Estatística para Ciência de Dados

Profa. Rebeca Valgueiro

Quem sou eu?

- Graduada em Engenharia Civil
- MBA em Gestão Empresarial
- Trabalho a mais de 4 anos no mercado de tecnologia atuando em projetos de:
    - Desenvolvimento web
    - Desenvolvimento desktop- windows
    - Data Science
    - IA e visão computacional

https://www.linkedin.com/in/rebecavalgueiro/

Probabilidade

Desempenha um papel central e indispensável na Ciência de Dados.

Fornece ferramentas matemáticas para quantificar e modelar a incerteza inerente aos dados e aos processos do mundo real.

Probabilidade

Ela é fundamental para:

- ➔ **Modelar a Incerteza: Lidar com dados imperfeitos e eventos aleatórios.**
- ➔ **Inferência Estatística: Generalizar a partir de amostras e testar hipóteses**
- ➔ **Tomada de Decisão: Analisar riscos e otimizar em ambientes incertos.**
- ➔ **Análise Exploratória: Identificar padrões, anomalias e visualizar a incerteza nos dados.**

Variável aleatória

É o resultado numérico da observação de um fenômeno aleatório

Discretas: Podem assumir um número finito ou contável de valores (ex: número de clientes que entram em uma loja, número de defeitos em uma peça produzida).
Contínuas: Podem assumir qualquer valor dentro de um intervalo (ex: temperatura diária, peso de um animal).

Probabilidade

- **Experimento Aleatório**: Algo que acontece com resultados incertos (ex: lançar uma moeda, medir a temperatura amanhã).
- **Espaço Amostral (Ω):** Todos os resultados possíveis desse experimento (ex: {Cara, Coroa}, todos os valores de temperatura possíveis).
- **Variável Aleatória (X):** Um valor que atribuímos a cada um desses resultados.

Exemplos

1º Experimento: Lançar uma moeda justa.

Espaço Amostral: Ω={Cara,Coroa}
Variável Aleatória X: número de "Caras" (ou "Coroas").
X(Cara)=1
X(Coroa)=0

2º Experimento: Medir a temperatura máxima de um dia em Recife.

Espaço Amostral: Ω = [20ºC a 40ºC]
Variável Aleatória T: o valor da temperatura máxima medida.
T =32.5 ºC
Neste caso, a variável aleatória já mapeia diretamente para um valor numérico.

Probabilidade

- **Evento (A):** É qualquer subconjunto do espaço amostral. Um evento representa um resultado específico ou um grupo de resultados de interesse.
  - Pode conter um único ponto amostral (evento elementar) ou múltiplos pontos amostrais.
  - Exemplos:
    - Ao lançar um dado, o evento "obter um número par" é A={2,4,6}.
    - Ao escolher um turista, o evento "o turista é de um país da América do Sul" seria um subconjunto do espaço amostral.
    - A temperatura máxima de um dia em Olinda ser "acima de 30°C" seria um intervalo dentro do espaço amostral contínuo.

Probabilidade

➔ **Probabilidade de um Evento (P(A)):**

- É uma medida numérica que quantifica a chance ou a frequência relativa com que um evento A ocorre.
- A probabilidade de um evento sempre está entre 0 e 1 (inclusive): $0 \le P(A) \le 1$.
- EXEMPLOS

1. Lançamento de um Dado Justo:

   - Experimento: Lançar um dado de seis faces (numeradas de 1 a 6).
   - Espaço Amostral: $\Omega=\{1,2,3,4,5,6\}$ (6 resultados igualmente prováveis).
   - Evento A: Obter um número par. $A=\{2,4,6\}$.

$$P(A) = \frac{\text{Número de resultados favoráveis}}{\text{Número total de resultados possíveis}} = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$$

Probabilidade

➔ **Probabilidade de um Evento** (P(A)):

2. Previsão do Tempo em Recife:

- Experimento: Observar o tempo em Recife amanhã.
- Espaço Amostral: Ω={Chove,Não Chove}.
- Evento C: Chover amanhã. C={Chove}.
- Probabilidade de C: A probabilidade de chover amanhã é geralmente dada em porcentagem nas previsões do tempo. Se a previsão indicar uma "chance de chuva de 70%", então: P(C)=0.7=70%

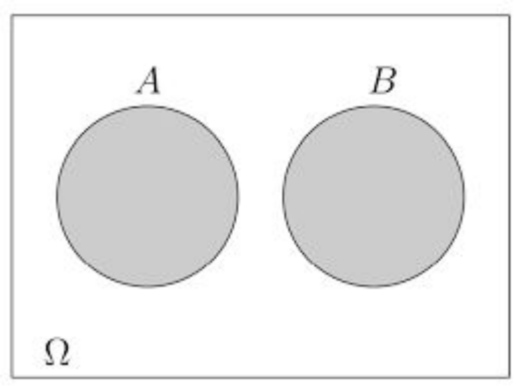
É importante notar que essa probabilidade é baseada em modelos meteorológicos e dados históricos, não em um simples espaço amostral de resultados igualmente prováveis.

