Modelo Normal

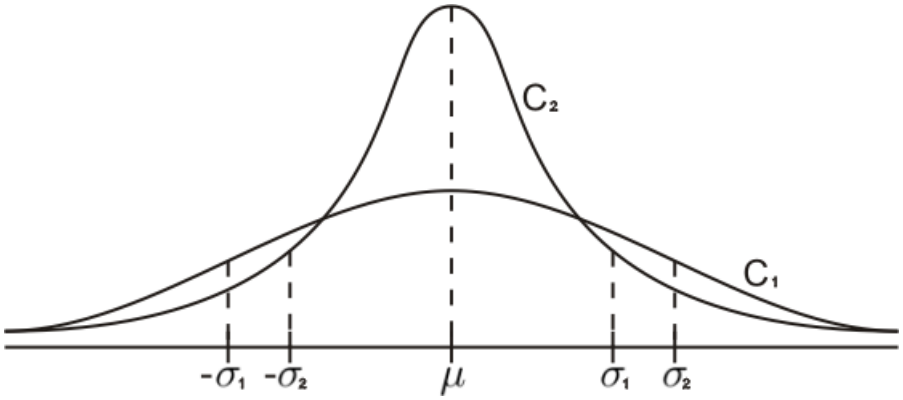
- Dizemos que uma v.a. X possui distribuição normal com parâmetros μ e σ^2, $\mu \in \mathbb{R}$ e $\sigma^2 > 0$, se a f.d.p. f_X é dada por:

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

- Notação: $X \sim N(\mu, \sigma^2)$
- Devido à impossibilidade de calcular a f.d.a. analiticamente, recorre-se à tabela da **Normal Padrão** ($N(0,1)$)

Modelo Normal

- Gráfico da f.d.p.



$E(X) = \mu$: representa o ponto de simetria de f_X

$Var(X) = \sigma^2$: representa a dispersão de f_X

Modelo Normal

Afirmação: $X \sim N(\mu, \sigma^2)$ se, e somente se $\frac{X-\mu}{\sigma} \sim N(0,1)$

- Φ denota a f.d.a. da Normal padrão:

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)} dx$$

- Temos então, $X \sim N(\mu, \sigma^2)$:

$$F_X(a) = P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Modelo Normal

- Exercitando com a tabela da Normal:
  - $\Phi(0.2) = 0.5793$
  - $\Phi(0.45) = 0.6736$
  - $\Phi(1.98) = 0.9761$
  - $\Phi(-0.45) = 1 - \Phi(0.45) = 0.3264$

Modelo Normal

Exemplos

- $Z \sim N(0,1)$
- $Y \sim N(4, 2^2)$
  - $F_Y(6) = P(Y \leq 6) = P\left(\frac{Y-4}{2} \leq \frac{6-4}{2}\right) = P(Z \leq 1) = \Phi(1) = 0.8413$
  - $P(2 < Y \leq 6) = P\left(\frac{2-4}{2} < \frac{Y-4}{2} \leq \frac{6-4}{2}\right) = P(-1 < Z \leq 1) = \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)] = 2\Phi(1) - 1 = 2()0.8413 - 1 = 0.6826$

Modelo Normal

- Valor Esperado:

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

- Variância

$$\begin{aligned} Var(X) &= E([X - E(X)]^2) \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2 \end{aligned}$$

Aplicação e Exemplos de Distribuição Amostral

- Planejamento: pretendemos avaliar a honestidade de uma moeda. Para isso, planeja-se lançar uma moeda 50 vezes, sendo cada lançamento independente dos demais. Definimos:

$$X_i = \begin{cases} 1, & \text{quando cara} \\ 0, & \text{quando coroa} \end{cases}$$

- Assim, a variável

  $Y = \sum_{i=1}^{50} X_i =$ número de caras nos 50 lançamentos

- Y é uma v.a. que pode assumir os valores $\{0, 1, 2, ..., 50\}$

Aplicação e Exemplos de Distribuição Amostral

- $Y = \sum_{i=1}^{50} X_i \sim B(50, p)$, pois é uma soma de bernoullis
- $p = P(X_i = 1) = P(\text{obter cara})$
- Y é a estatística usada para avaliar a honestidade da moeda
- se a moeda for honesta $p = \frac{1}{2}$
- Definimos por fim $T = \frac{Y}{n}$, e a partir de T, estimamos p
- Supondo que foram observadas 30 caras nos 50 lançamentos, $T = \frac{30}{50} = 0.6 = \hat{p}$

