6 - Coeficiente de Correlação de Spearman

Dispõe-se de uma amostra de n pares de observações (X_i, Y_i), $i = 1, 2 \ldots n$, amostra aleatória de uma v. a. bivariada (X, Y), com distribuição de probabilidades desconhecida, X e Y v. a. contínuas.

Sejam $R(X_i)$ os postos associados a X_i e $R(Y_i)$ os postos associados a Y_i, $i = 1, 2 \ldots n$.

O coeficiente de correlação amostral de Spearman é definido como

$$r_s = \frac{\sum R(X_i) R(Y_i) - n\left(\frac{n+1}{2}\right)^2}{\left[\sum R^2(X_i) - n\left(\frac{n+1}{2}\right)^2\right]^{1/2} \left[\sum R^2(Y_i) - n\left(\frac{n+1}{2}\right)^2\right]^{1/2}}$$

que é o coeficiente de correlação de Pearson calculado para os postos.

Se não houver empates ou um n.º moderado de empates, utilizar

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} \left[R(X_i) - R(Y_i) \right]^2}{n(n^2-1)}.$$

Se houver perfeita correlação positiva, $R(X_i) = R(Y_i)$ e $r_s = 1$.

Se houver perfeita correlação negativa

	1	2	3	4	5	6	7	8	9	10
	10	9	8	7	6	5	4	3	2	1
$\lvert R(X)-R(Y) \rvert$	9	7	5	3	1	1	3	5	7	9

Verifica-se que neste caso, para n qualquer

$$\sum \left[R(X_i) - R(Y_i) \right]^2 = \frac{n(n^2-1)}{3},$$ de modo que

$$r_s = 1 - \frac{6\, n(n^2-1)/3}{n(n^2-1)} = 1 - 2 = -1$$

Teste de Hipóteses

H_0: X e Y são não correlacionadas $(\rho = 0)$

H_a: X e Y são correlacionadas $(\rho \neq 0)$.

Rejeita-se H_0 se $r_s > w_{1-\alpha/2}$ ou $r_s < w_{\alpha/2}$,

w_p é obtido da Tabela A.10, para $p = 0{,}9$, $0{,}95$, $0{,}975$, $0{,}99$, $0{,}995$, $0{,}999$ e $n = 4, 5 \dots 30$,

e $w_p = -w_{1-p}$.

Ex: Deseja-se verificar se existe relação entre o nível de agressividade de gêmeos

Amostra de 12 gêmeos

X_i	86	71	77	68	91	72	77	91	70	71	88	87
Y_i	88	77	76	64	96	72	65	90	65	80	81	72
$R(X_i)$	8	3,5	6,5	1	11,5	5	6,5	11,5	2	3,5	10	9
$R(Y_i)$	10	7	6	1	12	4,5	2,5	11	2,5	8	9	4,5
$[R(X_i)-R(Y_i)]^2$	4	12,25	0,25	0	0,25	0,25	16	0,25	0,25	20,25	1	20,25

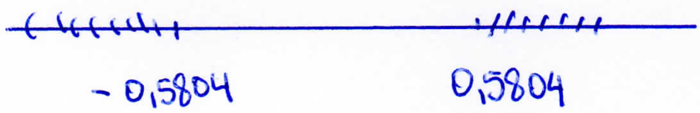
$$r_s = 1 - \frac{6.75}{12.143} = 0{,}7378$$

$H_0: \rho = 0$

$H_a: \rho \neq 0$

Para $\alpha = 0,05$ $\quad n = 12$ $\quad w_{0,975} = 0,5804$

$w_{0,025} = -w_{0,975} = -0,5804$



$r_s > 0,5804$

Rejeita-se H_0.

Para $H_0: \rho = 0$

$H_a: \rho > 0$

$\alpha = 0,05$ $\quad w_{0,95} = 0,4965$

Rejeita-se H_0. Os dados sugerem que existe correlação positiva entre o nível de agressividade dos gêmeos.

7- Testes do tipo Kolmogorov Smirnov

Testes de Aderência

O teste χ^2 de aderência é útil para dados na escala nominal, podendo ser usado para dados em escala superior. No entanto, exigiam grandes tamanhos de amostra, pois a distribuição da estatística de teste sob H_0 é assintoticamente qui-quadrado.

Teste de Aderência de Kolmogorov

Dados: Amostra aleatória $X_1, X_2 \dots X_n$ de uma função distribuição desconhecida $F(x)$.

Seja $F^*(x)$ uma função distribuição completamente especificada.

A) Teste Bicaudal

$H_0: F(x) = F^*(x), \quad -\infty < x < +\infty$

$H_a: F(x) \neq F^*(x)$ para pelo menos um valor de x.