Probabilidade

- **Probabilidade de um Evento** (P(A)):
  - Axiomas Fundamentais da Probabilidade (de Kolmogorov):
    - Para qualquer evento A, $P(A) \geq 0$.
    - A probabilidade do espaço amostral inteiro é 1: $P(\Omega)=1$ (algo no espaço amostral deve acontecer).

Probabilidade

➔ **Eventos Mutuamente Exclusivos (ou Disjuntos):**



- Dois eventos A e B são mutuamente exclusivos se eles não podem ocorrer ao mesmo tempo. Sua interseção é o conjunto vazio ($A \cap B = \varnothing$).
- Para eventos mutuamente exclusivos, a probabilidade da união é a soma das probabilidades: $P(A \cup B) = P(A) + P(B)$.
- Exemplos:
  - Ao lançar um dado, os eventos "obter um número par" e "obter um número ímpar" são mutuamente exclusivos.
  - Um turista escolhido aleatoriamente não pode ser simultaneamente brasileiro e argentino (assumindo nacionalidade única para simplificar).

Probabilidade

- **Eventos Independentes:**
  - Dois eventos A e B são independentes se a ocorrência de um não afeta a probabilidade de ocorrência do outro.
  - Para eventos independentes, a probabilidade da interseção é o produto das probabilidades: $P(A \cap B)=P(A) \times P(B)$.
  - Exemplos:
    - O resultado de dois lançamentos consecutivos de um dado justo são eventos independentes.
    - A nacionalidade de dois turistas escolhidos aleatoriamente
    - Se chover em Recife hoje e a Compesa realizar um corte de água amanhã (assumindo que não há relação causal direta).

Probabilidade

- **Probabilidade Condicional (P(A|B)):**
  - É a probabilidade de um evento A ocorrer, dado que o evento B já ocorreu.
  - A fórmula é: $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

    desde que P(B)>0.
  - Exemplos:
    - Qual a probabilidade de obter um número par ao lançar um dado (evento A), dado que o resultado foi maior que 3 (evento B)?
      - B={4,5,6}, P(B)=3/6=0.5
      - A∩B={4,6}, P(A∩B)=2/6=1/3
      - P(A|B)=(1/3)/(1/2)=2/3

EXERCÍCIO 1 - Espaço Amostral e Eventos

Considere um jogo de futebol entre os times A e B e. Os resultados possíveis são: Vitória do A (V), Empate (E) ou Derrota do A (D).

a) Qual é o espaço amostral (Ω) deste experimento?

b) Defina o evento X como "O A não perde o jogo". Quais resultados pertencem ao evento X?

c) Defina o evento Y como "O jogo tem um vencedor". Quais resultados pertencem ao evento Y?

d) Os eventos X e Y são mutuamente exclusivos? Justifique sua resposta.

a) $\Omega=\{V,E,D\}$

b) $X=\{V,E\}$

c) $Y=\{V,D\}$

d) Os eventos X e Y **não são mutuamente exclusivos**. Eles podem ocorrer simultaneamente se o resultado for V

EXERCÍCIO 2 - Probabilidade de eventos

Em uma pesquisa sobre hábitos culturais em Olinda, um pesquisador escolhe aleatoriamente um morador da cidade. Suponha que a distribuição da população por nível de escolaridade seja a seguinte:

Ensino Fundamental: 30%

Ensino Médio: 45%

Ensino Superior: 25%

Qual a probabilidade de o morador escolhido aleatoriamente ter:

a) Ensino Superior?

b) Ensino Fundamental ou Médio?

c) Não ter Ensino Fundamental?

Temos: $P(F)=0.30$ $P(M)=0.45$ $P(S)=0.25$

a) $P(S)=0.25$

b) $P(F \cup M)=P(F)+P(M)=0.30+0.45=0.75$

c) $P(\neg F)=1-P(F)=1-0.30=0.70$ (Isso significa ter Ensino Médio ou Superior).

