

## Exercícios de Estatística Básica 1

Professores: Sandro Fonseca, Maurício Thiel, Eliza Melo

Name: Isis Prazeres Mota

## EXERCICIO 1

TEXTO

1 CODIGO USADO PARA RESOLVER O PROBLEMA

TEXTO

## EXERCICIO 4

Eleger cinco partículas do Particle Data Group (PDG) e combinar os resultados para suas massas. Para combinar resultados entre medidas, usamos as seguintes estimativas para o valor esperado ( $\bar{x}$ ) e erro associado ( $\sigma$ ) definidas a seguir:

$$\bar{x} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}}$$

$$\sigma_{\bar{x}} = \sigma = \frac{1}{\sqrt{\sum_{i=1}^N \frac{1}{\sigma_i^2}}}$$

J	K	L	M	N	O	P	Q	R	S
Elétron		Muon		N Baryon p		N Baryon n		Mesons pi+-	
x_i	sigma_i	x_i	sigma_i	x_i	sigma_i	x_i	sigma_i	x_i	sigma_i
548,579909065	0,000000016	0,1134289259	0,0000000025	1,0072764666	0,000000000053	1,00866491595	0,00000000049	139,57039	0,00018
548,579909070	0,000000016	0,1134289257	0,0000000025	1,0072764666	0,000000000010	1,00866491588	0,00000000049	139,57039	0,00017
548,57990946	0,000000022	0,1134289267	0,0000000029	1,0072764666	0,000000000053	1,00866491600	0,00000000043	139,57021	0,00014
548,57990943	0,000000023	0,1134289256	0,0000000029	1,0072764666	0,000000000033	1,00866491597	0,00000000043	139,57077	0,00018
548,57990945	0,000000024	0,1134289264	0,0000000030	1,0072764666	0,000000000032	1,00866491560	0,00000000033	139,57071	0,00053
548,5799092	0,00000004	0,1134289168	0,0000000034	1,0072764669	0,000000000091	1,00866491578	0,00000000033	139,56995	0,00035
548,5799110	0,00000012	0,113428913	0,0000000017	1,007276467	0,000000000090	1,008665904	0,0000000014	139,57022	0,00014
548,5799111	0,00000012			1,007276467	0,000000000010			139,56782	0,00037
548,579903	0,0000013			1,007276467	0,000000000013			139,56996	0,00067
				1,007276467	0,000000000013			139,56752	0,00037
				1,007276470	0,0000000012			139,5704	0,00110
								139,5664	0,00090
								139,5686	0,0020
								139,5660	0,0024

Para encontrar o valor esperado e o erro associado de cada partícula, no excel, eu:

1. Liste todos os valores de  $x_i$  e todos os valores de  $\sigma$ .
2. Na sequência, peguei cada valor de  $x$  e dividi por cada valor de  $\sigma^2$ .
3. Fiz o somatório.
4. Dividi 1 por  $\sigma^2$ .
5. Fiz o somatório.

6. A partir dos somatórios, eu os dividi, achando, assim, os valores de  $\bar{x}$ .
7. Depois eu dividi 1 pela raiz de  $\sigma^2$ , achando, com isso, o erro associado.

E os valores encontrados foram:

- Elétron:

$$\bar{x} = 548,579909070693$$

$$\sigma_{\bar{x}} = 0,00000001126724218$$

- Muon:

$$\bar{x} = 0,113428924913$$

$$\sigma_{\bar{x}} = 0,00000000114791829$$

- N Baryon p:

$$\bar{x} = 1,007276466584$$

$$\sigma_{\bar{x}} = 0,00000000000882126$$

- N Baryon n:

$$\bar{x} = 1,008664915959$$

$$\sigma_{\bar{x}} = 0,00000000016326211$$

- Mesons pi+-:

$$\bar{x} = 139,570349154630$$

$$\sigma_{\bar{x}} = 0,00006899102836307$$

### EXERCICIO 7

Dentre os calouros de uma Universidade, 2587 são alunos e 2832 são alunas. Inscreveram-se nos cursos da área tecnológica 1291 alunos e 547 alunas.

Determine a probabilidade de se sortear aleatoriamente um estudante do sexo masculino da área tecnológica. Para encontrar a probabilidade de se sortear aleatoriamente um estudante do sexo masculino da área tecnológica, precisamos seguir alguns passos:

1. Calcular o total de estudantes: Para isso, somamos o número de alunos e alunas.  
Total de estudantes = 2587(alunos) + 2832(alunas) = 5419
2. Identificar o total de alunos da área tecnológica: Conforme o enunciado, temos 1291 alunos inscritos nos cursos da área tecnológica.
3. Calcular a probabilidade: A probabilidade de se sortear um estudante do sexo masculino da área tecnológica é dada pela razão entre o número de alunos da área tecnológica e o total de estudantes.

$$Probabilidade = \frac{\text{Alunos da área tecnológica}}{\text{Total de estudantes}}$$

$$Probabilidade = \frac{1291}{5419}$$

4. Agora, vamos calcular o valor:  $Probabilidade = \frac{1291}{5419} \approx 0,2383$

Portanto, a probabilidade de se sortear aleatoriamente um estudante do sexo masculino da área tecnológica é aproximadamente 0,2383, ou 23,83%.

### EXERCICIO 8

Em uma cidade, 15% dos táxis são azuis e o restante são verdes. Em uma noite, um táxi atropelou uma pessoa e fugiu.

Uma testemunha identificou como azul o táxi envolvido no acidente. A polícia constatou que, nas mesmas circunstâncias da noite do acidente, essa testemunha identificou cada cor corretamente em 80% das vezes, e confundiu as cores em 20% das vezes.

Determine a probabilidade de ter sido azul o táxi envolvido no acidente.

