Bioestatística

M.I. Eng. Biomédica

2015-2016

Aula Teórica 9

ANOVA

- Exemplo:
  - White e Froeb estudaram a influência do "fumo passivo" na função respiratória. Constituíram 6 grupos:

1. Não fumadores (NS)
2. Fumadores passivos (PS)
3. Fumadores não inaladores (NI)
4. Fumadores leves (LS)– menos de 10 cigarros/dia
5. Fumadores moderados (MS) – entre 11 a 39 cigarros/dia
6. Fumadores crónicos (HS) – mais de 40 cigarros/dia

ANOVA

FEF data for smoking and nonsmoking males

Group number, i	Group name	Mean FEF (L/s)	*sd* FEF (L/s)	n_i
1	NS	3.78	0.79	200
2	PS	3.30	0.77	200
3	NI	3.32	0.86	50
4	LS	3.23	0.78	200
5	MS	2.73	0.81	200
6	HS	2.59	0.82	200

Source: Reprinted by permission of *The New England Journal of Medicine*, *302*(13), 720–723, 1980.

FEF – forced mid-expiratory flow

in Fundamentals of Biostatistics, B. Rossner, pp 517

ANOVA

- Suponhamos que existem *k* grupos. O grupo i apresenta n_i observações. A observação j do grupo i, denotada por y_{ij}, pode ser escrita segundo o modelo:

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

em que μ é uma constante, α_i é uma constante especifica do grupo i e e_{ij} é um termo de erro associado a cada observação que segue uma distribuição normal de média 0 e variância σ^2.

- Não é possível estimar a partir das observações todos os parâmetros, assim é usual recorrer-se a restrições. Uma das restrições é considerar a soma dos αi igual a zero.

ANOVA

- A equação anterior traduz a análise de variância de uma só entrada ou o modelo ANOVA unifactorial (one way ANOVA).
- Com este modelo a média de um número arbitrário de grupos, os quais seguem uma distribuição normal, pode ser comparada.
- A variabilidade nos dados é verificada quanto à sua origem: variabilidade dentro dos grupos ou variabilidade entre os grupos.

ANOVA

- O valor μ representa a média geral de todos os grupos em conjunto
- O valor α_i representa a diferença entre a média do grupo i para a média geral
- O valor e_{ij} representa o erro aleatório relativamente $\mu+\alpha_i$ para uma observação individual do grupo i

ANOVA

- **Teste de hipótese para a análise de variância**

- A hipótese nula, H_0, neste caso é as médias de todos os grupos serem iguais. É equivalente a ter $\alpha_i = 0$ para qualquer i.

- A hipótese alternativa é, então, que pelo dois grupos apresentam médias diferentes. É equivalente a ter $\alpha_i \neq 0$ para pelo menos um i.

ANOVA

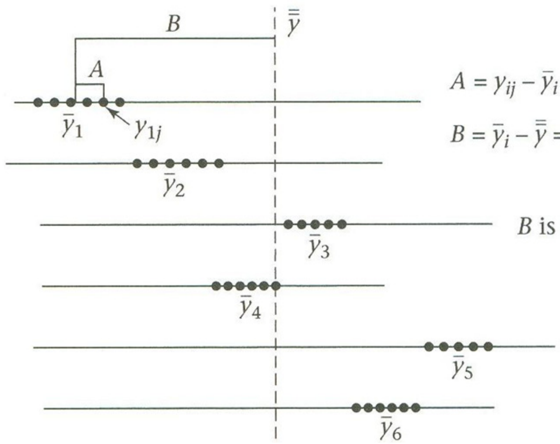
- O desvio de uma observação relativamente à media geral pode ser representada por:

$$y_{ij} - \overline{\overline{y}} = \left(y_{ij} - \overline{y_i} \right) + \left(\overline{y_i} - \overline{\overline{y}} \right)$$

- em que $\overline{\overline{y}}$ representa a média geral e $\overline{y_i}$ a média do grupo i.

- O primeiro termo do lado direito da equação é a variabilidade dentro do grupo e o segundo termo a variabilidade entre grupos.

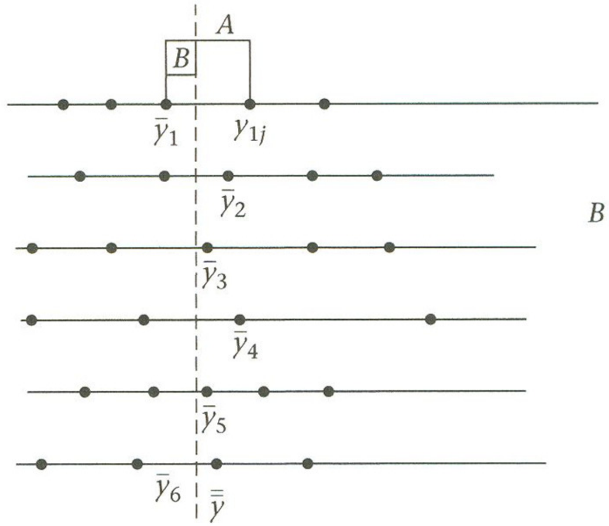
ANOVA



$A = y_{ij} - \bar{y}_i$ = within-group variability

$B = \bar{y}_i - \bar{\bar{y}}$ = between-group variability

B is large relative to *A*.



