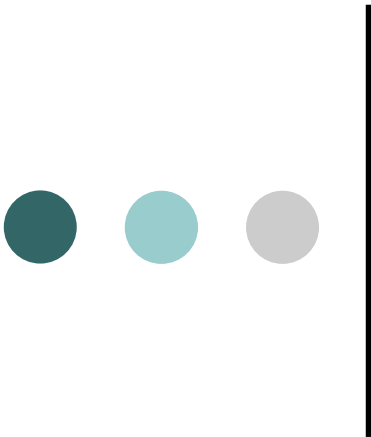
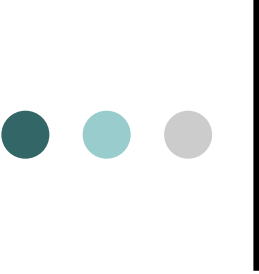




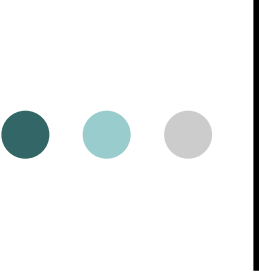
Estatística Não Paramétrica

- ▣ Como construir testes de aderência



Teste Qui-quadrado

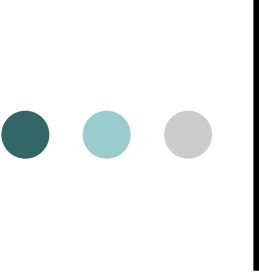
- Suposições
 - amostra aleatória
 - Dados nominais (sexo: M ou F) ou numéricos (idade: menor que 15, 15-24, 25-34, 35-44, 45-54, e maior que 55).
 - Probabilidades ou freqüências das categorias conhecidas
- A idéia é comparar as freqüências observadas com as freqüências esperadas.



Teste Qui-quadrado

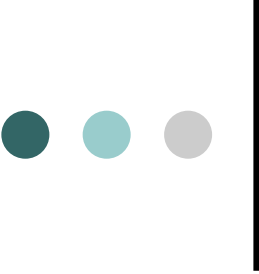
○ Hipóteses

- H_0 : a amostra é selecionada de uma população que segue uma determinada distribuição.
- H_a : a amostra não é selecionada de uma população que segue uma determinada distribuição.



Teste Qui-quadrado

- Estatística de teste
$$\chi^2 = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i}$$
- χ^2 tem uma distribuição aproximadamente qui-quadrado com g.l.=r-g-1 graus de liberdade onde r é o número de categorias e g é número de parâmetros da hipótese nula que foi estimado. Se um valor de χ^2 é maior ou igual que um valor tabelado de qui-quadrado com r-g-1 graus de liberdade com nível de significância α , rejeitamos a hipótese nula.



Exemplo 1

- A administração de um aeroporto queria conferir a alegação de um controlador de voo de que o número de mensagens de rádio recebidas por minuto é uma variável aleatória de distribuição de Poisson com média 1,5. Se alegação for correta isto pode implicar a contratação de mais funcionários. Em uma amostra de 200 intervalos de um minuto são registrados o número de mensagens por minuto.

Solução

$$0,2231 \times 200$$

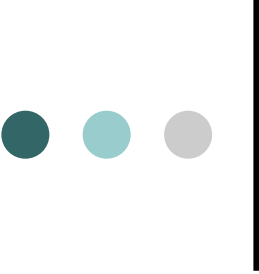
Probabilidade usando
distribuição de Poisson

Número de Mensagens de Rádio	Freqüência Observada
0	70
1	57
2	46
3	20
4	5
5	2

Número de Mensagens de Rádio	Probabilidade
0	0,2231
1	0,3347
2	0,2510
3	0,1255
4	0,0471
5	0,0141
6	0,0035
7	0,0008

Número de Mensagens de Rádio	Freqüência Esperada
0	44,6
1	66,9
2	50,2
3	25,1
4 ou mais	13,1

A freqüência esperada para 5 ou mais é menor que 5 e, portanto muito pequena. Isto é recomendado sempre que temos uma freqüência esperada pequena.

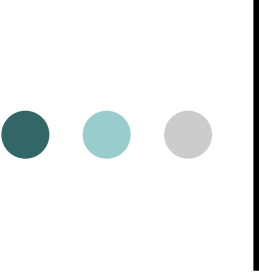


Solução

- Hipóteses:

- H_0 : a população amostrada segue uma distribuição de poisson com média 1,5.
- H_a : a população amostrada não segue uma distribuição de poisson com média 1,5.