Para resolver essa questão, vamos usar o Teorema de Bayes, que nos ajuda a calcular a probabilidade de um evento baseado nos conhecimentos que temos.

Aqui estão os eventos e probabilidades importantes para nossa análise:

- $P(\text{Azul})$ : a probabilidade de um táxi ser azul, que é 15% ou 0,15.
- $P(\text{Verde})$ : a probabilidade de um táxi ser verde. Como só existem táxis azuis e verdes, e 15% são azuis, então 85% ou 0,85 dos táxis são verdes.
- $P(\text{Testemunha}|\text{Azul})$ : a probabilidade de a testemunha identificar o táxi como azul quando ele é de fato azul, que é 80% ou 0,8.
- $P(\text{Testemunha}|\text{Verde})$ : a probabilidade de a testemunha identificar o táxi como azul quando ele é de fato verde, que é 20% ou 0,2 (confundindo as cores).

Queremos encontrar a probabilidade de o táxi ser azul dado que a testemunha disse que era azul, ou seja,

$P(\text{Azul}|\text{Testemunha})$ . Podemos usar o Teorema de Bayes para isso:

$$P(\text{Azul}|\text{Testemunha}) = \frac{P(\text{Testemunha}|\text{Azul}) * P(\text{Azul})}{P(\text{Testemunha})}$$

O denominador,  $P(\text{Testemunha})$ , pode ser calculado considerando ambas as formas de a testemunha ter visto um táxi azul - seja porque o táxi era realmente azul, ou porque ela erroneamente identificou um táxi verde como sendo azul.

Portanto:  $P(\text{Testemunha}) = P(\text{Testemunha}|\text{Azul}) * P(\text{Azul}) + P(\text{Testemunha}|\text{Verde}) * P(\text{Verde})$

Substituindo os valores:  $P(\text{Testemunha}) = (0,8 * 0,15) + (0,2 * 0,85) = 0,12 + 0,17 = 0,29$

Agora podemos calcular  $P(\text{Azul}|\text{Testemunha})$ :

$$P(\text{Azul}|\text{Testemunha}) = \frac{0,8 * 0,15}{0,29} = \frac{0,12}{0,29} \approx 0,4138$$

Isso significa que, dada a declaração da testemunha, a probabilidade de o táxi envolvido no acidente ser azul é aproximadamente 41,38%.

### EXERCICIO 9

Enquanto 7% das mamografias identificam um caso de câncer quando ele não existe (taxa de falsos-positivos), 10% não identificam a doença quando ela existe (taxa de falsos-negativos). Sabendo que a incidência de câncer sobre a população feminina é cerca de 0,8%, determine a probabilidade de que uma mulher esteja com câncer ao receber um resultado de teste positivo.

Para resolver essa questão, usaremos Bayes.

Primeiro, vamos definir as probabilidades:

- $P(\text{Câncer}) = 0,008$ : a probabilidade de uma mulher ter câncer, ou seja, a incidência do câncer na população feminina.
- $P(\text{Não Câncer}) = 1 - P(\text{Câncer}) = 0,992$ : a probabilidade de uma mulher não ter câncer.
- $P(\text{Positivo}|\text{Câncer}) = 1 - 0,1 = 0,9$ : a probabilidade de um teste dar positivo se a mulher realmente tem câncer, ou seja, 1 menos a taxa de falsos-negativos.
- $P(\text{Positivo}|\text{Não Câncer}) = 0,07$ : a probabilidade de um teste dar positivo se a mulher não tem câncer, ou seja, a taxa de falsos-positivos.

Queremos encontrar a probabilidade de uma mulher ter câncer dado que ela recebeu um resultado positivo no teste, ou seja,  $P(\text{Câncer}|\text{Positivo})$ .

- De acordo com o Teorema de Bayes, temos:

$$P(\text{Câncer}|\text{Positivo}) = \frac{P(\text{Positivo}|\text{Câncer}) * P(\text{Câncer})}{P(\text{Positivo})}$$

- Aqui,  $P(\text{Positivo})$  é a probabilidade de um teste dar positivo, que pode ocorrer tanto se a mulher tem câncer quanto se não tem. Então, podemos calcular  $P(\text{Positivo})$  da seguinte maneira:

$$P(\text{Positivo}) = P(\text{Positivo}|\text{Câncer}) * P(\text{Câncer}) + P(\text{Positivo}|\text{Não Câncer}) * P(\text{Não Câncer})$$

- Substituindo os valores fornecidos:

$$P(\text{Positivo}) = (0,9 * 0,008) + (0,07 * 0,992) = 0,0072 + 0,06944 = 0,07664$$

- Agora, podemos calcular  $P(\text{Câncer}|\text{Positivo})$ :

$$P(\text{Câncer}|\text{Positivo}) = \frac{(0,9 * 0,008)}{0,07664} = \frac{0,0072}{0,07664} \approx 0,0939$$

Isso significa que a probabilidade de uma mulher estar com câncer ao receber um resultado de teste positivo é aproximadamente 9,39%.

### EXERCICIO 10

Três urnas têm a seguinte composição: a primeira contém 5 bolas brancas e 6 pretas, a segunda contém 4 brancas e 5 pretas, e a terceira 4 brancas e 4 pretas.

Após escolher por acaso uma urna e se retirar uma bola preta, determine a probabilidade de que a sorteada tenha sido extraída da terceira urna.

Para determinar a probabilidade de que uma bola preta tenha sido extraída da terceira urna, dada que uma bola preta foi sorteada, usaremos o Teorema de Bayes. Vamos definir os eventos:

- $A_i$ : o evento de escolher a urna  $i$ , para  $i = 1, 2, 3$ . Como a urna é escolhida ao acaso, cada uma tem uma probabilidade de  $\frac{1}{3}$  de ser escolhida, ou seja,  $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$ .
- $B$ : o evento de sortear uma bola preta.

