CC0288 - Inferência Estatística I

Aula - 19/04/2023.

Prof. Maurício

1. Estimação de Máxima Verossimilhança: Vejam os seguintes exemplos:

   Em um cruzamento genético os descendentes podem ser de três tipos distintos. Foram realizados n cruzamentos sendo observados x_i descendentes do tipo i, $i = 1, 2, 3$. De acordo com o modelo genético, as proporções dos três tipos seriam $(2+p)/4$, $(1-p)/2$ e $p/4$ e as progênies são independentes.

   a. Mostre que $\hat{p}$ é uma raiz da equação quadrática
   $$np^2 + (2x_2 + x_3 - x_1)p - 2x_3 = 0.$$
   b. Suponha que $x_1 = 58$, $x_2 = 33$ e $x_3 = 9$. Ache $\hat{p}$, e calcule as frequências esperadas sob o modelo.
   c. Teste a qualidade do ajuste usando um nível de significância de 5%. Qual o nível descritivo do teste?

Solução: Seja X_i o número de cruzamentos do tipo i, $i = 1, 2, 3$, entre os n cruzamentos.

$$(X_1, X_2, X_3) \sim Trinomial(n, p_1, p_2, p_3),$$

com

$$p_1 = \frac{2+p}{4}, \;\; p_2 = \frac{1-p}{2} \quad e \quad p_3 = \frac{p}{4}$$

com $p_1 + p_2 + p_3 = 1$.

Note que:

$$X_i \sim B(n, p_i), \;\; i = 1, 2, 3.$$

Com

$$E(X_i) = n\, p_i \; i = 1, 2, 3.$$

Assim,

$$E(X_1) = n\,\frac{2+p}{4} \quad E(X_2) = n\,\frac{1-p}{2} \quad E(X_3) = n\,\frac{p}{4}$$

que são as frequências esperadas do nosso modelo genético.

Além disso temos:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!}\, p_1^{x_1} \;\; p_2^{x_2} \;\; p_3^{x_3},$$

com

$$x_1 + x_2 + x_3 = n.$$

A função de verossimilhança é dada por:

$$L(p) = \frac{n!}{x_1!x_2!x_3!} \left[\frac{2+p}{4}\right]^{x_1} \left[\frac{1-p}{2}\right]^{x_2} \left[\frac{p}{4}\right]^{x_3},$$

$$L(p) = \frac{n!}{x_1!x_2!x_3! \ 4^{x_1+x_3}2^{x_2}} [2+p]^{x_1} \ [1-p]^{x_2} \ [p]^{x_3},$$

$$L(p) = c \times (2+p)^{x_1} \ (1-p)^{x_2} \ (p)^{x_3},$$

Aplicando logaritmo natural temos:

$$l(p) = \log(c) + x_1 \ \log(2+p) + x_2 \log(1-p) + x_3 \ \log(p).$$

A derivada primeira de $l(p)$ é dada por:

$$l^{'}(p) = \frac{x_1}{2+p} - \frac{x_2}{1-p} + \frac{x_3}{p}.$$

A derivada segunda de $l(p)$ é dada por:

$$l^{''}(p) = -\frac{x_1}{(2+p)^2} - \frac{x_2}{(1-p)^2} - \frac{x_3}{p^2} < 0.$$

Fazendo

$$l^{'}(p) = 0$$

$$\frac{x_1}{2+p} - \frac{x_2}{1-p} + \frac{x_3}{p} = 0$$

Multiplicando por $(2+p)(1-p) \ p$ temos:

$$(p-p^2)x_1 - (2p+p^2)x_2 + (2-p-p^2)x_3 = 0$$

$$(-x_1 - x_2 - x_3)p^2 + (x_1 - 2x_2 - x_3)p + 2x_3 = 0$$

multiplicando por (-1)