Aplicação e Exemplos de Distribuição Amostral

- Qual a importância de saber a distribuição de Y? Para avaliar se a ocorrência de 30 caras em 50 lançamentos nos traz evidências se a moeda é honesta ou não.
- Assumindo que a moeda é honesta, dado $p = \frac{1}{2}$:

$$P(Y = 30|n = 50, p = 0.5) = \begin{pmatrix} 50 \\ 30 \end{pmatrix} (0.5)^{30}(0.5)^{20} = 0.042$$

- $P(Y \geq 30|n = 50, p = 0.5) = 0.08$

Aplicação e Exemplos de Distribuição Amostral

- Assim, se a moeda fosse honesta, a probabilidade de ocorrer mais de 30 caras em 50 lançamentos é de aproximadamente 0.08.
- Essa probabilidade é evidência suficiente contra a honestidade da moeda?

Distribuição Amostral

- **Resultado:** Seja X uma v.a. com média μ e variância σ^2. Seja $X_1, ..., X_n$ uma amostra casual simples de X.
  - $\bar{X}_n = \frac{X_1 + ... + X_n}{n}$
  - $E\left(\bar{X}_n\right) = \mu$
  - $Var\left(\bar{X}_n\right) = \frac{\sigma^2}{n}$
- Exemplo: X_1, X_2, X_3 ensaios *bernoulli*(0.3) independentes
  - $E\left(X_i\right) = 0.3 \quad \Rightarrow \quad E\left(\bar{X}_3\right) = 0.3$
  - $Var\left(X_i\right) = 0.3(0.7) = 0.21 \quad \Rightarrow \quad Var\left(\bar{X}_3\right) = \frac{0.21}{3} = 0.07$

Teorema Central do Limite

- **Resultado (T.C.L.):** Para amostras casuais simples $X_1, ..., X_n$ colhidas de uma população com média μ e variância σ^2, a distribuição amostral de $\bar{X}_n$ aproxima-se de uma distribuição Normal de média μ e variância $\frac{\sigma^2}{n}$, quando n for suficientemente grande.

Teorema Central do Limite

- Exemplo: $X_1, ..., X_n$ uma amostra aleatória
  - $exp(2)$: $f_{X_i}(x) = 2e^{-2x}\mathbb{I}_{(x>0)}$
  - $E(X_i) = \frac{1}{2}$
  - $Var(X_i) = \frac{1}{4}$

  Suponha que X_i modela o tempo de vida de um transistor em horas. Os tempos de 100 transistores são coletados. Desejamos estudar a variável aleatória $\bar{X}_{100}$, e pelo T.C.L., temos que:
  - $E\left(\bar{X}_{100}\right) = \frac{1}{2}$
  - $Var\left(\bar{X}_{100}\right) = \frac{1/4}{100} = \frac{1}{400}$
  - $\bar{X}_n \sim N\left(\frac{1}{2}, \frac{1}{400}\right)$

Teorema Central do Limite

- **Resultado do T.C.L.:** Se $X_1, ..., X_n$ é uma amostra casual simples com média μ e variância σ^2, e definimos $\bar{X}_n = \frac{X_1 + ... + X_n}{n}$, quando n for suficientemente grande:

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

- Retomando o exemplo dos transistores: $\frac{\bar{X}_{100} - (1/2)}{(1/2)/\sqrt{100}} \sim N(0,1)$