Precisamos calcular  $P(A_3|B)$ , a probabilidade de a urna escolhida ser a terceira, dado que uma bola preta foi sorteada.

Primeiro, vamos calcular  $P(B|A_i)$  para cada urna, ou seja, a probabilidade de sortear uma bola preta de cada urna:

- Para a primeira urna,  $P(B|A_1) = \frac{6}{11}$  pois há 6 bolas pretas em um total de 11 bolas.
- Para a segunda urna,  $P(B|A_2) = \frac{5}{9}$  pois há 5 bolas pretas em um total de 9 bolas.
- Para a terceira urna,  $P(B|A_3) = \frac{4}{8} = \frac{1}{2}$  pois há 4 bolas pretas em um total de 8 bolas.

Agora, usaremos Bayes e a Lei Total da Probabilidade para calcular  $P(A_3|B)$ :

$$P(A_3|B) = \frac{P(B|A_3) * P(A_3)}{P(B)}$$

Onde  $P(B)$  é a probabilidade total de sortear uma bola preta, calculada pela Lei Total da Probabilidade:

$$P(B) = P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + P(B|A_3) * P(A_3)$$

Substituindo os valores conhecidos:

$$P(B) = \frac{6}{11} * \frac{1}{3} + \frac{5}{9} * \frac{1}{3} + \frac{1}{2} * \frac{1}{3}$$

$$\text{Calculando } P(B): P(B) = \frac{6}{11} + \frac{5}{27} + \frac{1}{6} = \frac{6}{33} + \frac{5}{27} + \frac{11}{66} = \frac{12}{66} + \frac{10}{66} + \frac{11}{66} = \frac{33}{66} = \frac{1}{2}$$

Agora, calculamos  $P(A_3|B)$ :

$$P(A_3|B) = \frac{\frac{1}{2} * \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

Portanto, a probabilidade de que a bola preta tenha sido extraída da terceira urna, dado que uma bola preta foi sorteada, é  $\frac{1}{3}$  ou aproximadamente 33.33%.

### EXERCICIO 11

Seja  $x$  uma variável contínua, como as possíveis posições de uma partícula confinada em uma região de dimensão  $a$ , cuja densidade de probabilidade  $\rho(x)$  é proporcional a função  $\sin^2 \frac{\pi}{a} x$ . Determine o valor médio e o desvio padrão associados à variável  $x$ .

Para determinar o valor médio ( $\langle x \rangle$ ) e o desvio padrão ( $\sigma_x$ ) de uma variável contínua  $x$ , dada uma densidade de probabilidade proporcional a  $\sin^2 \frac{\pi}{a} x$  num intervalo de 0 a  $a$ , precisamos primeiro normalizar a função de densidade de probabilidade  $\rho(x)$  e depois usar as definições do valor médio e do desvio padrão na mecânica estatística e na teoria das probabilidades.

### Normalização da Função de Densidade de Probabilidade

A função de densidade de probabilidade  $\rho(x)$  é proporcional a  $\sin^2 \frac{\pi}{a} x$ , então podemos escrever:

$$\rho(x) = k \sin^2 \frac{\pi}{a} x$$

Onde  $k$  é a constante de proporcionalidade. Para normalizar  $\rho(x)$  no intervalo de 0 a  $a$ , usamos a condição de normalização:

$$\int_0^a \rho(x) dx = 1$$

Substituindo a expressão de  $\rho(x)$  e resolvendo a integral, temos:

$$\int_0^a k \sin^2 \frac{\pi}{a} x dx = 1$$

A integral de  $\sin^2$  pode ser resolvida utilizando a identidade trigonométrica  $\sin^2(\theta) = \frac{1 - \cos(2\theta)}{2}$ , então:

$$\begin{aligned} k \int_0^a \frac{1 - \cos\left(\frac{2\pi}{a} x\right)}{2} dx &= 1 \\ \frac{k}{2} \left[ x - \frac{a}{2\pi} \sin\left(\frac{2\pi}{a} x\right) \right]_0^a &= 1 \\ \frac{k}{2} [a - 0] &= 1 \\ k &= \frac{2}{a} \end{aligned}$$

Então, a função de densidade de probabilidade normalizada é:

$$\rho(x) = \frac{2}{a} \sin^2 \frac{\pi}{a} x$$

### Valor Médio ( $\langle x \rangle$ ):

O valor médio de  $x$  é dado por:  $\langle x \rangle = \int_0^a x \rho(x) dx = \frac{2}{a} \int_0^a x \sin^2 \frac{\pi}{a} x dx$

Usando python temos:

```
1 import sympy as sp
2
3 # Definindo as variaveis
4 x, a = sp.symbols('x a')
5
6 # Definindo a funcao de densidade de probabilidade normalizada
7 rho_x = (2/a) * sp.sin(sp.pi*x/a)**2
8
9 # Calculando o valor medio de x
10 valor_medio_x = sp.integrate(x * rho_x, (x, 0, a))
11
12 print(f"Valor medio de x: {valor_medio_x}")
```

Este código define  $x$  e  $a$  como símbolos matemáticos e especifica  $\rho(x)$ , a função de densidade de probabilidade normalizada. Em seguida, ele calcula o valor médio de  $x$  ao integrar  $x * \rho(x)$  de 0 a  $a$ .

A integração simbólica deve fornecer uma expressão para o valor médio de  $x$  em termos de  $a$ , que é o que esperamos baseando-se na formulação do problema.

