Estatítica para Ciência de Dados

Aula 4: Distribuição Normal e Teorema Central do Limite

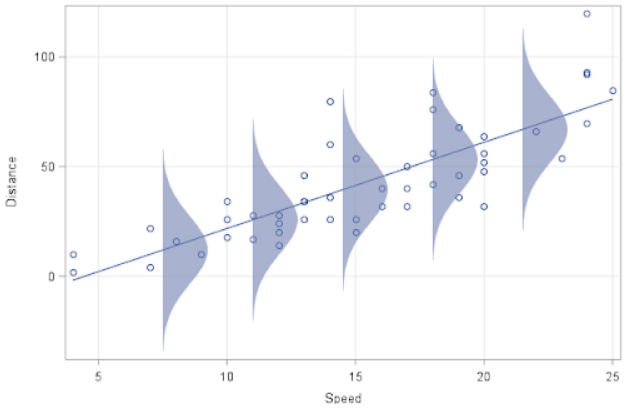
Francisco A. Rodrigues
ICMC/USP
francisco@icmc.usp.br

Aula 4: Distribuição Normal e Teorema Central do Limite

- **Distribuição Normal**
  Teorema Central do Limite
  Lei dos grandes números

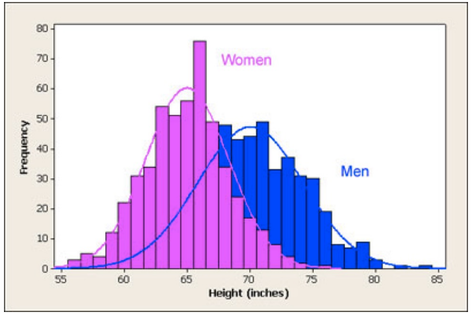
Distribuição Normal

- Em 1809 Gauss publicou o trabalho "Theoria motus corporum coelestium in sectionibus conicis solem ambientium" onde ele introduziu o método dos mínimos quadrados, o método da máxima verossimilhança e a distribuição normal, dentre outros conceitos estatísticos.