EXERCÍCIO 3 - Eventos Mutuamente Exclusivos e União de Eventos

Em um sorteio simples de um número inteiro de 1 a 20 na Loteria:

a) Qual a probabilidade de o número sorteado ser par?

b) Qual a probabilidade de o número sorteado ser maior que 15?

c) Os eventos "o número é par" e "o número é maior que 15" são mutuamente exclusivos? Por quê?

d) Qual a probabilidade de o número sorteado ser par ou maior que 15?

Espaço amostral: $\Omega=\{1,2,3,...,20\}$

a) P(Par)= 10/20 = 0.5

b)P(Maior15)= 5/20 = 0.25

c) Não. Os números 16, 18 e 20 pertencem a ambos os eventos.

d)$P(Par \cup Maior15)=P(Par)+P(Maior15)-P(Par \cap Maior15)$

$P(Par \cap Maior15)= 3/ 20 = 0.15$

$P(Par \cup Maior15)=0.5+0.25-0.15=0.6$

EXERCÍCIO 4 - Probabilidade Condicional

Uma pesquisa com frequentadores da Praia de Boa Viagem revelou que:

60% dos frequentadores gostam de nadar (N).

40% dos frequentadores gostam de tomar sol (S).

25% dos frequentadores gostam de nadar e tomar sol.

Se um frequentador é escolhido aleatoriamente, qual a probabilidade de que ele goste de:

a) Tomar sol, dado que ele gosta de nadar? (P(S|N))

b) Nadar, dado que ele gosta de tomar sol? (P(N|S))

P(N)=0.60 (gosta de nadar)

P(S)=0.40 (gosta de tomar sol)

P(N∩S)=0.25 (gosta de nadar e tomar sol)

$$P(S|N) = \frac{P(S \cap N)}{P(N)} = \frac{0.25}{0.60} \approx 0.417$$

$$P(N|S) = \frac{P(N \cap S)}{P(S)} = \frac{0.25}{0.40} = 0.625$$

Distribuições de Probabilidade

Descrevem como a probabilidade é distribuída para os diferentes valores que uma variável aleatória pode assumir.
Descreve a relação entre cada resultado possível para uma variável aleatória e suas probabilidades.

Distribuições de Probabilidade Discretas

Sumarizam as probabilidades de uma variável aleatória discreta. Através de uma função matemática, atribui uma probabilidade de saída a cada valor específico (Função massa de probabilidade)

Distribuições de Probabilidade Discretas

→ **Distribuição de Bernoulli:**

- Quando sucesso e fracasso são os únicos resultados possíveis
  - Sucesso = 1
  - Fracasso = 0
- Exemplo:
  - Peça produzida: produto com falha ou não
  - Assistir um Filme: ter no netflix ou não
  - Chutar uma questão na prova: acertar ou não

Distribuições de Probabilidade Discretas

- **Distribuição de Bernoulli - Exemplo**
  - Considere uma questão objetiva com 5 alternativas. Qual a probabilidade de , no chute, eu acertar essa questão?

    X = chutar a questão

    P(Sucesso) = P(X =1) = 1 / 5 = 0.2

    P(Falha) = P (X = 0) = 4/ 5 = 0.8

    ou P (X=0) = 1 - P(X=1)

Distribuições de Probabilidade Discretas

- **Distribuição de Bernoulli:**
  - Modela um experimento com apenas dois resultados possíveis, geralmente chamados de "sucesso" (com probabilidade p) e "fracasso" (com probabilidade 1−p).
  - Variável aleatória X pode assumir os valores 0 (fracasso) ou 1 (sucesso).
  - Função de Probabilidade de Massa (PMF):

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

Distribuições de Probabilidade Discretas

- **Distribuição Binomial:**
  - Considere agora que eu quero analisar a probabilidade de N sucessos
  - Teremos o Bernoulli N vezes!

  - Considere uma questão objetiva com 5 alternativas. Qual a probabilidade de , no chute, eu acertar 3 questões de 4 no total?
    - X = número de acertos totais
    - p = P(acertar 1 questão) = 0.2

    como calcular?

Distribuições de Probabilidade Discretas

- **Distribuição Binomial:**
  - Modela o número de sucessos (k) em um número fixo de tentativas independentes (n) de um experimento de Bernoulli, onde cada tentativa tem a mesma probabilidade de sucesso (p).
  - Variável aleatória X pode assumir valores de 0 a n.

$$\text{PMF: } P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Voltando pra questão e aplicando a fórmula, temos que PMF = 0.0256

E qual seria a probabilidade de acertar pelo menos 3 questões?