Rodando este código Python, o resultado será uma expressão para o valor médio  $\langle x \rangle$  em termos de  $a$ . Para este caso específico, o resultado esperado é que o valor médio seja  $\frac{a}{2}$ , pois a função de densidade de probabilidade é simétrica em torno de  $\frac{a}{2}$  dentro do intervalo  $[0, a]$ , indicando que a partícula tem igual probabilidade de ser encontrada em qualquer ponto dentro deste intervalo, resultando em um valor médio no ponto médio do intervalo.

### Desvio Padrão ( $\sigma_x$ ):

O desvio padrão é calculado pela raiz quadrada da variância, que é a média dos quadrados das diferenças do valor de cada ponto em relação à média,  $\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$ , onde  $\langle x^2 \rangle$  é o valor médio do quadrado de  $x$ .

Usando Python, temos:

```

1 import sympy as sp
2
3 # Definindo as variaveis
4 x, a = sp.symbols('x a')
5
6 # Definindo a funcao de densidade de probabilidade normalizada
7 rho_x = (2/a) * sp.sin(sp.pi*x/a)**2
8
9 # Calculando o valor medio de x
10 valor_medio_x = sp.integrate(x * rho_x, (x, 0, a))
11
12 # Calculando o valor medio de x^2
13 valor_medio_x2 = sp.integrate(x**2 * rho_x, (x, 0, a))
14
15 # Calculando o desvio padrao
16 sigma_x = sp.sqrt(valor_medio_x2 - valor_medio_x**2)
17
18 print(f"Valor medio de x: {valor_medio_x}")
19 print(f"Desvio padrao de x: {sigma_x}")

```

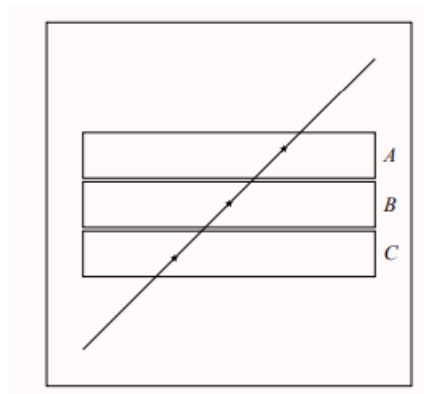
Este script realiza os seguintes passos:

1. Define  $x$  e  $a$  como símbolos matemáticos e a densidade de probabilidade  $\rho(x)$ .
2. Calcula  $\langle x \rangle$  integrando  $x * \rho(x)$  de 0 a  $a$ .
3. Calcula  $\langle x^2 \rangle$  integrando  $x^2 * \rho(x)$  de 0 a  $a$ .
4. Calcula o desvio padrão  $\sigma_x$  usando a fórmula  $\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$ .

A execução deste código fornecerá tanto o valor médio de  $x$  quanto o desvio padrão  $\rho(x)$  em termos de  $a$ .

## EXERCICIO 12

A figura abaixo representa um sistema de detecção de múons incidentes, constituído por três tubos de ionização  $A$ ,  $B$  e  $C$ .



Se a localização do ponto de ionização em cada tubo é determinada com 60% de eficiência (ou seja, com probabilidade igual a 0,6) e a reconstrução da trajetória de um múon requer a determinação de pelo menos três pontos em câmaras distintas, a eficiência do sistema é dada por  $B(3|3;0.6) \Rightarrow 21.6\%$ . Determine as eficiências para sistemas compostos por quatro e cinco câmaras.

A eficiência do sistema de detecção de múons é calculada usando a fórmula da probabilidade binomial:  $[B(k|n; p) = \binom{n}{k} p^k (1-p)^{n-k}]$  onde:

- $(n)$  é o número total de tentativas,
- $(k)$  é o número de tentativas bem-sucedidas,
- $(p)$  é a probabilidade de sucesso em uma única tentativa.

Para um sistema com quatro câmaras, precisamos de pelo menos três detecções bem-sucedidas para reconstruir a trajetória do múon. Calculamos para 3 e 4 sucessos:

$$[B(3|4; 0.6) = \binom{4}{3} (0.6)^3 (0.4)^1 = 0.3456]$$

$$[B(4|4; 0.6) = \binom{4}{4} (0.6)^4 (0.4)^0 = 0.1296]$$

Somando essas probabilidades, obtemos uma eficiência de 47,52%.

Para um sistema com cinco câmaras, calculamos para 3, 4 e 5 sucessos:

$$[B(3|5; 0.6) = \binom{5}{3} (0.6)^3 (0.4)^2 = 0.3456]$$

$$[B(4|5; 0.6) = \binom{5}{4} (0.6)^4 (0.4)^1 = 0.2592]$$

$$[B(5|5; 0.6) = \binom{5}{5} (0.6)^5 (0.4)^0 = 0.0576]$$

A eficiência para um sistema de cinco câmaras é de 5,76%.

### EXERCICIO 13

Qual a probabilidade de que dentre 720 pessoas duas aniversariem num mesmo dia? (compare binomial e Poisson)

#### Abordagem Binomial

A abordagem direta usando a distribuição binomial é impraticável aqui, porque estamos lidando com um grande número de tentativas (720 pessoas) e cada tentativa tem uma pequena probabilidade de sucesso (compartilhar o mesmo aniversário com outra pessoa específica). Além disso, a formulação binomial padrão não se aplica de forma simples a este problema, devido à complexidade de calcular diretamente as probabilidades de combinações específicas de aniversários. Portanto, a abordagem mais comum é calcular a probabilidade do evento complementar (nenhuma pessoa compartilhando o mesmo aniversário) e subtrair isso de 1.

#### Abordagem de Poisson

A aproximação de Poisson é útil para eventos raros em um grande número de tentativas, o que parece adequado dada a grande quantidade de pessoas. No entanto, a aplicação direta do modelo de Poisson para este problema específico também não é direta, pois estamos interessados na probabilidade de um evento raro (pelo menos duas pessoas compartilhando o mesmo aniversário) em um contexto específico de distribuição uniforme de aniversários ao longo do ano.

