AMS 7
Lecture 14

(case study cont. on page before)

1-way Analysis cont.

Makes SS_B Big even if null is true

1) SS_B gets bigger as I (#groups) increases

easy fix: divide by $(I-1)$ degrees freedom

DF_B = degrees freedom $\frac{SS_B}{DF_B} = MS_B$ mean sq

2) units can make SS_B arbitrarily ↑ or ↓

easy fix: $s_1^2 = \frac{1}{n_1-1} \sum_{j=1}^{n_1} (y_j - \bar{y}_i)^2$

noise estimate from each group = variance

overall noise/variance estimate =

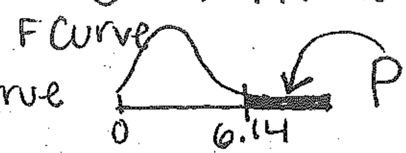
$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2 + \ldots + (n_I-1)s_I^2}{n_1-1 + \ldots + (n_I-1)}$$

SS_W = within group sum of squares

$\frac{SS_W}{DF_W} = MS_W$ within group mean square

$\frac{\text{signal}}{\text{noise}} = \frac{MS_B}{MS_W}$ F-Ratio $0 \leq$ F-Ratio

long run hist. of F if null true



Remember: We reject the null if P small

<.6%, null=bad

P282 Anova table

all it really says is that the null is wrong

Which groups differ?

Mult. comparisons: Bouferroni Method

1) decide how many comp. to make

$k =$

2) decide confidence

$100(1-a)\%$ 95% CI $a = .05$

3) $\hat{SE} = \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

root mean squared error $= \sqrt{MS_W}$

Group 1 vs	Group 2
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$
$\hat{SE}_{IID}$ 2 ind	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

1 way anova: assum $\sigma_i = \sigma$

$$\bar{y}_i - \bar{y}_j \pm t_{n-I}^{1-a \; .95} \hat{SE}(\bar{y}_i - \bar{y}_j) \quad 95\% \text{ CI}$$

Bonferroni: Correction for mult comp

Replace $(1-a)$ by $(1-\frac{a}{K})$

L-295 4) make K $100(1-a)\%$ CIs

in form .....

fix: use bigger t $\boxed{t_{n-I}^{1-\frac{a}{K}}}$

70

Last Section: Categorical Data Analysis

L-295 CDCP Case Study $n=692$ smokers

(G) Gum (P) Patch (I) Inhaler

Outcome: Smoking after 5 mo.

method: table (called $\underline{I \times J \text{ Contingency}}$)

	not Smoking	S	
(G)	59	191	250
(I)	27	95	122
(P)	57	263	320
	143	549	692

1=NS, 0=S

$\hat{p}_I = \frac{27}{122} = 22.1\% = \hat{p}(NS \mid I)$ $\hat{p}_G = \frac{59}{250} = 23.6\% = \hat{p}(NS \mid G)$

$\hat{p}_P = 57/320 = 17.8\%$ yes, large diff.

null: $p_G = p_I = p_P$ Alt: not so

Compare Observed freq vs Expected

Key est.: if null true, $p_G = p_I = p_P$

$$\left.\begin{array}{l} p_G = P(NS \mid G) \\ p_I = \text{''} \quad \text{''} \quad I \\ p_P = \text{''} \quad \text{''} \quad P \end{array}\right\} \text{all equal if null true}$$

In other words, if the null were true, method + smoking ~~status~~ Status would be independent in pop

$P(M \cap S) \overset{\text{indep.}}{=} P(M) \cdot P(S)$

$\hat{p}_{ij}$ if null true →

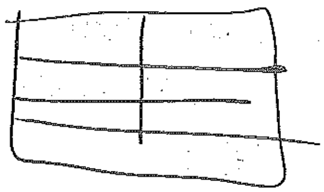
		NS	(S) S	
(M)	(G)	.075	.287	.361
	(I)	.036	.140	.176
	(P)	.096	.367	.462
		.207	.793	1

Mult. by the total to get the estimated data table

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

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

+7.3	−7.3	0
+1.8	−1.8	0
−9.1	+9.1	
0	0	0



Residuals ↑ Always 0

$$\frac{(\hat{O}_{11} - \hat{E}_{11})^2}{\hat{E}_{11}} + \ldots \frac{(\hat{O}_{32} - \hat{E}_{32})^2}{\hat{E}_{32}}$$

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