P(X>=3) = P(X=3) + P(X=4)

Distribuições de Probabilidade Discretas

- **Distribuição de Poisson:**
  - Numa central telefônica, o número de chamadas chega com uma frequência média de 8 chamadas por minuto. Qual a probabilidade da central receber 10 chamadas no próximo minuto?
  - Como calcular?

Distribuições de Probabilidade Discretas

- **Distribuição de Poisson:**
  - Modela o número de eventos que ocorrem em um intervalo de tempo ou espaço fixo, dado uma taxa média de ocorrência (λ). Os eventos devem ser independentes e ocorrer a uma taxa constante.
  - Variável aleatória X pode assumir valores inteiros não negativos (0, 1, 2, ...).

$$\text{PMF: } P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

e = 2,7182...

Distribuições de Probabilidade Discretas

➔ **Distribuição de Poisson:**

- Numa central telefônica, o número de chamadas chega com uma frequência média de 8 chamadas por minuto. Qual a probabilidade da central receber 10 chamadas no próximo minuto?
- Como calcular?

λ = 8 chamadas por minuto

PMF: $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$

P(X = 10) = 0.0993

Característica	Bernoulli	Binomial	Poisson
O que modela	Resultado de uma única tentativa binária	Número de sucessos em n tentativas	Número de eventos em um intervalo fixo
Número de Tentativas	Uma	Fixas (n)	Não aplicável (taxa por intervalo)
Independência das Tentativas	Implícita (uma única tentativa)	Sim	Sim (eventos independentes)
Probabilidade de Sucesso (p)	Fixa	Fixa para cada tentativa	Não diretamente um p (usa taxa λ)
Parâmetros	p (probabilidade de sucesso)	n (número de tentativas), p	λ (taxa média)
Suporte (Valores Possíveis)	{0,1}	{0,1,...,n}	{0,1,2,...}
PMF (Fórmula)	$P(X=1)=p$ $P(X=0)=1-p$	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$

Biblioteca - Scipy

É uma biblioteca fundamental e de código aberto em Python para computação científica e técnica. Ela se baseia na biblioteca NumPy e fornece uma vasta coleção de algoritmos matemáticos e funções de alto nível.

scipy.stats: Este é o submódulo crucial para probabilidade e estatística. Ele contém:

- Distribuições de Probabilidade: Implementações de diversas distribuições discretas (Bernoulli, Binomial, Poisson, etc.) e contínuas (Normal, Exponencial, Gama, Beta, etc.).
- Estatísticas Descritivas: Funções para calcular medidas como média, mediana, moda, variância, desvio padrão, percentis, etc.
- Testes Estatísticos: Uma ampla gama de testes estatísticos (t-testes, ANOVA, testes de qui-quadrado, testes de normalidade, etc.) para realizar inferência estatística e comparar grupos de dados.
- Correlação e Regressão: Funções para calcular coeficientes de correlação e realizar análises de regressão linear.

Biblioteca - SciPy

Instalação pip install scipy

Importação from scipy import stats

obs.: não é recomendado importar o scipy inteiro

Biblioteca - Matplotlib

Distribuição de Bernoulli

Função bernoulli

bernoulli.pmf(k, p): Calcula a Função de Probabilidade de Massa (PMF) para um dado valor k (que pode ser 0 para fracasso ou 1 para sucesso) e a probabilidade de sucesso p.

Biblioteca - Matplotlib

Distribuição de Bernoulli

Exemplo

Considere que a probabilidade de chuva para o dia de amanhã é 45%

```
from scipy import stats
p_chuva = 0.45
prob_chuva = stats.bernoulli.pmf(1, p_chuva)
prob_nao_chuva = stats.bernoulli.pmf(0, p_chuva)
```

Biblioteca - Matplotlib

Distribuição Binomial

Função binom

binom.pmf(k, n, p): Calcula a Função de Probabilidade de Massa (PMF) para um dado número de sucessos k, o número total de tentativas n, e a probabilidade de sucesso p. Retorna a probabilidade de obter exatamente k sucessos em n tentativas.

