Revisão de probabilidade

Variáveis aleatórias, funções de probab., esperança, covariância e correlação

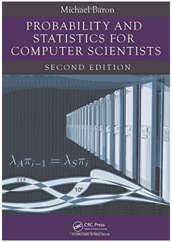
André Gustavo dos Santos

Departamento de Informática
Universidade Federal de Viçosa

INF222 - 2022/2

– Fonte do material

O conteúdo e as figuras são do livro complementar da disciplina (cap. 3):



Baron, Michael.
Probability and statistics for computer scientists.
Chapman and Hall/CRC, 2013.

Tópicos da aula

Variável aleatória

Definição

Uma variável aleatória é uma função de um resultado,

$$X = f(\omega)$$

ou seja, é uma quantidade que depende de um fator aleatório.

Lançamento de dado

- O resultado do lançamento de um dado é uma variável aleatória.
- Sabemos os possíveis resultados, um número inteiro entre 1 e 6, mas o resultado em si depende da sorte, só saberemos depois do dado ser lançado.

- O domínio de uma variável aleatória é o espaço amostral Ω
- O contradomínio pode ser o conjunto dos reais $\mathbb{R}$, apenas os positivos $(0, \infty)$, os inteiros $\mathbb{Z}$, um intervalo, etc, dependendo dos valores que a variável aleatória puder assumir
- Quando o experimento é realizado e o resultado ω é conhecido, o valor da variável aleatória $X(\omega)$ é determinado

Variável aleatória

Lançamento de moeda

- Considere um experimento de lançar uma moeda 3 vezes e contar o número de caras
- Seja X o número de caras
- Antes do experimento, seu valor é desconhecido
- O que sabemos é que será um valor entre 0 e 3
- Resultar em um valor é um evento, então podemos calcular probabilidades: (o = coroa, a = cara)

$$P\{X=0\} = P\{ooo\} = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$$

$$P\{X=1\} = P\{ooa\} + P\{oao\} + P\{aoo\} = \frac{3}{8}$$

$$P\{X=2\} = P\{oaa\} + P\{aoa\} + P\{aao\} = \frac{3}{8}$$

$$P\{X=3\} = P\{aaa\} = \frac{1}{8}$$

x	$\mathbf{P}\{X=x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

- A tabela resume tudo o que sabemos sobre X antes do experimento
- Não sabemos X antes do resultado ω, mas sabemos suas probabilidades

UFV

Funções de probabilidade

Definições

- A coleção de todas as probabilidades relacionadas a X é a distribuição de X
- A função probabilidade (função massa de probabilidade, fmp) é definida como

$$f(x) = \mathbf{P}\{X = x\}$$

- A função de distribuição (função de distribuição acumulada, fda) é definida como

$$F(x) = \mathbf{P}\{X \leq x\} = \sum_{y \leq x} f(y)$$

Propriedades

- Note que os eventos $\{X = x\}$ são disjuntos e exaustivos, então $\sum_x f(x) = 1$
- $F(x)$ é não decrescente, com valores entre 0 e 1

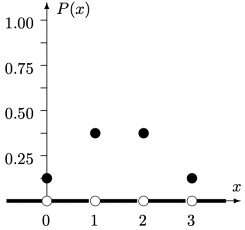
Obs.: alguns autores usam $p(x)$ ou $P(x)$ para representar a fmp em vez de $f(x)$.

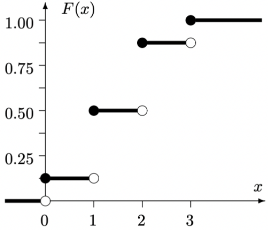
UFV

Funções de probabilidade

- Função probabilidade e função de distribuição para o exemplo da moeda

x	$\mathbf{P}\{X = x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1





- $F(x)$ é constante entre dois valores consecutivos de X
- $F(x)$ dá um salto de $f(x)$ em cada x de X
- $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$
- Dado um intervalo $A = (a, b]$, a probabilidade do evento A pode ser obtida da fda: $\mathbf{P}\{A\} = \mathbf{P}\{a < X \leq b\} = F(b) - F(a)$

Funções de probabilidade

Erros em módulos independentes

Um programa contém dois módulos. O número de erros X_1 do primeiro módulo tem a fmp $f_1(x)$ e o número de erros X_2 do segundo tem fmp $f_2(x)$, independente de $X1$:

x	$\mathbf{P_1}\{X_1 = x\}$	$\mathbf{P_2}\{X_2 = x\}$
0	0.5	0.7
1	0.3	0.2
2	0.1	0.1
3	0.1	0

Encontre a fmp e a fda para $Y = X_1 + X_2$, o número total de erros.