Teorema Central do Limite

- Utilidade do Resultado:

$$
\begin{aligned}
P\left(\bar{X}_{100} \leq x\right) &= P\left(\frac{\bar{X}_{100} - (1/2)}{(1/2)/\sqrt{100}} \leq \frac{x - (1/2)}{(1/2)/\sqrt{100}}\right) \\
&= P\left(Z \leq 10(2x - 10)\right)
\end{aligned}
$$

$$
\begin{aligned}
P\left(\bar{X}_{100} \geq x\right) &= 1 - P\left(\bar{X}_{100} \leq x\right) \\
&= 1 - P\left(\frac{\bar{X}_{100} - (1/2)}{(1/2)/\sqrt{100}} \leq \frac{x - (1/2)}{(1/2)/\sqrt{100}}\right) \\
&= 1 - P\left(Z \leq 10(2x - 10)\right)
\end{aligned}
$$

Aproximação da Distribuição Binomial pela Normal

- Consideremos uma população em que a proporção de indivíduos portadores de uma certa característica é p.
- Colhida uma amostra casual simples de indivíduos, podemos construir

$$X_i = \begin{cases} 1, & \text{se o indivíduo i possui a característica} \\ 0, & \text{caso contrário} \end{cases}$$

- $\Rightarrow X_i \sim ber(p);\ i = 1, 2, ..., n$
- Se os indivíduos são independentes: $S_n = X_1 + ... + X_n \sim Bin(n, p)$

Aproximação da Distribuição Binomial pela Normal

- $\hat{p} = \frac{S_n}{n}$ é uma média amostral
- Utilizando a distribuição exata (n pequeno)

$$P\left(\hat{p} = \frac{k}{n}\right) = P\left(\frac{S_n}{n} = \frac{k}{n}\right) = P\left(S_n = k\right) = \begin{pmatrix} n \\ k \end{pmatrix} p^k \left(1-p\right)^{n-k}$$

$k = 0, 1, ..., n$

- Utilizando a aproximação para a Normal (n grande)

$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$

Aproximação da Distribuição Binomial pela Normal

- **Exemplo:** Se p for a proporção de fumantes no estado de SP, $p = 0.2$ e tivermos coletado uma amostra casual simples de 500 indivíduos

$$X_i = \begin{cases} 1, & \text{se o indivíduo i é fumante} \\ 0, & \text{caso contrário} \end{cases}$$

  - $\hat{p} = \frac{\sum_{i=1}^{500} X_i}{500}$
  - $\hat{p} \sim N\left(0.2, \frac{0.2 \times 0.8}{500}\right) = N\left(0.2, 0.00032\right)$
  - $P\left(\hat{p} \leq 0.25\right) = P\left(Z \leq 2.795\right) = \Phi\left(2.795\right) = 0.9974$

Aproximação da Distribuição Binomial pela Normal

- $\hat{p} = \frac{S_n}{n} \Rightarrow S_n = n\hat{p}$
- Quando n é grande o suficiente $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$
- Qual a distribuição de S_n quando n é grande o suficiente?

Aproximação da Distribuição Binomial pela Normal

- Propriedade:
  - $X \sim N(a, b)$
  - $Y = \alpha X + \beta$
  - $\Rightarrow Y \sim N(\alpha a + \beta, \alpha^2 b)$
- Aplicação:
  - $S_n = X_1 + ... + X_n$
  - $\hat{p} = \frac{S_n}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$
  - $S_n = n\hat{p} \sim N\left(np, np(1-p)\right)$
- Portanto: $Bin(n, p) \approx N\left(np, np(1-p)\right)$ quando n é grande

Aproximação da Distribuição Binomial pela Normal

- **Exemplo:** $X \sim Bin(100, 0.4)$
  - $E(X) = 100 \times 0.4 = 40$
  - $Var(X) = 100 \times 0.4 \times 0.6 = 24$
  - $X \approx N(40, 24)$
  - $P(X \leq 50) = P\left(Z \leq \frac{50-40}{\sqrt{24}}\right) \approx \Phi\left(\frac{10}{\sqrt{24}}\right) = \Phi(2.04) \approx 0.9793$

Intervalo de Confiança como Estimativa de p

- $X \sim Bin(n, p)$
- n suficientemente grande, $X \sim N\left(np, np(1-p)\right)$
- $Z = \frac{X - np}{\sqrt{np(1-p)}} \approx N(0, 1)$
- $\gamma = 0.95$ é o grau de confiança

