ELEMENTOS DE PROBABILIDADES E ESTATÍSTICA

Ano Letivo 2018/2019

Exemplos de ajustamentos de modelos

Num estudo sobre a incidência de certa doença numa população de insectos, um grupo de biólogos registou ao longo de um ano o número de insectos contaminados em cada amostra de 5 insectos, tendo para tal recolhido 200 amostras. Os resultados obtidos foram os seguintes:

Nº de insectos contaminados	0	1	2	3	4	5
Nº de amostras	17	53	68	44	16	2

Seja X a v.a. que representa o n.º de insectos contaminados, numa amostra de 5 insectos

Será que os dados fornecem evidência estatística de que X tem distribuição Binomial?

H_0: $X \frown Bi(5, p)$ *vs* H_1: $X \not\frown Bi(5, p)$

p desconhecido $\Rightarrow$ estimar p: $\hat{p} = \overline{X} / 5$ com valor observado $\overline{x} / 5 = 1.975 / 5 = 0.395 \quad (\ell = 1)$

x_i	o_i	$\hat{p}_i$	$\hat{e}_i$	$(o_i - \hat{e}_i)^2 / \hat{e}_i$
0	17	0.0811	16.2109	0.0384
1	53	0.2646	52.9198	0.0001
2	68	0.3455	69.1020	0.0176
3	44	0.2256	45.1161	0.0276
4	16	0.0736	14.7280	0.1099
5	2	0.0096	1.9232	0.0031
	200		200	$\mathbf{x^2_0 = 0.1966}$

$\hat{p}_i = P(Bi(5, 0.395) = x_i)$; $\hat{e}_i = n\,\hat{p}_i$

E.T.: $X^2 \underset{H_0}{\sim} \chi^2_{(6-1-1)}$

Valor-p: $P(\chi^2_{(4)} \geq 0.1966) = 0.9955$

Decisão: Rejeita-se H_0 para níveis de significância $\alpha \geq 0.9955 \Rightarrow$ <u>Não se rejeita</u> H_0 para todo o α usual

Conclusão: Os dados fornecem forte evidência estatística de que X tem distribuição Binomial.

Exemplos de ajustamentos de modelos

Num estudo sobre o gorgulho do feijão, introduziram-se larvas nos feijões, que lhes serviram de alimento. As crisálidas saíram através de um buraco feito no feijão e, como tal, o número de buracos por feijão indica-nos o número de insectos adultos que saíram. Na tabela abaixo são apresentados os resultados de uma amostra de 100 feijões:

Nº de gorgulhos saídos por feijão:	0	1	2	3	4	7
Nº de feijões:	53	32	12	1	1	1

Seja X a v.a. que representa o n.º de gorgulhos que saem de um feijão

Será que os dados fornecem evidência estatística de que X tem distribuição Poisson?

$\boxed{H_0: X \frown P(\lambda) \quad vs \quad H_1: X \not\frown P(\lambda)}$

λ desconhecido $\Rightarrow$ estimar λ: $\hat{\lambda} = \overline{X}$ com valor observado $\overline{x} = 0.7$ $\quad (\ell = 1)$

C_i	o_i	$\hat{p}_i$	$\hat{e}_i$
0	53	0.4966	49.6585
1	32	0.3476	34.7610
2	12	0.1217	12.1663
3	1	0.0284	2.8388
4	1	0.0050	0.4968*
5	0	0.0007	0.0696*
6	0	8.11E-05	0.0081*
≥ 7**	1	8.88E-06	0.0009*
	100		100

$\hat{p}_i = P(P(0.7) \in C_i)$; $\hat{e}_i = n\,\hat{p}_i$

**A classe 8 teve de ser modificada de {7} para {7, 8, ...} por forma a que o conjunto de todas as classes constitua uma partição do suporte da distribuição Poisson ($\mathbb{N}_0$)

*Valores esperados inferiores a 1 $\Rightarrow$ Deve juntar-se classes até que $\hat{e}_i \geq 5, \forall i$

C_i	o_i	$\hat{p}_i$	$\hat{e}_i$	$(o_i - \hat{e}_i)^2 / \hat{e}_i$
0	53	0.4966	49.6585	0.2248
1	32	0.3476	34.7610	0.2193
2	12	0.1217	12.1663	0.0023
≥ 3	3	0.0341	3.4142	0.0502
	100		100	$\mathbf{x^2_0 = 0.4967}$

E.T.: $X^2 \underset{H_0}{\sim} \chi^2_{(4-1-1)}$

Valor-p: $P(\chi^2_{(2)} \geq 0.4967) = 0.7801$

Decisão: Rejeita-se H_0 para níveis de significância $\alpha \geq 0.7801 \Rightarrow$ <u>Não se rejeita</u> H_0 para todo o α usual

Conclusão: Os dados fornecem forte evidência estatística de que X tem distribuição Poisson.

