

# TP 547- Princípios de Simulação de Sistemas de Comunicação

Prof. Samuel Baraldi Mafra

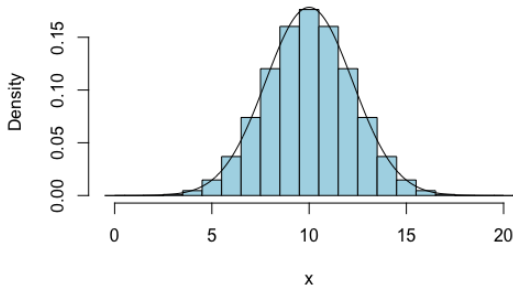


## Aproximações da binomial e poisson com a distribuição Normal

## Aproximação Binomial

- A distribuição normal pode ser usada como uma aproximação da distribuição binomial, em certas circunstâncias, a saber:
- Se  $X \approx B(n, p)$  e se  $n$  for grande e / ou  $p$  estiver próximo de  $\frac{1}{2}$ , então  $X$  é aproximadamente  $N(np, npq)$  (onde  $q = 1 - p$ ).
- Em alguns casos, resolver um problema usando a distribuição Normal pode ser mais fácil do que usar um Binomial.

## Normal Approximation to a Binomial Distribution



Para fazer os cálculos precisamos pegar os valores inferior e superior dos retângulos, para  $x=10$ , seria  $x_{\min}=9.5$  e  $x_{\max}=10.5$ .

Sessenta por cento dos alunos da 5ª série frequentam a escola em um determinado distrito escolar urbano. Se uma amostra de 50 crianças da 5ª série for selecionada, encontre a probabilidade de que 29 estejam realmente matriculadas na escola.

File: aproxbinomiallex.ipynb

## Aproximação de Poisson

- A distribuição normal também pode ser usada para aproximar a distribuição de Poisson para grandes valores de  $\lambda$  (a média da distribuição de Poisson).
- Se  $X \approx \text{Po}(\lambda)$ , então para grandes valores de  $\lambda$ ,  $X \approx N(\lambda, \lambda)$  aproximadamente.

Um departamento de polícia recebe em média 50 solicitações por dia. Qual a probabilidade de receber 35 solicitações num dia selecionado aleatoriamente?

File: `aproxpoissonnormalex.ipynb`

## Correlação

Variáveis dentro de um conjunto de dados podem ser relacionadas por vários motivos.

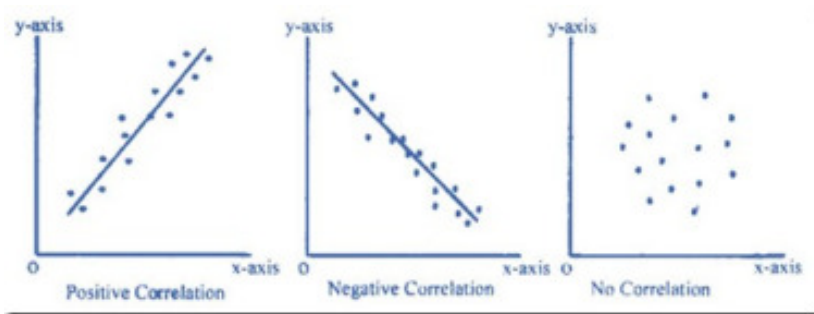
Por exemplo:

- Uma variável pode depender dos valores de outra variável.
- Uma variável pode ser ligeiramente associada a outra variável.
- Duas variáveis podem depender de uma terceira variável desconhecida.



# Variáveis Correlacionadas

- Correlação positiva: ambas as variáveis mudam na mesma direção.
- Correlação Neutra: Sem relação na mudança das variáveis.
- Correlação negativa: as variáveis mudam em direções opostas.



- A covariância de duas variáveis  $X$  e  $Y$  é uma medida da variabilidade conjunta destas variáveis aleatórias.
- Se as variáveis tem covariância positiva tendem a mostrar um comportamento semelhante, ou seja, os menores(maiores) valores da variável  $X$  corresponde aos menores(maiores) da variável  $Y$ .
- Se a covariância é negativa então as variáveis tendem a mostrar um comportamento oposto, ou seja, os menores(maiores) valores da variável  $X$  corresponde aos maiores(menores) da variável  $Y$ .

$$\text{Cov}[X,Y] = \sum [(X - \mu_X)(Y - \mu_Y)] / (N - 1) \quad (1)$$

Exemplo: Covariância

Calcular a covariância para os dados abaixo:

x: [2.1, 2.5, 3.6, 4.0]

y: [8, 10, 12, 14]

file: covarianceex.ipynb

A covariância apenas mostra a tendência de relação entre as variáveis. Não têm como quantificar a relação por esta métrica, para isto é usado o coeficiente de correlação.