Intervalo de Confiança como Estimativa de p

$$
\begin{aligned}
0.95 &= P\left(-1.96 \leq Z \leq 1.96\right) \\
&= P\left(-1.96 \leq \frac{X - np}{\sqrt{np(1-p)}} \leq 1.96\right) \\
&= P\left(-1.96\sqrt{np(1-p)} \leq X - np \leq 1.96\sqrt{np(1-p)}\right) \\
&= P\left(\frac{-1.96\sqrt{np(1-p)}}{n} \leq \frac{X - np}{n} \leq \frac{1.96\sqrt{np(1-p)}}{n}\right) \\
&= P\left(\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}}\right)
\end{aligned}
$$

Intervalo de Confiança como Estimativa de p

- p é desconhecido
- $p(1-p) \leq \frac{1}{4}$
  - $\Rightarrow \sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{1}{4n}}$
  - $\Rightarrow -\sqrt{\frac{p(1-p)}{n}} \geq -\sqrt{\frac{1}{4n}}$
- $0.95 \approx P\left(\hat{p} - 1.96\sqrt{\frac{1}{4n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{1}{4n}}\right)$
- Caso geral:

  $\left[\hat{p} - z_\gamma\sqrt{\frac{1}{4n}}; \hat{p} + z_\gamma\sqrt{\frac{1}{4n}}\right]$ é um IC de $\gamma \times 100\%$ para p

Intervalo de Confiança como Estimativa de p

- **Exemplo:** Numa pesquisa de mercado, $n = 400$ pessoas foram entrevistadas sobre determinado produto, e 60% destas pessoas preferiam a marca A. $\hat{p} = 0.6$, logo, o IC com grau de confiança $\gamma = 0.95$ é dado por:

$$\left[0.6 - 1.96\frac{1}{\sqrt{1600}}; 0.6 + 1.96\frac{1}{\sqrt{1600}}\right] = [0.551; 0.649]$$

Intervalo de Confiança como Estimativa de p

- **Exemplo:** Suponha que em $n = 400$ provas, obtemos $k = 80$ sucesso. Vamos obter um intervalo de confiança para p, com $\gamma = 0.9$:
  - $\hat{p} = \frac{80}{400} = 0.2$
  - $z_{0.9} = 1.645$

$$\left[0.2 - 1.645\frac{1}{\sqrt{1600}}; 0.2 + 1.645\frac{1}{\sqrt{1600}}\right] = [0.159; 0.2411]$$

  - Usando $\hat{p}$

$$\left[\hat{p} - z_\gamma\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_\gamma\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0.167; 0.233]$$

Intervalo de Confiança como Estimativa de p

- O intervalo que utiliza $\hat{p}$ como estimativa tem menor amplitude do que o intervalo que utiliza $p(1-p) \leq \frac{1}{4}$
  - $[0.159; 0.2411]$: $0.2411 - 0.159 = 0.082$
  - $[0.167; 0.233]$: $0.233 - 0.167 = 0.066$
- Finalmente,os intervalos de confiança para p podem então ser de duas formas:

$$I_1 = \left[\hat{p} - z_\gamma \sqrt{\frac{1}{4n}}; \hat{p} + z_\gamma \sqrt{\frac{1}{4n}}\right]$$

$$I_2 = \left[\hat{p} - z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Intervalo de Confiança como Estimativa de p

- z_γ é tal que $\gamma = P\left(-z_\gamma \leq Z \leq z_\gamma\right)$; $Z \sim N(0,1)$
- Como determinar então, z_γ?

$$\begin{aligned}
\gamma &= P\left(-z_\gamma \leq Z \leq z_\gamma\right) = P(Z \leq Z_\gamma) - P(Z \leq -Z_\gamma) \\
&= P(Z \leq Z_\gamma) - P(Z \geq Z_\gamma) = P(Z \leq Z_\gamma) - [1 - P(Z \leq -Z_\gamma)] \\
&= 2P(Z \leq Z_\gamma) - 1 = 2\Phi(Z_\gamma) - 1
\end{aligned}$$

$$\Rightarrow \frac{\gamma + 1}{2} = \Phi(z_\gamma)$$

$$\Rightarrow \Phi^{-1}\left(\frac{\gamma + 1}{2}\right) = z_\gamma$$