Probabilidade e Estatística

para Engenharia e Engenharia Química

Códigos: MAD201 / MAD469 (MAD231)

Oferecido pelo:

DEPARTAMENTO DE MÉTODOS ESTATÍSTICOS - DME

INSTITUTO DE MATEMÁTICA - UFRJ

Introdução à disciplina

- Site do curso ProbEst (todas as informações sobre a disciplina)
- Bibliografia:
  - Pinheiro, J. I. D. et al. *Probabilidade e Estatística: quantificando a incerteza*. Editora Campus, 2012.
  - Bussab, W. O. e Morettin, A. M. *Estatística Básica*. Editora Saraiva, 2009.
  - Bertsekas, D. P. and Tsitsiklis, J. N. *Introduction to Probability*, 2nd Edition. Athena Scientific, 2008.
  - Blitzstein, J. K. & Hwang, J. - *Introduction to Probability*, CRC Press, 2015.

Ementa do Curso

- Parte I - Probabilidade:
  - Cálculo das Probabilidades.
  - Variáveis aleatórias discretas.
  - Variáveis aleatórias contínuas.
  - Função de uma variável aleatória.
  - Variáveis aleatórias bidimensionais.
  - Vetores aleatórios multidimensionais.
- Parte II - Estatística:
  - Análise exploratória de dados amostrais.
  - Amostragem e estimação pontual.
  - Estimação por intervalo
  - Introdução à teoria dos testes de hipóteses.

Probabilidade

Modelos determinísticos e probabilísticos

- **Modelo determinístico**: Pode ser descrito facilmente através de uma fórmula
  Exemplo: Corpo em queda livre no vácuo. A velocidade final, em cm/s, é dada por $v = \sqrt{2gh}$, onde g é a aceleração da gravidade (em cm/s^2) e h é a altura inicial (em cm).
- **Modelo probabilístico**: Sabemos os possíveis resultados do experimento, mas não sabemos (ou é extremamente difícil) precisá-lo.
  Exemplo: Lançamento de uma moeda ou lançamento de um dado.

Nesse curso estudaremos mais a fundo os modelos probabilísticos.

Definições básicas

- Um *experimento* é qualquer processo, real ou hipotético, no qual os resultados podem ser identificados ao longo do tempo
- O *espaço amostral* é a coleção de todos os possíveis resultados de um experimento. Usualmente denotado pela letra Ω. Elementos do espaço amostral são usualmente denotados pela letra ω.
- Um *evento* A é um subconjunto do espaço amostral Ω.

Exemplo 1

Lançamento de 10 moedas: se $Cara \leftrightarrow 0$ e $Coroa \leftrightarrow 1$, então

$$\begin{aligned}\Omega &= \{\omega = (\omega_1, \omega_2, \ldots, \omega_{10}) : \omega_i \in \{0,1\} \text{ para } 1 \leq i \leq 10\} \\ &= \{0,1\}^{10}\end{aligned}$$

Alguns eventos de interesse:

- "o j-ésimo lançamento foi coroa": $A_j = \{\omega \in \{0,1\}^{10} : w_j = 1\}$
- "pelo menos um lançamento foi coroa": $B = \bigcup_{j=1}^{10} A_j$
- "todos lançamentos foram cara": $C = B^c = \{(0, 0, \ldots, 0)\}$

Exemplo 2

Selecione uma carta de um baralho de 52 cartas:

$$\Omega = \{\text{A}\spadesuit, \ldots, K\spadesuit, \text{A}\heartsuit, \ldots, K\heartsuit, \text{A}\diamondsuit, \ldots, K\diamondsuit, \text{A}\clubsuit, \ldots, K\clubsuit\}$$

Alguns eventos de interesse:

- "A carta é Ás": $A = \{\text{A}\spadesuit, \text{A}\heartsuit, \text{A}\diamondsuit, \text{A}\clubsuit\}$
- "A carta é de cor escura": $B = \{\text{A}\spadesuit, \ldots, K\spadesuit, \text{A}\clubsuit, \ldots, K\clubsuit\}$
- "A carta é de copas": $C = \{\text{A}\heartsuit, \ldots, K\heartsuit\}$
- "A carta é de ouro": $D = \{\text{A}\diamondsuit, \ldots, K\diamondsuit\}$
- "A carta é de cor branca": $E = B^c$

Operações em conjuntos e diagramas de Venn

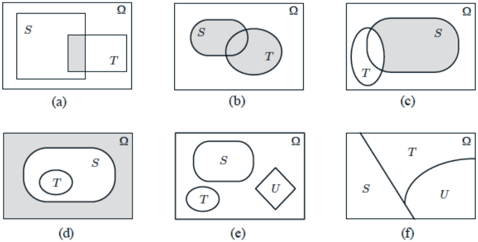


Figura 1: (Bertsekas, D.P. e Tsitsiklis, J.N. Introduction to probability, p.5.)

Em todos os diagramas Ω é o espaço amostral, e S, T e U são eventos.

a) A região sombreada é $S \cap T$
b) A região sombreada é $S \cup T$
c) A região sombreada é $S \cap T^c$
d) Aqui $T \subset S$. A região sombreada é S^c
e) Os conjuntos T, S e U são *disjuntos* (ou *mutuamente exclusivos*)
f) Os conjuntos T, S e U formam uma *partição* de Ω.

Algumas propriedades úteis

- Denotamos o conjunto vazio por $\emptyset$
- Para subconjuntos $A, B, C \subseteq \Omega$ vale que:
  - $A^c \cup A = \Omega$
  - $A \cap A^c = \emptyset$
  - $(A^c)^c = A$
  - $A \cup \Omega = \Omega$
  - $A \cap \Omega = A$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- Leis de de Morgan: se $A_1, A_2, \ldots, A_n$ são eventos em Ω, então

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c \quad \text{e} \quad \left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

Definição clássica de probabilidade

Seja A um evento de um espaço amostral Ω finito. A *probabilidade* do evento A é definida por

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{número de resultados favoráveis a } A}{\text{número total de resultados em } \Omega}$$

Notação: $|A|$ denota a cardinalidade (i.e., quantidade de elementos) do evento A

Hipóteses:

- Resultados equiprováveis!
- Espaço amostral Ω finito!

Exemplo 3

Se jogarmos dois dados honestos, qual dos resultados é mais provável: a soma de 11 ou a soma de 12?

Solução: O espaço amostral do lançamento de dois dados é

$$\begin{aligned}\Omega &= \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \ldots, 6\} \text{ para } 1 \leq i \leq 2\} \\ &= \{1, \ldots, 6\}^2.\end{aligned}$$

Para cada resultado individual possível $w \in \Omega$:

$$\mathbb{P}(\omega) = \mathbb{P}(\{\omega\}) = \frac{1}{36}.$$

Os eventos de interesse são:

- "A soma de 11": $A = \{(5,6), (6,5)\}$
- "A soma de 12": $B = \{(6,6)\}$

Logo, $\mathbb{P}(A) = 2/36 > 1/36 = \mathbb{P}(B)$.

De volta ao exemplo 1

Lançamento de 10 moedas honestas. Lembre que $Cara \leftrightarrow 0$ e $Coroa \leftrightarrow 1$, de modo que

$$\Omega = \{0,1\}^{10} \text{ e } |\Omega| = 2^{10}.$$

- Como $|A_j| = |\{\omega \in \{0,1\}^{10} : \omega_j = 1\}| = 2^9 : \mathbb{P}(A_j) = \frac{2^9}{2^{10}} = 1/2$
- Como $|C| = 1 : \mathbb{P}(C) = 1/2^{10}$
- Como $|B| = |\cup_{j=1}^{10} A_j| = 2^{10} - 1$: $\mathbb{P}(B) = 1 - (1/2)^{10}$

Note que $B = C^c$ e $\mathbb{P}(B) = 1 - P(C)$.
Em geral, se A é um evento:

$$A^c \text{ é um evento e } \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

Exemplo 4

Full House em uma mão de Poker: Uma mão de 5 cartas é extraída de um baralho de 52 cartas. Se o baralho foi embaralhado de forma que cada carta tenha a mesma chance de aparecimento, qual é a probabilidade de se obter um Full House?
Obs.: Uma mão é chamada Full House se há 3 cartas do mesmo valor e 2 outras cartas do mesmo valor.

Solução: Seja A o evento "a mão é um Full House". Temos que

$$|A| = 13\binom{4}{3}12\binom{4}{2},$$

logo,

$$\mathbb{P}(A) = \frac{13\binom{4}{3}12\binom{4}{2}}{\binom{52}{5}} \approx 0.00144.$$

Exemplo 5

O problema do aniversário: Existem k pessoas numa sala. Supondo que o aniversário de cada pessoa é igualmente provável e que não há gêmeos na sala, qual a probabilidade de que duas ou mais pessoas celebrem o aniversário no mesmo dia?

