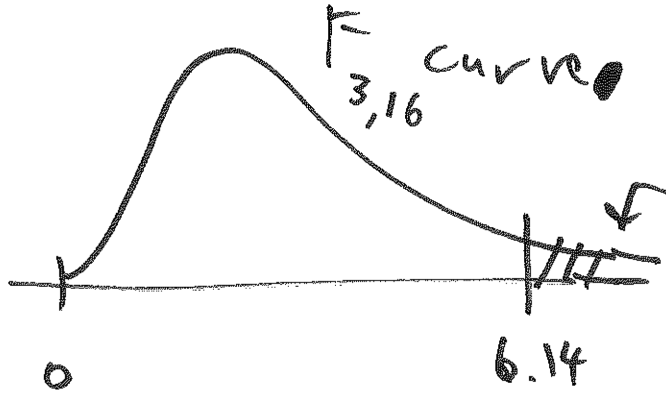
$$n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + n_3(\bar{y}_3 - \bar{y})^2 + n_4(\bar{y}_4 - \bar{y})^2$$

I ← 4

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

$$\text{noise estimate} = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 + \cdots + (n_I - 1) s_I^2}{(n_1 - 1) + \cdots + (n_I - 1)}$$

$$\underbrace{(n_1 + \cdots + n_I)}_{n} - I = n - I$$



long run hist of F if null true

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

$0.56\% = 0.0056 = p$

$p < 5\% \leftrightarrow$ do the differences among the sample means are both statsig & practsig

group 1 vs. group 2

estimate of $(\mu_1 - \mu_2)$: $(\bar{Y}_1 - \bar{Y}_2)$

$$\widehat{SE}_{IID}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

2 indep

this is based on group 1 { group 2

pop. SD σ_1 | pop SD σ_2

$$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma^2}{n_2}}$$

if $\sigma_1 = \sigma_2 = \sigma$

$$SE(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

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

in 1-way ANOVA assume

$\sigma_1 = \sigma_2 = \ldots = \sigma$ our best estimate of σ^2

is $MS_w = \hat{\sigma}^2$ = mean squared error

our best estimate of σ is

$\sqrt{MS_w} = \hat{\sigma}$ = root mean squared error so in 1-way ANOVA

$$\hat{SE}(\bar{Y}_1 - \bar{Y}_2) = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

and in general

$$\hat{SE}(\bar{Y}_i - \bar{Y}_j) = \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

95% CI for $(\mu_i - \mu_j)$

$$(\bar{Y}_i - \bar{Y}_j) \pm t_{n-I}^{(1-\alpha)} \, \hat{SE}(\bar{Y}_i - \bar{Y}_j)$$

ignoring multiple comparisons

Bonferroni correction for mult. comp.: replace $(1-\alpha)$ by $\left(1 - \frac{\alpha}{k}\right)$

ex. $\begin{pmatrix} k = 6 \\ \alpha = .05 \end{pmatrix}$

$(1-\alpha) = 95\%$ but $1 - \frac{\alpha}{k} = 1 - \frac{.05}{6} = 99.2\%$

Karl Pearson (1890s) key observation:

if null true, $p_G = p_I = p_P$

$$p_G = p(NS \mid G)$$
$$p_I = p(NS \mid I)$$
$$p_P = p(NS \mid P)$$

all equal if null true

in other words, if null true

(method) M & (smoking status) S

would be <u>independent</u> in pop.

indep

$$p(M \,\&\, S) = p(M) \cdot p(S)$$

⑤ $\hat{p}_{ij}$ if null true

	NS	S	
G	.075	.287	$\frac{250}{692} = .361$
I	.036	.140	.176
P	.096	.367	.462
	$\frac{143}{692} = .207$	.793	1

⑭

$\hat{p}_{null}(G \,\&\, NS) =$

$\hat{p}(G) \cdot p(NS)$

$\left(\frac{250}{692}\right)\left(\frac{143}{692}\right)$

⑥ $\hat{O}_{ij}$

59	191	250
27	95	122
57	263	320
143	549	692

$\hat{E}_{ij} = \hat{p}_{ij} \cdot 692$

51.7	198.3	250
25.2	96.8	122
66.1	253.9	320
143	549	692

$(.075) \cdot 692$

$(\hat{O}_{ij} - \hat{E}_{ij})$

residuals

+7.3	-7.3	0
+1.8	-1.8	0
-9.1	+9.1	0
0	0	0

$$\frac{(\hat{O}_{11} - \hat{E}_{11})^2}{\hat{E}_{11}} + \frac{(\hat{O}_{12} - \hat{E}_{12})^2}{\hat{E}_{12}} + \dots$$

$$+ \frac{(\hat{O}_{32} - \hat{E}_{32})^2}{\hat{E}_{32}} = \cancel{0}$$

chi-squared

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(\hat{O}_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

$$= 3.062$$