

Pontos de dicotomização para a obtenção do coeficiente de correlação tetracórico

Sachiko Araki Lira
Instituto Paranaense de Desenvolvimento Econômico e Social – IPARDES
CEP 82630-900 – Santa Cândida – Curitiba – PR
sachiko@onda.com.br

Anselmo Chaves Neto
Universidade Federal do Paraná -PPGMNE/DEST-UFPR
Centro Politécnico – CEP 81531-0990 – Curitiba – PR
anselmo@est.ufpr.br

(Recebido: 27 de dezembro de 2004)

Resumo: Existem situações em que a característica de interesse, tratada como variável, não é medida diretamente embora seja possível observar o seu efeito. Essas características são tratadas como variáveis latentes. Assim, uma pessoa pode ser considerada inteligente ou não, ou ainda ser ansiosa ou não, por exemplo. O Coeficiente de Correlação Tetracórico é um estimador do grau de associação entre duas variáveis latentes X_L e Y_L , subjacentes às variáveis dicotômicas X e Y efetivamente observadas. Uma das suposições para a utilização do Coeficiente de Correlação Tetracórico é a de que os pontos de dicotomização das variáveis medidas X e Y estejam o mais próximo possível das respectivas medianas (percentil 50). Foram geradas amostras através de simulações e utilizados o percentil 25 (P_{25}), percentil 50 (P_{50} = mediana) e percentil 75 (P_{75}) como pontos de dicotomização das variáveis X e Y . Se obteve os desvios absolutos médios, comparando-se os Coeficientes de Correlação Tetracórico e Linear de Pearson. Os resultados comprovam que as estimativas dos coeficientes são melhores, quando se adota a mediana como pontos de dicotomização.

Palabras-chave: correlação tetracórico, correlação de Pearson, coeficiente de correlação.

Abstract: There are situations in which the interest characteristic, treated as a variable, is not directly measured, although it is possibly to observe its effect. These characteristics are treated as latent variables. To exemplify, a person can be considered intelligent or not, or even be anxious or not. The Tetrachoric Correlation Coefficient is an estimative measure of Pearson Linear Coefficient regarding the latent variables X_L and Y_L (both continuous and

normal), underlying the dichotomic variables X and Y really observed. One of the assumptions for using the Tetrachoric Correlation Coefficient is that the dichotomization points of the measured variables X and Y be as close as possible of their respective medians (percentile 50). Samples were generated by simulations and used percentile 25 (P_{25}), percentile 50 (P_{50} = median) and percentile 75 (P_{75}) as X and Y variables dichotomization points. The average absolute deviations were obtained by comparing the Tetrachoric Correlation Coefficient to the Pearson Linear Coefficient. The results showed that the coefficient estimates are better when the median is used as dichotomy points.

Key-words: tetrachoric correlation, Pearson coefficient, correlation coefficient.

1 Introdução

O método usualmente conhecido para estimar a correlação entre duas variáveis é o Coeficiente de Correlação Linear de Pearson, também conhecido como Coeficiente de Correlação do Produto-Momento. Este estimador da correlação, foi estudado por Francis Galton e seu aluno Karl Pearson, em 1897⁸ (SCHULTZ e SCHULTZ, 1992).

Uma das suposições para a utilização deste estimador é a de que as variáveis envolvidas na análise sejam medidas na escala intervalar, além de que o vetor aleatório das variáveis, $[X, Y]$, tenha distribuição normal bivariada com parâmetro de correlação ρ . Entretanto, em muitas características psicológicas, educacionais ou sociológicas, a característica tratada como variável não é medida diretamente, embora ela possa ser facilmente observada indiretamente, por isso são chamadas de variáveis latentes. Assim, tem-se a inteligência, a ansiedade, a satisfação do consumidor, o nível de compreensão de texto, o conhecimento em matemática, entre outras, como exemplos. Para medi-las é necessário criar uma escala de medida nominal que indica a presença ou ausência da característica.