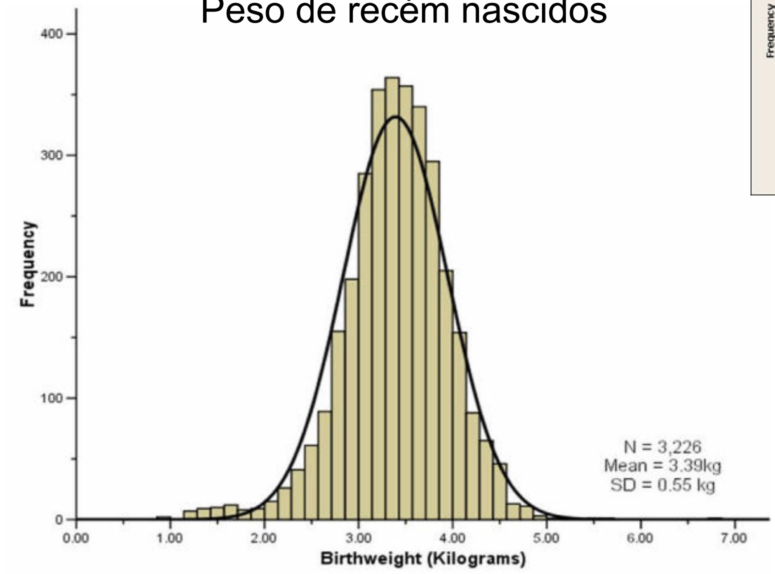


Distribuição Normal

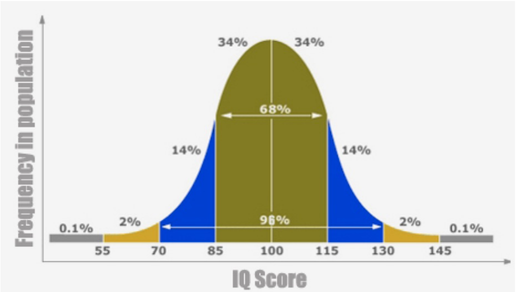
Alturas



Peso de recém nascidos



QI



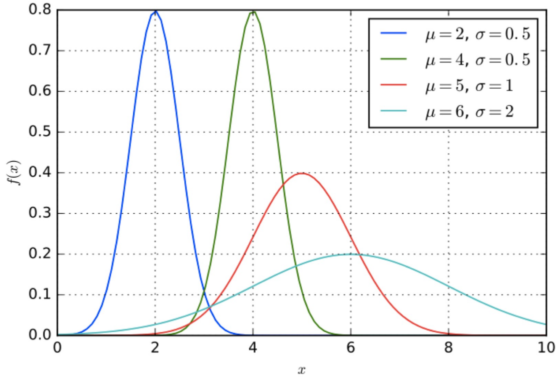
Distribuição Normal

- **Definição**: A v.a. contínua X que tome todos os valores na reta real $-\infty < x < \infty$ segue uma distribuição normal (ou Gaussiana) se sua função densidade de probabilidade é dada por:

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

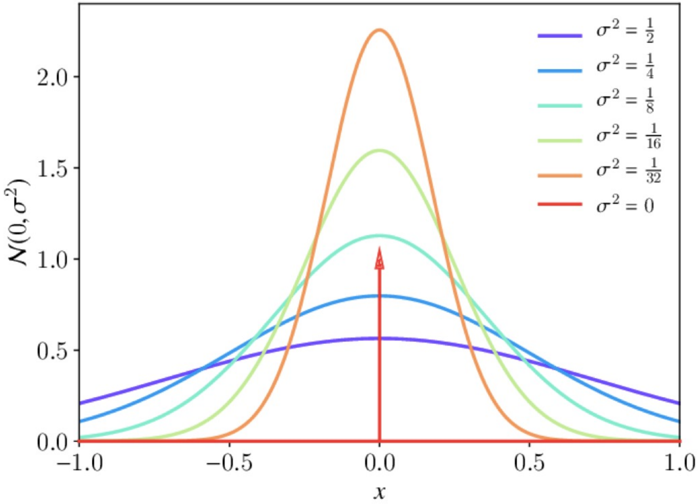
Onde

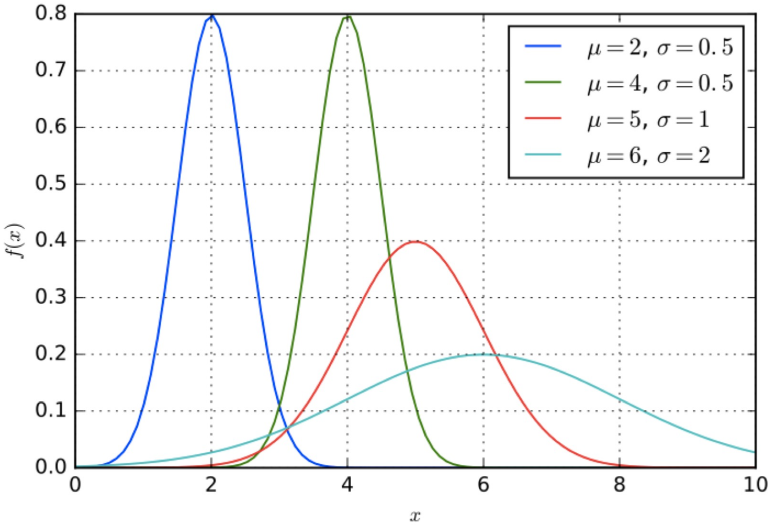
- $E[X] = \mu$
- $V(X) = \sigma^2$



Distribuição Normal

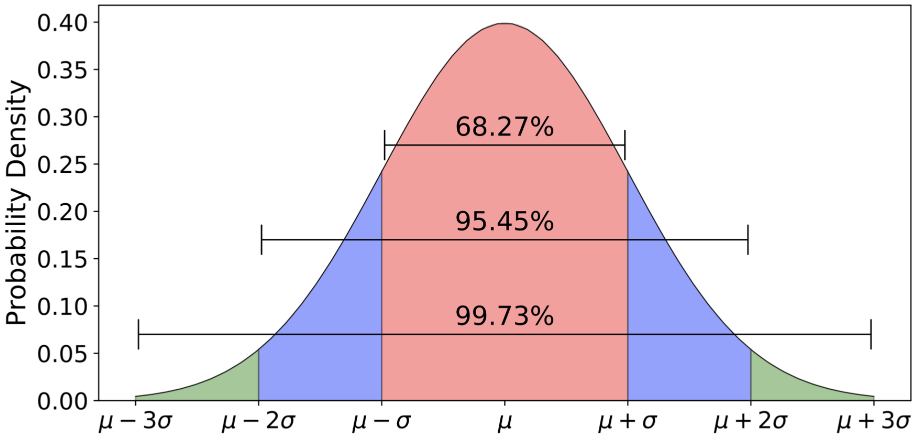
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$





Distribuição Normal

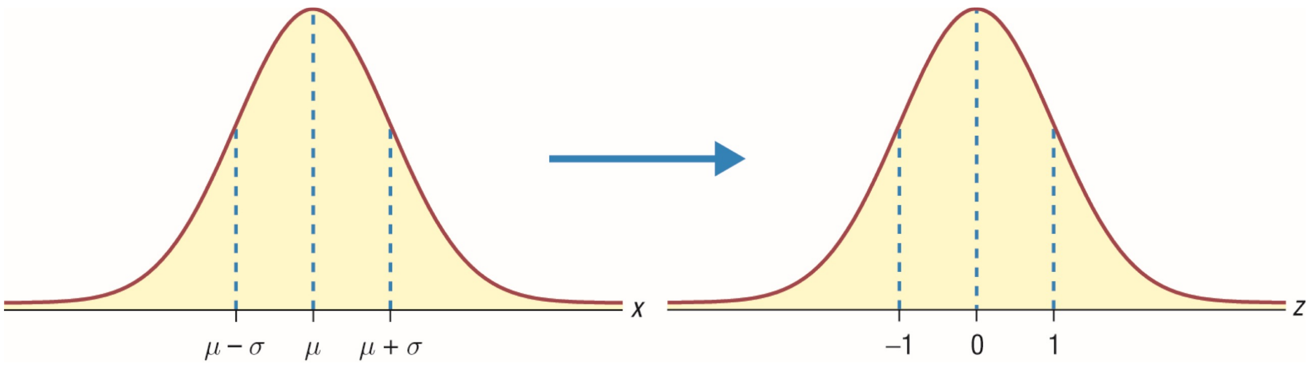
- **Propriedades:**
- f(x) é simétrica em relação à μ.
- $f(x) \to 0$ quando $x \to \pm\infty$.
- O valor máximo de X ocorre em $x = \mu$.



Distribuição Normal

- **Tabulação:** $X \sim N(\mu, \sigma)$

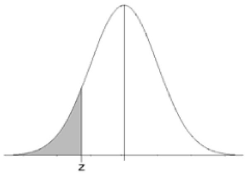
$$Z = \frac{X - \mu}{\sigma} \rightarrow Z \sim N(\mu = 0, \sigma = 1)$$



Distribuição Normal

- **Tabulação:**

Standard Normal Cumulative Probability Table



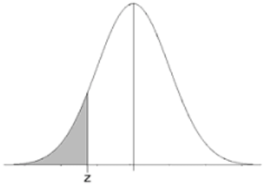
Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455

Distribuição Normal

- **Exemplo:**
- Se $X \sim N(\mu = 165, \sigma^2 = 9)$, calcule $P(X < 162)$.
- $P(X < 162) = P\left(\frac{X-\mu}{\sigma} < \frac{162-165}{3}\right)$ = P(Z <-1) =0.158