B is small relative to *A*.

in Fundamentals of Biostatistics, B. Rossner, pp 519

ANOVA

- Se a variabilidade entre grupos é grande e a variabilidade dentro do grupo é pequena então H_o é rejeitada.
- Se a variabilidade entre grupos é pequena e a variabilidade dentro do grupos é grande então H_o é aceite.
- Elevando ao quadrado a equação anterior e somando para todas as observações, tem-se:

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{\overline{y}}\right)^2}_{SS_T}=\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y_i}\right)^2}_{SS_D}+\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\overline{y_i}-\overline{\overline{y}}\right)^2}_{SS_G}$$

ANOVA

- Seja

$$MS_G = \frac{SS_G}{k-1}$$

- E

$$MS_D = \frac{SS_D}{n-k}$$

- O teste efectua-se com base no quociente entre MS_G e MS_D. Se este quociente é muito grande rejeita-se a hipótese nula, caso contrário aceita-se.

- Sob a hipótese nula o quociente MS_G/MS_D segue uma distribuição $F_{k-1,n-k}$.

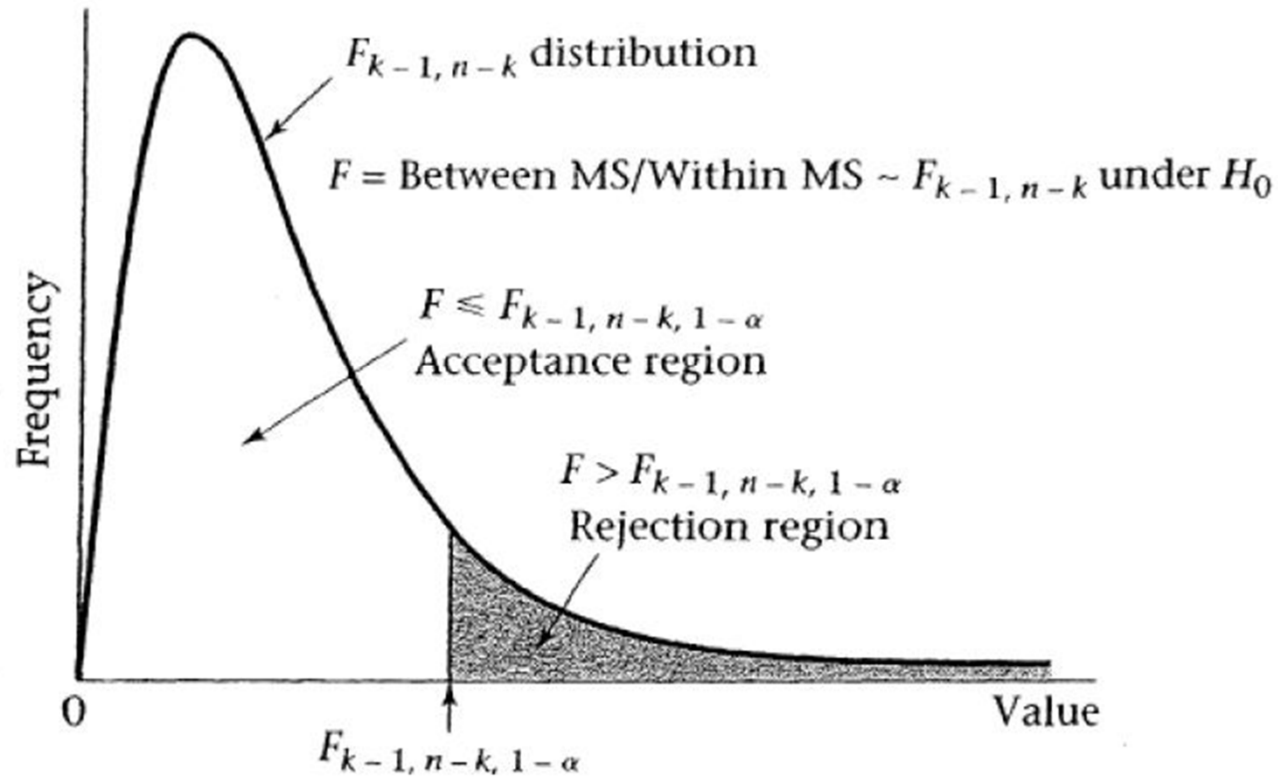
ANOVA

- Os passos que envolvem o teste que admite:

$$H_0 : \alpha_i = 0 \qquad H_1 : \alpha_i \neq 0$$

- são:
- 1) determinar MSG e MSD;
- 2) calcular o teste-estatístico F= MSG/MSD
- 3) se $F > F_{k-1,n-k,1-\alpha}$, então rejeita-se H_o
- se $F \leq F_{k-1,n-k,1-\alpha}$, então aceita-se H_o

ANOVA



in Fundamentals of Biostatistics, B. Rossner, pp 521

ANOVA

- **Comparação de grupos específicos**

- Suponhamos que desejamos comparar dois grupos cujas médias diferem entre si. Temos então:

$$\bar{Y}_1 \sim N\left[\mu + \alpha_1, \sigma^2 / n_1\right]$$

$$\bar{Y}_2 \sim N\left[\mu + \alpha_2, \sigma^2 / n_2\right]$$

- Pelo que:

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left[\alpha_1 - \alpha_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

- E sob H_o

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left[0, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

ANOVA

- A variância σ² pode ser estimada a partir das variâncias amostrais dos grupos. Assim:

$$s^2 = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i - 1)} = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{n-k} = MS_D$$

- A comparação entre grupos é então efectuada com recurso ao teste t.

$$t = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n-k}$$