Karl Pearson introduziu o Coeficiente de Correlação Tetracórico (FACHEL, 1986). Esse coeficiente é utilizado, segundo Downie e Heath (1959), McNemar (1969) e Bunchaft e Kellner (1999), para estimar o grau de associação existente entre as variáveis X_L e Y_L contínuas, mas dicotomizadas (ao serem medidas) pelo pesquisador, por alguma razão.

A estimativa obtida com o Coeficiente de Correlação Tetracórico pode ser considerada como uma estimativa do Coeficiente de Correlação Linear de Pearson considerando as variáveis latentes X_L e Y_L (ambas contínuas e normais), subjacentes às variáveis dicotômicas X e Y efetivamente observadas com a

⁸ Esta informação foi obtida no *site*: www.ime.usp.br/~abe/cronologiajaneiro02.doc

dicotomização (LORD e NOVICK, 1967), (FERGUSON, 1976) e (WHERRY, 1984).

Algumas restrições são apresentadas quanto à utilização desse coeficiente. A primeira delas se, refere à existência das variáveis latentes X_L e Y_L , normalmente distribuídas, subjacentes às variáveis observadas X e Y . A segunda é que as variáveis observadas, ao serem medidas, devem ser dicotomizadas o mais próximo possível da mediana.

Este trabalho tem por objetivo mostrar por meio de simulações, os desvios entre os coeficientes de Correlação Linear de Pearson e Tetracórico estimados, utilizando os percentis 25, 50 e 75, como pontos de dicotomização para a obtenção deste último coeficiente.

2 Metodologia

2.1 Coeficiente de correlação linear de Pearson e seu erro padrão

O coeficiente de correlação ρ , entre as componentes do vetor aleatório $[X, Y]$, é um parâmetro da distribuição normal bivariada cuja função densidade de probabilidade é dada por:

$$f_{X,Y}(X, Y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \quad (1)$$

onde ρ é o parâmetro de correlação e μ_X , μ_Y , σ_X e σ_Y são os outros parâmetros do modelo.

Como é bem conhecido, ρ é definido por

$$\rho = \rho_{X,Y} = \frac{\text{COV}(X, Y)}{\sigma_X\sigma_Y} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y} \quad (2)$$

O estimador desse coeficiente é mais conhecido como Coeficiente de Correlação Linear de Pearson ou do Produto-Momento e é idêntico ao Estimador de Máxima Verossimilhança cuja expressão é a seguinte:

$$\hat{\rho}_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

Em que: $\hat{\rho}$ é o coeficiente de correlação linear estimado;

\bar{X} é a média amostral da variável X ;

\bar{Y} é a média amostral da variável Y .

Quando $\rho = 0$, o erro padrão do estimador pode ser obtido pela expressão:

$$\sigma_{\hat{\rho}} = \frac{1}{\sqrt{n-1}} \quad (4)$$

Para testar a hipótese nula $H_0 : \rho = 0$ contra a alternativa $H_1 : \rho \neq 0$, usa-se a estatística (5) que é obtida com base em uma amostra aleatória de tamanho n . A estatística, como se sabe, tem uma distribuição t com $n - 2$ graus de liberdade, ou seja:

$$t = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} \sim t_{n-2} \quad (5)$$

2.2 Estimador do coeficiente de correlação tetracórico e seu erro padrão

A equação tetracórica, conforme apresentada em ELDERTON (1953, p. 175), é dada por:

$$\frac{ad-bc}{yy'n^2} = \hat{\rho}_t + \hat{\rho}_t^2 \frac{zz'}{2} + \hat{\rho}_t^3 \frac{(z^2-1)(z'^2-1)}{6} + \hat{\rho}_t^4 \frac{z(z^2-3)(z'^2-3)}{24} + \hat{\rho}_t^5 \frac{(z^4-6z^2+3)(z'^4-6z'^2+3)}{120} + \hat{\rho}_t^6 \frac{z(z^4-10z^2+15)z'(z'^4-10z'^2+15)}{720} + \hat{\rho}_t^7 \frac{(z^6-15z^4+45z^2-15)(z'^6-15z'^4+45z'^2+15)}{5040} + \dots \quad (6)$$