B) Teste Unicaudal

$H_0: F(x) \geq F^*(x) \quad -\infty < x < \infty$

$H_a: F(x) < F^*(x)$ para pelo menos um valor de x

C) Teste Unicaudal

$H_0: F(x) \leq F^*(x) \quad -\infty < x < \infty$

$H_a: F(x) > F^*(x)$ para pelo menos um valor de x

Seja $F_n(x) = \dfrac{n^{\underline{o}} \text{ de obs} \leq x}{n}$ estima $F(x)$

Estatística de teste e regra de decisão:

A) $T_1 = \sup\limits_{x} |F^*(x) - F_n(x)|$

Rejeita-se H_0 ao nível α se $T_1 > w_{1-\alpha}$, onde $w_{1-\alpha}$ é o quantil de ordem $1-\alpha$ dado na Tabela A.13 [Quantis da estatística do teste de Kolmogorov], para o teste bicaudal

B) $T_1^+ = \sup\limits_x \left[F^*(x) - F_n(x) \right]$

Rejeita-se H_0 ao nível α se $T_1^+ > w_{1-\alpha}$, Tabela A.13, teste unicaudal.

C) $T_1^- = \sup\limits_x \left[F_n(x) - F^*(x) \right]$

Rejeita-se H_0 ao nível α se $T_1^- > w_{1-\alpha}$, Tabela A.13, teste unicaudal.

Exemplo

Amostra aleatória de tamanho $n = 10$ da v.a. X

0,621 0,503 0,203 0,477 0,710 0,581 0,329
0,480 0,554 0,382

Deseja-se testar a hipótese que $X \sim \mathcal{U}[0,1]$

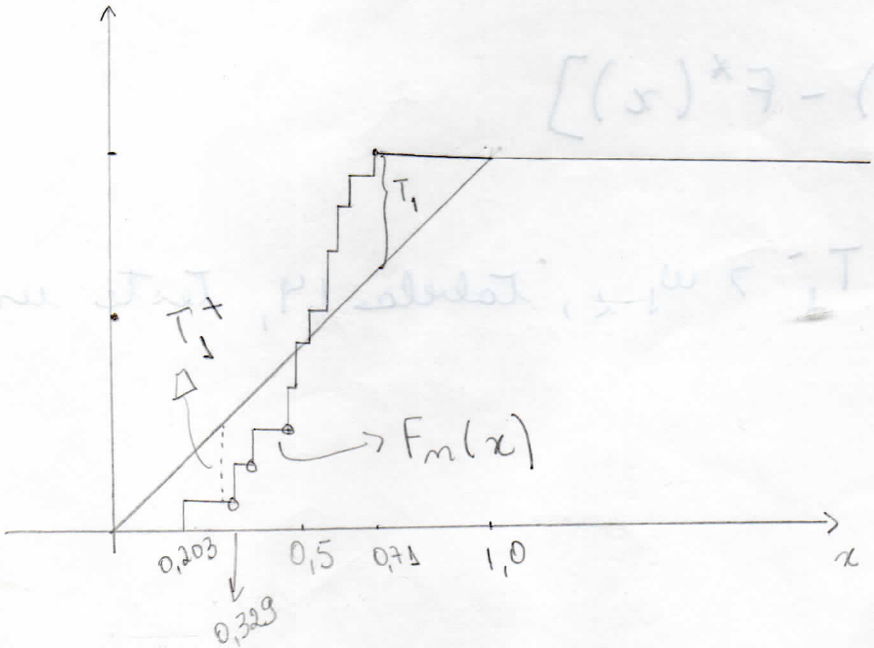
$$F^*(x) = \begin{cases} 0 & \text{se } x < 0 \\ \int_0^x dx = x & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

$\lvert F^* - F_n \rvert$	x	$F_n(x)$	$F^*(x)$
	0,203	0,1	0,203
	0,329	0,2	0,329
	0,382	0,3	0,382
	0,477	0,4	0,477
	0,480	0,5	0,480
	0,503	0,6	0,503
	0,554	0,7	0,554
0,219	0,581	0,8	0,581
0,279	0,621	0,9	0,621
0,290	0,710	1,0	0,710

$H_0: F(x) = F^*(x)$

$H_a: F(x) \neq F^*(x)$

$\alpha = 0{,}05 \qquad n = 10 \qquad w_{1-\alpha} = w_{0,95} = 0{,}409$



$T_1 = \sup_x |F^*(x) - F_n(x)| = |F^*(0{,}71) - F_n(0{,}71)| =$

$= 1 - 0{,}71 = 0{,}29$

$T_1 < 0{,}409$, não rejeitamos H_0

Se quiséssemos testar

$H_0: F(x) \geq F^*(x) \quad \forall x$

$H_a: F(x) < F^*(x)$ para algum x

$$T_1^+ = \sup_x \left[F^*(x) - F_n(x) \right]$$ só procuro onde $F^* > F_n$

T_1^+ será o calculado à esquerda do 2º salto

$$T_1^+ = \lim_{x \to 0,329^-} \left[F^*(x) - F_n(x) \right] = 0,329 - 0,1 = 0,229$$

$0,329$ → 2º elemento da amostra

$T_1^- = T_1$

No geral, $T_1 = \max \{ T_1^+ , T_1^- \}$

Obs:

- O teste de Kolmogorov não pode ser aplicado a algumas situações (por ex., dados na escala nominal) nas quais os testes χ^2 de aderência pode ser utilizado.

- Quando os dois testes puderem ser aplicados, o teste de Kolmogorov é preferível para pequenas amostras porque é um teste exato (mesmo para pequenas amostras) enquanto o teste χ^2 assume grandes amostras.