Variáveis aleatórias contínuas

FAE

Rodrigo Hermont Ozon

Banca de admissão de professor
FAE Business School

04 de novembro de 2019

Conteúdo

Plano de Aula

Pré-Requisitos de conhecimento: Conhecimento prévio de cálculo de esperança, variância e desvio-padrão para variáveis discretas. Para as contínuas será necessário um conhecimento prévio de cálculo de integrais definidas.

Por quê distinguir variáveis aleatórias e contínuas e momentos da distribuição (valor esperado e variância, desvio-padrão) ?

Objetivo da aula de hoje:

- Definir variáveis aleatórias contínuas e discretas;
- Como medimos esperança (valor médio), variância e desvio-padrão para variáveis discretas e como isso pode ser extendido para as contínuas;
- Por quê utilizamos cálculo integral para as contínuas ?
- Como calculamos valor esperado, variância e desvio-padrão para as contínuas ?
- Fazer alguns exercícios de *cases* reais demonstrando a aplicabilidade dos conceitos ensinados;
- Introdução para cálculos de momentos para FDPs de distribuições contínuas conhecidas (próximas aulas)

O material para exercícios dos exemplos usados na aula de hoje e exercícios propostos estão disponíveis no meu repositório do GitHub:

Acesse e baixe os arquivos de dados no formato .xls e também essa apresentação no formato LaTeX (para os nerds de plantão)

Estatística é muitíssimo interessante e nada complicada, pois hoje temos os computadores a nosso favor!!!

Definido Variáveis Aleatórias

Uma **variável aleatória** é aquela cujos valores estão associadas a probabilidades (ou seja razão de chance).

- **Variável Aleatória Discreta** $\Rightarrow$ Ela pode assumir um número finito de diferentes valores dentro de um intervalo finito. *"Qualquer função definida sobre o espaço amostral* Ω *que atribui um valor real a cada elemento do espaço amostral."*
- **Variável Aleatória Contínua** $\Rightarrow$ É aquela que pode assumir um número infinito de diferentes valores dentro de um intervalo finito. Dizemos que x é uma v.a. contínua que poderá assumir qualquer valor real entre seus valores de mínimo e máximo. *"Seja* X *uma variável aleatória. Suponha que o contradomínio* ($\mathbb{R}_x$) *de* X *seja um intervalo ou uma coleção de intervalos. Então diremos que* X *é uma variável aleatória contínua."*

Variável aleatória discreta x contínua

Exemplos de variáveis aleatórias discretas:

- Lançamento de uma moeda ou um dado
- Contagem do número de pessoas de uma família
- Quantidade de alunos aprovados em econometria

Exemplos de variáveis aleatórias contínuas:

- As estaturas dos participantes aqui
- Cotação de preços de ações
- Distância percorrida

Formalização matemática

Propriedades da Discreta:
Seja X uma v.a. discreta e sejam $\{x_1, x_2, x_3, ...x_n\}$ os valores que esta variável pode tomar. A função de probabilidade de X, é uma função

$$fdp(x_i) = p(X = x_i)$$

que permite calcular as probabilidades de todos os valores de X, tal que:

- $fdp(x_i) \geq 0 \quad \forall x \in \mathbb{R}$ pois $p(\text{Evento}) = \frac{n(\text{Evento})}{n(\Omega)}$
- $\sum_x = fdp(x) = 1$ (a soma de todas as probabilidades ou frequências tem que fechar em 100%)

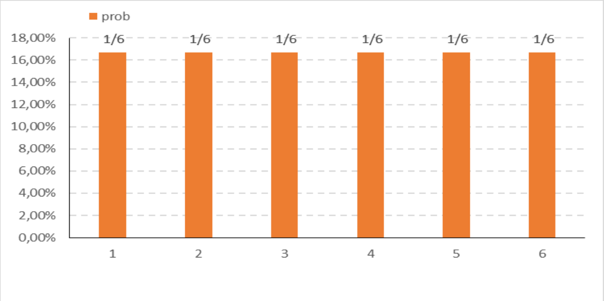
Exemplo 1:
Em um lançamento de um dado quais as probabilidades (frequências) que podem ocorrer de cair em cada uma das faces possíveis ?

face	**prob. (espaço amostral Ω)**
1	1/6=16,67%
2	1/6=16,67%
3	1/6=16,67%
4	1/6=16,67%
5	1/6=16,67%
6	1/6=16,67%