- $f_Y(0) = \mathbf{P}\{Y = 0\} = \mathbf{P}\{X_1 = 0 \cap X_2 = 0\}$
  $= f_1(0)f_2(0) = (0.5)(0.7) = 0.35$
- $f_Y(1) = \mathbf{P}\{Y = 1\} = f_1(0)f_2(1) + f_1(1)f_2(0) = 0.31$
- $f_Y(2) = \mathbf{P}\{Y = 2\} = f_1(0)f_2(2) + f_1(1)f_2(1) + f_1(2)f_2(1) = 0.18$
- $f_Y(3) = \mathbf{P}\{Y = 3\} = f_1(1)f_2(2) + f_1(2)f_2(1) + f_1(3)f_2(0) = 0.12$
- $f_Y(4) = \mathbf{P}\{Y = 4\} = f_1(2)f_2(2) + f_1(3)f_2(1) = 0.03$
- $f_Y(5) = \mathbf{P}\{Y = 5\} = f_1(3)f_2(2) = 0.01$

- $F_Y(0) = f_Y(0) = 0.35$
- $F_Y(1) = F_Y(0) + f_Y(1) = 0.66$
- $F_Y(2) = F_Y(1) + f_Y(2) = 0.84$
- $F_Y(3) = F_Y(2) + f_Y(3) = 0.96$
- $F_Y(4) = F_Y(3) + f_Y(4) = 0.99$
- $F_Y(5) = F_Y(4) + f_Y(5) = 1.00$

Tipos de variáveis aleatórias

- Variáveis discretas assumem valores contáveis, que podem ser listados em uma sequência
- Ex.: número de erros em um programa, número de arquivos na fila de impressão
- Não precisam ser inteiras. Ex.: a proporção de componentes com defeito em um lote de 100 pode ser 0/100, 1/100, ..., 100/100, são 101 valores diferentes

- Variáveis contínuas assumem qualquer valor em um intervalo
- Ex.: tempo de execução de um programa, tempo de instalação de um software, tempo de espera, peso e altura de uma pessoa, consumo de km por litro, ...
- Podem estar num intervalo limitado (a, b) ou ilimitado $(a, +\infty)$, $(-\infty, b)$, $(-\infty, +\infty)$, ou na união de vários intervalos

Tipos de variáveis aleatórias

Salto nas olimpíadas

- O resultado de um salto em distância, formalmente, é uma variável aleatória contínua
- Já o resultado do salto em altura é discreto, porque a barra é colocada em um número finito de alturas
- Na prática, o resultado do salto em distância é discreto porque o valor real é arrendondado para uma precisão de centímetros, por exemplo, 8,95 m

Vetor aleatório

- Muitas vezes lidamos com várias variáveis aleatórias simultaneamente
- Ex: uso de memória e processador;
  quantidade de memória, velocidade da CPU e preço de um computador

Definições

- Se X e Y são variáveis aleatórias, o par (X, Y) é um vetor aleatório
- Sua distribuição é chamada distribuição conjunta de X e Y
- As distribuições individuais de X e Y são chamadas distribuições marginais
- Os conceitos se aplicam a mais de duas variáveis, $(X_1, X_2, \ldots, X_n)$

Vetor aleatório

- Assim como com uma variável, a distribuição conjunta de um vetor é uma coleção de probabilidades do vetor (X, Y) assumir um valor (x, y)
- Note que $(X, Y) = (x, y)$ se $X = x$ e $Y = y$
- Esse "e" significa interseção:

$$f(x, y) = \mathbf{P}\{(X, Y) = (x, y)\} = \mathbf{P}\{(X = x) \cap (Y = y)\}$$

- $\{(X, Y) = (x, y)\}$ forma eventos mutuamente exclusivos e exaustivos, então:

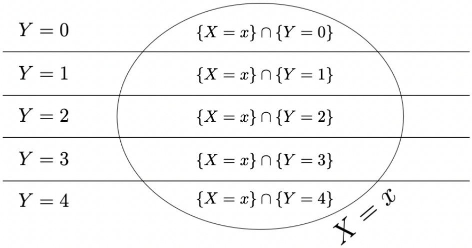
$$\sum_x \sum_y f(x, y) = 1$$

Vetor aleatório

- A função de distribuição conjunta f_{XY} contém informação completa sobre o comportamento do vetor aleatório; em particular, podemos calcular as distribuições marginais pela regra da soma

$$f_X(x) = \mathbf{P}\{X = x\} = \sum_y f_{XY}(x, y)$$

$$f_Y(y) = \mathbf{P}\{Y = y\} = \sum_x f_{XY}(x, y)$$



Vetor aleatório

Erros em um programa

Um programa contém dois módulos. A distribuição conjunta do número de erros X no primeiro módulo e Y no segundo módulo é $f(0,0) = f(0,1) = f(1,0) = 0.2$, $f(1,1) = f(1,2) = f(1,3) = 0.1$, $f(0,2) = f(0,3) = 0.05$. Encontre:

- as distribuições marginais de X e Y
- a probabilidade de não haver erro no primeiro módulo
- a distribuição do total de erros no programa

Solução:

organizar a distribuição conjunta em uma tabela; as distribuições marginais são a soma de linha e coluna

$f_{XY}(x,y)$		y = 0	1	2	3	$f_X(x)$
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$f_Y(y)$		0.40	0.30	0.15	0.15	1.00

a probabilidade de não haver erro no primeiro módulo é $f_X(0) = 0.50$

a distribuição do total de erros $Z = X + Y$ é dada ao lado

$$
\begin{aligned}
f_Z(0) &= \mathbf{P}\{X + Y = 0\} \\
&= \mathbf{P}\{X = 0 \cap Y = 0\} \\
&= f_{XY}(0,0) = 0.20 \\
f_Z(1) &= \mathbf{P}\{X = 0 \cap Y = 1\} \\
&\quad + \mathbf{P}\{X = 1 \cap Y = 0\} \\
&\quad f_{XY}(0,1) + f_{XY}(1,0) = 0.40 \\
f_Z(2) &= f_{XY}(0,2) + f_{XY}(1,1) = 0.15 \\
f_Z(3) &= f_{XY}(0,3) + f_{XY}(1,2) = 0.15 \\
f_Z(4) &= f_{XY}(1,3) = 0.10
\end{aligned}
$$

Independência de variáveis aleatórias

- Da distribuição conjunta de um vetor aleatório podemos calcular as distribuições marginais
- O contrário geralmente não é possível, pois a distribuição marginal não traz informações sobre a inter-relação entre as variáveis
- Por exemplo, distribuições marginais não podem dizer se duas variáveis aleatórias X e Y são independentes

Definição

Duas variáveis aleatórias X e Y são independentes se, para todo x e y,

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

- Isso significa que os eventos $\{X = x\}$ e $\{Y = y\}$ são independentes para todo x e y, ou seja, X e Y assumem valores independentemente um do outro
- Note que para mostrar independência de X e Y, precisamos verificar se a distribuição conjunto é o produto das distribuições marginais para todo par x e y
- Para mostrar dependência, basta um contraexemplo, um par (x, y) com $f_{XY}(x,y) \neq f_X(x) f_Y(y)$

Independência de variáveis aleatórias

Erros em um programa (continuação)

No exemplo anterior (erros no programa com dois módulos), verifique se os erros nos módulos ocorrem de forma independente

Solução: verificar se a distribuição conjunta bate com as distribuições marginais

- $f_{XY}(0,0) = 0.2$ e $f_X(0)f_Y(0) = (0.50)(0.40) = 0.2$ – Ok!
- $f_{XY}(1,0) = 0.2$ e $f_X(1)f_Y(0) = (0.50)(0.40) = 0.2$ – Ok!
- $f_{XY}(0,1) = 0.2$ e $f_X(0)f_Y(1) = (0.50)(0.30) = 0.15$ – diferente

não é necessário verificar mais; como há pelo menos um caso em que a fórmula para variáveis aleatórias independentes não vale, concluímos que X e Y não são independentes, ou seja, o número de erros nos módulos é dependente.