Solução: Se A denota o evento "nenhuma das k pessoas fazem aniversário no mesmo dia", então queremos calcular $\mathbb{P}(A^c)$. Nesse exemplo,

$$\Omega = \{1, \ldots, 365\}^k.$$

Note que $|\Omega| = 365^k$ e

$$|A| = \frac{365!}{(365-k)!} = 365 \cdot 364 \cdots (365 - k + 1).$$

Logo,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - \frac{|A|}{|\Omega|} = 1 - \frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k}.$$

Exemplo 5

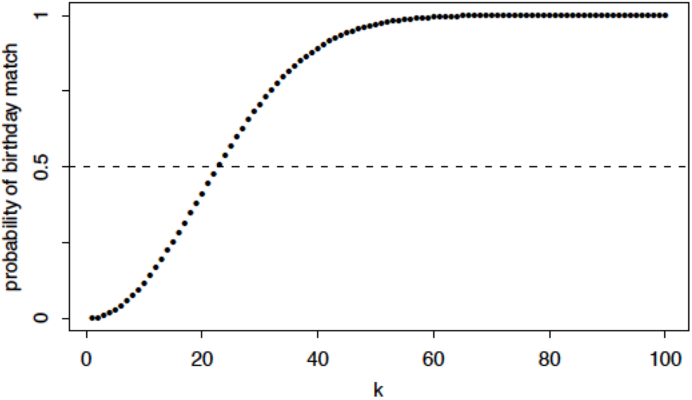


Figura 2: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.12.) Probabilidade de que num grupo de k pessoas duas ou mais celebrem aniversário no mesmo dia. Essa probabilidade excede 1/2 para $k = 23$.

Observação

Lembre que para Ω infinito, a definição clássica não se aplica.

Exemplos de espaços amostrais Ω infinitos incluem:

- tempo de vida útil de um aparelho eletrônico: $\Omega = [0, \infty)$
- o número de clientes atendidos em uma agência bancária durante um período de tempo fixado: $\Omega = \{0, 1, 2, 3, \dots\}$

Como definir probabilidade para espaço amostral Ω infinito?

Definição axiomática de probabilidade

Seja Ω um espaço amostral. Uma *probabilidade* é uma função $\mathbb{P}$ que atribui a eventos $A \subseteq \Omega$ um número $\mathbb{P}(A) \in [0,1]$ e satisfaz os seguintes *axiomas*:

1. $\mathbb{P}(\Omega) = 1$ e $\mathbb{P}(\emptyset) = 0$
2. Para eventos $(A_i)_{i\geq 1}$ disjuntos (isto é $A_i \cap A_j = \emptyset$ se $i \neq j$):

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Dos Axiomas 1 e 2 segue que para eventos $A_1, \ldots, A_n$ disjuntos:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

De fato, considere $A_{n+1} = A_{n+2} = \ldots = \emptyset$.

Propriedades da probabilidade

Teorema: Sejam $A, B \subseteq \Omega$ eventos. Temos que:

1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
3) $0 \leq \mathbb{P}(A) \leq 1$
4) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

O item 4 é um caso particular do princípio da *Inclusão-Exclusão*:
Para eventos A, B e C, também vale que:

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) \\ - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).\end{aligned}$$

Probabilidade condicional

Considere uma probabilidade $\mathbb{P}$ definida em um espaço amostral Ω. Para eventos A e B com $\mathbb{P}(B) > 0$, a *probabilidade condicional* de A dado B é definida como

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

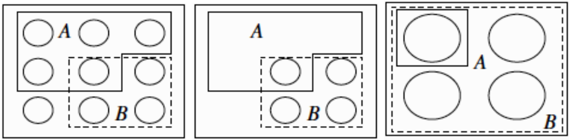


Figura 3: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.44.) Intuição para $\mathbb{P}(A|B)$. Da esquerda para direita: a) Eventos A e B do espaço amostral. b) Como sabemos que B ocorreu, descartamos todos resultados em B^c. c) No espaço amostral restrito, renormalizamos de modo que a massa total seja 1.

Exemplo 1

Duas cartas são escolhidas aleatoriamente, uma de cada vez e sem reposição, de um baralho de 52 cartas. Seja A o evento "a primeira carta é de copas" e B o evento "a segunda carta é vermelha". Determine $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$.

Solução: Note que:

- $\mathbb{P}(A \cap B) = (13)(25)/(51)(52)$
- $\mathbb{P}(A) = 1/4$
- $\mathbb{P}(B) = (26)(51)/(51)(52)$

Logo, temos $\mathbb{P}(A|B) = 25/102$ e $\mathbb{P}(B|A) = 25/51$.

Note que $\mathbb{P}(B|A)$ é $2 \times \mathbb{P}(A|B)$.

Exemplo 2

Lançamos uma moeda honesta três vezes. Sejam A o evento "mais caras do que coroas" e B o evento "o primeiro lançamento é cara ". Determine $\mathbb{P}(A|B)$.

Solução: Denotando $Cara \leftrightarrow 0$ e $Coroa \leftrightarrow 1$, temos que o espaço amostral $\Omega = \{0,1\}^3$, e os eventos A e B são dados por

$$A = \{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\}$$
$$B = \{(0,0,0),(0,1,0),(0,0,1),(0,1,1)\}.$$

Logo,

$$\mathbb{P}(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{3}{8},$$

e $\mathbb{P}(B) = |B|/|\Omega| = 1/2$, de forma que

$$\mathbb{P}(A|B) = 3/4.$$

Regra da Multiplicação

Teorema: (Regra da multiplicação) Se A e B são eventos com probabilidade positiva, então

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B).$$

Mais geralmente, se $A_1, \ldots, A_n$ são eventos tais que $\left(\bigcap_{i=1}^{k} A_i\right)$ tem probabilidade positiva para todo $1 \leq k \leq (n-1)$, então

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\ldots\mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Exemplo 3

Três cartas são extraídas, sem reposição, de um baralho com 52 cartas. Assumindo que em cada extração, cada uma das cartas restantes têm a mesma chance de ser escolhida, determine a probabilidade de nenhuma das três cartas seja de copas.

Solução: Defina para cada $i = 1, 2, 3$, o evento

$$A_i = \text{“a } i\text{-ésima carta não é de copas”}$$

Note que $\mathbb{P}(A_1) = 39/52$, $\mathbb{P}(A_2|A_1) = 38/51$ e $\mathbb{P}(A_3|A_1 \cap A_2) = 37/50$. Logo, pela regra da multiplicação,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) = \frac{39}{52}\frac{38}{51}\frac{37}{50}.$$

Partição de um espaço amostral

Os eventos $A_1, \ldots A_n$ formam uma *partição* do espaço amostral Ω se as seguintes condições são satisfeitas:

1. $\Omega = \bigcup_{i=1}^{n} A_i$
2. $A_i \cap A_j = \emptyset$ para todos $1 \leq i, j \leq n$ com $i \neq j$.

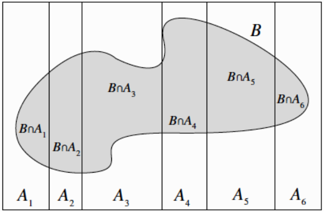


Figura 4: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.50.) Partição $A_1, \ldots, A_6$ do espaço amostral.

Para qualquer evento $B \subseteq \Omega : B = \cup_{i=1}^{n} (B \cap A_i)$.

Lei da Probabilidade Total

Teorema: (Lei da Probabilidade Total) Considere $A_1, \ldots, A_n$ formam uma partição de Ω tal que $P(A_i) > 0$ para todo $1 \leq i \leq n$. Para todo evento B, temos que

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n).$$

Exemplo 4

João decide participar de um torneio de xadrez, onde a metade dos jogadores é profissional (tipo 1), um quarto é semi-profissional (tipo 2) e um quarto é amador (tipo 3). Suponha que as probabilidades do João vencer uma partida sejam 0.3, 0.4 e 0.5, se o adversário for respectivamente do tipo 1, tipo 2 e tipo 3. Se um adversário é escolhido ao acaso, determine a probabilidade do João sair vitorioso.