Como 1/6=16,67% a chance de cair o dado com face 1 ou 2 será dada pela soma 16,67%+16,67%=33,33%

***Pergunta:** Qual a distribuição de probabilidade que melhor descreve o lançamento de um dado ?

Histograma da função densidade de probabilidade do lançamento de um dado (não-viciado):



*Calcule a variância de face do dado $= \{1, 2, 3, 4, 5, 6\}$ e compare com o valor esperado (esperança) dele. (Lembre-se que Variância amostral da discreta $= var(X) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$. Explique com o resultado obtido, porque isso ocorre.(Verifique que a média também será igual a $E(x) = \sum_{i=1}^{n} x_i prob_{xi}$)

(Hoffmann, 2006, p. 8 - 9) [2] :

Denominamos frequência relativa do evento A ao quociente entre o número de vezes em que A ocorreu (n_A) e o número total de eventos observados (n).

Podemos, então, definir probabilidade de A como o limite da frequência relativa de A, quando o número de eventos observados tende ao infinito, isto é:

$$p(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

O conceito de probabilidade como limite de uma frequência relativa é bastante útil na prática. As casas de jogos e as companhias de seguros dependem da estabilidade, a longo prazo, de frequências relativas. Entretanto, a definição não é satisfatória do ponto de vista matemático, porque não é possível uma interpretação rigorosa sem usar o próprio conceito de probabilidade.

Afirmações como "a probabilidade de que haja uma guerra nuclear no próximo ano é 0,05", que envolvem uma probabilidade subjetiva, não podem ser interpretadas como limites de frequências relativas. *[grifo meu, sic]*

Exemplo 2: Um economista aplicou um questionário via formulários do Google para pesquisar o nível de renda dos clientes de uma empresa pela qual ele presta consultoria. Para avaliar se os dados coletados são coerentes com suas expectativas e se existe um padrão estatístico (distribuição de probabilidades das respostas) ele deverá identificar as características da distribuição que gera esses dados e para isso precisará calcular o valor médio, variância e desvio-padrão...

Renda (em Salários Mínimos)	**Freq. Absoluta**
5 a 10	5
10 a 15	8
15 a 20	4
20 a 25	3

(Lembrando que $E(X) = \sum_{i=1}^{n} p(x) \cdot$ ponto médio e a $Var(X) = E(X^2) - [E(X)]^2$ sendo o desvio padrão a raiz quadrada da variância). Você também poderia avaliar a assimetria e curtose dessa distribuição

***Pergunta**: É possível encontrar a probabilidade (freq. relativa) de um indivíduo ganhar específicamente 5,3 salários mínimos ?

(Sartoris, 2013 p. 3) [1]

Sendo o conjunto X definido por $X = \{x \in \mathbb{R}\,|0 < x < 2\}$, *qual a probabilidade de, ao sortearmos um número qualquer deste conjunto este número pertença ao intervalo* $[0,5;1,5]$ *?*

O conjunto X é um conjunto contínuo, já que contém todos os números reais que sejam maiores do que 0 e menores do que 2.

Tem, por exemplo, o número 1; o número 0,5; o número 0,4; mas também tem o 0,45; o 0,475; o 0,46. Dados dois elementos deste conjunto, sempre é possível encontrar um número que esteja entre estes dois. Não há "saltos" ou "buracos", daí a idéia de continuidade. ***Ao contrário do dado em que os valores possíveis são 1, 2, 3, 4, 5 e 6 (não existe 1,5 ou 2,1), que é um conjunto discreto.***

Neste caso, a probabilidade de sortearmos qualquer número entre 0,5 e 1,5 (inclusive), que é um intervalo de comprimento igual a 1 (= 1,5 – 0,5), de um intervalo possível que tem comprimento igual a 2 (= 2 – 0) será dada por: $p(0,5 \leq x \leq 1,5) = 1/2$ *(ou seja, 1 chance em 2 possíveis)*

