Biostatistics - ANOVA

Req. Reading Rosner: pg. 516-537, Pg. 543 eq. 12.20, 555-562, 576-77

Pulmonary Disease Example w/ Forced Mid-Expiratory Flow (FEF)

Consider K groups

Non Smokers (NS) Light Smokers (LS) Moderate Smokers (MS)

we have n_i observations of each Group..

200 200 200

ANOVA ; Wilcox-Rank Sum ; Mann-Whitney U ;

We will fit the model

$$Y_{ij} = \mu + \alpha_i + e_{ij}$$, to , Multiple input single output MISO

Multiple input multiple output MIMO ; multivariate analysis. vs. variable

$$E[X_i] = \mu + \alpha_i$$

Can not solve for $\mu + \alpha_i$ variables. ; Guage?

Rosner Constraint; $\sum \alpha_i = 0$ ← makes

or SAS $\alpha_{last} = 0$

Null Hypothesis: all $\alpha_i = 0$, or $\prod \alpha_i = 0$

More math.

for each sample constant term in group

$$Y_{ij} - \mu_{all} = Y_{ij} - \alpha_i + \alpha_i - \mu_{all}$$

( ) ( ) ( )

Test Statistics

BMSS $= \sum_i \sum_j (y_{ij} - \alpha_i)^2$; Between Mean Square ; df=n-k

WMSS $= \sum_i \sum_j (\alpha_i - M_{all})^2$; Within Mean Square df=k-1

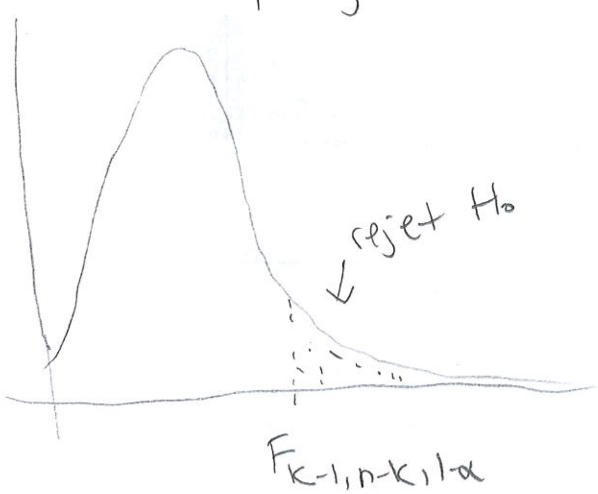
$$F = \frac{BMSS}{WMSS}$$

If $F \geq F_{k-1, n-k, 1-\alpha}$ reject H_0

$F \leq F_{k-1, n-k, 1-\alpha}$ accept H_0

Sum Squres is a check of computations.

$$SS = \sum_i \sum_j (y_{ij} - M_{all})^2$$



before looking at data.

LSD Procedure.; used pooled estimate of variance s^2 for t-tests.

Suppose there are 10 groups $\binom{10}{2} = 45$

$E[.05\% \cdot 45] = 2$ significant comparisons.

Bonferroni Adjustment

Chang α for significant finding

$$\alpha^* = \alpha \Big/ \binom{K}{2}$$

if $t \geq t_{n-k, 1-\alpha^*}$ reject H_0

Bonferroni much more conservative than LSD

Scheffe's Multiple Comparisons.

False Discovery Rate Testing Procedure.; less conservative. (used in genetic tests)

→ instead of controling for type I error. reject $H_0 \mid H_0 = $ true

used in genetic studies

① Rank P-Values in k tests

② $q_i \triangleq \frac{K P_i}{i}$ $P_i < P_K$

$FDR_i = \min(q_i, \ldots, q_K)$

Relation Between Linear Regression & ANOVA

(1) Can perform an ANOVA as described, $Y_{ij} = \mu + \alpha_i + e_{ij}$

(2) Can perform a regression on a model of form

$$Y = \alpha + \sum_{j=1}^{k-1} \beta_j X_j + e,$$

where $X_j = \begin{cases} 1 & \text{subject in } j \\ 0 & \text{otherwise} \end{cases}$

; Regression SS = Between SS
Residual SS = Within SS

Can use LSD procedure.

Residual Component $\triangleq Y_i - \overline{Y}_i$
Regression Component $\triangleq \overline{Y} - \overline{Y}_i$
Sum Squares defined similarly.

test $\beta = 0$ vs. $\beta \neq 0$ F-test w/ same procedure.

( )

Biostatistics

$(Y_{ii} - \mu_{all})^2 = (Y_{ij} - \alpha_i)^2 + (\alpha_i - \mu_{all})^2$; not easy proof

$Y_{ij}^2 - 2Y_{ij}\mu_{all} + \cancel{\mu_{all}^2} = Y_{ij}^2 - 2Y_{ij}\alpha_i + \alpha_i^2 + \alpha_i^2 - 2\alpha_i\mu_{all} + \cancel{\mu_{all}^2}$

$+2Y_{ij}\mu_{all} = +2\alpha_i\mu_{all} + 2Y_{ij}\alpha_i - 2\alpha_i^2$

interested if expectation values are same. in summation

$\sum_i^{n_i}\sum_j^{k} 2Y_{ij}\mu_{all} = \sum_i\sum_j 2\alpha_i\mu_{all} + 2Y_{ij}\alpha_i - 2\alpha_i^2$

$2\mu_{all}\sum_i^{n_i}\sum_j^{k} Y_{ij} =$

is $\sum\sum \alpha_i\mu_{all} = \sum\sum \alpha_i^2$?

$\mu_{all}\underbrace{\sum\sum \alpha_i}_{=0} = \sum_i^{j}\sum_j^{k}\alpha_i^2$?

$\alpha_1^2 + \alpha_2^2 + \alpha_3^2$

$E[(X-E[X])^2] = E[X^2] - 2E[X$

$X^2 - 2XE[X] + E[X]^2$

$\mu \qquad \mu^2$

$E[\, X^2 - 2X\mu \qquad]$

$E[X^2] - 2E[X]E[X] + E[X]^2$

$= E[X^2] + E[X]^2$

$E[X] = \mu$, it's a #

$E[2X] = 2E[X]$