Sumarização

- A distribuição de uma variável aleatória ou de um vetor aleatório, isto é, a coleção completa das respectivas probabilidades, contém informação completa sobre seu comportamento
- Pode ser resumida em poucas características essenciais que descrevem seu valor médio, valor mais provável, sua amplitude, variabilidade, etc
- Os mais comuns são esperança, variância, desvio padrão, covariância e correlação

Esperança matemática

Definição (preliminar)

A esperança matemática (valor esperado) de uma variável aleatória é seu valor médio

- É o valor médio que esperaríamos obter, se pudéssemos realizar o experimento indefinidamente

Exemplo I

Considere uma variável aleatória X que assume valores 0 ou 1 com probabilidade $\mathbf{P}\{0\} = \mathbf{P}\{1\} = 0.5$.

Observando X muitas vezes, notaremos $X = 0$ cerca de 50% das vezes e $X = 1$ cerca de 50%. A média de X será próxima de 0.5, que é o valor médio esperado.

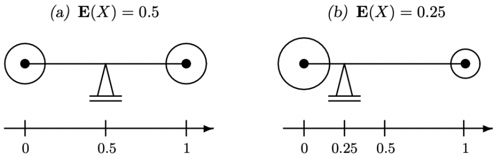
Exemplo II

Considere agora que variável aleatória X assume valores 0 ou 1 com probabilidades $\mathbf{P}\{0\} = 0.75$ e $\mathbf{P}\{1\} = 0.25$.

Se ganhamos R$1 toda vez que observamos $X = 1$, em média ganhamos R$1 a cada 4 observações, ou R$0.25 por observação, que é o valor médio esperado.

Esperança matemática

- A esperança matemática pode ser vista como o centro de gravidade
- No exemplo I temos pesos iguais nos pontos 0 e 1, o equilíbrio está no ponto 0.5
- No exemplo II há peso 0.75 no ponto 0 e 0.25 no 1, o equilíbrio está no ponto 0.25



- A fórmula geral é então

Definição

A esperança matemática (valor esperado) de uma variável aleatória é seu valor médio (no caso, a média dos resultados ponderada por suas probabilidades de ocorrência)

$$\mu = \mathbf{E}(X) = \sum_x x f(x)$$

Esperança matemática

Exemplo

A esperança matemática nos exemplos anteriores, I e II, vale:

- I) $\mathbf{E}(X) = (0)(0.5) + (1)(0.5) = 0.5$
- II) $\mathbf{E}(X) = (0)(0.75) + (1)(0.25) = 0.25$

- Note que bate com os valores esperados das observações
- De certa forma, a esperança matemática é a melhor previsão que temos de X
- A variável em si é aleatória, toma valores diferentes com diferentes probabilidades
- A esperança matemática $\mathbf{E}(X)$, no entanto, é única, não é aleatória

Esperança matemática de uma função

- Muitas vezes estamos interessados em outra variável, Y, que é uma função de X
- Por exemplo, o tempo de download depende da velocidade da conexão, o lucro de uma loja de computadores depende do número de computadores vendidos, a comissão do gerente depende do lucro da loja, etc.
- O valor esperado de $Y = g(X)$ é obtido por uma fórmula similar

$$\mathbf{E}(g(X)) = \sum_x g(x)f(x)$$

Propriedades

- Para quaisquer variáveis aleatórias X e Y e constantes não-aleatórias a, b, c

$$\mathbf{E}(aX + bY + c) = a\mathbf{E}(X) + b\mathbf{E}(Y) + c$$

- Em particular

$$\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$$

$$\mathbf{E}(aX) = a\mathbf{E}(X)$$

$$\mathbf{E}(c) = c$$

- Para variáveis independentes X e Y

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$$

Obs: isso vale também para algumas variáveis dependentes, não pode ser usado para provar independência

