TEMA XVII: ESTIMACIÓN NON PARAMÉTRICA

Os contrastes non paramétricos refíren-se a tres características maix xerais:

[1] Contraste de Bondade de Axuste Supoñemos unha distribución determinada, e tráta-se de ver se os dados son consistentes con esa distribución.

[2] Contraste de Independéncia Tráta-se de ver se as observacións son independentes, dentro dunha única povoación.

[3] Contraste de Homoxeneidade Tráta-se de ver se todas as observacións proveñen da mesma povoación.

Contrastes de Bondade de Axuste

- **Contraste de Chi Cuadrado de Pearson**
- **Contraste de Kolmogorov-Smirnov**

1.- Test Chi-Cuadrado [Pearson] Aplíca-se para ver se un conxunto de dados observados coincide ou non con un conxunto de dados agardados ou teóricos. Para unha serie de dados podemos ver até que ponto esta mostra se pode considerar como pertencente a unha distribución teórica coñecida.

Como se aplica o test X^2

H_0: F_0 é a distribución
H_1: F_0 non é a distribución

Divíden-se as observacións en k clases mutuamente excluintes que denotaremos como A_1, A_2, ...,A_k de maneira que estas clases cobren todo o percorrido da función.

Tóma-se unha mostra de tamaño n e sexa n_i a frecuéncia observada da clase i (nº de observacións mostrais en cada unha das clases). Sexa e_i a frecuéncia esperada da clase i. Sexa P_i a probabilidade da clase i baixo a distribución que estamos supoñendo en H_0.

$P_i=P(A_i)$ $e_i=n\cdot p_i$ n=20, k=5, p=0.2

Tratará-se de ver que as discrepáncias entre os dados sexa suficiente para aceitar ou rexeitar, para isto utilizará-se un estatístico.

$\Sigma_{i=1}^{k} e_i=n=\Sigma_{i=1}^{k} n_i$

	A_1	A_2	A_3	...	A_n
frecuéncia observada n_i					
frecuéncia esperada e_i					

$$Q = \frac{\Sigma_{i=1}^{k} (n_i-e_i)^2}{e_i} \sim X_{n-1}^2$$

$$Q = \frac{\Sigma_{i=1}^{k} n_i}{e_i} - n$$

Fixado un α calcúla-se $\rightarrow X_{\alpha,k-1}^2$

$Q>X_{\alpha,k-1}^2 \Rightarrow$ rexeitamos H_0

A distribución suposta pode que dependa duns parámetros, se estes parámetros son descoñecidos e hai que estima-los por máxima verosimilitude $\Rightarrow Q\rightarrow X_{k-\mu-1}^2$ onde μ=nº de parámetros estimados ao aplicar o MMV.

Se a variábel é contínua ps dados agrúpan-se en clases para cubrir a totalidade da distribución.

▪ Exercicio *Os aficionados ás carreiras de cabalos de cote afirman que nunha carreira arredor dunha pista circular existen vantaxes significativas para os cabalos que ocupan certas posicións de pista. A posición de pista de calquera cabalo é asignada na liña de saída; nunha carreira de 8 cabalos a posición 1 é a mais achegada á barandilla e a 8 é a mais alonxada. Podemos probar o efeito da posición de pista analisando os resultados da carreira dados de acordo con dita posición dirante o primeiro ano de carreiras de 1955. Os dados son:*

	1	2	3	4	5	6	7	8	
nºveces que o cabalo gañou	29	19	18	25	17	10	15	11	N=144

H_0: $P_i=1/8$
H_1: $P_i\neq 1/8$
é unha distribución discreta
$e_i=144/8=18$

$$Q=16.33 \text{ sendo } Q=\frac{\Sigma(n_i-e_i)^2}{e_i}\sim X^2_7$$

—se $\alpha=0.01 \Rightarrow X^2_{0.01,7}=18.475 \Rightarrow$ aceitamos H_0
—se $\alpha=0.05 \Rightarrow X^2_{0.05,7}=14.07 \Rightarrow$ rexeitamos H_0

Este contraste non decide.

▪ Exercicio *A vida de 70 motores tivo a seguinte distribución:*

Anos funcionando	[0,1)	[1,2)	[2,3)	[3,4)	³4	
n_i	30	13	6	5	6	N=70
e_i	33.15	17.458	9.184	4.879	5.32	

Pode supoñer-se que a vida dos motores segue unha distribución expoñencial.