Em que:

a, b, c, d são as freqüências da tabela de contingência 2x2, conforme apresentada adiante;

n = a + b + c + d é o total de observações da amostra

Z é o escore normal padronizado correspondente à área menor ou igual a p . Por exemplo, se $p = 0,50$, então tem-se que $z = 0$ (tabela de áreas sob a curva normal);

z' é o escore normal padrão correspondente à área menor ou igual a p' . Se $p' = 0,50$, então tem-se que $z' = 0$;

y é o valor da ordenada no ponto p da função $y = f(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$ $z \in \mathbb{R}$. Para o exemplo citado, se $z = 0$, então $y = f(0) = \frac{e^{-0}}{\sqrt{2\pi}} = 0,39894$ (tabela de ordenadas da curva normal);

y' é o valor da ordenada no ponto p' da função $y' = f(z') = \frac{e^{-\frac{z'^2}{2}}}{\sqrt{2\pi}}$ $z \in \mathbb{R}$.

Tabela 1. Dicotomização das variáveis X e Y

Variável X			
	1	0	Total
Variável 1	a	b	a+b
Y 0	c	d	c+d
Total	a+c	b+d	n

Guilford (1950) apresenta uma solução aproximada para o cálculo do Coeficiente de Correlação Tetracórico, ignorando os termos de grau superior a 2, na expressão (6):

$$\frac{ad - bc}{yy'n^2} = \hat{\rho}_t + \hat{\rho}_t^2 \frac{zz'}{2} \quad (7)$$

Em que: $\hat{\rho}_t$ é o Coeficiente de Correlação Tetracórico estimado;

a, b, c, d são as frequências da tabela 2x2;

z é o escore normal padrão correspondente à área menor ou igual a **p**;

z' é o escore normal padrão correspondente à área menor ou igual a **p'**;

y é o valor da ordenada no ponto **p**;

y' é o valor da ordenada no ponto **p'**;

n = a + b + c + d é o total de observações da amostra.

Chamando o primeiro termo da expressão (7) de **c**; o coeficiente de $\hat{\rho}_t$ de **b**; e $\frac{zz'}{2}$ de **a**, pode-se formar a equação do 2.º grau:

$$a\hat{\rho}_t^2 + b\hat{\rho}_t + c = 0 \quad (8)$$

$$\text{que é resolvida em } \hat{\rho}_t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (9)$$

Uma outra solução é apresentada em Guilford (1950), utilizando o cosseno⁹:

⁹Conforme demonstrado em Wonnacott e Wonnacott (1978), existe uma relação entre o Coeficiente de Correlação $\hat{\rho}$ e o $\cos \theta$, $\hat{\rho} = \cos \theta$ e $-1 \leq \cos \theta \leq +1$.

$$\hat{\rho}_t = \cos\left(\frac{180\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}\right) \quad (10)$$

Quando o produto bc é igual a ad , o ângulo é 90° e o cosseno é igual a zero, conseqüentemente $\hat{\rho}_t = 0$.

O erro padrão aproximado desse estimador (Tetracórico) é dado por:

$$\hat{\sigma}_{\hat{\rho}_t} = \frac{\sqrt{p \times q \times p' \times q'}}{y' \times y \times \sqrt{n}} \sqrt{(1 - \hat{\rho}_t^2) \times \left[1 - \left(\frac{\text{sen}^{-1} \hat{\rho}_t}{90^\circ}\right)^2\right]} \quad (11)$$

Em que: $\hat{\sigma}_{\hat{\rho}_t}$ é o erro padrão do estimador do Coeficiente de Correlação Tetracórico;

$\hat{\rho}_t$ é o Coeficiente de Correlação Tetracórico estimado;

$\text{sen}^{-1} \hat{\rho}_t$ é o arco seno de $\hat{\rho}_t$;

$n = a + b + c + d$ é o total de observações da amostra .