Biblioteca - Matplotlib

Distribuição Binomial

Exemplo

Suponha que a probabilidade de um cliente que visita uma loja online comprar determinado produto seja de 15%. Se 20 clientes visitam a loja em um dia:
Qual a probabilidade de exatamente 3 clientes comprarem?

```
from scipy import stats
n_clientes = 20
p_compra = 0.15

# Probabilidade de exatamente 3 clientes comprarem (PMF)
prob_exato_3 = stats.binom.pmf(3, n_clientes, p_compra)
```

Biblioteca - Matplotlib

Distribuição de Poisson

Função poisson

poisson.pmf(k, lambda): Calcula a Função de Probabilidade de Massa (PMF) para um dado número de eventos k e a taxa média lambda. Retorna a probabilidade de ocorrerem exatamente k eventos no intervalo.

Biblioteca - Matplotlib

Distribuição de Poisson

Exemplo

Em média, 8 clientes chegam por hora em uma loja de artesanato. Assumindo que o número de chegadas segue uma distribuição de Poisson qual a probabilidade de exatamente 5 clientes chegarem em uma hora?

```
from scipy import stats
taxa_media = 8  # Clientes por hora
# Probabilidade de exatamente 5 clientes chegarem (PMF)
prob_exato_5 = stats.poisson.pmf(5, taxa_media)
```

Distribuições de Probabilidade Contínuas

Funções que permitem calcular a probabilidade de uma variável aleatória, pertencente a determinado intervalo (Função de densidade de probabilidade)
A probabilidade de variáveis aleatórias discretas é dada pela área abaixo da curva de função de densidade de probabilidade - Integral.

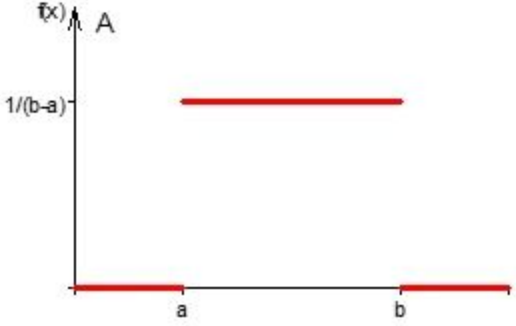
Distribuições de Probabilidade Contínuas

➔ **Distribuição Uniforme contínua:**

- Todos os valores dentro de um intervalo [a,b] têm a mesma probabilidade.
- Função de Densidade de Probabilidade (PDF):

• $f(x) = \frac{1}{b-a}$ para $a \leq x \leq b$

• $f(x) = 0$ caso contrário



Distribuições de Probabilidade Contínuas

- **Distribuição Uniforme:**
  - Helena vai todos os dias de ônibus até a faculdade. Há somente um ônibus que lhe serve, e ele passa no ponto entre as 7h e 7h30. Se Helena chegar no ponto às 7h24, qual a probabilidade de que ela consiga pegar o ônibus?

    X: instante de chegada do ônibus (0 < X < 30)
    P(X>24) = ??

- $f(x) = \frac{1}{b-a}$ para $a \leq x \leq b$
- $f(x) = 0$ caso contrário

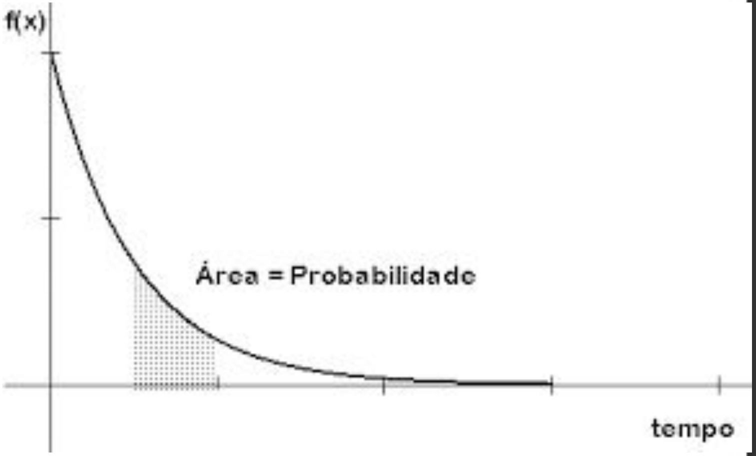
a = 0
b = 30
f(x) = 1/ 30, p/ 0 <x <30
P(x>24) = P(24<x<30) = 6*1/ 30 = 1/ 5 = 20%
(Integral da área do gráfico)

Distribuições de Probabilidade Contínuas

➔ **Distribuição Exponencial:**

- Modela o tempo entre eventos em um processo de Poisson (onde os eventos ocorrem a uma taxa constante λ).

