Universidade Federal da Bahia
Programa de Pós-Graduação em Ciência de Dados e Big Data
Carlos Magno Santos Ribeiro de Brito

---

Prof Dr. Paulo Canas Rodriques

**Respostas da segunda lista de exercício**

Questão 01) Por engano misturaram-se quatro pilhas novas com três pilhas usadas. Escolhendo ao acaso, e sem reposição, duas dessas pilhas, determine a probabilidade uma ser nova e outra usada.

Quando não há reposição (eventos dependentes), a probabilidade da primeira tentativa ser nova é dada como $P(N) = \frac{4}{7}$ e a probabilidade da segunda ser usada é $P(U|N) = \frac{3}{6}$, pois o espaço amostral total diminuiu. A probabilidade de uma ser nova e outra usada pode ser definida como $P(N \cap U) = P(N) \times P(U|N) = \frac{4}{7} \times \frac{3}{6} = \frac{2}{7}$.

Igualmente, considerando também o caso inverso, onde a primeira tentativa seria uma pilha usada e a segunda uma nova, temos $P(U \cap N) = \frac{3}{7} \times \frac{4}{6} = \frac{2}{7}$.

Logo, para o caso geral, soma-se ambos $P(G) = \frac{2}{7} + \frac{2}{7} = \frac{4}{7}$

Questão 02) Sabe-se que 5% das pessoas que começam uma dieta têm disturbio alimentar. Ao selecionar, ao acaso, 50 pessoas em dieta, determine:

a) A probabilidade de que pelo menos uma pessoa sofra de distúrbio alimentar;

Para resolver este problema, podemos tomar a distribuição binomial[1] como modelo a ser utilizado. A probabilidade de disturbio alimentar é dada como $p = \frac{5}{100} = \frac{1}{20}$ e $n = 50$. Definindo $X$ como o número de pessoas com distúrbio alimentar, $X \sim b(n = 50; p = \frac{1}{20})$.

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(X \geqslant 1) = \binom{50}{x} \left(\frac{1}{20}\right)^x \left(1 - \frac{1}{20}\right)^{50-x}, \quad x = 1, 2, 3, \ldots, 50$$

Isso é o mesmo que $P(X \geqslant 1) = 1 - P(X = 0)$. Para o valor de $P(X = 0)$, temos:

[1] Alternativamente, esse valor porderia ser encontrado pela distribuição de poisson, para comparações ($np < 7$)

$$P(X=0) = \binom{50}{0}\left(\frac{1}{20}\right)^0\left(1-\frac{1}{20}\right)^{50} \approx 0,0769$$
$$P(X \geqslant 1) = 1 - 0,0769 \approx 0,9231$$

Ou seja, $92,31\%$.

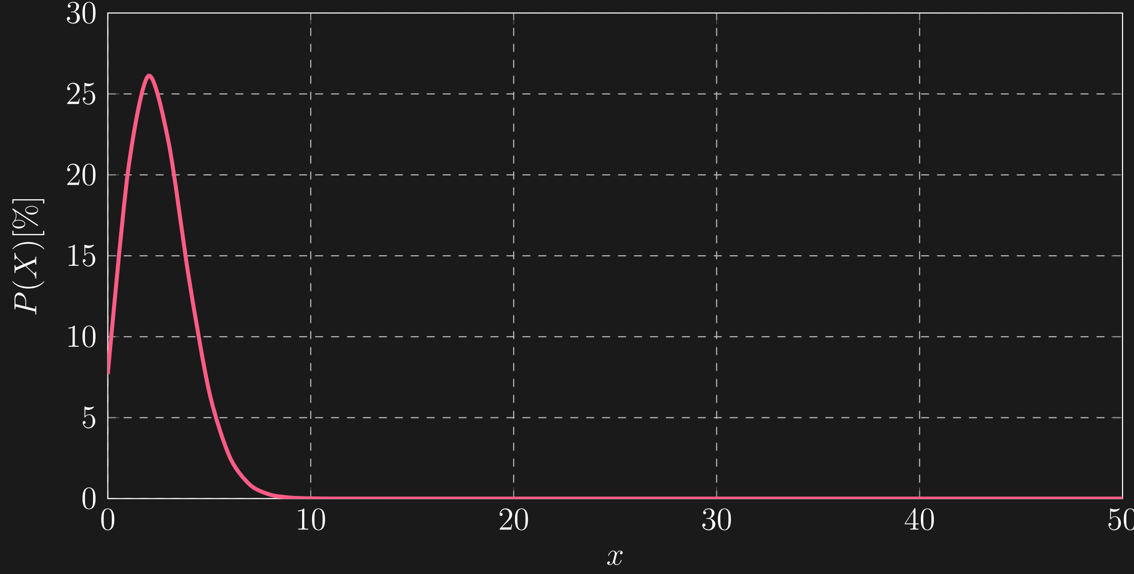
Obtém-se com Python os valores para a distribuição binomial com o seguinte código e plota-se o gráfico dessa distribuição:

```
from scipy.stats import binom

n = 50
p = 0.05
x = range(0, n + 1)

a = binom.pmf(x, n, p)
for i, o in enumerate(a):
    print("{} {:.4f}".format(i,o*100))
```

Gráfico da f.p. $p(x)$ para $n = 50$ e $p = 1/20$




b) O número médio e o desvio padrão das pessoas com distúrbio alimentar.