Standard Normal Cumulative Probability Table



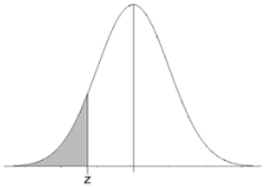
Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
					...					
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379

Distribuição Normal

- **Exemplo:**
- Se $X \sim N(\mu = 10, \sigma^2 = 4)$, calcule $P(X > 13)$.
- $P(X > 13) = P\left(\frac{X-\mu}{\sigma} > \frac{13-10}{2}\right) = \mathrm{P}(\mathrm{Z} > 1{,}5) = 1 - \mathrm{P}(\mathrm{Z} < 1{,}5) = 1 - 0.93 = 0.07$

Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
					...					
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

Distribuição Normal

- **Exemplo:**
- Se $X \sim N(\mu = 165, \sigma^2 = 9)$, calcule $P(X < 162)$.

```python
import scipy.stats as st

media = 165
dp = 3
z = (162-media)/dp
print(st.norm.cdf(z))
```

```
0.15865525393145707
```

- Se $X \sim N(\mu = 10, \sigma^2 = 4)$, calcule $P(X > 13)$.

```python
import scipy.stats as st

media = 10
dp = 2
z = (13-media)/dp
print(1-st.norm.cdf(z))
```

```
0.06680720126885809
```

Distribuição Normal

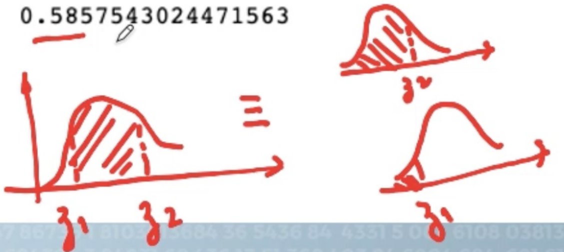
- **Exemplo:** O peso médio de 500 estudantes do sexo masculino de uma determinada universidade é 75,5 Kg e o desvio padrão é 7,5 Kg. Admitindo que os pesos são normalmente distribuídos, determine a percentagem de estudantes que pesam entre 60 e 77,5 Kg.

$$P(60 \le X < 77{,}5) = P\left(\underbrace{\frac{60-\mu}{\sigma}}_{z_1} < \frac{X-\mu}{\sigma} < \underbrace{\frac{77{,}5-\mu}{\sigma}}_{z_2}\right).$$

$$= P(Z \le z_2) - P(Z \le z_1) = \boxed{0{,}58}$$

```python
import scipy.stats as st
media = 75.5
dp = 7.5
z1 = (60-media)/dp
z2 = (77.5-media)/dp
st.norm.cdf(z2)-st.norm.cdf(z1)
```

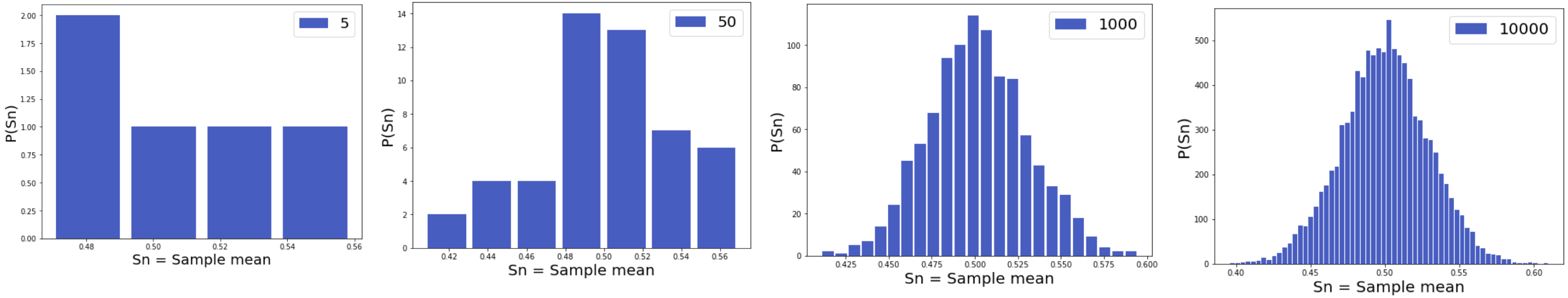
0.5857543024471563



Teorema Central do Limite

- **Teorema:** Seja uma amostra da população X com média e variância finita. Então:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(\mu = 0, \sigma^2 = 1)$$