Solução: Defina para cada $i = 1, 2, 3$, o evento

$$A_i = \text{“o adversário escolhido é do tipo } i\text{”}.$$

Temos que $\mathbb{P}(A_1) = 0.5$, $\mathbb{P}(A_2) = 0.25$ e $\mathbb{P}(A_3) = 0.25$. Defina B o evento “João saiu vitorioso”. Temos que $\mathbb{P}(B|A_1) = 0.3$, $\mathbb{P}(B|A_2) = 0.4$ e $\mathbb{P}(B|A_3) = 0.5$ Assim, pela lei da probabilidade total

$$P(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) + \mathbb{P}(A_3)\mathbb{P}(B|A_3) = 0.375$$

Teorema de Bayes

Teorema: (Teorema de Bayes) Se A e B são eventos com probabilidade positiva, então

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}.$$

Além disso, se $A_1, \ldots, A_n$ formam uma partição de Ω talque $P(A_i) > 0$ para todo $1 \leq i \leq n$, então

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)}.$$

Exemplo 5

Um paciente é testado para uma doença que atinge 1% da população. O resultado do teste é positivo. Seja D o evento "o paciente testado tem a doença" e T o evento "o resultado do teste para esse paciente é positivo". Suponha que a precisão do teste é 95%, i.e.
$\mathbb{P}(T|D) = 0.95 = \mathbb{P}(T^c|D^c)$. Calcule $\mathbb{P}(D|T)$.

Solução: Pelo Teorema de Bayes e a Lei da Probabilidade total:

$$\mathbb{P}(D|T) = \frac{\mathbb{P}(D)\mathbb{P}(T|D)}{\mathbb{P}(T)} = \frac{\mathbb{P}(D)\mathbb{P}(T|D)}{\mathbb{P}(D)\mathbb{P}(T|D) + \mathbb{P}(D^c)\mathbb{P}(T|D^c)} \approx 0.16$$

Exemplo 6

Considere duas moedas, sendo uma honesta e a outra *viesada* cuja probabilidade de cara é 3/4. Escolhemos uma das moedas ao acaso e então efetuamos três lançamentos. Em todos os lançamentos o resultado foi cara. Dado essa informação, qual é a probabilidade de a moeda honesta ter sido escolhida?

Solução: Sejam A o evento " em todos os lançamentos o resultado foi cara" e F o evento "a moeda honesta foi escolhida". Queremos calcular $\mathbb{P}(F|A)$. É imediato calcular $\mathbb{P}(A|F)$ e $\mathbb{P}(A|F^c)$:

$$\mathbb{P}(A|F) = 1/8 = (1/2)^3 \text{ e } \mathbb{P}(A|F^c) = (3/4)^3.$$

Também temos que $\mathbb{P}(F) = \mathbb{P}(F^c) = 1/2$. Logo, pelo Teorema de Bayes e a Lei da Probabilidade total:

$$\mathbb{P}(F|A) = \frac{\mathbb{P}(F)\mathbb{P}(A|F)}{\mathbb{P}(F)\mathbb{P}(A|F) + \mathbb{P}(F^c)\mathbb{P}(A|F^c)} \approx 0{,}23.$$

Falácia do Procurador

Em 1999, Sally Clark foi condenada pelo assassinato dos seus dois filhos, ambos recém nascidos. A decisão foi tomada com base no depoimento de um especialista. Segundo ele, a probabilidade de uma bebê recém nascido morrer da síndrome de morte súbita infantil (SMSI) era de $1/8500$, de forma que a probabilidade de duas mortes devida a SMSI na mesma família era $(1/8500)^2 \approx 1/(73 \times 10^6)$. Com base nisso, o especialista conclui que a probabilidade da inocência de Clark era de $1/(73 \times 10^6)$. Qual é o problema com a linha de raciocínio do especialista?

1. Hipótese de independência não justificada
2. Usou incorretamente $\mathbb{P}(\text{“evidência”}|\text{“Inocência”})$ ao invés de $\mathbb{P}(\text{“Inocência”}|\text{“evidência”})$.

Mensagem: não confunda $\mathbb{P}(A|B)$ com $\mathbb{P}(B|A)$!

Independência

Considere uma probabilidade $\mathbb{P}$ definida em um espaço amostral Ω. Lembre que para quaisquer eventos A e B,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A).$$

Dois eventos A e B são *independentes* se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Independência é o oposto de mutuamente exclusivos (disjuntos)!

Teorema. Seja $\mathbb{P}$ uma probabilidade definida em um espaço amostral Ω. Se A e B são eventos independentes, então também o são os eventos A^c e B, A e B^c, e A^c e B^c.

Independência

Os eventos $A_1, \ldots, A_n$ são *independentes dois a dois*, se os eventos A_i e A_j são independentes para todos $i \neq j$.

Os eventos $A_1, \ldots, A_n$ são *independentes* se

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$$

para qualquer subconjunto $J \subseteq \{1, \ldots, n\}$.

Exemplo 1

Independência Dois a Dois não implica independência.

De fato, no lançamento de 2 moedas honestas independentes, os eventos $A =$"primeira lançamento é cara", $B =$"segundo laçamento é cara" e $C =$"os dois lançamentos são iguais " são independentes dois a dois mas não são independentes.

Exemplo 2

*Q*1-2017/01: Três atiradores, Alberto, Bernardo e Carlos, atiram independentemente uma vez num alvo. Sabe-se que as probabilidades de acertar o alvo são 2/3 para Alberto, 3/4 para Bernardo e 1/2 para Carlos. Calcule a probabilidade

a) de que o alvo seja atingido;
b) de que somente um dos três acerte o alvo;
c) de no máximo um dos três acertar o alvo;
d) condicional de Alberto ter acertado, dado que o alvo foi atingido por no máximo um dos três atiradores.

Solução: Considere os eventos

$$A = \text{“Alberto acerta o alvo”},\ B = \text{“Bernardo acerta o alvo”}$$
$$C = \text{“Carlos acerta o alvo”},\ D = \text{“o alvo ser atingido”}$$
$$E = \text{“o alvo ser atingido por somente um dos atiradores”}$$
$$F = \text{“no máximo um dos três atiradores acertar o alvo”}$$

a) Queremos calcular $\mathbb{P}(D)$. Para isso, use a independência de A^c, B^c e C^c para calcular $\mathbb{P}(D^c)$ e lembre

$$\mathbb{P}(D) = 1 - \mathbb{P}(D^c) \approx 0.958.$$

b) Queremos calcular $\mathbb{P}(E)$. Escreva o evento E como

$$E = \{A^c \cap B^c \cap C\} \cup \{A^c \cap B \cap C^c\} \cup \{A \cap B^c \cap C^c\},$$

use a lei de probabilidade total e independência dos eventos A, B e C, para obter $\mathbb{P}(E) = \frac{1+3+2}{24} = 0.25$

c) Queremos calcular $\mathbb{P}(F)$. Nesse caso, note que

$$F = (A^c \cap B^c \cap C^c) \cup E,$$

use a lei da probabilidade total e a independência dos eventos A, B e C, para obter $\mathbb{P}(E) = \frac{1+1+3+2}{24} \approx 0.292$

d) Queremos calcular $\mathbb{P}(A|F)$. Para isso, note que

$$\mathbb{P}(A|F) = \frac{\mathbb{P}(A \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(A \cap B^c \cap C^c)}{\mathbb{P}(F)} = \frac{2/24}{7/24} \approx 0.28.$$

Variável aleatória

Seja Ω um espaço amostral. Uma *variável aleatória* (v.a.) é uma função de Ω assumindo valores em $\mathbb{R}$.

V.a.s fornecem "resumos numéricos" de um experimento aleatório.

Variáveis aleatórias serão denotadas em geral por letras maiúsculas do final do alfabeto:

$$X, Y, Z, \ldots$$

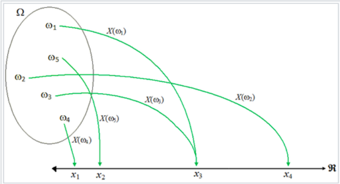


Figura 5: (Variável Aleatória. In: Wikipédia, a enciclopédia livre.) Visualização de uma variável aleatória X. A v.a. X atribuí um valor numérico $X(w)$ para cada resultado possível $w \in \Omega$.

Exemplos de variáveis aleatórias

1. Em um experimento envolvendo 10 lançamentos de uma moeda, o número de caras obtidos é uma variável aleatória.
2. Em um experimento envolvendo 5 lançamentos de um dado, os seguintes são variáveis aleatórias:
   a) A soma das 5 faces.
   b) O número de faces 4 nos 5 lançamentos.
   c) O valor da maior face obtida.
   d) O valor da menor face obtida.
3. Em um experimento envolvendo a transmissão de uma mensagem, o tempo necessário para transmitir a mensagem, o número de símbolos recebidos com erro e o tempo que leva para a mensagem ser recebida são variáveis aleatórias.