O cálculo exato para este problema é mais comumente abordado pelo princípio do complemento, como mencionado. Vejamos primeiro essa abordagem e, em seguida, discutiremos como as aproximações podem ser consideradas:

**Cálculo via Princípio do Complemento** A probabilidade de duas pessoas não compartilharem o mesmo aniversário é:

1. Para a primeira pessoa, qualquer dia serve, então a probabilidade é de 365/365.
2. Para a segunda pessoa, apenas 364 dias estão disponíveis para evitar o aniversário da primeira, então a probabilidade é de 364/365.
3. Continuando assim, a probabilidade de que 720 pessoas tenham aniversários únicos é:  

$$P(\text{não compartilharemos o mesmo aniversário}) = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365-719}{365}$$

A probabilidade de pelo menos duas pessoas compartilharem o mesmo aniversário é o complemento disso:  
 $P(\text{pelo menos dois compartilharem}) = 1 - P(\text{não compartilharem})$

**Comparação com Poisson e Binomial** A natureza deste problema torna a comparação direta com modelos binomiais e de Poisson não intuitiva, pois ambos os modelos exigem simplificações ou ajustes específicos para aplicação direta. O modelo binomial exige um número fixo de tentativas com probabilidade constante de sucesso, o que não se aplica claramente aqui devido ao aumento progressivo da probabilidade de sobreposição à medida que mais aniversários são adicionados. A aproximação de Poisson, embora útil para eventos raros, não captura facilmente a dependência entre as probabilidades de diferentes pessoas compartilharem o mesmo aniversário sem uma reformulação significativa do problema.

Porém, exemplificando em python, temos:

```

1 def probabilidade_compartilhar_aniversario(n):
2     prob_nao_compartilhar = 1.0
3     for i in range(n):
4         prob_nao_compartilhar *= (365 - i) / 365
5     return 1 - prob_nao_compartilhar
6
7 # Numero de pessoas
8 n = 720
9
10 # Calculando a probabilidade
11 prob = probabilidade_compartilhar_aniversario(n)
12 print(f"A probabilidade de pelo menos duas pessoas dentre {n} compartilharem o mesmo
    aniversario e aproximadamente {prob:.10f}")

```

Esse script define uma função que calcula a probabilidade de que pelo menos duas pessoas em um grupo de  $n$  pessoas compartilhem o mesmo aniversário. Ele faz isso calculando primeiro a probabilidade de que ninguém compartilhe um aniversário (todos tenham aniversários únicos) e então subtrai esse valor de 1.

Com 720 pessoas, essa probabilidade será praticamente 1 (ou 100%) devido ao grande número de pessoas, muito além do "Paradoxo do Aniversário" clássico, que já mostra uma probabilidade significativa com 23 pessoas.

#### EXERCICIO 14

Cada uma das 15 questões de um teste tem 4 alternativas e apenas uma delas é correta. Desse modo, a probabilidade ( $p$ ), a priori, de acerto ao acaso de uma questão é  $1/4$ .

- Determine a distribuição de probabilidades de acertos ao acaso de  $m$  das 15 questões.
- Represente em um histograma.
- Se 1000 alunos fizerem o teste respondendo as questões ao acaso, quantos, em média, acertarão pelo menos 3 questões?

##### a) Distribuição de Probabilidades

A probabilidade de acertar exatamente  $m$  questões de um total de  $n = 15$  questões, onde a probabilidade de acerto por questão é  $p = 1/4$  e a de erro é  $q = 1 - p = 3/4$ , é dada pela fórmula da distribuição binomial:

$$P(X = m) = \binom{n}{m} p^m q^{(n-m)}$$

onde  $\binom{n}{m}$  é o coeficiente binomial, que calcula o número de maneiras de escolher  $m$  sucessos de  $n$  tentativas.

##### b) Histograma

Podemos usar Julia para calcular estas probabilidades para cada valor de  $m$  de 0 a 15 e depois plotar um histograma.

##### c) Média de Alunos Acertando pelo Menos 3 Questões

Para encontrar quantos alunos, em média, acertarão pelo menos 3 questões, precisamos somar as probabilidades de um aluno acertar 3 ou mais questões e multiplicar pelo total de alunos (1000 neste caso).

