Cap. 8 Testes de Hipóteses (cont.)

8.7 Teste de ajustamento do qui-quadrado de Pearson para hipótese nula simples

Objetivo: verificar a hipótese de que um conjunto de observações é proveniente de uma determinada distribuição, i.e. $X \sim F$.

Dados:

- Amostra aleatória $(X_1, \dots, X_n)$ de dimensão n, com n "grande", concretizada numa amostra observada $(x_1, \dots, x_n)$.
- Partição em k classes ou categorias, $C_1, C_2, \dots, C_k$, do conjunto dos valores possíveis de X (R_X), i.e.:
  a) $C_i \cap C_j = \emptyset, \ \forall i \neq j, \ i,j = 1,2,\dots,k$;
  b) $\bigcup_{i=1}^{k} C_i = R_X$.

Frequências Observadas: O_i, $i = 1, \dots, k$.

O_i = número de observações da amostra que pertencem a C_i $\left[\sum_{i=1}^{k} O_i = n\right]$

Frequências Esperadas: E_i, $i = 1, \dots, k$

E_i = número esperado de observações da amostra que devem pertencer a C_i assumindo $X \sim F$.

$$\left[\sum_{i=1}^{k} E_i = n\right]$$

<u>Hipóteses</u> : $H_0: X \sim F$ vs $H_1: X \not\sim F$

$$p_i^0 = P(X \in C_i \mid H_0) = P(X \in C_i \mid X \sim F)$$

Hipóteses são então formuladas da seguinte forma:

$$\boxed{H_0: p_i = p_i^0,\ \forall i = 1, \dots, k \quad \text{vs} \quad H_1: \exists\, i : p_i \neq p_i^0.}$$

<u>Estatística de teste</u>

$$Q_{H_0} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \underset{\text{sob } H_0}{\overset{a}{\sim}} \chi^2_{(k-\beta-1)}$$

O_i = frequência absoluta na amostra aleatória

$E_i = n p_i^0$ = frequências absolutas esperadas sob Ho

k = nº de classes

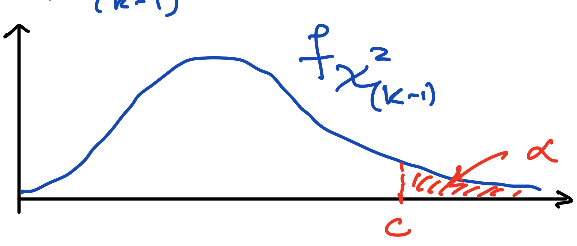
$\beta = 0$ porque o nº de parâmetros a estimar é zero (<u>só vamos considerar este caso</u>).

<u>Tomada de decisão</u>

- Com base na região crítica ao nível de significância α : $RC_\alpha \approx \left[c = F^{-1}_{\chi^2_{(k-1)}}(1-\alpha), +\infty\right[$

$q_{obs} \in RC_\alpha \Rightarrow$ rejeitar Ho

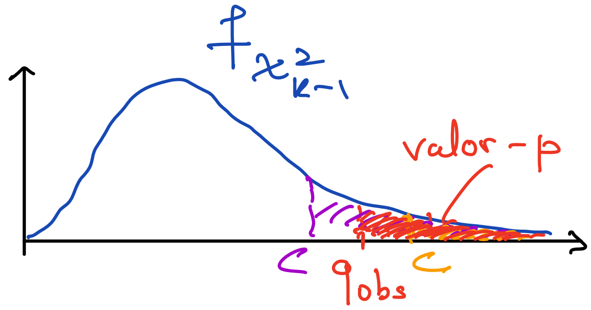
$q_{obs} \notin RC_\alpha \Rightarrow$ ñ rejeitar Ho



- Com base no valor-p :

valor-p $\approx P(Q_{H_0} > q_{obs}) =$

$= 1 - F_{Q_{H_0}}(q_{obs})$



$\alpha \geq$ valor-p $\Rightarrow$ rejeitar H_0

$\alpha <$ valor-p $\Rightarrow$ ñ rejeitar H_0

Observações :

- Antes de calcular $q_{obs} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ deve averiguar-se se

  $E_i \geq 1, \forall i$ e $E_i \geq 5$ em pelo menos 80% dos i's

  pois caso contrário é necessário agrupar classes.

- Quando a v.a. X é contínua ou discreta com muitos valores distintos, é usual formar as classes usando as regras para a construção de histogramas com classes de amplitude constante.