PDF: $f(x) = \lambda e^{-\lambda x}$ para $x \geq 0$



Exemplo

- ➔ Imagine que você está analisando o tempo que os clientes esperam na linha telefônica antes de serem atendidos por um representante. Após coletar dados por um longo período, você observa que, em média, chegam 5 chamadas por minuto ao sistema.

Se chegam 5 chamadas por minuto, a taxa média de chegadas é lambda = 5 chamadas/minuto.

Qual a probabilidade de que o tempo entre duas chamadas consecutivas seja menor que 0.5 minutos (30 segundos)? 0.918 - 91,8%

Qual a probabilidade de que o tempo entre duas chamadas consecutivas seja maior que 1 minuto? 0.0067 - 0.67%

Distribuições de Probabilidade Contínuas

➔ **Distribuição Normal (Gaussiana):**

- Uma das distribuições mais importantes, caracterizada por sua forma de sino simétrica em torno da média (μ) e com dispersão controlada pelo desvio padrão (σ).

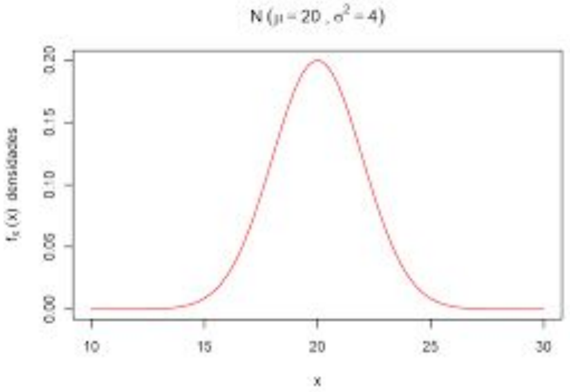
$$\text{PDF: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Distribuições de Probabilidade Contínuas

➔ **Distribuição Normal (Gaussiana):**

- Uma das distribuições mais importantes, caracterizada por sua forma de sino simétrica em torno da média (μ) e com dispersão controlada pelo desvio padrão (σ).
- Moda = Mediana = Média

PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$



Distribuições de Probabilidade Contínuas

➔ **Distribuição Normal (Gaussiana) - Características**

- Forma de Sino: A representação gráfica da distribuição normal tem um formato característico de sino, simétrico em torno da sua média. O ponto mais alto da curva corresponde à média

- Simetria: A distribuição é perfeitamente simétrica em relação à sua média. Isso significa que metade dos dados está abaixo da média e a outra metade está acima.

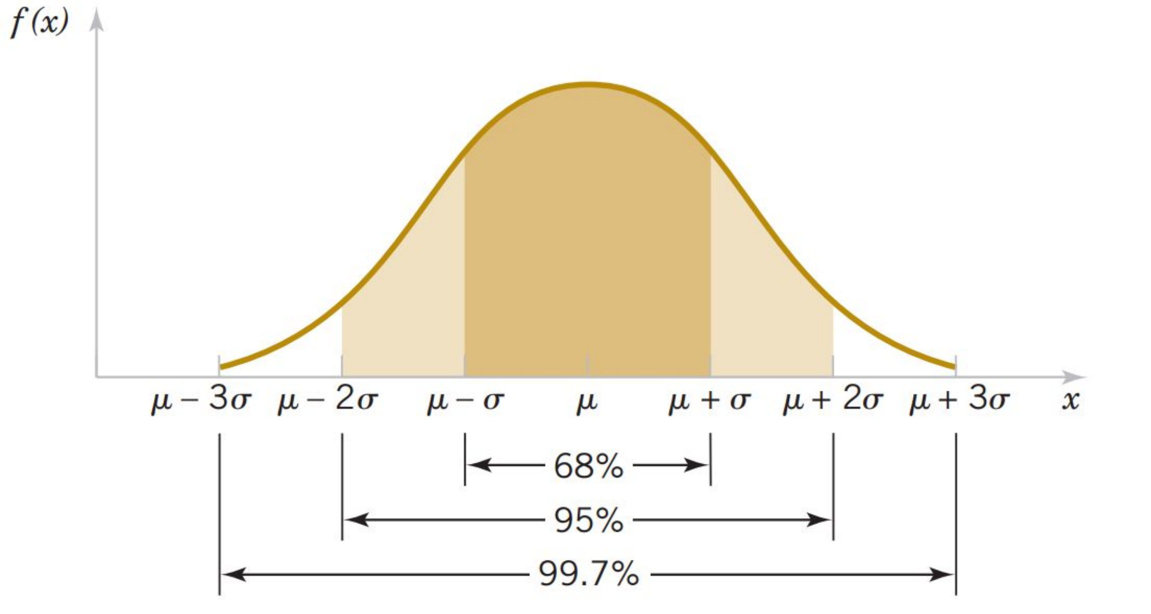
- Unimodal: A distribuição normal tem apenas um pico, que ocorre na média.