Para testar a hipótese de que $\rho_t = 0$, é possível utilizar a estatística $t = \frac{\hat{\rho}_t}{\hat{\sigma}_{\hat{\rho}_t}}$. O

erro padrão poderá ser calculado considerando apenas a primeira parte da expressão (11), como apresenta McNemar (1969):

$$\hat{\sigma}_{\hat{\rho}_t} = \frac{\sqrt{p \times q \times p' \times q'}}{y' \times y \times \sqrt{n}} \quad (12)$$

2.3 Suposições básicas para a utilização do coeficiente de correlação tetracórico

As suposições básicas para a utilização do estimador do Coeficiente de Correlação Tetracórico são as de que as variáveis X_L e Y_L (latentes) devam ser contínuas, normalmente distribuídas, relacionadas linearmente e, ainda, X e Y devem ser dicotomizadas (ao serem medidas) o mais próximo possível à mediana.

O Coeficiente de Correlação Tetracórico é menos confiável (possui erro padrão maior) do que o de Pearson, sendo que sua variabilidade é cerca de 50% maior (GUILFORD, 1950), quando $\rho = 0$. Para obter a mesma confiabilidade¹⁰ para

¹⁰A confiabilidade, aqui, é usada como sinônimo de erro padrão.

o Coeficiente de Correlação Tetracórico que a obtida no Coeficiente de Correlação de Pearson, é necessário o dobro do tamanho da amostra.

É possível obter maiores detalhes sobre o Coeficiente de Correlação Tetracórico em Lira (2004).

2.4 Desvios absolutos

A comparação entre as estimativas dos Coeficientes de Correlação Linear de Pearson e o Tetracórico foi feita utilizando-se diferentes pontos de dicotomização e calculando os desvios absolutos apresentados a seguir:

$$DAP_{25} = \sum_{i=1}^n |\hat{\rho}_{t25_i} - \hat{\rho}_i| \quad (13)$$

Em que: DAP_{25} é o desvio absoluto entre o Coeficiente Linear de Pearson e o Tetracórico utilizando o percentil 25 como ponto de dicotomização;

$\hat{\rho}_{t25_i}$ é o Coeficiente Tetracórico estimado utilizando o percentil 25 como ponto de dicotomização da i -ésima amostra, $i = 1, 2, \dots, n$;

$\hat{\rho}_i$ é o Coeficiente Linear de Pearson estimado da i -ésima amostra, $i = 1, 2, \dots, n$.

$$DAP_{50} = \sum_{i=1}^n |\hat{\rho}_{t50_i} - \hat{\rho}_i| \quad (14)$$

Em que: DAP_{50} é o desvio absoluto entre o Coeficiente Linear de Pearson e o Tetracórico utilizando o percentil 50 como ponto de dicotomização;

$\hat{\rho}_{t50_i}$ é o Coeficiente Tetracórico estimado utilizando o percentil 50 como ponto de dicotomização da i -ésima amostra, $i = 1, 2, \dots, n$;

$\hat{\rho}_i$ é o Coeficiente Linear de Pearson estimado da i -ésima amostra, $i = 1, 2, \dots, n$.

$$DAP_{75} = \sum_{i=1}^n |\hat{\rho}_{t75_i} - \hat{\rho}_i| \quad (15)$$

Em que: DAP_{75} é o desvio absoluto entre o Coeficiente Linear de Pearson e o Tetracórico utilizando o percentil 75 como ponto de dicotomização;

$\hat{\rho}_{t75_i}$ é o Coeficiente Tetracórico estimado utilizando o percentil 75 como ponto de dicotomização da i -ésima amostra, $i = 1, 2, \dots, n$;

$\hat{\rho}_i$ é o Coeficiente Linear de Pearson estimado da i -ésima amostra, $i = 1, 2, \dots, n$.

2.5 Simulação Monte Carlo

Foram geradas amostras das variáveis X e Y , normalmente distribuídas, utilizando-se o programa disponibilizado pelo Statistical Analysis Software (SAS), que se encontra no apêndice. Apresenta-se a seguir o algoritmo adotado para o desenvolvimento do programa.