A média e o desvio padrão são calculados como:

$$E(X) = np = 50 \times \frac{1}{20} = 2,5$$

$$\sigma = \sqrt{Var(X)} = \sqrt{npq} = \sqrt{50 \times \frac{1}{20} \times \left(1 - \frac{1}{20}\right)} \approx 1,54$$

Questão 03) Para a população masculina de um determinado país, com idades entre 18 e 74 anos, a pressão sistólica tem distribuição aproximadamente normal com média 129 mmHg e desvio padrão 19.8 mmHg. Considere que os níveis da pressão são normais quando menores que 130 mmHg.

a) Qual a probabilidade de um homem dessa população possuir pressão sistólica normal?

Para isso, esse homem deve ter níveis de pressão abaixo de 130 mmHg. Usando a distribuição normal como modelo, tem-se:

$$X \sim N(\mu = 129; \sigma^2 = 19,8^2)$$

Onde $X$ é a pressão sistólica. Para ela sendo $P(X < 130)$, tem-se:

$$P(X < 130) = P\left(\frac{X - 129}{19,8} < \frac{130 - 129}{19,8}\right) = P\left(Z < \frac{1}{19,8}\right) \approx 0,5201$$

O resultado foi obtido com python, com o seguinte código:

```
from scipy.stats import norm


x = norm(129,19.8).cdf(130) # Sem padronizar
z = norm(0,1).cdf(1/19.8) # Padronizada em Z

print(x," ", z)

# 0.5201400375890497
```

b) Qual a probabilidade de um homem dessa população possuir hipertensão moderada (pressão sistólica entre 160 e 179 mmHg)?

Para este caso, tem-se:

$$P(160 \leqslant X \leqslant 179) = P\left(\frac{160 - 129}{19,8} \leqslant \frac{X - 129}{19,8} \leqslant \frac{179 - 129}{19,8}\right)$$

$$P\left(\frac{31}{19,8} \leqslant Z \leqslant \frac{50}{19,8}\right) = P\left(0 \leqslant Z \leqslant \frac{50}{19,8}\right) - P\left(0 \leqslant Z \leqslant \frac{31}{19,8}\right) \approx 0,0529$$

O resultado foi obtido com python, com o seguinte código:

```
from scipy.stats import norm

# A partir de menos infinito
z1 = norm(0, 1).cdf(50 / 19.8)
z2 = norm(0, 1).cdf(31 / 19.8)

print(z1 - z2)

# 0.05293376095968105
```

c) Selecionando ao acaso 1000 homens dessa população, quantos seriam diagnosticados com hipertensão moderada (pressão sistólica entre 160 e 179 mmHg)?

Tendo já obtido os valores de pressão sistólica para um conjunto ao qual pertece esses 1000 homens, basta apenas realizar a multiplicação:

$$n = 1000 \times P(160 \leqslant X \leqslant 179) \approx 53$$

d) Qual a probabilidade de um homem dessa população possuir pressão sistólica igual a 180 mmHg?

Como o modelo de distribuição é contínuo $\int_a^b f(x)\,dx$, onde as probabilidades são dadas como áreas sob a curva gaussiana, um ponto infinitesimal $dx$ específico terá uma área aproximadamente igual a 0. Ou seja, $P(X = 180) = 0$.

e) Qual a pressão sistólica que deixa acima 80% dos indivíduos?

Para $P(Z < z_0) = 0,2$, onde $0,2$ corresponde a área acumulada na curva normal. Então, precisamos encontrar o valor de $z_0$, para em seguida achar o de $x_0$. Fazendo a transformação inversa da curva normal, tem-se:

$$z_0 = \frac{x_0 - \mu}{\sigma} \rightarrow x_0 = z_0\sigma + \mu$$

O resultado pode ser obtido com o python e é igual a $112,34$ mmHg.

```
from scipy.stats import norm
mu = 129
sigma = 19.8
z0 = norm.ppf(0.2)

print(z0*sigma+mu)

# 112.33589957525629
```

Curva normal parametrizada