E qual a probabilidade deste número ser exatamente igual a 1 ?
*Ou seja, de sortear um único número entre um total de números presente no conjunto X de... **infinitos!** A probabilidade será dada, então por:*

$$p(x = 1) = \lim_{n \to \infty} \frac{1}{n} = 0$$

*Portanto, embora seja possível de ocorrer, a probabilidade de ser igual a 1 (ou igual a qualquer número) é igual a zero, se estivermos falando de um **conjunto contínuo**. A probabilidade só será diferente de zero se estivermos falando de um intervalo contido neste conjunto.*
Como consequência disso, não fará diferença se o intervalo para o qual encontramos inicialmente a probabilidade (entre 0,5 e 1,5) fosse fechado ou aberto (isto é, incluísse ou não os extremos), pois a probabilidade de ser exatamente 0,5 ou 1,5 é zero. Portanto, como X é um conjunto contínuo:

$$p(0,5 \leq x \leq 1,5) = p(0,5 < x < 1,5) = \frac{1}{2}$$

Integrando x Somando para variáveis aleatórias

O uso da integral...

Um pesquisador deseja medir a profundidade de um rio em locais aleatórios, onde x será a medida nesse local específico. Sabemos que existem diferentes pontos do rio onde a profundidade varia. Assim o pesquisador precisa delimitar uma faixa (área) para um provável ponto A (menos profundo) e ponto B (mais profundo). O espaço delimitado pela área compreendida entre A e B será a variável aleatória x de interesse.

Ele sabe que a área delimitada por essa curva tem que ser igual a 1 (100%) e isso lhe concederá sua função densidade de probabilidade, portanto temos:

$$p(a \leq x \leq b) = \int_a^b f(x)dx = 1$$

Após o pesquisador tirar várias faixas de área de igual amplitude (mesmo tamanho), ele estabelece um ponto médio ($x = x_i = (b - a)/n$) para cada uma das áreas e agora ele pensa em calcular uma média desses valores:

$$\sum_{i=1}^{n} x_i p_i \approx \sum_{i=1}^{n} \text{ponto médio} f(x) h$$

Ele sabe que quanto menor for a amplitude ($h = x_n - x_i$) mais precisa será sua média, portanto ele acrescenta um limite:

$$\lim_{n \to \infty} \sum_{i=1}^{n} \text{ponto médio} f(x) h$$

E se o limite existir então integramos para obter:

$$\int_a^b \text{ponto médio} f(x) dx$$

Onde dx substituiria o h da amplitude.
Assim ele descobre como chegar no conceito de valor médio de uma variável aleatória contínua...

Propriedades da Contínua

Seja X uma variável aleatória tal que:

- $fdp(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $fdp(x)$ tem no máximo um número finito de descontinuidades em qualquer subintervalo finito de $\mathbb{R}$.
- Seja $A \subset \mathbb{R}$, a probabilidade (p) de $p(x \in A) = p(A) = \int_A fdp(x)dx$
- $$\int_{-\infty}^{+\infty} fdp(x)dx = 1$$
- $$\int_a^b fdp(x)dx = p(a \leq x \leq b)$$
- $\int_a^\infty fdp(x)dx = \lim_{k\to\infty} \int_a^k fdp(x)dx$ ou $\int_{-\infty}^b fdp(x)dx = \lim_{k\to\infty} \int_k^b fdp(x)dx$ desde que o limite seja finito.

Esperança (valor médio), variância e desvio-padrão

Uma distribuição de probabilidade pode ser muitas vezes resumida em termos de algumas características, conhecidas como ***momentos*** *da distribuição. Dois dos momentos mais utilizados são a* ***média****, ou* ***valor esperado****, e a* ***variância****.* (GUJARATI, 2000 p. 769) [3]

Esperança, valor esperado e valor médio são a mesma coisa ?
Como medir a esperança de uma discreta e uma contínua ?
Qual a probabilidade de obtermos um valor específico numa amostra de dados discretos e contínuos ?

- Na discreta é dada por $p(x) = f(x_i) = \dfrac{x_i}{\sum_{i=1}^{n}}$
- Na contínua, como vimos é nula, pois $p(X = X_0) = 0$

A esperança ou valor esperado é a média ponderada dos valores que a variável aleatória ou função assume, usando-se, como pesos para ponderação, as probabilidades correspondentes a cada valor.

$$\boxed{\text{E(X)}=\sum_x x \cdot p(x) \quad \text{se v.a. discreta}}$$

$$\boxed{\text{E(X)}=\int_{-\infty}^{\infty} x \cdot f(x)dx \quad \text{se v.a. contínua}}$$

Nessa definição supõe-se que somatório e a integral convergem. Em caso contrário dizemos que o valor esperado da variável aleatória X não existe. *"A única diferença entre este caso e o valor esperado de uma va discreta é que substituímos o símbolo do somatório pelo símbolo da integral."* (Gujarati, 2000 p. 770) [3]

1. O valor esperado de uma constante é a própria constante. Assim, se b for uma constante, $E(b) = b$

[1] Na Teoria das Probabilidades, duas variáveis aleatórias são independentes quando a ocorrência duma não é influenciada pel[illegible] a da outra.