Usando Julia para fazer os cálculos:

```

1 using Plots
2 using Distributions
3

```



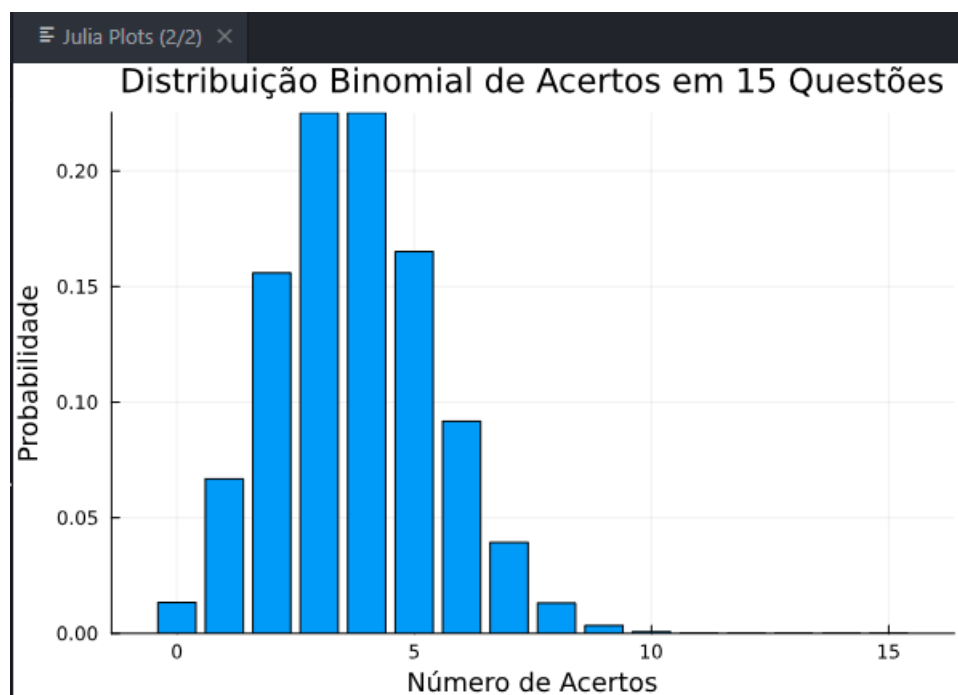
```

4 # Parametros da distribuicao
5 n = 15 # numero de questoes
6 p = 1/4 # probabilidade de acerto por questao
7
8 # Distribuicao binomial
9 distribuicao = Binomial(n, p)
10
11 # Calculando a distribuicao de probabilidades para cada numero de acertos de 0 a 15
12 probabilidades = [pdf(distribuicao, k) for k in 0:n]
13
14 # Parte b: Plotando o histograma
15 bar(0:n, probabilidades, legend=false,
16     xlabel="Numero de Acertos", ylabel="Probabilidade",
17     title="Distribuicao Binomial de Acertos em 15 Questoes")
18
19 # Parte c: Calculando a media de alunos que acertam pelo menos 3 questoes
20 alunos_total = 1000
21 prob_pelo_menos_3 = sum([pdf(distribuicao, k) for k in 3:n])
22 media_alunos_pelo_menos_3 = alunos_total * prob_pelo_menos_3
23
24 println("Media de alunos acertando pelo menos 3 questoes: ",
25         media_alunos_pelo_menos_3)

```

Este script Julia faz o seguinte:

1. Define os parâmetros do problema.
2. Usa a função pdf do tipo Binomial para calcular as probabilidades de acerto de 0 a 15 questões.
3. Plota um histograma dessas probabilidades.



4. Calcula e imprime a média de alunos que acertam pelo menos 3 questões, respondendo ao acaso.

```
Média de alunos acertando pelo menos 3 questões: 763.9121888205418
```

Um problema clássico envolvendo a distribuição de Poisson é o experimento de Rutherford-Geiger, da contagem do número de partículas  $\alpha$  emitidas por uma amostra de polônio, em intervalos de 7,5s, num total de 2608 intervalos.

A tabela abaixo, mostra as frequências ( $f_m$ ) correspondentes ao número de contagens ( $m$ ) em cada intervalo.

$m$	$f_m$
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	2
13	0
14	0

- Determine o número médio de contagens em cada intervalo de 7,5s.
- Compare a distribuição de frequências das contagens do experimento com a distribuição de Poisson de média igual ao número médio de contagens.

**a) Determinar o número médio de contagens em cada intervalo de 7,5s**

O número médio de contagens ( $\mu$ ) pode ser encontrado usando a fórmula para a média ponderada:

$$\mu = \frac{\sum(m \cdot f_m)}{\sum f_m}$$

onde  $m$  representa o número de contagens e  $f_m$  é a frequência de cada  $m$ .

Vamos calcular isso:

```
1 # Dados do experimento
2 contagens = 0:14
3 frequencias = [57, 203, 383, 525, 532, 408, 273, 139, 45, 27, 10, 4, 2, 0, 0]
4
5 # Calculando a media
6 mu = sum(contagens .* frequencias) / sum(frequencias)
7
8 println("Numero medio de contagens em cada intervalo de 7,5s: ", mu)
```

Dando o resultado: Número médio de contagens em cada intervalo de 7,5s: 3.870398773006135

- b) Comparar a distribuição de frequências com a distribuição de Poisson** Agora que temos a média ( $\mu$ ), podemos comparar a distribuição real com a distribuição de Poisson de parâmetro  $\mu$ . A probabilidade de obter  $k$  eventos em um intervalo para uma distribuição de Poisson é dada por:

$$P(k; \mu) = \frac{e^{-\mu} \mu^k}{k!}$$

Vamos calcular as probabilidades esperadas para cada  $k$  e depois comparar as frequências observadas com as esperadas.

```
1 using Distributions
2
3 # Criando uma distribuicao de Poisson com media mu
4 poisson_dist = Poisson(mu)
5
6 # Calculando as probabilidades esperadas para cada numero de contagens
7 # usando a distribuicao de Poisson
```

```

8 probabilidades_poisson = [pdf(poisson_dist, k) for k in contagens]
9
10 # Calculando as frequencias esperadas
11 frequencias_esperadas = [p * sum(frequencias) for p in probabilidades_poisson]
12
13 # Imprimindo a comparacao
14 println("m | Frequencia Observada | Frequencia Esperada (Poisson)")
15 for (m, f_obs, f_esp) in zip(contagens, frequencias, frequencias_esperadas)
16     println("$m | $f_obs | $(round(f_esp, digits=2))")
17 end

```

Dando o resultado:

```

m | Frequência Observada | Frequência Esperada (Poisson)
0 | 57 | 54.38
1 | 203 | 210.46
2 | 383 | 407.28
3 | 525 | 525.45
4 | 532 | 508.42
5 | 408 | 393.56
6 | 273 | 253.87
7 | 139 | 140.37
8 | 45 | 67.91
9 | 27 | 29.2
10 | 10 | 11.3
11 | 4 | 3.98
12 | 2 | 1.28
13 | 0 | 0.38
14 | 0 | 0.11

```

### EXERCICIO 16

Em um grupo de pessoas a altura média é de 170cm com desvio padrão de 5cm. Calcule a altura acima da qual estão os 10% mais altos.