Distribuições de Probabilidade Contínuas

- **Distribuição Normal (Gaussiana) - Características**
  - Definida por Dois Parâmetros: Uma distribuição normal específica é completamente definida por dois parâmetros:
    - **Média:** Determina o centro da distribuição. O sino é centrado neste valor.
    - **Desvio Padrão:** Determina a dispersão ou a largura da distribuição. Um desvio padrão maior resulta em uma curva mais larga e achatada, indicando maior variabilidade nos dados. Um desvio padrão menor resulta em uma curva mais estreita e alta, indicando menor variabilidade.

Distribuições de Probabilidade Contínuas

➔ Distribuição Normal (Gaussiana) - Regra de ouro



Exemplo

- Imagine que você está conduzindo um estudo em escolas de ensino médio e decide medir a altura de todos os estudantes. Depois de coletar e organizar todos os dados, você cria um histograma para visualizar a distribuição das alturas.
- O que você provavelmente observaria é o seguinte:
  - A maioria dos estudantes teria alturas em torno de um valor central. Por exemplo, muitas alturas estariam próximas de 1,65 metros.
  - À medida que você se afasta desse valor central em ambas as direções (alturas maiores e menores), o número de estudantes com essas alturas diminuiria gradualmente. Poucos estudantes seriam muito baixos e poucos seriam muito altos
  - O histograma das alturas formaria uma forma aproximadamente simétrica, com um pico no meio e as barras diminuindo suavemente em ambos os lados.

Exemplo

- Média: O pico da curva (a altura mais comum) representaria a média da altura dos estudantes. Digamos que, neste exemplo, a média seja de 1,65 metros.
- Desvio Padrão: A "largura" da curva de sino nos diria sobre a variabilidade das alturas. Um desvio padrão menor significaria que as alturas dos estudantes estão mais agrupadas em torno da média. Um desvio padrão maior indicaria uma maior dispersão nas alturas. Vamos supor que o desvio padrão das alturas seja de 0,08 metros (8 centímetros).

Exemplo

Com isso podemos fazer algumas afirmações sobre as alturas dos estudantes:

- Regra dos 68-95-99.7:
  - Aproximadamente 68% dos estudantes teriam alturas entre (1.65 - 0.08 = 1.57) metros e (1.65 + 0.08 = 1.73) metros.
  - Aproximadamente 95% dos estudantes teriam alturas entre (1.65 - (2 *0.08) = 1.49) metros e (1.65 + (2 *0.08) = 1.81) metros.
  - Aproximadamente 99.7% dos estudantes teriam alturas entre (1.65 - (3* 0.08) = 1.41) metros e (1.65 + (3 *0.08) = 1.89) metros.

Biblioteca - Matplotlib

Distribuição Uniforme

Função uniform

uniform(loc=0, scale=1): Define a distribuição uniforme

loc = a e scale = b-a.

rvs(size=x): Gera amostras aleatórias da distribuição.

size = número de amostras a serem geradas.

.cdf(x, loc=0): Calcula a função de distribuição cumulativa (CDF) no ponto x. Retorna a probabilidade de que a variável aleatória seja menor ou igual a x.

Biblioteca - Matplotlib

Distribuição Uniforme

Exemplo

Defina valores de distribuição uniforme entre 5 a 15 minutos e calcule:

1 - A probabilidade de esperar menos de 8 minutos

2 - A probabilidade de esperar entre 7 e 12 minutos

Biblioteca - Matplotlib

Distribuição Uniforme

Exemplo

```python
from scipy import stats

# Definindo os parâmetros da distribuição uniforme
a = 5  # Limite inferior do intervalo (minutos)
b = 15 # Limite superior do intervalo (minutos)

# Criando a distribuição uniforme
dist_uniforme = stats.uniform(loc=a, scale=b- a)

# A probabilidade de esperar menos de 8 minutos
probabilidade_menos_8 = dist_uniforme.cdf(8)

# A probabilidade de esperar entre 7 e 12 minutos
probabilidade_entre_7_e_12 = dist_uniforme.cdf(12) - dist_uniforme.cdf(7)

# Imprimindo o resultado
print("A probabilidade de esperar menos de 8 minutos é: ",probabilidade_menos_8)
print("A probabilidade de esperar entre 7 e 12 minutos é: ",probabilidade_entre_7_e_12)
```