Sejam U e V , variáveis aleatórias independentes com distribuição normal padrão, ou seja, $U \sim N(0, 1)$ e $V \sim N(0, 1)$ e os seguintes parâmetros: μ_X , μ_Y , σ_X^2 , σ_Y^2 e ρ . As variáveis aleatórias X e Y são definidas por:

$$X = \mu_X + \sqrt{\sigma_X^2} U \quad (16)$$

$$Y = \mu_Y + \rho \left(\frac{\sqrt{\sigma_Y^2}}{\sqrt{\sigma_X^2}} \right) (X - \mu_X) + \sqrt{\sigma_Y^2 (1 - \rho^2)} V \quad (17)$$

Em que: μ_X e $\mu_Y \in \mathbf{R}$; σ_X^2 e $\sigma_Y^2 \in \mathbf{R}^+$; $-1 \leq \rho \leq 1$

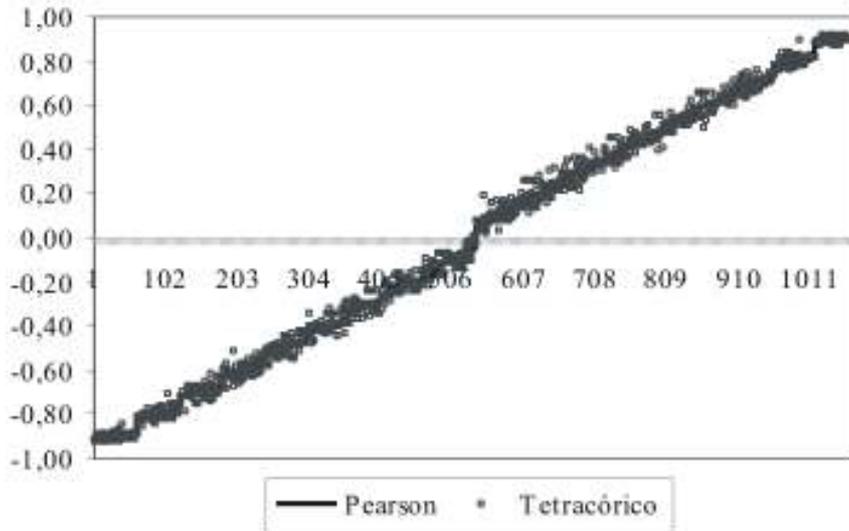
Então tem-se que distribuição conjunta de (X, Y) tem distribuição normal bivariada com parâmetros μ_X , μ_Y , σ_X^2 , σ_Y^2 e ρ .

3 Resultados e Discussão

Foram geradas amostras aleatórias normais bivariadas com coeficientes de correlação ρ (parâmetro) variando entre $-0,90$ e $+0,90$, sendo um total de 60 amostras compostas de 1.000 pares de observações $[X, Y]$. Utilizou-se como pontos de dicotomização para o cálculo do Coeficiente Tetracórico os percentis 25 (P_{25}), 50 (mediana) e 75 (P_{75}). Os gráficos a seguir mostram os coeficientes de Correlação Linear de Pearson e os coeficientes de Correlação Tetracórico, com diferentes pontos de dicotomização.

O gráfico 1, adiante, apresenta a comparação entre os coeficientes Linear de Pearson e o Tetracórico utilizando a mediana (P_{50}), como ponto de dicotomização.

GRÁFICO 1 - COEFICIENTES DE CORRELAÇÃO LINEAR DE PEARSON E TETRACÓRICO

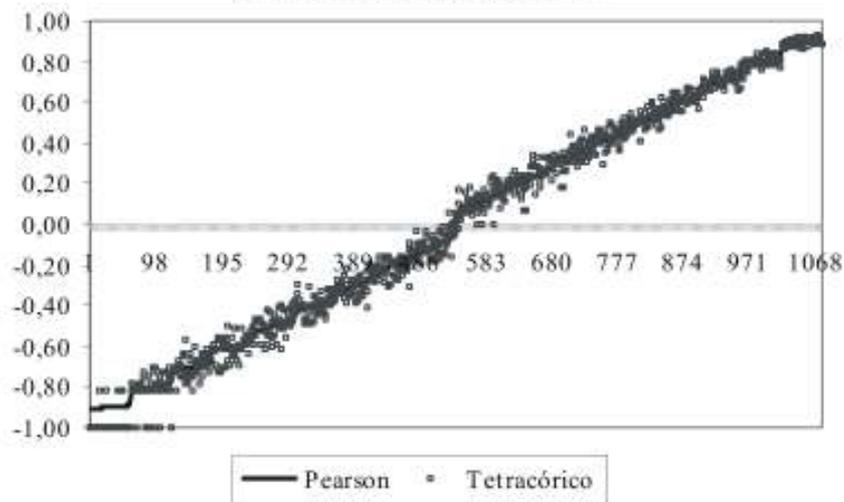


NOTA: Utilizou-se a mediana como ponto de dicotomização.