$$(x_1 + x_2 + x_3)p^2 + (2x_2 + x_3 - x_1)p - 2x_3 = 0$$

Sejam

$$a = (x_1 + x_2 + x_3) = n \ ; b = x_3 + 2x_2 - x_1 \quad e \quad c = -2x_3.$$

Como

$$x_1 = 58, \ \ x_2 = 33 \ \ \ x_3 = 9, \quad n = x_1 + x_2 = x_3 = 100.$$

Assim

$$100p^2 + 17p - 18 = 0$$

Vamos resolver usando o R:

```
> x_1=58;x_2=33;x_3=9
> n=x_1+x_2+x_3;n
[1] 100
>
> a=n;a
[1] 100
> b=-2*x_2+x_3-x_1;b
[1] 17
> c=-2*x_3;c
[1] -18
>
> #### 100p^2 -17p -18=0
>
> delta=b^2-4*a*c;delta
[1] 7489
> aux=sqrt(delta);aux
[1] 86.53901
>
> p=(-b +c(-1,1)*aux)/(2*a);p
[1]  0.347695 -0.517695
>
>
> raiz=polyroot(c(c,b,a));raiz
[1]  0.347695-0i -0.517695+0i
>
> abs(raiz)
[1] 0.347695 0.517695
> p_est=abs(raiz)[1];p_est
[1] 0.347695
>
> p_1_est=(2+p_est)/4;p_1_est
[1] 0.5869238
> p_2_est=(1-p_est)/2;p_2_est
[1] 0.3261525
> p_3_est=p_est/4;p_3_est
[1] 0.08692376
>
>
> p_1_est+p_2_est+p_3_est ###teste!!!!!!
[1] 1
>
> e_1=n*p_1_est;e_1
[1] 58.69238
> e_2=n*p_2_est;e_2
[1] 32.61525
> e_3=n*p_3_est;e_3
[1] 8.692376
>
> e_1+e_2+e_3
[1] 100
>
> o=c(x_1,x_2,x_3);o
```

```
[1] 58 33  9
> e=c(e_1,e_2,e_3);e
[1] 58.692376 32.615248  8.692376
> d=o-e
> d2=d^2
>
> v=d^2/e;v
[1] 0.008167749 0.004538804 0.010886837
> q_cal=sum(v);q_cal
[1] 0.02359339
>
> tipo=1:3
>
> tab=cbind(tipo,o,e,d,d2,v);tab
tipo  o         e          d         d2           v
[1,]    1 58 58.692376 -0.6923761 0.47938462 0.008167749
[2,]    2 33 32.615248  0.3847521 0.14803421 0.004538804
[3,]    3  9  8.692376  0.3076239 0.09463248 0.010886837
>
>
> gl=3-1-1;gl
[1] 1
> alfa=0.05
> q_tab=qchisq(1-alfa,gl);q_tab
[1] 3.841459
>
> q_cal< q_tab
[1] TRUE
> nd=1-pchisq(q_cal,gl);nd
[1] 0.877924
>
> nd>alfa  ##Não rejeitar H_0!!!!
[1] TRUE
>
```

Vemos que a estimativa de MV saiu quando resolvemos uma equação do segundo grau.

Nem sempre isto acontece. Por isso precisamos de métodos numéricos para achar a estimativa de MV procurada.

Note que:

Seja $L(\theta)$ a função de verossimilhança e $l(\theta) = \log(\theta)$.

Seja a função escore

$$U(\theta) = \frac{\partial l}{\partial \theta}.$$

Temos que o estimador de MV de θ, $\hat{\theta}$ satisfaz:

$$U(\hat{\theta}) = 0,$$

de modo que expandindo $U(\hat{\theta})$ em série de Taylor, em torno de um ponto θ_0, obtemos

$$0 = U(\hat{\theta}) \approx U(\theta_0) + (\hat{\theta} - \theta_0)U^{'}(\theta_0)$$

Assim

$$\hat{\theta} - \theta_0 \approx \frac{U(\theta_0)}{U'(\theta_0)}$$

e

$$\hat{\theta} \approx \theta_0 - \frac{U(\theta_0)}{U'(\theta_0)}$$

Esta última equação gera o procedimento iterativo conhecido como Newton-Raphson.

$$\theta_{j+1} = \theta_j - \frac{U(\theta_j)}{U'(\theta_j)}$$

que é iniciado com um valor θ_0 e daí um novo valor θ_1 é obtido
através de

$$\theta_1 = \theta_0 - \frac{U(\theta_0)}{U'(\theta_0)}$$

e assim por diante até a estabilização do processo, ou seja, dado $\epsilon > 0$ pequeno

$$|\theta_{j+1} - \theta_j| < \epsilon.$$

Vamos mostrar agora o método Scoring de Fisher.
Vamos colocar a equação de Newton-Raphson na forma:

$$\theta_{j+1} = \theta_j + \frac{U(\theta_j)}{-U'(\theta_j)}$$

Vamos substituir $U^{'}(\theta_j)$ por $E\left(U^{'}(\theta_j)\right) = -I_F(\theta)$. Assim

$$\theta_{j+1} = \theta_j + \frac{U(\theta_j)}{I_F(\theta_j)}$$

Este método apresenta uma significativa simplificação no procedimento.