- **Exemplo 1** O lançamento de um dado 1000 vezes conduziu à seguinte tabela de frequências observadas (O_i) :

i =	1	2	3	4	5	6	$K=6$
O_i =	174	174	154	179	154	165	$n=1000$

Será que os resultados obtidos sustentam a hipótese de que "o dado está equilibrado"?

$H_0: P(X=i) = p_i^0 = 1/6, \; i=1,\dots,6 \; (\text{ou } X \sim \text{Unif.}(1,\dots,6))$

$H_1: \exists i: P(X=i) = p_i^0 \neq 1/6 \; (\text{ou } X \not\sim \text{Unif}(1,\dots,6))$

$$E_i = n \cdot p_i^0 = 1000 \times 1/6 = 166.67$$

$$Q_{H_0} = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} \underset{\text{sob } H_0}{\overset{a}{\sim}} \chi^2_5$$

i	O_i	p_i^0	$E_i = np_i^0$	$\frac{(O_i-E_i)^2}{E_i}$
1	174	1/6	166.67	0.322
2	174	1/6	166.67	0.322
3	154	1/6	166.67	0.963
4	179	1/6	166.67	0.912
5	154	1/6	166.67	0.963
6	165	1/6	166.67	0.017
Total	1000	1	~1000	3.499 ($= q_{obs}$)

Decisão:

- Região crítica: para $\alpha = 0.05$ temos
  $c = F^{-1}_{\chi^2_5}(0.95) = 11.07 \Rightarrow RC_{0.05} = [11.07, +\infty[$
  Como $q_{obs} \approx 3.5 \notin RC_{0.05}$, concluímos que não há evidência para rejeitar H_0 ao nível de significância de 5%.
- Valor-p $\approx P(Q_{H_0} > 3.499) \in]0.6, 0.7[$ (tabela) é bastante alto (i.e. $P(\text{rejeitar } H_0 \mid H_0 \text{ verd.}) > 0.6$).

Exemplo 2: O número de defeitos por circuito, num certo tipo de circuitos, deve seguir uma distribuição de Poisson com valor médio 0.75. Numa a.a. de 60 circuitos obtiveram-se os resultados seguintes:

Nº de defeitos	0	1	2	3	
O_i	32	15	9	4	$n=60$

Os dados suportam a hipótese apresentada a um nível de significância de 5%?

- $H_0: X \sim \text{Poisson}(0.75)$, i.e. $P(X=x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$$E(X) = Var(X) = \lambda = 0.75, \quad x = 0, 1, \ldots$$

$H_1: X \not\sim \text{Poisson}(0.75)$

- Temos que acrescentar a classe adicional nº de defeitos ≥ 4 com freq. obs. $= 0$.

- Cálculo dos E_i sob H_0:

E_i: $p_i^0 = P(X=i-1 | H_0) = \dfrac{e^{-0.75} \times 0.75^{i-1}}{(i-1)!}$

e $\quad E_i = 60 \times p_i^0, \quad i = 1, \ldots, 4$

$\Rightarrow p_1^0 = 0.472, \; p_2^0 = 0.354, \; p_3^0 = 0.133, \; p_4^0 = 0.033$

$E_1 = 28.32, \quad E_2 = 21.24, \quad E_3 = 7.98, \quad E_4 = 1.98$

e $\quad p_5^0 = 1 - (p_1^0 + p_2^0 + p_3^0 + p_4^0) = 0.008, \quad E_5 = 0.48$

- Temos $E_5 < 1$ e só 60% dos E_i's ≥ 5, pelo que é necessário agrupar classes.
-

Classes	Nº de def.	O_i	p_i^0	$E_i = np_i^0$	$(O_i - E_i)^2/E_i$
C_1	0	32	0.472	28.32	0.478
C_2	1	15	0.354	21.24	1.833
C_3	≥ 2	13	0.174	10.44	0.628
Totais	$\bigcup_{i=1}^{3} C_i = \mathbb{N}_0$	60	1.000	60.00	2.939 $(= q_{obs})$

$k = 3$

- $RC_{0.05} \approx \left[c = F^{-1}_{\chi^2_2}(0{,}95) = 5.991, +\infty\right[$

- Decisão: $2.939 \notin RC_{0.05} \Rightarrow$ não rejeitar H_0, $\alpha = 5\%$

Ou seja, não há evidência para rejeitar a hipótese de $X \sim$ Poisson (0.75) ao nível de significância de 5%.

- Pergunta 9 do 2º exame de PE de 24/02/2022 às 10h30.