Para resolver esse problema, podemos usar o conceito de escore Z na distribuição normal. O escore Z é definido como:  $Z = \frac{X - \mu}{\sigma}$   
Onde:

- X é o valor na distribuição (neste caso, a altura que estamos tentando encontrar),
- $\mu$  é a média da distribuição (170 cm, no caso),
- $\sigma$  é o desvio padrão da distribuição (5 cm, no caso).

Queremos encontrar a altura X acima da qual estão os 10% mais altos. Isso significa que estamos procurando o percentil 90, ou o valor de X para o qual  $P(X < x) = 0,90$ .

Para encontrar o valor de Z correspondente ao percentil 90, podemos usar uma tabela de distribuição normal padrão ou uma função de calculadora/código que forneça essa informação. Vamos fazer isso em Julia como no exemplo:

```

1 using Distributions
2
3 # Media (\mu) e desvio padrao (\sigma)
4 mu = 170
5 sigma = 5
6
7 # Criar uma distribuicao normal com media \mu e desvio padrao \sigma
8 dist = Normal(mu, sigma)
9

```

```

10 # Encontrar o valor correspondente ao percentil 90
11 altura_90_percentil = quantile(dist, 0.90)
12
13 println("A altura acima da qual estao os 10% mais altos e $(round(altura_90_percentil
    , digits=2)) cm.")

```

Dando como Resultado: A altura acima da qual estão os 10% mais altos é 176.41 cm.

Esse código utiliza a função `quantile` da biblioteca `Distributions` para encontrar o valor da altura que corresponde ao percentil 90 na distribuição normal especificada pela média ( $\mu$ ) e desvio padrão ( $\sigma$ ). O resultado é arredondado para duas casas decimais para facilitar a leitura.

### EXERCICIO 17

A média dos diâmetros dos rolamentos de esfera produzidos por uma determinada máquina é de 0.482cm com desvio padrão de 0.004cm. Uma peça é considerada defeituosa se tiver mais que 0.491cm ou menos que 0.473cm. Qual a porcentagem de peças defeituosas produzidas?

Dada a distribuição normal com média  $\mu = 0.482\text{cm}$  e desvio padrão  $\sigma = 0.004\text{cm}$ , podemos calcular a probabilidade de uma peça ser defeituosa da seguinte forma:

1. Calcular o escore Z para os limites de defeito (0.473 cm e 0.491 cm).
2. Usar o escore Z para encontrar as probabilidades correspondentes.
3. A porcentagem de peças defeituosas será a soma das probabilidades de as peças estarem fora dos limites especificados.

Escore Z é calculado como:

$$Z = \frac{(X - \mu)}{\sigma}$$

onde  $X$  é o valor do limite,  $\mu$  é a média e  $\sigma$  é o desvio padrão.

**Para o limite inferior (0.473 cm):**

$$Z_{inferior} = \frac{(0,473 - 0,482)}{0,004} = \frac{-0,009}{0,004} = -2,25$$

**Para o limite superior (0.491 cm):**

$$Z_{superior} = \frac{(0,491 - 0,482)}{0,004} = \frac{0,009}{0,004} = 2,25$$

Vamos calcular as probabilidades usando uma tabela Z ou uma função de uma biblioteca de estatísticas, e somar as probabilidades para os valores abaixo de 0.473 cm e acima de 0.491 cm.

```

1  using Distributions
2
3  # Media e desvio padrao
4  mu = 0.482
5  sigma = 0.004
6
7  # Definir a distribuicao normal com a media e o desvio padrao
8  distribuicao_normal = Normal(mu, sigma)
9
10 # Calcular a probabilidade de uma peca ser menor que o limite inferior
11 prob_inferior = cdf(distribuicao_normal, 0.473)
12
13 # Calcular a probabilidade de uma peca ser maior que o limite superior
14 prob_superior = 1 - cdf(distribuicao_normal, 0.491)
15
16 # A porcentagem total de pecas defeituosas e a soma das duas probabilidades
17 porcentagem_defeituosa = (prob_inferior + prob_superior) * 100
18
19 println("A porcentagem de pecas defeituosas produzidas e $(round(
    porcentagem_defeituosa, digits=2))%.")

```

Resultado: A porcentagem de peças defeituosas produzidas é 2.44%.

Esse código em Julia utiliza a função `cdf` da biblioteca `Distributions` para calcular a probabilidade cumulativa até os limites especificados de 0.473 cm e 0.491 cm. A soma dessas probabilidades, multiplicada por 100, dá a porcentagem total de peças defeituosas.

### EXERCÍCIO 18

Mostre que no ajuste de uma função linear:

a)  $\chi^2 = \left(\frac{\sigma_y}{\sigma}\right)^2 (1 - r^2)$  (para uma amostra heterocedástica);

b)  $\chi^2 = \frac{\sigma_y^2}{(\epsilon_y^2/N)} (1 - r^2)$  (para uma amostra homocedástica).

Definindo alguns termos:

- $\chi^2$ : Este termo não é padrão em estatísticas e pode estar se referindo a alguma medida específica de erro ou dispersão. Normalmente, em estatística,  $\chi^2$  é usado para representar uma variável ao quadrado ou a estatística de teste Qui-quadrado, mas o contexto aqui sugere que pode ser uma representação de erro ou variância na previsão.
- $\sigma_y^2$ : Variância dos resíduos ou erros na previsão do modelo linear.
- $\sigma$ : Desvio padrão dos preditores ( $x$ ).
- $r$ : Coeficiente de correlação linear entre as variáveis  $x$  e  $y$ .
- $N$ : Número de observações.
- $\epsilon_y^2$ : Esta notação é atípica. Pode ser uma tentativa de representar a soma dos quadrados dos erros (SSE) ou uma notação específica para um contexto particular. Vamos assumir que possa se referir à soma dos quadrados dos erros, dada a estrutura da equação.