Propriedades dos valores esperados

1. O valor esperado de uma constante é a própria constante. Assim, se b for uma constante, $E(b) = b$
2. Se a e b forem constantes,

$$E(aX + b) = aE(X) + b$$

[1] Na Teoria das Probabilidades, duas variáveis aleatórias são independentes quando a ocorrência duma não é influenciada pel[illegible] a da outra.

Propriedades dos valores esperados

1. O valor esperado de uma constante é a própria constante. Assim, se b for uma constante, $E(b) = b$
2. Se a e b forem constantes,

$$E(aX + b) = aE(X) + b$$

3. Se X e Y forem variáveis aleatórias independentes[1], então

$$E(XY) = E(X)E(Y)$$

[1] Na Teoria das Probabilidades, duas variáveis aleatórias são independentes quando a ocorrência duma não é influenciada pel[illegible]a da outra.

1. O valor esperado de uma constante é a própria constante. Assim, se b for uma constante, $E(b) = b$
2. Se a e b forem constantes,

$$E(aX + b) = aE(X) + b$$

3. Se X e Y forem variáveis aleatórias independentes[1], então

$$E(XY) = E(X)E(Y)$$

4. Se X for uma variável aleatória com FDP denotada por $f(x)$ e se $g(X)$ for uma função qualquer de X, então

$$E[g(X)] = \sum_x g(X) \cdot f(x) \cdot \quad \text{se x for uma va discreta}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) \cdot x \cdot f(x) dx \quad \text{se for contínua}$$

[1]Na Teoria das Probabilidades, duas variáveis aleatórias são independentes quando a ocorrência duma não é influenciada pela [...]a da outra.

Considere a seguinte função densidade de probabilidade:

$$f(x) = \frac{1}{9}x^2; \quad 0 \leq x \leq 3$$

Obtenha seu valor esperado.
Resolução: Primeiramente avaliamos se $f(x)$ é uma FDP

$$fdp(x) = \int_0^3 \frac{1}{9}x^2 dx = 1$$

Resolvendo a integral temos:

$$\frac{1}{9} \cdot \int_0^3 x^2 dx \Rightarrow \frac{1}{9}\left[\frac{x^{2+1}}{2+1}\right]_0^3 \Rightarrow \frac{1}{9}\left[\frac{x^3}{3}\right]_0^3 \Rightarrow \frac{1}{9} \cdot \left[\frac{3^3}{3}\right] - \frac{1}{9} \cdot \left[\frac{0^3}{3}\right] \Rightarrow \frac{1}{9} \cdot 9 \Rightarrow 1$$

ou seja, a $f(x)$ é uma FDP!
Se quisermos avaliar a FDP entre 0 e 1, fazemos

$$\int_0^1 \frac{1}{9}x^2 dx = \frac{1}{9} \cdot \left[\frac{x^3}{3}\right]_0^1 = 1/27 = 3,7\%$$

Ou seja, a probabilidade de x se encontrar entre 0 e 1 é de 1/27

Agora, para obtermos o $\boxed{\text{valor esperado}}$ da FDP contínua fazemos:

$$E(X) = \int_0^3 x\left(\frac{1}{9}x^2\right)$$

$$\frac{1}{9}\cdot\int_0^3 xx^2 = \frac{1}{9}\cdot\int_0^3 x^3dx$$

$$= \frac{1}{9}\left[\frac{x^4}{4}\right]_0^3 = \frac{1}{9}\cdot\frac{3^4}{4} - \frac{1}{9}\cdot\frac{0^4}{4} = \frac{9}{4} = \boxed{2{,}25}$$

Prove que o valor esperado da função contínua a seguir não existe:

$$f(x) = 5x \quad 0 \leq x \leq 2$$

Por qual motivo ? Justifique sua resposta.