Exemplos de ajustamentos de modelos

Fizeram-se 50 medições do comprimento (em cm) de determinada espécie animal no estado adulto, obtendo-se os seguintes valores:

10.16	10.36	10.43	10.45	10.55	10.57	10.65	10.71	10.75	10.98
10.41	11.10	10.50	10.58	10.81	10.39	10.62	10.15	10.47	10.77
10.51	10.53	10.45	10.65	10.52	10.55	10.85	10.97	10.23	10.25
10.76	10.78	10.55	10.30	10.94	10.87	10.54	10.73	10.74	10.42
10.56	10.38	10.35	10.33	10.79	10.72	10.35	10.65	10.59	10.58

Seja X a v.a. que representa o comprimento (em cm) de um animal adulto

Será que os dados fornecem evidência estatística de que X tem distribuição Normal?

$\boxed{H_0: X \cap N(\mu, \sigma) \quad vs \quad H_1: X \not\cap N(\mu, \sigma)}$

μ desconhecido $\Rightarrow$ estimar μ: $\hat{\mu} = \overline{X}$ com valor observado $\overline{x} = 10.577$
σ desconhecido $\Rightarrow$ estimar σ: $\hat{\sigma} = S$ com valor observado $s = 0.217$ $\quad (\ell = 2)$

Tomando por base as classes construídas para a representação em histograma, tem-se:

C_i	o_i	$\hat{p}_i$	$\hat{e}_i$	$(o_i - \hat{e}_i)^2 / \hat{e}_i$
$(-\infty, 10.309)$*	5	0.1084	5.4205	0.0326
[10.309, 10.468)	11	0.1993	9.9658	0.1073
[10.468, 10.627)	15	0.2834	14.1694	0.0487
[10.627, 10.786)	11	0.2411	12.0572	0.0927
[10.786, 10.945)	5	0.1228	6.1392	0.2114
$[10.945, +\infty)$*	3	0.0450	2.2478	0.2517
	50		50	$\mathbf{x^2_0 = 0.7444}$

$\hat{p}_i = P(N(10.577, 0.217) \in C_i)$; $\hat{e}_i = n\,\hat{p}_i$

*As classes 1 e 6 foram modificadas por forma a que o conjunto de todas as classes constitua uma partição do suporte da distribuição Normal ($\mathbb{R}$)

E.T.: $X^2 \underset{H_0}{\sim} \chi^2_{(6-1-2)}$

Valor-p: $P(\chi^2_{(3)} \geq 0.7444) = 0.8627$

Decisão: Rejeita-se H_0 para níveis de significância $\alpha \geq 0.8627 \Rightarrow$ <u>Não se rejeita</u> H_0 para todo o α usual

Conclusão: Os dados fornecem forte evidência estatística de que X tem distribuição Normal.

Exemplos de ajustamentos de modelos

Considere a seguinte amostra (já ordenada) de 20 tempos de vida de uma determinada espécie de insectos:

0.10	0.19	0.21	0.30	0.45	0.55	0.63	0.80	1.22	1.31
1.40	1.57	1.88	2.10	2.48	2.77	3.54	4.60	5.42	6.20

Seja X a v.a. que representa o tempo de vida de um insecto

Será que os dados fornecem evidência estatística de que X tem distribuição Exponencial, com valor esperado igual a 2?

$E(X) = 2 \Leftrightarrow \lambda = ½ = 0.5$

H_0: X $\frown$ Exp(0.5)   *vs* H_1: X $\not\frown$ Exp(0.5)

Não existem parâmetros a estimar $\Rightarrow \ell = 0$

Tomando por base as classes construídas para a representação em histograma, tem-se:

C_i	o_i	p_i	e_i	$(o_i - e_i)^2 / e_i$
[**0.00**, 1.32)*	10	0.4831	9.6630	0.0118
[1.32, 2.54)	5	0.2360	4.7204	0.0166
[2.54, 3.76)	2	0.1282	2.5648	0.1244
[3.76, 4.98)	1	0.0697	1.3936	0.1112
[4.98, +∞)*	2	0.0829	1.6582	0.0705
	20		20	$\mathbf{x^2_0 = 0.3343}$

$p_i = P(Exp(0.5) \in C_i)$; $e_i = n\,p_i$

*As classes 1 e 5 foram modificadas por forma a que o conjunto de todas as classes constitua uma partição do suporte da distribuição Exponencial ($\mathbb{R}_0^+$)

E.T.: $X^2 \underset{H_0}{\sim} \chi^2_{(5-1)}$

Valor-p: $P(\chi^2_{(4)} \geq 0.3343) = 0.9875$

Decisão: Rejeita-se H_0 para níveis de significância $\alpha \geq 0.9875 \Rightarrow$ <u>Não se rejeita</u> H_0 para todo o α usual

Conclusão: Os dados fornecem forte evidência estatística de que X tem distribuição Exponencial.

<u>**ATENÇÃO**</u>: A dimensão da amostra é “pequena” (n = 20), pelo que se deve olhar para os resultados com algum cepticismo adicional.