Para todo $x \in \mathbb{R}$, escrevemos $\{X = x\}$, $\{X \leq x\}$ e $\{X \geq x\}$ para denotar respectivamente os eventos

$$\{w \in \Omega : X(w) = x\},\ \{w \in \Omega : X(w) \leq x\} \text{ e } \{w \in \Omega : X(w) \geq x\}.$$

Se a v.a. X conta o número de coroas em 10 lançamentos de uma moeda honesta, então usando $Cara \leftrightarrow 0$ e $Cara \leftrightarrow 1$:

- $\{X = x\} = \{w \in \{0,1\}^{10} : w_1 + \ldots + w_{10} = x\}$.
- $\{X \leq x\} = \{w \in \{0,1\}^{10} : w_1 + \ldots + w_{10} \leq x\}$.
- $\{X \geq x\} = \{w \in \{0,1\}^{10} : w_1 + \ldots + w_{10} \geq x\}$.

Função massa de probabilidade de v.a.s discretas

Uma v.a. X é *discreta* se existe um conjunto *enumerável* $\Omega_X \subset \mathbb{R}$ t.q.

$$X \in \Omega_X.$$

A *função massa de probabilidade* (f.m.p.) de uma v.a. X discreta é a função p_X dada por

$$p_X(x) = \mathbb{P}(X = x), x \in \Omega_X.$$

Note que a f.m.p. p_X de uma v.a. X satisfaz

$$\begin{cases} p_X(x) \geq 0 \text{ para todo } x \in \Omega_X \\ \sum_{x \in \Omega_X} p_X(x) = 1 \end{cases}$$

Variável aleatória Bernoulli

Uma v.a. X é *Bernoulli* de parâmetro $p \in [0,1]$, se:

1. X assume valores 0 ou 1, isto é $\Omega_X = \{0,1\}$.
2. $p_X(1) = \mathbb{P}(X=1) = p$ e $p_X(0) = \mathbb{P}(X=0) = 1-p$.

Notação: $X \sim \text{Ber}(p)$.

Observações:

- Codificação: $0 \leftrightarrow$ "fracasso" e $1 \leftrightarrow$ "sucesso".
- O espaço amostral Ω não aparece explicitamente na definição.
- *Ensaio de Bernoulli*: experimento aleatório que resulta em "sucesso" ou "fracasso", mas NÃO ambos.
- Parâmetro p: probabilidade de sucesso.

Variável aleatória Binomial

Uma v.a. X é *Binomial* de parâmetros n e p, se:

1. X assume valores $0, 1, \ldots, n$, isto é $\Omega_X = \{0, 1, \ldots, n\}$.
2. Para cada $x \in \{0, 1, \ldots, n\}$:

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

Notação: $X \sim \mathrm{Bin}(n, p)$.

Observações:

- X é o número de sucessos em n ensaios independentes de Bernoulli com probabilidade p de sucesso $\Rightarrow X \sim \mathrm{Bin}(n, p)$.

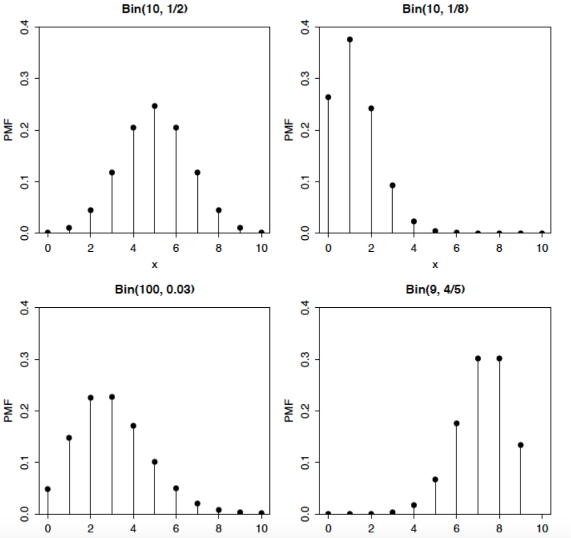


Figura 6: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.102.) Gráfico da f.m.p. para quatro variáveis aleatórias Binomiais.

Variável aleatória Poisson

Uma v.a. X é *Poisson* de parâmetro $\lambda > 0$, se:

1. X assume valores $0, 1, 2, \ldots$, isto é $\Omega_X = \mathbb{N}$.
2. Para cada $x \in \mathbb{N}$:
$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Notação: $X \sim \text{Poi}(\lambda)$.

Observações:

- **Relação Binomial e Poisson**: fixe $\lambda > 0$ e denote $p = \lambda/n$. Se n for grande o suficiente, então o número de sucessos em n ensaios de Bernoulli independentes com probabilidade p de sucesso pode ser bem aproximada por uma v.a. $\text{Poi}(\lambda)$.
- p_X é uma f.m.p. válida: $e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$.

Exemplo 2:

*Q*2-2017/01. Um aluno de educação física, conhecido na escola por ser "bom de pontaria", é desafiado pelos amigos num jogo de basquete. O desafio consiste em fazer 20 lances livres de forma independente ao cesto de basquete e avaliar sua performance nessas jogadas. Sabe-se que em média a cada 20 lançamentos ele acerta 15. Com base nesta informação, responda as perguntas a seguir.

a) Calcule a probabilidade do aluno acertar ao menos 17 das 20 jogadas.

b) O professor de basquete da escola decide entrar na brincadeira e fazer o desafio. Sabe-se que a probabilidade do professor errar o arremesso ao cesto é de 0.01. Usando a aproximação da Binomial pela Poisson, calcule a probabilidade do professor errar no máximo 2 vezes em 300 lançamentos ao cesto.

Solução:

a) Seja X a v.a. que representa o número de acertos ao cesto em 20 jogadas. Temos que $X \sim \text{Bin}(20, 0.75)$ e queremos calcular $\mathbb{P}(X \geq 17)$. Para isso, basta lembrar

$$\mathbb{P}(X \geq 17) = \sum_{x=17}^{20} \binom{20}{x} (0.75)^x (0.25)^{20-x} \approx 0.225.$$

b) Seja Y o número de erros ao cesto em 300 jogadas, temos que $Y \sim \text{Bin}(300, 0.01)$, que pode ser aproximada por uma v.a. Poisson de parâmetro 3. Logo,

$$\mathbb{P}(Y \leq 2) \approx \sum_{x=0}^{2} e^{-3} \frac{(3)^x}{x!} \approx 0.423.$$

Exemplo 3

*Q*2-2017/02. Uma empresa faz transporte de material por meio de caminhões e o volume de encomendas que ela recebe oscila ao longo do tempo. Admita que, escolhendo ao acaso um dia de trabalho:

- O número de entregas a serem feitas segue uma distribuição de Poisson com "média" de 5 entregas/dia.
- A empresa pode contratar, por empreitada, trabalhadores autônomos que constam de uma lista de 6 nomes. Todos eles tem a mesma chance p de virem a ser contratados para servir à empresa naquele dia, e há independência entre as decisões de se contratar ou não relativas aos diversos trabalhadores da lista;

Pergunta-se:

a) Qual é a probabilidade de que nesse dia sejam feitas no máximo 4 entregas, dado que ocorrerão pelo menos 3 entregas?

b) Quais são os valores possíveis de p, se a probabilidade de serem recrutados exatamente 3 empregados da lista de 6 é igual a 0.27648?

Solução:

a) Seja E o número de entregas do dia. Queremos calcular $\mathbb{P}(E \leq 4 | E \geq 3)$. Como $E \sim \text{Poi}(5)$, temos que

$$\mathbb{P}(E \leq 4 | E \geq 3) = \frac{\mathbb{P}(3 \leq E \leq 4)}{\mathbb{P}(E \geq 3)} = \frac{\mathbb{P}(E = 3) + \mathbb{P}(E = 4)}{1 - \mathbb{P}(E \leq 2)} \approx 0.361$$

b) Seja T o número de trabalhadores autônomos contratados no dia. Então, $T \sim \text{Bin}(6, p)$ e queremos achar p tal que

$$\mathbb{P}(T = 3) = 0.27648.$$

Como, $\mathbb{P}(T = 3) = 20p^3(1-p)^3 = 20[p(1-p)]^3$ temos que

$$p(1-p) = (0.2764820)^{1/3} = 0.24,$$

o que nos leva à equação do 2^o grau: $p^2 - p + 0.24 = 0$, cujas soluções são $p = 0.6$ e $p = 0.4$, ambas válidas.