#### a) Para uma amostra heterocedástica:

Na regressão linear, a heterocedasticidade refere-se à variabilidade desigual dos erros ou resíduos ao longo do intervalo dos preditores. Neste contexto, parece que  $\chi^2$  está tentando expressar alguma forma de erro ajustado para a variabilidade dos dados. A expressão dada é:

$$\chi^2 = \left(\frac{\sigma_y}{\sigma}\right)^2 (1 - r^2)$$

Este formato não é uma formulação padrão na análise de regressão, sugerindo uma medida de dispersão dos resíduos normalizada pelo desvio padrão dos preditores e ajustada pelo coeficiente de determinação ( $r^2$ ).

#### b) Para uma amostra homocedástica:

Na presença de homocedasticidade, a variabilidade dos erros ou resíduos é constante ao longo do intervalo dos preditores. A equação proposta é:

$$\chi^2 = \frac{\sigma_y^2}{(\epsilon_y^2/N)} (1 - r^2)$$

Aqui, parece haver uma confusão na notação, mas interpretando  $\epsilon_y^2$  como a soma dos quadrados dos erros (SSE), a equação parece querer expressar uma relação entre a variância dos resíduos e a média dos quadrados dos erros ajustada pela força da relação linear ( $1 - r^2$ ).

#### Discussão:

Ambas as expressões tentam relacionar a dispersão ou variabilidade dos erros em uma regressão linear com a força da relação linear ( $1 - r^2$ ), mas a notação e interpretação exata são um tanto atípicas e não se alinham diretamente com as formulações padrão em estatística. Normalmente, em regressão linear, discute-se a variância dos resíduos e como ela se relaciona com o coeficiente de determinação ( $R^2$ ) de maneira a avaliar a qualidade do ajuste. A primeira equação sugere uma tentativa de ajustar a variância dos resíduos pela variabilidade dos

preditores e pela força da relação linear, enquanto a segunda parece uma tentativa de normalizar essa variância pela média dos quadrados dos erros, ajustada também pela força da relação linear.

### EXERCICIO 19

Ao colidir com a superfície terrestre, um meteoro provoca uma cratera. A relação esperada entre o diâmetro ( $D$ ) da cratera e a energia cinética ( $E$ ) do meteoro no instante do impacto é dada por

$$D = kE^{1/4}$$

em que  $k$  é uma constante.

A tabela abaixo mostra os diâmetros das depressões causadas pelo impacto de diversas esferas de aço sobre a areia contida em uma caixa, e as correspondentes incertezas ( $\epsilon_D$ ) e energias cinéticas das esferas ao colidirem com a areia da caixa. As esferas são utilizadas para simularem a queda de meteoros.

$E(J)$	$D(cm)$	$\epsilon_D(cm)$
0,07	4,9	0,3
0,18	6,7	0,3
0,30	7,3	0,4
0,45	8,1	0,4
0,69	9,2	0,4

A partir de um ajuste linear, determine uma estimativa para o expoente da relação esperada entre a energia e o diâmetro.

Para determinar uma estimativa para o expoente da relação esperada entre a energia ( $E$ ) e o diâmetro ( $D$ ) da cratera causada por um impacto, vamos primeiramente transformar a relação dada  $D = kE^{1/4}$  para uma forma que possa ser ajustada linearmente. Para isso, aplicamos o logaritmo natural ( $\ln$ ) em ambos os lados da equação, resultando em:

$$\ln(D) = \ln(k) + \frac{1}{4}\ln(E)$$

Agora, podemos ver que temos uma equação da forma  $y = a + bx$ , onde:

- $y = \ln(D)$
- $x = \ln(E)$
- $a = \ln(k)$
- $b = 1/4$  é o expoente que queremos estimar, relacionado à inclinação da reta no ajuste linear dos logaritmos dos dados.

Podemos utilizar os dados fornecidos para realizar um ajuste linear e estimar o valor de  $b$ , que representa o expoente da relação entre a energia e o diâmetro.

Vamos calcular o logaritmo natural dos valores de  $E$  e  $D$ , e depois realizar um ajuste linear para encontrar a inclinação ( $b$ ).

Em Julia, temos:

```

1  using Plots
2  using GLM
3  using DataFrames
4
5  # Dados fornecidos
6  E = [0.07, 0.18, 0.30, 0.45, 0.69] # Energia em Joules
7  D = [4.9, 6.7, 7.3, 8.1, 9.2]      # Diametro em cm
8
9  # Convertendo para DataFrames
10 df = DataFrame(E = E, D = D)
11
12 # Aplicando o logaritmo natural
13 df[!, :logE] = log.(df.E)
14 df[!, :logD] = log.(df.D)
15
16 # Realizando um ajuste linear
17 ols = lm(@formula(logD ~ logE), df)
18
19 # Exibindo os resultados
20 println(coeftable(ols))
21
22 # Plotando os dados e a linha de ajuste
23 scatter(df.logE, df.logD, label="Dados Logarítmicos", xlabel="log(E)", ylabel="log(D)", legend=:topleft)
24 plot!(df.logE, predict(ols), label="Ajuste Linear", color="red")

```

Este script calcula o logaritmo dos valores de energia e diâmetro, realiza um ajuste linear e plota os resultados. A inclinação da reta ajustada aos dados logarítmicos deve fornecer uma estimativa para o expoente da relação entre a energia e o diâmetro.