$E(x)=1/\lambda=\bar{x}$ (estimado por MMV)

H_0: Expoñencial
H_0: Non Expoñencial

$$\bar{x}=\frac{0.5\cdot 30+1.5\cdot 13+2.5\cdot 6+3.5\cdot 5+4.5\cdot 6}{60}=94/60=1.567$$

k=5>5?

Fixamos un α $X^2_{0.01,3}=0.115$ $F(x)=1-e^{-\lambda x}$

$$Q=\frac{\Sigma(n_i-e_i)^2}{e_i}=(\Sigma n_i)/e_i-n$$

Calcúlan-se as probabilidades sendo $\lambda=1.567$

$P_1(0<x<1)=F(1)-F(0)=1-e^{-1.5}-(1-e^0)=0.7768$

$P_2(1<x<2)=F(2)-F(1)=1-e^{-1.5\cdot 2}-0.7768=0.9502-0.7768=0.1734$
$P_3(2<x<3)=F(3)-F(2)=1-e^{-1.5\cdot 3}-0.9502=0.9889-0.9502=0.03869$
$P_4(3<x<4)=F(4)-F(3)=1-e^{-1.5\cdot 4}-0.9889=0.9975-0.9889=0.0086$

H_0: Exp(1.567) $F(x)=1-e^{-\lambda x}$
H_1: No Exp

$P_i=P(A_i)=F(1)-F(0)=1-e^{-1/1.567}=0.4727$
$P_2=F(2)-F(1)=1-e^{-2/1.567}-(1-e^{-1/1.567})=0.7232-0.473=0.2494$
$P_3=F(3)-F(2)=1-e^{-3/1.567}-0.7232=0.8543-0.7232=0.1312$
$P_4=F(4)-F(3)=1-e^{-4/1.567}-0.8526=0.9233-0.8543=0.0697$
$P_5=1-F(4)=1-0.9233=0.0767$

$Q=(\Sigma n_i)/e_i-n = 3.2560$

$X^2_{0.05,3}=7.81 \Rightarrow$ aceitamos H_0

Critérios para criación de clases:
—hai que tomar que k≥5
—hai que tomar que $e_i \geq 3\ \forall i$
Se os e_i non son ≥3, a frecuéncia de várias clases agrúpan-se para dar valores ≥3

2.- Contraste de Kolmogorov-Smirnov

H_0: x ~ F
H_0: x + F

Este contraste calcula a distáncia entre a función de distribución teórica F e a función de distribución empírica Fn

Función de distribución empírica

$x_1 \leq x_2 \leq \ldots \leq x_n$

$$F_n(x)=\begin{cases}0 & \text{se } x<x_i \\ i/n & \text{se } x_i \leq x \leq x_{i+1} \\ 1 & \text{se } x \geq x_n\end{cases}$$

$D_n=\max(x)\ |F_n(x)-F(x)|$ D_n é un estatístico

$D_n=\max(x)\{\ |F_n(x_{i-1})-F(x_i)|,|F_n(x_i)-F(x_i)|\}$

Se $D_n \geq D_{\alpha,n} \Rightarrow$ rexeitamos H_0

Vantaxes e inconvintes (con respeito de X^2)
i) Para mostras pequenas o contraste mais axeitado é o de KS
ii) O contraste KS non necesita que os dados estexan agrupados en intervalos
iii) Non é tan exacto estimar os parámetros e despois utilizar o contraste de KS

▪ Exercicio ***Contrastar se a seguinte mostra de duracións de vida pode supoñer-se unha expoñencial de média 11.5.***

16, 8, 10, 12, 6, 10, 20, 7, 2, 24 *n=10*

H_0*:x~Exp(11.5)* *l =1/11.5* $F(x_i)=1-exp^{-lxi}$

x_i	$F_n(x_i)$	$F_n(x_{i-1})$	$F(x_i)$	$\lvert F_n(x_i)-F(x_i)\rvert$	$\lvert F_n(x_{i-1})-F(x_i)\rvert$
2	0.1	0	0.15963	0.05963	0.15963
6	0.2	0.1	0.4065	0.2065	0.3065