Valor esperado (esperança, média) para v.a.s discretas

O *valor esperado* de uma v.a. X discreta é definido por

$$\mathbb{E}(X) = \sum_{x \in \Omega_X} x\mathbb{P}(X = x) = \sum_{x \in \Omega_X} x p_X(x),$$

onde estamos supondo, no caso Ω_X infinito, que

$$\sum_{x \in \Omega_X} |x| p_X(x) < \infty.$$

O valor esperado é uma *medida de centralidade.* Esse valor depende somente da *distribuição* da v.a. X, isto é, da f.m.p. p_X.

Em particular se v.a.s X e Y têm a mesma f.m.p. então $\mathbb{E}(X) = \mathbb{E}(Y)$.

Exemplo 1

Se $X \sim \text{Ber}(p)$ então $\mathbb{E}(X) = p$. Em particular, se A é um evento e X é a v.a. dada por

$$X(w) = 1_{\{A\}}(w) = \begin{cases} 1, & \text{se } w \in A \\ 0, & \text{caso contrário} \end{cases}, \text{ (função indicadora)}$$

então $\mathbb{E}(X) = \mathbb{P}(A)$.

Exemplo 2

Exemplo 3: Se $X \sim \text{Bin}(n,p)$ então

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=1}^{n} x\binom{n}{x} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1}(1-p)^{n-x} \\
&= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} = np,
\end{aligned}$$

onde na quarta igualdade usamos o Binômio de Newton.

Exemplo 3

Se $X \sim \text{Poi}(\lambda)$, então

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} = \lambda. \end{aligned}$$

Linearidade da esperança

Teorema 2. (Linearidade da esperança) Se X é v.a, então para todos números reais a e b,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

Mais geralmente, X e Y são v.a.s, então para todos números reais a e b,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Exemplo 4

Considere uma urna com b bolas brancas e v bolas vermelhas. Qual é o número médio de bolas brancas se fizermos n extrações com reposição? E se as extrações forem sem reposição?

Solução: Seja X_i o indicador de que na i-ésima extração a bola é vermelha. Como as extrações são com reposição, as v.a.s $X_1, \ldots, X_n$ são todas Bernoulli de parâmetro $b/(b+v)$. O total de bolas vermelhas é $X = \sum_{i=1}^{n} X_i$. Pela linearidade da esperança,

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\frac{b}{b+v}.$$

O número médio de bolas brancas sem reposição também é igual $nb/(b+v)$, por que nesse caso as v.a.s $X_1, \ldots, X_n$ continuam sendo Bernoulli de parâmetro $b/(b+v)$. (Obviamente isso só faz sentido se $n \leq b+v$.)

Exemplo 5

Paradoxo de São Petersburgo: Suponha que uma moeda honesta é jogada repetidamente até que a primeira cara apareça. O jogo paga 2^n reais se a primeira cara aparecer na n-ésima jogada. Qual o lucro médio ao jogar esse jogo? Qual o preço que um indivíduo pagaria para entrar neste jogo?

Solução: Seja X o lucro obtido ao jogar esse jogo. Então $X = 2^N$, onde N é o número total de vezes que a moeda honesta é jogada e assim

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} 2^i \frac{1}{2^i} = \infty.$$

Note que

$$\mathbb{E}(N) = \sum_{i=1}^{\infty} i \frac{1}{2^i} = 2,$$

de forma que $\mathbb{E}(X) = \mathbb{E}(2^N) = \infty \neq 4 = 2^{\mathbb{E}(N)}$!

Mensagem: Em geral $\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$ para g não-linear!

Esperança de função de v.a.

Como calcular $\mathbb{E}(g(X))$ quando X é uma v.a. discreta?

Teorema 2: Se $g : \mathbb{R} \to \mathbb{R}$ e X é v.a. discreta, então

$$\mathbb{E}(g(X)) = \sum_{x \in \Omega_X} g(x)\mathbb{P}(X = x).$$

Variância e Desvio Padrão

Seja X uma v.a. A *variância* de X é definida como

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

A variância é uma medida de dispersão da v.a. X. Note que

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Em particular, se $\mathbb{E}(X) < \infty$, então $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$.

O *desvio padrão* de X é definido por

$$\mathrm{DP}(X) = \sqrt{\mathrm{Var}(X)}.$$

O desvio padrão é uma medida de dispersão com a mesma unidade que v.a. X (ao contrário da variância cuja unidade é o quadrado da unidade da v.a. X).

Exemplo 1

Em que situação a variabilidade é maior: em lançamentos de uma moeda honesta ou de uma moeda viesada?

Solução: Devemos calcular a variância de uma v.a. Bernoulli de parâmetro p e determinar o valor de p para o qual a variância é máxima. Se $X \sim \text{Ber}(p)$, então $\text{Var}(X) = p(1-p)$.

De fato, sabemos que $\mathbb{E}(X) = p$. Como $X^2 = X$ temos $\mathbb{E}(X^2) = \mathbb{E}(X)$. Logo,

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p(1-p).$$

Maximizando $\text{Var}(X)$ como função de p, temos que o valor máximo de $\text{Var}(X)$ é $1/4$ e esse valor é atingido para $p = 1/2$.

Assim a variabilidade é maior em lançamentos de uma moeda honesta.

Exemplo 2

Variância da Poisson: se $X \sim \text{Poi}(\lambda)$, então $\text{Var}(X) = \lambda$.

Solução: Precisamos calcular $\mathbb{E}(X^2)$, pois já vimos que $\mathbb{E}(X) = \lambda$. Para isso, derive duas vezes a expansão de Taylor de e^λ para obter

$$\sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} = \lambda e^\lambda \text{ e } \sum_{x=1}^{\infty} x^2 \frac{\lambda^x}{x!} = (\lambda + 1)\lambda e^\lambda.$$

Disso segue que,

$$\mathbb{E}(X^2) = e^{-\lambda} \sum_{x=1}^{\infty} x^2 \frac{\lambda^x}{x!} = \lambda^2 + \lambda,$$

e daí temos

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Independência de variáveis aleatórias discretas

Duas v.a.s discretas X e Y são *independentes* se

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y),$$

para todo $x \in \Omega_X$ e $y \in \Omega_Y$.

A função $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ é chamada *função massa de probabilidade conjunta* de X e Y.

Estudaremos em detalhe as f.m.p. conjunta mais adiante.

Exemplo 3:

Em um lançamento de dois dados honestos, se X é a face do primeiro dado e Y a face do segundo, então as v.a.s $X+Y$ e $X-Y$ não são independentes.

Por um lado temos que

$$\mathbb{P}(X+Y=12, X-Y=0) = \mathbb{P}(X=6, Y=6) = 1/36.$$

Por outro lado, $\mathbb{P}(X+Y=12) = \mathbb{P}(X=6, Y=6) = 1/36$ e

$$\mathbb{P}(X-Y=0) = \mathbb{P}(X=1, Y=1) + \ldots + \mathbb{P}(X=6, Y=6) = 1/6$$

de modo que

$$\mathbb{P}(X+Y=12, X-Y=0) = \frac{1}{36} \neq \frac{1}{36}\frac{1}{6} = \mathbb{P}(X+Y=12)\mathbb{P}(X-Y=0).$$

Logo as v.a.s $X+Y$ e $X-Y$ são dependentes.

Exemplo 4:

Variância da Binomial: Se $X \sim \text{Bin}(n, p)$, então $\text{Var}(X) = np(1-p)$.

Solução: Como $X = \sum_{i=1}^{n} X_i$ com $X_1, \ldots, X_n$ i.i.d. Bernoulli de parâmetro p, temos $X_i^2 = X_i$ para todo i e assim

$$\mathbb{E}(X^2) = \mathbb{E}\left(\sum_{i=1}^{n} X_i + \sum_{i,j=1:i\neq j}^{n} X_i X_j\right).$$

Da linearidade da esperança segue que

$$\mathbb{E}(X^2) = n\mathbb{E}(X_1) + n(n-1)\mathbb{E}(X_1X_2) = np + n(n-1)\mathbb{E}(X_1X_2).$$

Usando a independência de X_1 e X_2, temos

$$\mathbb{E}(X_1X_2) = \mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) = p^2,$$

de forma que

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = np - np^2 = np(1-p).$$

Exemplo 5:

*Q*2-2016/02. Um supermercado faz a seguinte promoção: o cliente, ao passar pelo caixa, lança um dado. Se sair face 6 tem um desconto de 30% sobre o total de sua conta. Se sair face 5 o desconto é de 20%. Se sair face 4 o desconto é de 10% e se ocorrerem faces 1, 2 ou 3, o desconto é de 5%.

a) Encontre a função de probabilidade do desconto concedido.

b) Calcule o desconto médio concedido.

c) Calcule o desvio padrão do desconto médio concedido.