Exemplo 3

Um economista deseja construir um modelo de projeção dos preços futuros da opção de uma ação. Ele sabe que os valores (em centavos) das cotas da opção variaram entre 0 e 1 dólar durante um bom tempo e essa é a expectativa do mercado. Ele estima um modelo econométrico para a função da curva de uma faixa de tempo e obtém a seguinte função:

$$\widehat{Y} = f(x) = 1,5(1-x)^2 \quad \text{com os valores oscilando numa faixa entre 0 e 1}$$

Ele deseja saber qual o valor médio (provável) dos valores da cotação se encontrarem na faixa entre 0 e 1 dólar...

1. Verifica se $f(x)$ é uma FDP igualando a 1:

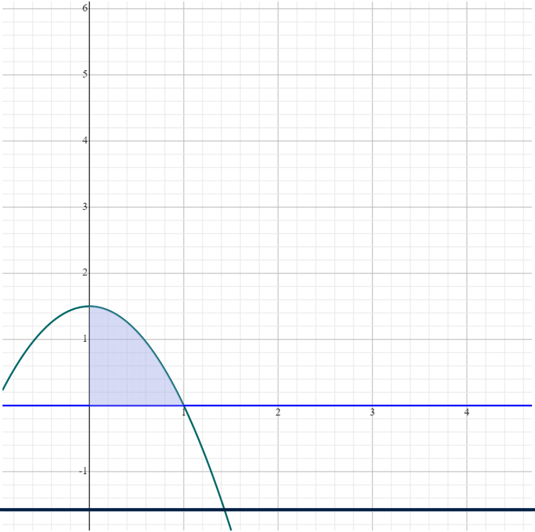
$$f(x) = \int_0^1 1.5\left(1 - x^2\right)dx = 1$$

$$= 1.5 \cdot \int_0^1 1 - x^2 dx \Rightarrow 1.5\left(\int_0^1 1dx - \int_0^1 x^2 dx\right)$$

Integral de uma constante $\int adx = ax$

$$= 1.5\left([x]_0^1 - \left[\frac{x^3}{3}\right]_0^1\right) \Rightarrow 1.5\left(1 - \frac{1}{3}\right) \Rightarrow 1 \quad \text{logo a f(x) é uma FDP!}$$

Área sobre a curva da FDP: $f(x) = \int_0^1 1.5\left(1 - x^2\right) dx = 1$



2. Como f(x) é uma FDP podemos obter o valor médio simplesmente inserindo x antes da função dentro da integral definida:

$$f(x) = \int_0^1 x1.5\left(1 - x^2\right)dx = 1$$

$$= 1.5 \cdot \int_0^1 x\,(1 - x)^2\,dx$$

Aplicando a fórmula do quadrado perfeito

$$(a - b)^2 = a^2 - 2ab + b^2 \Rightarrow 1^2 - 2 \cdot\ 1 \cdot\ x + x^2 \Rightarrow x\left(x^2 - 2x + 1\right)$$

$$= x^3 - 2x^2 + x$$

Aplicando a regra da soma:

$$= 1.5\left(\int_0^1 xdx - \int_0^1 2x^2dx + \int_0^1 x^3dx\right) \Rightarrow 1.5\left(\left[\frac{x^2}{2}\right]_0^1 - \left[2\frac{x^3}{3}\right]_0^1 + \left[\frac{x^4}{4}\right]_0^1\right)$$

Resolução (continuação)

$$1.5\left(\left[\frac{x^2}{2}\right]_0^1 - \left[2\frac{x^3}{3}\right]_0^1 + \left[\frac{x^4}{4}\right]_0^1\right) =$$

$$= 1.5\left(\frac{1}{2} - \frac{2}{3} + \frac{1}{4}\right) \Rightarrow 0,125$$

Ou seja, o valor médio correspondente a um intervalo de probabilidade compreendido entre 0 e 1 para a oscilação do preço da opção é de 0,125 centavos de dólar.