Biblioteca - Matplotlib

Distribuição Normal

Função norm

norm(loc=media, scale=desvio_padrao) define a distribuição normal com base na média e desvio padrão

rvs(loc=media, scale=desv_padrao, size=tam): Gera amostras aleatórias da distribuição normal.

size = número de amostras

cdf(x, loc=media): Calcula a função de distribuição cumulativa (CDF) no ponto x. Retorna a probabilidade de que a variável aleatória seja menor ou igual a x.

Biblioteca - Matplotlib

Distribuição Normal

Exemplo

Considere um exemplo onde a altura dos homens adultos aqui em Recife segue aproximadamente uma distribuição normal com uma média de 1,75 metros e um desvio padrão de 0,08 metros.

Biblioteca - Matplotlib

Distribuição Normal

Exemplo

```
from scipy import stats

# Parâmetros da distribuição normal
media = 1.75  # metros
desvio_padrao = 0.08  # metros

# Criando a distribuição normal
dist_altura = stats.norm(loc=media, scale=desvio_padrao)

# Calculando a probabilidade de um homem ser menor que 1,70m
probabilidade = dist_altura.cdf(1.70)
print("A probabilidade de um homem ser menor que 1,70m é: ", probabilidade)

# Calculando a probabilidade de um homem ter entre 1,70m e 1,80m
probabilidade = dist_altura.cdf(1.80) - dist_altura.cdf(1.70)
print("A probabilidade de um homem ter entre 1,70m e 1,80m é: ", probabilidade)
```

Biblioteca - Matplotlib

Distribuição Exponencial

Função expon

expon(loc=0, scale=lambda): Define a distribuição exponencial

scale = lambda (frequencia)

loc = ponto de início da distribuição

rvs(size=tam): Gera amostras aleatórias da distribuição normal.

size = número de amostras

Biblioteca - Matplotlib

Distribuição Exponencial

Exemplo

Vamos considerar um exemplo aqui em Recife onde o tempo entre a chegada de ônibus em um ponto específico segue uma distribuição exponencial com uma taxa média de lambda = 0.1 ônibus por minuto (ou seja, em média, chega um ônibus a cada (1 / 0.1 = 10) minutos).

Biblioteca - Matplotlib

Distribuição Exponencial

Exemplo

```
from scipy.stats import expon

# Definindo o parâmetro de taxa (lambda) e o parâmetro de escala (1/lambda)
taxa_lambda = 0.1
escala_lambda = 1 / taxa_lambda

# Criando o objeto da distribuição exponencial
dist_exponencial = expon(loc=0, scale=escala_lambda)

# 1. Gerar 5 amostras aleatórias do tempo entre chegadas
amostras_tempo = dist_exponencial.rvs(size=100)

# 2. Calcular a probabilidade de o próximo ônibus chegar em menos de 5 minutos
prob_menor_5 = dist_exponencial.cdf(5)
print("Probabilidade de o próximo ônibus chegar em < 5 minutos: ",prob_menor_5)

# 3. Calcular a probabilidade de esperar mais de 15 minutos pelo próximo ônibus
prob_maior_15 = 1 - dist_exponencial.cdf(15)
print("Probabilidade de esperar > 15 minutos: ", prob_maior_15)
```

Distribuição	Tipo	Suporte	Parâmetros	Função Principal (SciPy)
Bernoulli	Discreta	{0,1}	p (probabilidade de sucesso)	binom.pmf(k, 1, p)
Binomial	Discreta	{0,1,...,n}	n (número de tentativas), p (prob. sucesso)	binom.pmf(k, n, p)
Poisson	Discreta	{0,1,2,...}	λ (taxa média de ocorrência)	poisson.pmf(k, mu)
Uniforme (Contínua)	Contínua	[a,b]	a (mínimo), b (máximo)	uniform.pdf(x, a, b)
Normal (Gaussiana)	Contínua	(−∞,∞)	μ (média), σ (desvio padrão)	norm.pdf(x, loc=mu, scale=sigma)
Exponencial	Contínua	[0,∞)	λ (taxa) ou scale (1/λ)	expon.pdf(x, scale=1/lambda)

Obrigada!
Bons estudos
💪