Solução: Seja X a v.a. que representa o desconto dado pelo supermercado.

a) Temos que $\Omega_X = \{0.05, 0.1, 0.2, 0.3\}$ e a f.m.p p_X é dada por

$$p_X(x) = \begin{cases} 1/2, & \text{se } x = 0.05 \\ 1/6, & \text{se } x = 0.1 \\ 1/6, & \text{se } x = 0.2 \\ 1/6, & \text{se } x = 0.3 \end{cases}$$

b) Queremos calcular $\mathbb{E}(X)$. Por definição,

$$\mathbb{E}(X) = 0.05 \times \frac{1}{2} + 0.1 \times \frac{1}{6} + 0.2 \times \frac{1}{6} + 0.3 \times \frac{1}{6} = 0.125$$

Continuação da solução:

c) Queremos calcular $\mathrm{DP}(X) = \sqrt{\mathrm{Var}(X)}$. Para isso, note que

$$\mathbb{E}(X^2) = (0.05)^2 \times \frac{1}{2} + (0.1)^2 \times \frac{1}{6} + (0.2)^2 \times \frac{1}{6} + (0.3)^2 \times \frac{1}{6} \approx 0.02458,$$

e lembre que

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \approx 0.02458 - (0.125)^2 \approx 0.009.$$

Logo, $\mathrm{DP}(X) = \sqrt{0.009} \approx 0.094$.

Variáveis aleatórias contínuas

Enquanto v.a.'s discretas *contam* fenômenos de interesse, v.a.'s contínuas vêm de um processo de *medição*. Alguns exemplos são:

- Tempo que leva para atendimento em uma fila
- Tempo de vida de um equipamento eletrônico
- Tamanho de uma peça em uma fábrica

Em todos esses casos note que $\Omega = [0, \infty)$, um conjunto *não-enumerável*

Dizemos que X é uma v.a. *contínua* se existe uma função $f_X : \mathbb{R} \to \mathbb{R}$, denominada *função densidade de probabilidade* (f.d.p.) satisfazendo:

1) $f_X(x) \geq 0$

2) $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$

Variáveis aleatórias contínuas

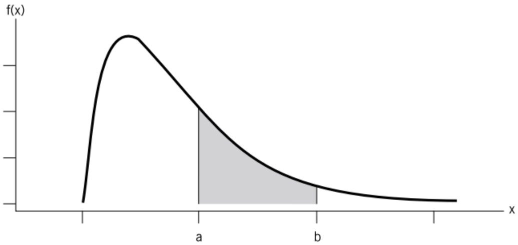


Figura 7: (Pinheiro, J. I. D., et. al. Probabilidade e Estatística, p. 83) Função densidade de probabilidade de uma variável aleatória contínua.

Note que $f_X(x)$ NÃO é $\mathbb{P}(X = x)$! Esse último valor é sempre zero:

$$\mathbb{P}(X = x) = \int_x^x f_X(t)\, dt = 0.$$

Portanto, se $a < b$ é verdade que

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b).$$

Variável aleatória uniforme

Uma v.a. contínua é dita *uniforme* no intervalo $[a, b]$ se a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \leq x \leq b \\ 0, & \text{caso contrário.} \end{cases}$$

Notação: $X \sim \text{Unif}([a, b])$.

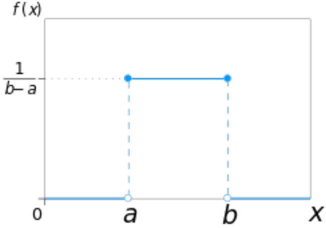


Figura 8: Distribuição Uniforme. In: Wikipédia, a enciclopédia livre.

Exemplo 1

Ao arredondar números reais para os inteiros mais próximos, os erros de arredondamento assumem valores entre $-0{,}5$ e $0{,}5$. É razoável modelá-los como uma v.a. uniforme nesse intervalo, de modo que a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} 1, & \text{se } -0{,}5 \leq x \leq 0{,}5 \\ 0, & \text{caso contrário.} \end{cases}$$

Dessa forma, a probabilidade de que o erro de arredondamento, em módulo, seja menor que ou igual a $0,2$ é dada por

$$P(|X| \leq 0{,}2) = \int_{-0{,}2}^{0{,}2} f_X(x)\, dx = 0{,}4.$$

Variável aleatória exponencial

Uma v.a. contínua é dita *exponencial* de parâmetro λ se a sua f.d.p. é dada por

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Notação: $X \sim \text{Exp}(\lambda)$.

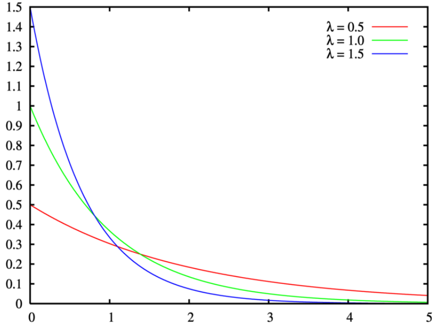


Figura 9: Distribuição Exponencial. In: Wikipédia, a enciclopédia livre.

Variável aleatória exponencial - aplicações

- Tempo entre emissões de partículas por uma substância radioativa
- Tempo entre chamadas telefônicas em uma central de auto-atendimento
- Distância entre mutações em uma fita de DNA
- Altura de moléculas de gás em um campo gravitacional uniforme a temperatura e pressão fixos
- Em Meteorologia, precipitação máxima em um determinado intervalo de tempo
- Tempo de vida útil de aparelhos eletrônicos

Exemplo 2

Seja T o tempo entre duas falhas consecutivas de um aparelho eletrônico, medido em dias, e suponha que $T \sim \text{Exp}(2)$.

a) Qual a probabilidade de termos pelo menos dois dias entre duas falhas consecutivas?

b) Sabendo que não ocorre uma falha há pelo menos cinco dias, qual a probabilidade de que precisemos esperar pelo menos mais dois dias até a próxima falha?

Solução:

a) Sendo $f_T(t) = 2e^{-2t}$, para $t \geq 0$, temos que:

$$\mathbb{P}(T > 2) = \int_2^\infty 2e^{-2t}\, dt = \frac{1}{e^4} \approx 0{,}0183.$$

b)

$$\mathbb{P}(T > 7 \mid T > 5) = \frac{\mathbb{P}(T > 7)}{\mathbb{P}(T > 5)} = \frac{\int_7^\infty 2e^{-2t}\ dt}{\int_5^\infty 2e^{-2t}\ dt} = \frac{1/e^{14}}{1/e^{10}} = \frac{1}{e^4} \approx 0{,}0183.$$

Variável aleatória exponencial - perda de memória

Teorema: Uma variável aleatória contínua X satisfaz a propriedade da *perda de memória*, i.e.,

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s), \forall s, t > 0$$

se e somente se $X \sim \text{Exp}(\lambda)$.

Prova: Provemos somente que se $X \sim \text{Exp}(\lambda)$ então X tem a propriedade da perda de memória. De fato,

$$\begin{aligned}\mathbb{P}(X > s + t \mid X > t) &= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s).\end{aligned}$$

A prova da implicação contrária foge do escopo do curso.

Função de distribuição acumulada

Independente de qual seja o tipo de uma v.a. X, sempre podemos definir a sua *função de distribuição acumulada* (f.d.a.) como

$$F_X(x) = \mathbb{P}(X \leq x), \text{ para } x \in \mathbb{R}.$$

Teorema: A f.d.a. de uma v.a. X satisfaz às propriedades abaixo:

- F_X é não-decrescente, ou seja, $x < y \Rightarrow F_X(x) \leq F_X(y)$
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- A f.d.a. caracteriza a distribuição de modo único: se Y é outra v.a. tal que $F_X = F_Y$, então a distribuição de X e Y é a mesma
- Se X é discreta, então $F_X(x) = \sum_{y \leq x} p_X(y)$
- Se X é contínua, então $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$, e $f_X(x) = F'_X(x)$

Exemplo 3

Seja $X \sim \text{Bin}(4, 1/2)$. Então as suas f.m.p. e f.d.a. são dadas, respectivamente, por

$$p_X(x) = \begin{cases} 1/16, & \text{se } x = 0 \\ 1/4, & \text{se } x = 1 \\ 3/8, & \text{se } x = 2 \\ 1/4, & \text{se } x = 3 \\ 1/16, & \text{se } x = 4 \end{cases} \quad \text{e} \quad F_X(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1/16, & \text{se } 0 \leq x < 1 \\ 5/16, & \text{se } 1 \leq x < 2 \\ 11/16, & \text{se } 2 \leq x < 3 \\ 15/16, & \text{se } 3 \leq x < 4 \\ 1, & \text{se } x \geq 4. \end{cases}$$

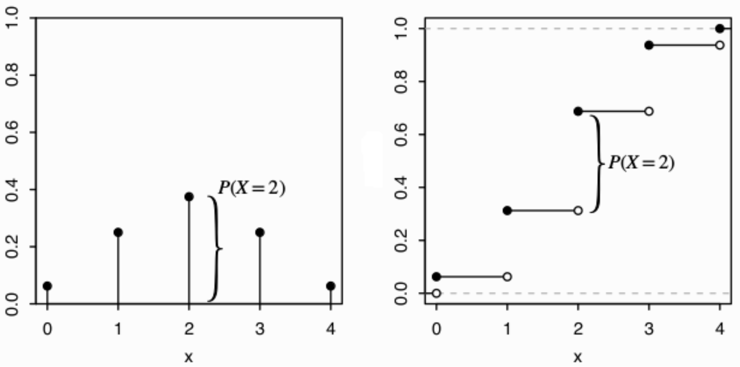


Figura 10: (Blitzstein, J.K. e Hwang, J. Introduction to Probability, p.109.) Função massa de probabilidade e função de distribuição acumulada de uma variável aleatória $X \sim \text{Bin}(4, 1/2)$.