Seja X uma variável aleatória e $E(X) = \mu$. A distribuição ou dispersão, dos valores X em torno do valor esperado pode ser medida pela variância, que é definida como:

$$var(X) = \sigma_X^2 = E(X - \mu)^2$$

A raiz quadrada positiva de σ_X^2,σ_X é definida como desvio-padrão de X. A variância ou desvio-padrão dão indicação de quão próximo ou dispersamente os valores X individuais se espalham em torno do seu valor médio. A variância definida anteriormente é calculada como segue:

$$var(X) = \sum_x (X - \mu)^2 f(x) \quad \text{se X for uma va discreta}$$

$$var(X) = \int_{-\infty}^{\infty} (X - \mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx \quad \text{se X for uma va contínua}$$

Por conveniência de cálculo, a fórmula da variância dada anteriormente também pode ser expressa como

$$var(X) = \sigma_X^2 = E(X - \mu)^2 \Rightarrow E(X^2) - \mu^2 \Rightarrow var(X) = E(X^2) - [E(X)]^2$$

Propriedades da variância

1. $E(X-\mu)^2 = E(X^2) - \mu^2$
2. a variância de uma constante é zero
3. Se a e b forem constantes, então

$$var(aX + b) = a^2 var(X)$$

4. Se X e Y forem variáveis aleatórias independentes, então

$$var(X + Y) = var(X) + var(Y)$$

$$var(X - Y) = var(X) + var(Y)$$

5. Se X e Y forem va independentes e a e b forem constantes, então

$$var(aX + bY) = a^2 var(X) + b^2 var(Y)$$

Exemplo 4 Considere a seguinte FDP:

x	$f(x)$
-2	0,625
1	0,125
2	0,25

Para sabermos a variância dessa discreta fazemos:

$$var(X) = E(X^2) - [E(X)]^2$$

$$E(X) = -2 \cdot 0,625 + 1 \cdot 0,125 + 2 \cdot 0,25 = -0,625$$

Então temos $E(X) = -0625$ e $E(X^2) = 3,625$ logo
$var(X) = 3,625 - (-0,625)^2 \Rightarrow \boxed{3{,}23}$

x	**f(x)**	**x$\hat{2}$**
-2	0,625	4
1	0,125	1
2	0,25	4

Como o desvio-padrão é dado pela raiz quadrada da variância obtemos
$\sqrt{var(X)} \Rightarrow \sqrt{3,23} \Rightarrow \boxed{\sigma_X = 2,06}$

Obtenha a variância para a função densidade:

$$f(x) = \frac{1}{9}x^2; \qquad 0 \leq x \leq 3$$

Solução:
Como vimos no Exemplo 26 que essa função é uma FDP, pois o valor da integral se iguala a 1, partimos aplicando a fórmula:
$$Var(X) = \int_a^b x^2 f(x)dx$$
Obtemos $var(X) = \int_0^3 x^2 \left(\frac{1}{9}x^2\right) dx$ temos que:

$$var(X) = \frac{1}{9} \cdot \int_0^3 x^2 x^2 dx \Rightarrow \frac{1}{9} \cdot \int_0^3 x^4 dx \Rightarrow \frac{1}{9}\left[\frac{x^{4+1}}{4+1}\right]_0^3 \Rightarrow \frac{1}{9}\left[\frac{x^5}{5}\right]_0^3$$

$$var(X) = \frac{1}{9} \cdot \frac{243}{5} \Rightarrow \frac{27}{5} \Rightarrow \boxed{5{,}4}$$

Clique no link abaixo e baixe o e-book com as páginas e exercícios indicados a seguir:

Morettin & Bussab, Estatística Básica. 6a edição, São Paulo, 2010.

- pág 167 (Fazer os exercícios 1, 3 e 4) (encontre as FDP)
- pág 172 (Fazer o exercício 9)
- pág 173 (Fazer os exercícios 10 e 12)

Esta apresentação (no formato $\LaTeX$) e o material recomendado está disponível no meu repositório do GitHub

Você também poderá acessar meu CV e portfólio em:

https://rhozon.github.io/

- Meu repositório no GitHub
- Meu perfil no LinkeDin
- Meu Currículo na Plataforma Lattes
- Minha página no Tableau
- me mande um zapzap ... ou então me envie um e-mail

Mutíssimo obrigado a todos! ☺

Referências

SARTORIS, A. **Estatística e Introdução à Econometria**. Quarta edição. São Paulo, 2013

HOFFMANN, R. **Estatística para economistas**. Quarta edição, São Paulo, 2006.

GUJARATI, D.N., PORTER, D.N. **Econometria Básica.** Quinta Edição, São Paulo, 2011. Disponível em github.com/rhozon/