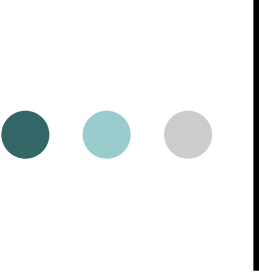
- Temos $r=5$ categorias (termos).



Solução

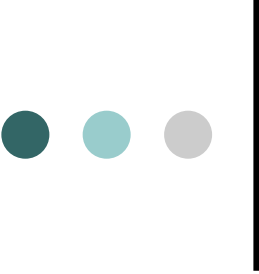
$$\chi^2 = \frac{(70 - 44,6)^2}{44,6} + \frac{(57 - 66,9)^2}{66,9} + \frac{(46 - 50,2)^2}{50,2} \\ + \frac{(20 - 25,1)^2}{25,1} + \frac{(7 - 13,1)^2}{13,1} = 20,2$$

- Como g.l.=5-0-1=4, o valor crítico é $\chi^2_{0,01} = 13,277$. Temos que $\chi^2 = 20,2 > 13,277$ e a hipótese nula deve ser rejeitada.



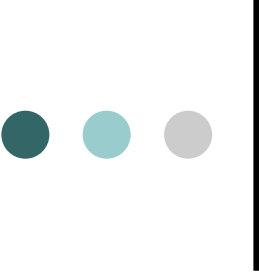
Teste Kolmogorov-Smirnov para uma amostra

- O teste qui-quadrado foi introduzido para dados nominais.
- O teste de Kolmogorov foi introduzido para dados contínuos. Este teste foi introduzido em 1933 por um matemático.
- A idéia é avaliar a função de distribuição acumulada: $F(x)=P(X\leq x)$.
- Suponha que foi selecionada uma amostra aleatória de uma função de distribuição desconhecida $F(x)$. O teste avalia se $F(x)=F_0(x)$ para todo x comparando $F_0(x)$ com $S(x)$ que é obtido da amostra.



Teste Kolmogorov-Smirnov para uma amostra

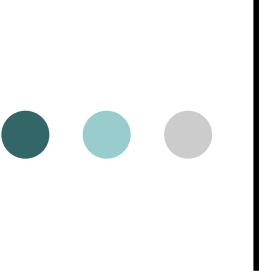
- Hipóteses
 - $H_0: F(x)=F_0(x)$ para todo o valor de x .
 - $H_a: F(x) \neq F_0(x)$ para pelo menos um valor de x .
- Estatística de teste
 - $D = \sup_x |S(x) - F_0(x)|$
- Regra de decisão
 - Rejeita H_0 se as estatísticas de teste D é maior que o quantil $1 - \alpha$ para o teste Kolmogorov-Sminorv



Exemplo 2

- Em um estudo sobre pesos (em gramas) do rim de 36 cachorros vira-lata obteve-se a tabela abaixo. Deseja-se saber se os dados são normalmente distribuídos com média 85gr e desvio padrão 15 gr. Use $\alpha=5\%$.

58	78	84	90	97	70	90	86	82
59	90	70	74	83	90	76	88	84
68	93	70	94	70	110	67	68	75
80	68	82	104	92	112	84	98	80



Exemplo

- $H_0: F(x)=F_0(x)$ para todo o valor de x .
- $H_a: F(x) \neq F_0(x)$ para pelo menos um valor de x .
- $F_0(x)$ é a função distribuição acumulada de uma normal com média 85gr e desvio padrão 15 gr.

Exemplo

Calculo de $S(x)$

x_i	Freq	Freq Ac	$S(x_i)$
58	1	1	0,0278
59	1	2	0,0556
67	1	3	0,0833
68	3	6	0,1667
:	:	:	:
112	1	36	1,000

Calculo de $F_0(x)$

x	z	$P(0 \leq Z \leq z)$	$F_0(x_i)$
58	-1,80	0,4641	$0,0359=0,5-0,4641$
59	-1,73	0,4582	$0,0418=0,5-0,4582$
67	-1,20	0,3849	$0,1151=0,5-0,3849$
68	-1,13	0,3708	$0,1292=0,5-0,3708$
:	:	:	:
112	1,80	0,4641	$0,9641=0,5+0,4641$

A maior distância é encontrada para $x=90$. $D=0,1485$. Contudo pode ser que $D=0,1485$ não seja a maior distância. $D=0,1485$. Usando a tabela de teste Kolmogorov-Sminorv temos que o valor crítico para D é 0,221 e portanto a Hipótese nula não deve ser rejeitada.