Note que $\mathbb{P}(X = x) = F_X(x) - \lim_{y \to x^-} F_X(y)$, ou seja, a probabilidade do valor x ser assumido é o tamanho do saldo da f.d.a. no ponto x.

Variável aleatória exponencial

Lembremos que $X \sim \text{Exp}(\lambda)$ se a sua f.d.p. é dada por $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$, de modo que $F_X(x) = 0$ para $x < 0$, e

$$F_X(x) = \mathbb{P}(X \leq x) = \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}, \text{ se } x \geq 0.$$

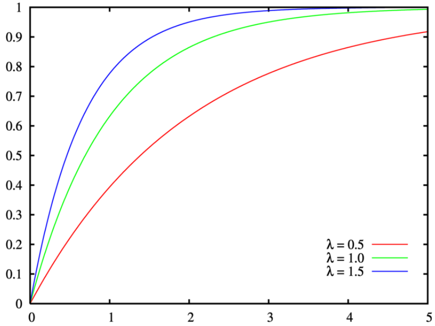


Figura 11: Distribuição Exponencial. In: Wikipédia, a enciclopédia livre.

Relação entre exponencial e Poisson

Um processo de "chegadas"(ou ocorrências) a tempo contínuo é chamado de *Processo de Poisson* com taxa $\lambda > 0$ se:

1. O número de chegadas que ocorrem em um intervalo de comprimento t é uma v.a. Poisson de parâmetro λt.
2. O número de chegadas que ocorrem em intervalos disjuntos são independentes uns dos outros.

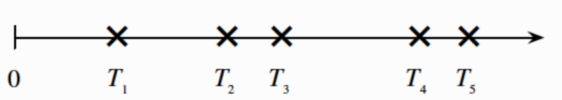


Figura 12: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.223.) Processo de Poisson. Cada marca X indica a ocorrência de uma chegada.

- Em um processo de Poisson, denote por N_t o número de chegadas no intervalo $[0, t]$. Temos que $N_t \sim \text{Poi}(\lambda t)$.
- Seja T_1 o tempo da primeira chegada. Qual é a distribuição de T_1? Para cada $t > 0$, observe que

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t},$$

  de modo que $F_{T_1}(t) = 1 - e^{-\lambda t}$ e portanto $T_1 \sim \text{Exp}(\lambda)$.
- Qual a distribuição de $T_2 - T_1$, o tempo entre a primeira e segunda chegadas? Como o número de chegadas em intervalos disjuntos são independentes, $T_2 - T_1$ é independente de T_1.[1] Além disso, usando o mesmo argumento acima, podemos deduzir que $T_2 - T_1 \sim \text{Exp}(\lambda)$.
- De modo similar, $T_3 - T_2 \sim \text{Exp}(\lambda)$, independentemente de $T_2 - T_1$ e T_1.

[1] Isso requer demonstração. Envolve a falta de memória da exponencial.

- De modo geral, o tempo entre duas chegadas sucessivas é uma v.a. exponencial de parâmetro λ, independentemente dos tempos entre as chegadas anteriores.
- **Conclusão:** em um Processo de Poisson de taxa $\lambda > 0$,
    - o número de chegadas em um intervalo de comprimento t é $\text{Poi}(\lambda t)$, e
    - os tempos entre chegadas são $\text{Exp}(\lambda)$ independentes.
- Assim, o Processo de Poisson relaciona as v.a.s Poisson (discreta) com as exponenciais (contínua).

Média e variância de uma variável aleatória contínua

Analogamente ao caso discreto, se X é uma v.a. contínua com f.d.p. f_X, definimos a esperança de X por

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx,$$

e a variância de X por

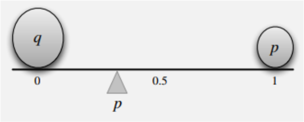
$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

Também é verdade que $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}[X]^2$, onde

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx.$$

Média como centro de massa

Note que podemos fazer uma analogia entre a média de uma v.a. e o centro de massa de um corpo cuja densidade de massa é a f.m.p. ou f.d.p. da v.a. em questão:



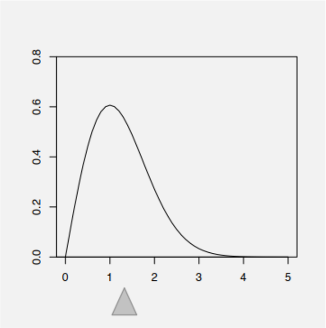


Figura 13: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.139/201.) Média como centro de massa, casos discreto e contínuo, respectivamente.

Média e variância da uniforme

Seja $X \sim \text{Unif}([a,b])$. É fácil verificar que:

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a}\ dx = \frac{a+b}{2}$$
$$\mathbb{E}[X^2] = \int_a^b x^2 \frac{1}{b-a}\ dx = \frac{a^2+ab+b^2}{3}$$

Logo, deduzimos que

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{(b-a)^2}{12}.$$

Média e variância da exponencial

Seja $X \sim \text{Exp}(\lambda)$. É fácil verificar que (integração por partes):

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$$
$$\mathbb{E}[X^2] = \int_0^\infty x^2\lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}$$

Assim, temos que

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{\lambda^2}.$$

Exemplo 1

*Q*1-2018/01. O Chile encontra-se em uma das regiões do mundo que apresentam os maiores índices de ocorrência de terremotos. Suponha que o intervalo de tempo T entre a ocorrência de dois terremotos sucessivos possa ser modelado por uma variável aleatória exponencial com média desconhecida $1/\lambda$, em meses. Com base em dados históricos sabe-se que a probabilidade de que o intervalo de tempo entre dois terremotos exceda um ano é de 0,001%.

a) Usando a informação contida nos dados históricos, calcule o valor de λ e em seguida escreva a expressão matemática da função densidade de probabilidade da variável aleatória T.

b) Determine o número médio de terremotos no Chile durante um período de 6 meses e em seguida calcule a probabilidade de que nesse período de tempo ocorram no máximo dois terremotos.

Solução:

a) Temos que $10^{-5} = \mathbb{P}(T > 12) = e^{-12\lambda}$, de modo que $\lambda = \ln(10^5)/12 \approx 0{,}9594$. Portanto, a f.d.p. de T é dada por

$$F_T(t) = \begin{cases} 0{,}9594e^{-0{,}9594t}, & \text{se } t \geq 0 \\ 0, & \text{caso contrário.} \end{cases}$$

b) Seja X a v.a. que representa o número de terremotos no Chile em um período de 6 meses. Então $X \sim \text{Poi}(6 \times 0{,}9594)$, de modo que $\mathbb{E}(X) = 6 \times 0{,}9594 \approx 5{,}76$, e

$$\mathbb{P}(X \leq 2) = \sum_{x=0}^{2} \mathbb{P}(X = x) = \sum_{x=0}^{2} e^{-5{,}76} \frac{(5{,}76)^x}{x!} \approx 0{,}0736.$$

Outras medidas de centralidade e dispersão

- O quantil q de uma v.a. X com f.d.a. F_X é o menor número ζ_q que satisfaz $F_X(\zeta_q) \geq q$
- O primeiro quartil ou quartil inferior da v.a. X, denotado por $q_1(X)$, é o quantil $\zeta_{0,25}$
- O segundo quartil ou mediana da v.a. X, denotado por $q_2(X)$, é o quantil $\zeta_{0,5}$
- O terceiro quartil ou quartil superior da v.a. X, denotado por $q_3(X)$, é o quantil $\zeta_{0,75}$
- A distância interquartil, denotada por $\mathrm{DIQ}(X)$, é a diferença entre os quartis superior e inferior, isto é $\mathrm{DIQ}(X) = q_3(X) - q_1(X)$.