Teorema Central do Limite

- **Exemplo:** Seja a variável aleatória com distribuição de probabilidade dada abaixo ($\mu = 5{,}4;\ \sigma^2 = 4{,}44$). Uma amostra com 40 observações é sorteada. Qual é a probabilidade de que a média amostral ser maior do que 5?

$$E[X] = 5.4; V(X) = 4.43$$

X	3	6	8
P(X=x)	0,4	0,3	0,3

$$P(\bar{X} > 5) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{5-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > \frac{5-5{,}4}{\sqrt{4{,}44}/\sqrt{40}}\right) = 1 - P\left(Z < \frac{5-5{,}4}{\sqrt{4{,}44}/\sqrt{40}}\right) = \ldots$$

Distribuição Normal

```python
def esperanca(X,P):
    E = 0
    for i in range(0, len(X)):
        E = E + X[i]*P[i]
    return E

def variancia(X,P):
    E = 0; E2 = 0
    for i in range(0, len(X)):
        E = E + X[i]*P[i]
        E2 = E2 + (X[i]**2)*P[i]
    V = E2-E**2
    return V

X = [3,6,8]
P = [0.4,0.3,0.3]
E = esperanca(X,P)
V = variancia(X,P)
print("Esperança:", E, "Variância:",V)
```

```
Esperança: 5.4 Variância: 4.439999999999991
```

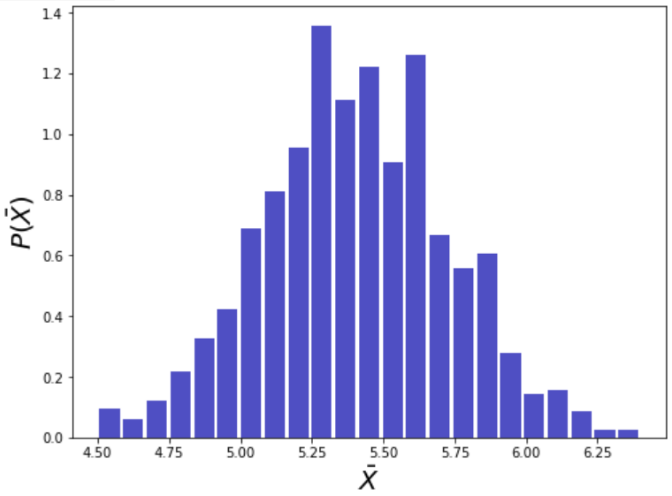
Distribuição Normal

```python
import scipy.stats as st
import numpy as np

mu = E
sigma = np.sqrt(V)
n = 40
x = 5
Zt = (x - mu)/(sigma/np.sqrt(n))
pt = 1-st.norm.cdf(Zt)
print('Probabilidade:',pt)
```

```
Probabilidade: 0.885046886863795
```

Distribuição Normal

```
import matplotlib.pyplot as plt

n = 40
ns = 1000 #numero de simulacoes
vx = [] # armazena a media amostral
for s in range(0,ns):
    A = np.random.choice(X, n, p=P)
    vx.append(np.mean(A))
plt.figure(figsize=(8,6))
plt.hist(x=vx, bins='auto',color='#0504aa',
         alpha=0.7, rwidth=0.85, density = True)
plt.xlabel(r'$\bar{X}$', fontsize=20)
plt.ylabel(r'$P(\bar{X})$', fontsize=20)
plt.show(True)
print("Media das amostras:", np.mean(vx), "Media da população:", E)
```




```
Media das amostras: 5.40425 Media da população: 5.4
```

Sumário

- **Distribuição Normal**
  Teorema Central do Limite

Leitura Complementar