Exemplo 3

Considere os dados abaixo, resultantes da observação de 40 valores de uma variável aleatória qualquer X . Deseja-se testar a hipótese de que esta variável aleatória tenha distribuição normal com média μ igual a 3,6 e variância σ^2 igual a 0,8.

2,2	4,1	3,5	4,5	5,0	3,7	3,0	2,6	3,4	1,6
3,1	3,3	3,8	3,1	4,7	3,7	2,5	4,3	4,9	3,6
2,9	3,3	3,9	3,1	4,8	3,1	3,7	4,4	3,2	4,1
1,9	3,4	4,7	3,8	3,0	2,6	3,9	3,0	4,2	3,5

$$H_0 : X \sim N(\mu = 3,6; \sigma^2 = 0,8)$$

$$H_1 : X \sim ?$$



$$H_0 : (X - 3,6)/0,8944 = Z \sim N(0,1)$$

$$H_1 : (X - 3,6)/0,8944 \sim ?$$

Valores padronizados:

-1,57	0,56	-0,11	1,01	1,57	0,11	-0,67	-1,12	-0,22	-2,24
-0,56	-0,34	0,22	-0,56	1,23	0,11	-1,23	0,78	1,45	0,00
-0,78	-0,34	0,34	-0,56	1,34	-0,56	0,11	0,89	-0,45	0,56
-1,90	-0,22	1,23	0,22	-0,67	-1,12	0,34	-0,67	0,67	-0,11

Exemplo 3

Valores padronizados ordenados:

-2,24	-1,9	-1,57	-1,23	-1,12	-1,12	-0,78	-0,67	-0,67	-0,67
-0,56	-0,56	-0,56	-0,56	-0,45	-0,34	-0,34	-0,22	-0,22	-0,11
-0,11	0,00	0,11	0,11	0,11	0,22	0,22	0,34	0,34	0,56
0,56	0,67	0,78	0,89	1,01	1,23	1,23	1,34	1,45	1,57

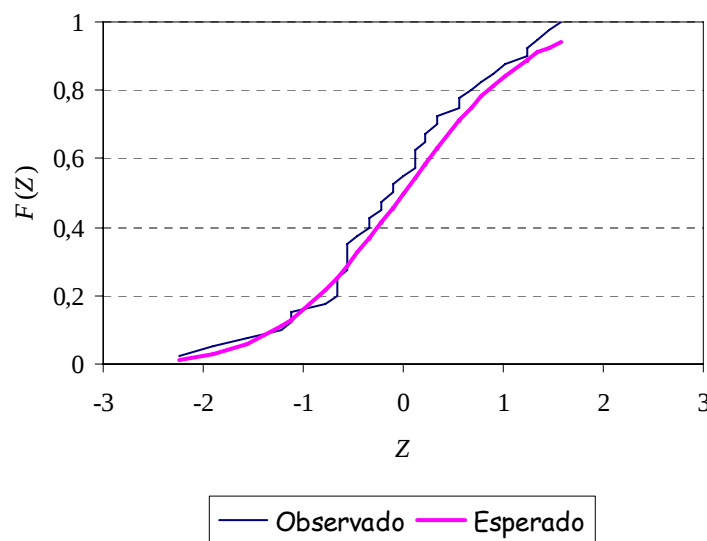
$$F_{obs}(Z_i) = \frac{i}{n} \Rightarrow F_{obs}(-1,57) = \frac{3}{40}$$

$$F_{esp}(Z_i) = P(Z < Z_i) \Rightarrow F_{esp}(Z < -1,57) = 0,0582$$

$$D = \max |F_{obs}(Z_i) - F_{esp}(Z_i)|$$

→ valores críticos tabelados!

$$D = 0,0919 \quad D_{crít} = 0,2150 \quad (\alpha = 5\%)$$



Conclusão: não rejeita a hipótese de que os dados provenham de uma normal, a 5% de significância.

Tabela

Tamanho da amostra (N)	Nível de significância para $D = \max F_{obs}(X) - F_{esp}(X) $				
	0,20	0,15	0,10	0,05	0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
Mais de 35	$\frac{1,07}{\sqrt{N}}$	$\frac{1,14}{\sqrt{N}}$	$\frac{1,22}{\sqrt{N}}$	$\frac{1,36}{\sqrt{N}}$	$\frac{1,63}{\sqrt{N}}$