- O coeficiente de correlação de Pearson pode ser usado para resumir a força da relação linear entre duas amostras de dados.
- O coeficiente de correlação de Pearson é calculado como a covariância das duas variáveis dividida pelo produto do desvio padrão de cada amostra de dados. É a normalização da covariância entre as duas variáveis para dar uma pontuação interpretável.

$$\rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}} = \frac{\text{Cov}[X,Y]}{\sigma_X \sigma_Y}, \quad -1 \leq \rho_{X,Y} \leq 1 \quad (2)$$

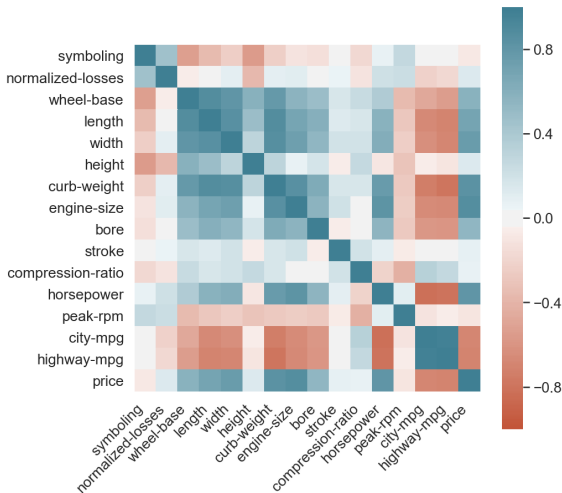
Exemplo: Coeficiente de correlação

Calcular o coeficiente de correlação para os dados abaixo:

x: [2.1, 2.5, 3.6, 4.0]

y: [8, 10, 12, 14] File: correlationex.ipynb

## Dataset de informações de veículos.



<http://archive.ics.uci.edu/ml/datasets/automobile>

### Exemplos:

- Calcular o coeficiente de correlação entre duas variáveis normais geradas aleatoriamente.
- Verificar a influência da semente no coeficiente de correlação.



```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(10)

X=np.random.normal(0,1,10)
np.random.seed(40)
Y=np.random.normal(0,1,10)

C = np.corrcoef(X,Y)
print(C)
```

## Exemplo: Correlação com variáveis dependentes

- Fazer testes para mostrar a influência de uma variável na coeficiente.
- $\text{data1} = 20 * \text{NV} + 100$
- $\text{data2} = \text{data1} + (10 * \text{NV} + 50)$
- NV: Variável normal padrão.

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
# seed random number generator
np.random.seed(1)
# prepare data
data1 = 20 * np.random.normal(0,1,1000) + 100
data2=data1+(10 * np.random.normal(0,1,1000) + 50)
C = np.corrcoef(data1 , data2)

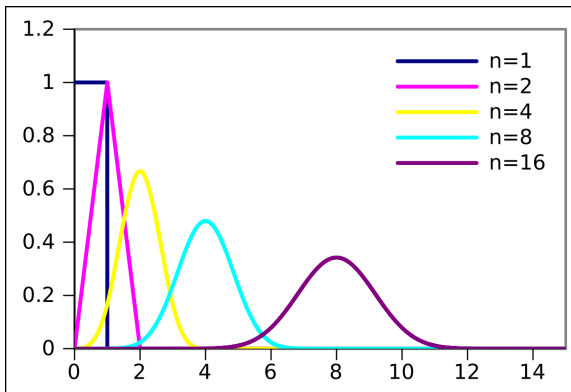
print(C)
```

file: corrvardependent.ipynb

# Variáveis Correlacionadas

## Cuidado!!!!!!!!!!!!

As operações com diferentes variáveis modificam a distribuição dos dados, com poucas exceções como a distribuição normal, a soma de duas variáveis de uma mesma distribuição não gera uma terceira com a mesma distribuição.



# Variáveis Correlacionadas

- Para gerar uma variável normal correlacionada são usados os seguintes passos:
- Gerar duas sequencias de variáveis normais independentes  $X_1, X_2$ ;
- Sendo  $X_1 = N(0,1)$ ,  $X_2 = N(0,1)$  e a variavel gerada  $Z = N(0,1)$
- Calcular  $Z = \rho X_1 + \sqrt{1 - \rho^2} X_2$ ;
- $Z$  terá uma correlação  $\rho$  com respeito a  $X_1$ .

The proof is simple. It is clear that  $Z$  is a Gaussian random variable, since it is a linear combination of Gaussian random variables. It also follows that  $Z$  is zero mean if  $X$  and  $Y$  are zero mean. The variance of  $Z$  is

$$\begin{aligned}\sigma_Z^2 &= E \left\{ [\rho X + \sqrt{1 - \rho^2} Y]^2 \right\} \\ &= \rho^2 E \{X^2\} + 2\rho\sqrt{1 - \rho^2} E \{XY\} + (1 - \rho^2) E \{Y^2\}\end{aligned}\quad (7.86)$$

Since  $E \{XY\} = E \{X\} E \{Y\} = 0$  and  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ , the preceding becomes

$$\sigma_Z^2 = \rho^2 \sigma^2 + (1 - \rho^2) \sigma^2 = \sigma^2 \quad (7.87)$$

The covariance  $E \{XZ\}$  is

$$\begin{aligned}E \{XZ\} &= E \{X[\rho X + (1 - \rho)Y]\} \\ &= \rho E \{X^2\} + (1 - \rho) E \{XY\} \\ &= \rho E \{X^2\} = \rho \sigma^2\end{aligned}\quad (7.88)$$

where the last step follows because  $X$  and  $Y$  are independent and zero mean. The correlation coefficient  $\rho_{XZ}$  is

$$\rho_{XZ} = \frac{E\{XZ\}}{\sigma_X \sigma_Z} = \frac{\rho \sigma^2}{\sigma^2} = \rho \quad (7.89)$$

as desired.

- Fazer um código para gerar  $Z$  em função de  $X_1$  e  $X_2$  e plotar  $Z$ .  
Usar  $\rho = 0.9$ .



```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(10)
rho=-0.9
X=np.random.normal(0,1,1000)
Y=np.random.normal(0,1,1000)

z=rho*X+pow(1-rho**2,0.5)*Y

C = np.corrcoef(X,z)
print(C)

plt.plot(X,z,'o')
plt.show()
```

File: varcorrrho.ipynb