Propriedades

Erros em um programa (continuação)

No exemplo dos erros no programa com dois módulos:

- $\mathbf{E}(X) = (0)(0.5) + (1)(0.5) = 0.5$
- $\mathbf{E}(Y) = (0)(0.4) + (1)(0.3) + (2)(0.15) + (3)(0.15) = 1.05$

O número total esperado de erros é então

- $\mathbf{E}(X + Y) = 0.5 + 1.05 = 1.65$

Observações

- Obviamente o programa não terá 1.65 erros, pois o número de erros é inteiro
- Mas nem por isso devemos arrendondar o valor para 2
- X e Y são inteiros, mas a esperança matemática (valor médio) pode não ser

Variância

- A esperança matemática mostra o valor médio de uma variável aleatória, o valor esperado da variável, mais ou menos algum erro.
- Quão grande pode ser este "erro"?
- O quanto a variável pode variar em torno deste valor esperado?

E-mails

Certo usuário recebe 48 ou 52 e-mails por dia, com chances iguais, 50-50%.
Outro recebe 0 ou 100 e-mails, também com chance 50-50%.

- Em ambos os casos, a esperança matemática é a mesma, 50
- Mas no primeiro, o valor real sempre está perto de 50, no segundo sempre longe
- A primeira variável aleatória é mais estável, tem menos variabilidade

- A variabilidade de uma variável aleatória pode ser medida pela sua distância da média $\mu = \mathbf{E}(X)$
- Mas essa medida é uma variável aleatória também
- Para usá-la como medida da distribuição, precisamos calcular sua esperança

Variância

Definição

A variância de uma variável aleatória é o valor esperado do quadrado de seu desvio da média:

$$\sigma^2 = \text{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \sum_x (x - \mu)^2 f(x)$$

Note que se a distância não fosse elevada ao quadrado, o resultado seria $\mu - \mu = 0$

- Note que a variância é sempre não-negativa
- E que ela vale 0 somente se todos os resultados são iguais à média
- A variância pode ser calculada também pela fórmula

$$\text{Var}(X) = \mathbf{E}(X^2) - \mu^2$$

Desvio padrão

- Uma dificuldade com a variância é que, se a variável aleatória é medida em alguma unidade, a variância será medida por esta unidade ao quadrado
- A variância de um lucro seria medida por R$^{2}, do espaço em disco por Gb^2, a variância do número de matriculados em uma turma seria medida por alunos^2
- A variância não pode ser comparada com o resultado ou a esperança, pois tem unidade diferente

Definição

O desvio padrão é a raiz quadrada da variância:

$$\sigma = \text{Std}(X) = \sqrt{\text{Var}(X)}$$

- O desvio padrão tem a mesma unidade da variável aleatória
- Esta é a principal razão da introdução de mais uma medida de variabilidade

UFV

Covariância

- Esperança, variância e desvio padrão dão características da distribuição de uma variável aleatória
- Quando há duas variáveis, precisamos de medidas da associação entre elas

Definição

A covariância de duas variáveis aleatórias é o valor esperado do produto das distâncias de cada uma em relação à respectiva esperança:

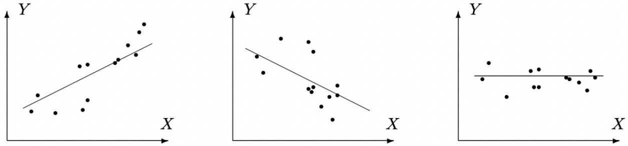
$$\begin{aligned}\sigma_{XY} = \mathrm{Cov}(X, Y) &= \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y)) \\ &= \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)\end{aligned}$$

- Ela mede a inter-relação das duas variáveis, a sua variância conjunta

Covariância

Pela fórmula da covariância, $\text{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))$, pode-se concluir:

- Se $Cov(X, Y) > 0$
  - desvios positivos $(X - \mathbf{E}(X))$ são mais prováveis de serem multiplicados por desvios positivos $(Y - \mathbf{E}(Y))$ e desvios negativos $(X - \mathbf{E}(X))$ são mais prováveis de serem multiplicados por desvios negativos $(Y - \mathbf{E}(Y))$
  - valores altos de X implicam valores altos de Y e baixos de X implicam baixos de Y
  - as variáveis têm uma correlação positiva
- Se $Cov(X, Y) < 0$
  - ocorre o contrário, valores altos de X estão geralmente relacionados a valores baixos de Y e baixos de X a altos de Y
  - as variáveis têm uma correlação negativa
- Se $Cov(X, Y) = 0$
  - as variáveis são não correlacionadas



(a) $\text{Cov}(X, Y) > 0$ (b) $\text{Cov}(X, Y) < 0$ (c) $\text{Cov}(X, Y) = 0$

Correlação

Definição

O coeficiente de correlação entre duas variáveis aleatórias X e Y é definido por

$$\rho = \frac{\text{Cov}(X, Y)}{\text{Std}(X)\text{Std}(Y)}$$

- é a covariância em outra escala, normalizada

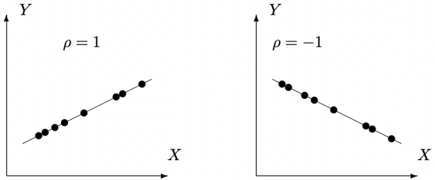
- Note que a unidade da covariância é o produto das unidades de medida de X e Y
- Apesar de informar se estão relacionadas, não informa o grau da relação
- Para saber se são forte ou fracamente relacionadas, deve-se comparar com a magnitude de X e Y
- É isto que é feito pelo coeficiente de correlação, que não tem dimensão

Correlação

- Como interpretar o valor de ρ? Que valores pode ter?
- É possível mostrar que

$$-1 \leq \rho \leq 1$$

e que $|\rho| = 1$ somente quando os valores de X e Y estão numa linha reta



Assim

- valores de ρ próximos de 1 indicam forte correlação positiva
- valores próximos de -1 indicam forte correlação negativa
- e valores próximos de 0 indicam fraca correlação (ou sem correlação)

Correlação

Erros em um programa (continuação)

- Esperança matemática de erros nos módulos

x	$f_X(x)$	$x\,f_X(x)$
0	0.5	0
1	0.5	0.5
	$\mathbf{E}(X) = \mu_X = 0.5$	

y	$f_Y(y)$	$y\,f_Y(y)$
0	0.4	0
1	0.3	0.3
2	0.15	0.3
3	0.15	0.45
	$\mathbf{E}(Y) = \mu_Y = 1.05$	

- Variância (um exemplo de cada fórmula) e desvio padrão

x	$f_X(x)$	$x - \mu_X$	$(x - \mu_X)^2 f_X(x)$
0	0.5	-0.5	0.125
1	0.5	0.5	0.125
			$\mathrm{Var}(X) = \sigma_X^2 = 0.25$
			$\mathrm{Std}(X) = \sigma_X = \sqrt{0.25} = 0.5$

y	$f_Y(y)$	y^2	$y^2\,f_Y(y)$
0	0.4	0	0
1	0.3	1	0.3
2	0.15	4	0.6
3	0.15	9	1.35
			$\mathbf{E}(y^2) = 2.25$

$$\mathrm{Var}(Y) = \sigma_Y^2 = \mathbf{E}(y^2) - \mu_Y^2 = 1.1475$$
$$\mathrm{Std}(Y) = \sigma_Y = \sqrt{1.1475} \approx 1.0712$$

- Covariância e coeficiente de correlação

$$\mathbf{E}(XY) = (0)(0)(0.2) + \cdots + (1)(0)(0.2) + (1)(1)(0.1) + (1)(2)(0.1) + (1)(3)(0.1) = 0.6$$

$$\mathrm{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = 0.6 - (0.5)(1.05) = 0.075$$

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Std}(X)\mathrm{Std}(Y)} = \frac{0.075}{(0.5)(1.0712)} = 0.14 \Rightarrow \text{correlação positiva mas não tão forte}$$