ELEMENTOS DE PROBABILIDADES E ESTATÍSTICA

Ano Letivo 2018/2019

Exemplos de ajustamentos de modelos

Considere os seguintes dados, que dizem respeito ao peso de 37 crianças de uma determinada classe etária:

16.7	16.8	16.9	17.1	17.2	17.3	17.4	17.5	17.6	17.7
17.8	17.8	17.9	18.2	18.3	18.4	18.6	18.7	18.8	18.9
18.9	19.0	19.1	19.3	19.6	19.6	19.6	19.6	19.8	19.9
20.1	20.4	20.5	20.6	20.6	20.7	20.8			

Seja X a v.a. que representa o peso de uma criança

Será que os dados fornecem evidência estatística de que X tem distribuição Uniforme?

$\boxed{H_0: X \cap U[a, b] \quad vs \quad H_1: X \not\cap U[a, b]}$

a desconhecido $\Rightarrow$ estimar *a*: $\hat{a} = X_{(1)}$ com valor observado $x_{(1)} = 16.7$

b desconhecido $\Rightarrow$ estimar *b*: $\hat{b} = X_{(n)}$ com valor observado $x_{(37)} = 20.8$ $\quad (\ell = 2)$

Tomando por base as classes construídas para a representação em histograma, tem-se:

C_i	o_i	$\hat{p}_i$	$\hat{e}_i$	$(o_i - \hat{e}_i)^2 / \hat{e}_i$
[16.70, 17.39)	6	0.1683	6.2268	0.0083
[17.39, 18.08)	7	0.1683	6.2268	0.0960
[18.08, 18.77)	5	0.1683	6.2268	0.2417
[18.77, 19.46)	6	0.1683	6.2268	0.0083
[19.46, 20.15)	7	0.1683	6.2268	0.0960
[20.15, **20.80**]*	6	0.1585	5.8659	0.0031
	37		37	$\mathbf{x^2_0 = 0.4533}$

$\hat{p}_i = P(U[16.7, 20.8] \in C_i)$; $\hat{e}_i = n\,\hat{p}_i$

*A classe 6 foi modificada por forma a que o conjunto de todas as classes constitua uma partição do suporte da distribuição U[16.7, 20.8]

E.T.: $X^2 \underset{H_0}{\sim} \chi^2_{(6-1-2)}$

Valor-p: $P(\chi^2_{(3)} \geq 0.4533) = 0.929$

Decisão: Rejeita-se H_0 para níveis de significância $\alpha \geq 0.929 \Rightarrow$ <u>Não se rejeita</u> H_0 para todo o α usual

Conclusão: Os dados fornecem forte evidência estatística de que X tem distribuição Uniforme.

Exemplos de ajustamentos de modelos

Uma amostra observada numa certa população X conduziu aos resultados e à classificação que se apresentam em baixo.

n	x(1)	x(60)	$\overline{x}$	s
60	0.846	20.033	6.98	3.783

Ci	oi
[0.846, 4.044)	13
[4.044, 7.242)	28
[7.242, 10.44)	11
[10.44, 13.638)	4
[13.638, 16.836)	2
[16.836, 20.034)	2
	60

Será que os dados fornecem evidência estatística de que X tem distribuição Normal?

H_0: $X \frown N(\mu, \sigma)$ *vs* H_1: $X \not\frown N(\mu, \sigma)$

μ desconhecido $\Rightarrow$ estimar μ: $\hat{\mu} = \overline{X}$ com valor observado $\overline{x} = 6.98$

σ desconhecido $\Rightarrow$ estimar σ: $\hat{\sigma} = S$ com valor observado $s = 3.783$ $(\ell = 2)$

Tomando por base as classes acima, tem-se:

C_i	o_i	$\hat{p}_i$	$\hat{e}_i$	$(o_i - \hat{e}_i)^2 / \hat{e}_i$
$(-\infty$, 4.044)*	13	0.2188	13.1306	0.0013
[4.044, 7.242)	28	0.3088	18.5258	4.8451
[7.242, 10.44)	11	0.2922	17.5318	2.4335
[10.44, 13.638)	4	0.1410	8.4594	2.3508
[13.638, 16.836)	2	0.0346	2.0770	0.0029
[16.836, $+\infty$)*	2	0.0046	0.2753	10.8023
	50			$\mathbf{x^2_0 = 20.436}$

$\hat{p}_i = P(N(6.98, 3.783) \in C_i)$; $\hat{e}_i = n\,\hat{p}_i$

*As classes 1 e 6 foram modificadas por forma a que o conjunto de todas as classes constitua uma partição do suporte da distribuição Normal ($\mathbb{R}$)

E.T.: $X^2 \underset{H_0}{\sim} \chi^2_{(6-1-2)}$

Valor-p: $P(\chi^2_{(3)} \geq 20.436) = 0.0001$

Decisão: Rejeita-se H_0 para níveis de significância $\alpha \geq 0.0001 \Rightarrow$ <u>Rejeita-se</u> H_0 para todo o α usual

Conclusão: Os dados fornecem forte evidência estatística de que X não tem distribuição Normal.

<u>Nota</u>: É natural que se tenha rejeitado a hipótese nula já que a representação gráfica dos dados, de acordo com a classificação fornecida, conduz a um histograma significativamente assimétrico à direita, com uma forma mais próxima de uma distribuição como, por exemplo, o Qui-quadrado.