Fonte: Os autores

O gráfico 2 apresenta os coeficientes de Correlação Linear de Pearson e Tetracórico, tendo o percentil 25 (P_{25}) como ponto de dicotomização.

GRÁFICO 2 - COEFICIENTES DE CORRELAÇÃO LINEAR DE PEARSON E TETRACÓRICO

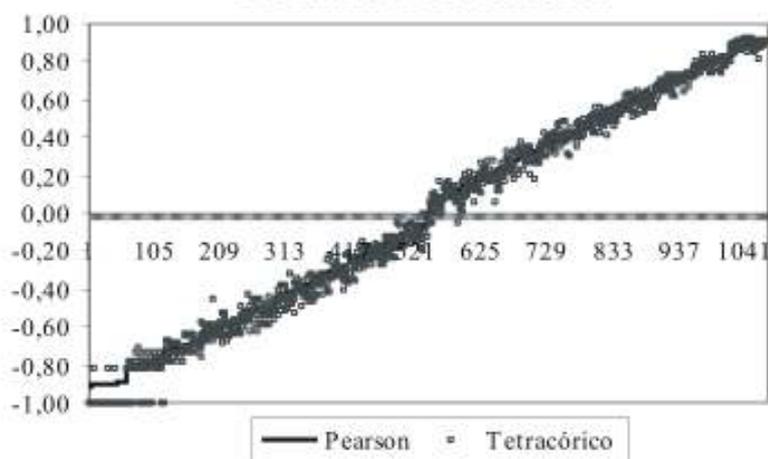


NOTA: Utilizou-se o percentil 25 como ponto de dicotomização.

Fonte: Os autores

Finalmente, o gráfico 3 apresenta os mesmos coeficientes, porém utilizando o percentil 75 (P_{75}) como ponto de dicotomização para o cálculo do Coeficiente de Correlação Tetracórico.

GRÁFICO 3 - COEFICIENTES DE CORRELAÇÃO LINEAR DE PEARSON E TETRACÓRICO



NOTA: Utilizou-se o percentil 75 como ponto de dicotomização.

Fonte: Os autores

O gráfico 1 indica que os coeficientes de Correlação Tetracóricos obtidos, tendo como os pontos de dicotomização a mediana, se aproximam mais do Coeficiente de Correlação de Pearson. Os pontos estão mais concentrados em torno do Coeficiente de Pearson. Já, os gráficos 2 e 3 indicam que, principalmente para valores dos coeficientes negativamente altos (próximos de $-0,90$), os coeficientes Tetracóricos se aproximam de $-1,0$, indicando que tanto o percentil 25 quanto percentil 75 superestimam o Coeficiente Linear de Pearson.

As tabelas, a seguir, apresentam os desvios absolutos médios dos coeficientes Tetracóricos com diferentes pontos de dicotomização, percentil 25, mediana e percentil 75 em relação ao Coeficiente Linear de Pearson.

Tabela 1. Coeficientes Linear de Pearson e Tetracórico, Número de Amostras e Desvios Absolutos Segundo Valores Médio, Máximo e Mínimo.

Valores	Coef. Linear de Pearson	Coeficiente Tetracórico			Número de Amostras	Desvio absoluto P ₂₅	Desvio absoluto P ₅₀	Desvio absoluto P ₇₅
		Ponto de Dicotomização						
		P ₂₅	P ₅₀	P ₇₅				
Médio	-0,8999	-0,9843	-0,8958	-0,9902	60	0,0972	0,0131	0,0981
Máximo	-0,8824	-0,8227	-0,8443	-0,8227				
Mínimo	-0,9103	-0,9990	-0,9251	-0,9990				
Médio	-0,8003	-0,8147	-0,7972	-0,8319	60	0,0523	0,0181	0,0638
Máximo	-0,7687	-0,7088	-0,7027	-0,7088				
Mínimo	-0,8190	-0,9990	-0,8477	-0,9990				
Médio	-0,6995	-0,7046	-0,6965	-0,7115	60	0,0411	0,0226	0,0342
Máximo	0,6340	-0,5711	-0,6179	-0,5711				
Mínimo	-0,7270	-0,8227	-0,7824	-0,7847				
Médio	-0,6006	-0,6036	-0,5965	-0,6089	60	0,0406	0,0264	0,0346
Máximo	-0,5537	-0,5004	-0,5144	-0,4478				
Mínimo	-0,6306	-0,7309	-0,6800	-0,6884				
Médio	-0,5002	-0,4984	-0,4932	-0,5191	60	0,0436	0,0248	0,0420
Máximo	-0,4432	-0,3977	-0,4258	-0,4225				
Mínimo	-0,5394	-0,6177	-0,5827	-0,6342				
Médio	-0,4017	-0,3978	-0,3948	-0,4105	60	0,0381	0,0277	0,0397
Máximo	-0,3501	-0,2913	-0,2800	-0,3143				
Mínimo	-0,4407	-0,4869	-0,4818	-0,5278				
Médio	-0,3008	-0,2999	-0,2967	-0,3204	60	0,0378	0,0271	0,0384
Máximo	-0,2482	-0,1914	-0,1936	-0,2131				
Mínimo	-0,3452	-0,4100	-0,3856	-0,4350				
Médio	-0,2011	-0,2039	-0,1944	-0,2177	60	0,0355	0,0260	0,0361
Máximo	-0,1504	-0,0858	-0,0941	-0,1169				
Mínimo	-0,2465	-0,3028	-0,2548	-0,3495				
Médio	-0,1017	-0,0921	-0,0947	-0,1135	60	0,0404	0,0244	0,0379
Máximo	-0,0446	0,0538	0,0063	-0,0050				
Mínimo	-0,1502	-0,1914	-0,1688	-0,2351				
Médio	0,0962	0,1008	0,1029	0,0820	60	0,0377	0,0292	0,0413
Máximo	0,1447	0,2385	0,1936	0,2029				
Mínimo	0,0272	-0,0249	0,0063	-0,0450				

Tabela 2. Coeficientes Linear de Pearson e Tetracórico, Número de Amostras e Desvios Absolutos Segundo Valores Médio, Máximo e Mínimo.

Valores	Coef. Linear de Pearson	Coeficiente Tetracórico			Número de Amostras	Desvio absoluto P ₂₅	Desvio absoluto P ₅₀	Desvio absoluto P ₇₅
		Ponto de Dicotomização						
		P ₂₅	P ₅₀	P ₇₅				
Médio	0,1964	0,2003	0,2055	0,1830	60	0,0379	0,0300	0,0373
Máximo	0,2459	0,3411	0,3150	0,2820				
Mínimo	0,1448	0,0730	0,1191	0,0634				
Médio	0,2971	0,3016	0,3049	0,2869	60	0,0353	0,0288	0,0366
Máximo	0,3428	0,4452	0,4087	0,3739				
Mínimo	0,2459	0,1849	0,2181	0,1849				
Médio	0,3971	0,3990	0,4052	0,3921	60	0,0352	0,0291	0,0341
Máximo	0,4413	0,5053	0,4927	0,4906				
Mínimo	0,3483	0,2991	0,3209	0,2991				
Médio	0,4975	0,4979	0,5019	0,4900	60	0,0298	0,0247	0,0297
Máximo	0,5400	0,5966	0,5673	0,5625				
Mínimo	0,4502	0,3739	0,4029	0,4060				
Médio	0,5977	0,6013	0,6018	0,5867	60	0,0269	0,0246	0,0294
Máximo	0,6375	0,6735	0,6779	0,6612				
Mínimo	0,5447	0,4756	0,5036	0,5126				
Médio	0,6981	0,7017	0,6945	0,6910	60	0,0253	0,0237	0,0231
Máximo	0,7289	0,7590	0,7584	0,7483				
Mínimo	0,6518	0,6230	0,6079	0,6033				
Médio	0,7987	0,7999	0,8000	0,7927	60	0,0192	0,0187	0,0203
Máximo	0,8192	0,8539	0,8990	0,8411				
Mínimo	0,7665	0,7205	0,7351	0,7373				
Médio	0,8994	0,8971	0,8990	0,8929	60	0,0128	0,0097	0,0164
Máximo	0,9093	0,9288	0,9227	0,9317				
Mínimo	0,8834	0,8580	0,8733	0,8197				

Fonte: Os autores

4 Conclusão

Por meio da análise efetuada, foi possível concluir que, de fato, obtém-se a melhor estimativa do Coeficiente Linear de Pearson, através do Coeficiente de Correlação Tetracórico, quando se utiliza a mediana como ponto de dicotomização das variáveis normais bivariadas. Conforme apresentados na tabela 1, os desvios

médios dos coeficientes Tetracórico utilizando a mediana como ponto de dicotomização (desvio P_{50}) são sempre menores quando comparados aos desvios P_{25} e P_{75} . Ainda, é possível observar que para valores altos dos coeficientes linear de Pearson, sejam negativos ou positivos ($\pm 0,90$), os desvios médios são menores.

Conclui-se, pelos resultados das simulações, que a utilização da mediana, como de ponto de dicotomização para o cálculo do Coeficiente Tetracórico, resulta em melhores estimativas do Coeficiente Linear de Pearson.

5 Referências

- BUNCHAFT, Guenia; KELLNER, Sheilah R.O. *Estatística sem mistérios*. 2.ed. Petrópolis: Vozes, 1999. v.2, 303p.
- DOWNIE, N. M.; HEATH, R. W. *Basic statistical methods*. New York: Harper & Brothers, 1959. 289p.
- ELDERTON, William P. *Frequency curves and correlation*. 4.ed. Washington: Harren Press, 1953. 272p.
- FACHEL, Jandyra M. G. *The C-type distribution as an underlying model for categorical data and its use in factor analysis*. London, 1986. 235p. Tese (Doutorado).
- FERGUSON, G. A. *Statistical analysis in psychology and education*. Tokyo: McGraw-Hill Kogagusha, 1976.
- GUILFORD, J. P. *Fundamental statistics in psychology and education*. 4.ed. New York: McGraw-hill Book, 1950. 605p.
- LIRA, Sachiko A. *Análise de correlação: abordagem teórica e de construção dos coeficientes com aplicações*. Curitiba, 2004. 196p. Dissertação (mestrado). Setores de Ciências Exatas e de Tecnologia, UFPR.
- LORD, F. e NOVICK, M. R. *Statistical theories of mental test scores*. Reading: Addison-Wesley, 1967. 568p.
- McNEMAR, Quinn. *Psychological statistics*. 4. ed. New York: J. Wiley & Sons, 1969. 529p.
- SCHULTZ, Duane P.; SCHULTZ, Sydney Ellen. *História da psicologia moderna*. 16. ed. São Paulo: Cultrix, 1992. 439 p.
- WHERRY, R. J. *Contributions to correlational analysis*. Orlando: Academic Press, 1984. 463p.
- WONNACOTT, Ronald J.; WONNACOTT, Thomas H. *Econometria*. 2. ed. São Paulo: Livros Técnicos e Científicos, 1978. 424 p.

Apêndice

Programa para gerar as variáveis normais X e Y

O programa que deu origem a este, foi obtido no site:
[www.sas.com_>service and support_>technical support_](http://www.sas.com/_>service_and_support_/technical_support_)

```
data arq; keep x y;  
  m1=30; m2=80; v1=15; v2=35; ro=0.90;  
do i=1 to 1000; /* tamanho da amostra */  
  x=m1+sqrt(v1)*rannor(12765);  
  y=(m2+ro*(sqrt(v2)/sqrt(v1))*(x-m1))+ sqrt(v2*(1-ro**2))*rannor(123);  
  output;  
end;  
run;
```