Observação: A mediana é uma outra medida de centralidade, e a distância interquartil é uma outra medida de dispersão.

Exemplo 2

Seja $X \sim \text{Exp}(2)$. Calcule $\mathbb{E}(X)$, $\text{Var}(X)$, $\text{DP}(X)$, $q_1(X)$, $q_2(X)$, $q_3(X)$ e $\text{DIQ}(X)$.

Solução: Sabemos que $\mathbb{E}(X) = 1/2$, $\text{Var}(X) = 1/4$ e $\text{DP}(X) = 1/2$. Para calcular o quantil q, note que ele é o valor ζ_q tal que

$$q = F_X(\zeta_q) = \int_0^{\zeta_q} 2e^{-2x}\,dx = 1 - e^{-2\zeta_q} \Rightarrow \zeta_q = -\frac{1}{2}\ln(1-q).$$

Portanto, temos que:

$$q_1(X) \approx 0{,}1438 \quad q_2(X) \approx 0{,}3466 \quad q_3(X) \approx 0{,}6931 \quad \text{DIQ}(X) = 0{,}5493$$

$\Rightarrow$ Não é verdade que $\mathbb{E}[X] = q_2(X)$ e $\text{Var}(X) = \text{DP}(X)$!

Variável aleatória normal

Uma v.a. X é *normal* ou *gaussiana* de parâmetros μ e σ^2 se a sua f.d.p. é dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \text{ para } x \in \mathbb{R}.$$

Notação: $X \sim N(\mu, \sigma^2)$.

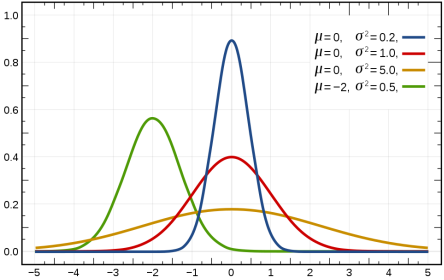


Figura 14: (Normal Distribution. In: Wikipédia, a enciclopédia livre.) Visualização de uma variável aleatória normal, com diversos valores de μ e σ^2.

Variável aleatória normal

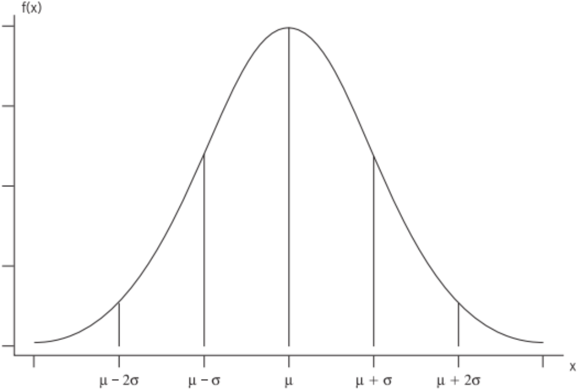


Figura 15: (Pinheiro, J. I. D., et. al. Probabilidade e Estatística, p. 99) Função densidade de probabilidade de uma variável aleatória normal, exibindo alguns valores importantes: 68,3% de probabilidade encontra-se entre $\mu - \sigma$ e $\mu + \sigma$; 95,4% de probabilidade encontra-se entre $\mu - 2\sigma$ e $\mu + 2\sigma$.

Variável aleatória normal

- Uma das distribuições mais importantes da Probabilidade e Estatística
- Teorema Central do Limite: "Se $E_1, \ldots, E_n$ são quantidades aleatórias independentes, então $E_1 + \ldots E_n$ tem distribuição aproximadamente normal, se n é suficientemente grande". Voltaremos a isso em breve!
- Pode ser usada para modelar fenômenos que são descritos pela sobreposição de diversos fenômenos independentes de origem desconhecida (e.g., interferências e ruído de medição de equipamentos)
- É impossível calcular $\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)\, dx$ analiticamente $\Rightarrow$ uso de tabelas para o cálculo de probabilidades

Variável aleatória normal

Dizemos que Z tem distribuição *normal padrão* se $Z \sim N(0,1)$. Denotamos por φ e Φ as suas f.d.p. e f.d.a., respectivamente:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

$$\Phi(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)\, dz$$

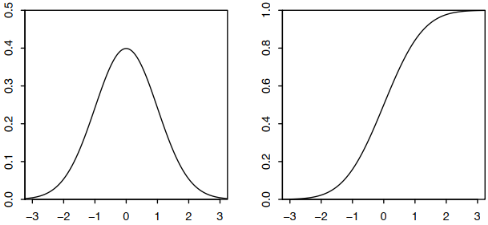


Figura 16: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.212.) Gráficos das funções φ e Φ, respectivamente.

Variável aleatória normal - propriedades

Se $X \sim N(\mu, \sigma^2)$ então vale que:

- $\mathbb{E}[X] = \mu$ e $\text{Var}(X) = \sigma^2$
- A f.d.p. de X é simétrica com respeito a $x = \mu$
- Em particular, $\Phi(-z) = 1 - \Phi(z)$, para todo $z \in \mathbb{R}$
- $\dfrac{X - \mu}{\sigma}$ tem distribuição normal padrão

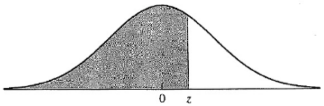
Essa última propriedade nos permite calcular qualquer probabilidade usando uma tabela da normal padrão:

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

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

$$\mathbb{P}(X \geq a) = \mathbb{P}\left(Z \geq \frac{a-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

A tabela da normal padrão



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319

Figura 17: Excerto de uma tabela de probabilidades da normal padrão. Note que, pela simetria, não é necessário informar probabilidades para os valores negativos de z.

Exemplo 1

Suponha que o tempo X, em minutos, corresponde ao tempo que uma pessoa leva para executar determinada tarefa e varia conforme uma distribuição normal com parâmetros μ (média) e σ^2 (desvio padrão). Suponha também que a probabilidade de que a tarefa seja executada em 70 minutos no máximo é 0,75, e a probabilidade de que a tarefa seja executada em no máximo 50 minutos é 0,25.

a) Determine os valores de μ e σ^2.

b) De todas as pessoas que necessitam de pelo menos 75 minutos para executar a tarefa, qual a porcentagem correspondente das pessoas que precisarão de mais de 85 minutos?

c) Qual a probabilidade de que o tempo que uma pessoa leva para executar a tarefa se desvie da média por menos que dois minutos?

Solução:

a) Sabemos que:

$$0{,}75 = \mathbb{P}(X \leq 70) = \Phi\left(\frac{70-\mu}{\sigma}\right) \Rightarrow \frac{70-\mu}{\sigma} = 0{,}67$$

$$0{,}25 = \mathbb{P}(X \leq 50) = \Phi\left(\frac{50-\mu}{\sigma}\right) \Rightarrow \frac{50-\mu}{\sigma} = -0{,}67$$

Temos então um sistema com duas equações e duas incógnitas para μ e σ, que ao ser resolvido, resulta em $\mu = 60$ e $\sigma = 14{,}9$ minutos.

b) Queremos calcular:

$$\mathbb{P}(X \geq 85 | X \geq 75) = \frac{\mathbb{P}(X \geq 85)}{\mathbb{P}(X \geq 75)} = \frac{1-\Phi(1{,}68)}{1-\Phi(1{,}01)} \approx 0{,}2977.$$

c) Temos que:

$$\begin{aligned}\mathbb{P}(|X-\mu| \leq 2) &= \mathbb{P}(-2 \leq X-\mu \leq 2) = \mathbb{P}(-2/\sigma \leq Z \leq 2/\sigma) \\ &= \mathbb{P}(-0{,}13 \leq Z \leq 0{,}13) = 2\Phi(0{,}13) - 1 \approx 0{,}1034.\end{aligned}$$

Função de variável aleatória

Se X é uma v.a. e $g : \mathbb{R} \to \mathbb{R}$ é uma função, então podemos estar interessados na distribuição de $Y = g(X)$:

- Se D é o diâmetro de um rolamento esférico, então seu volume é dado por $\dfrac{\pi D^3}{6}$
- Se V é a voltagem em um circuito elétrico onde passa uma corrente i (valor fixo conhecido), então a potência é dada por $W = Vi$
- Se X e Y são peso e altura, respectivamente, de indivíduos em uma população, então o IMC é dado por X/Y^2

Em certos casos é possível encontrar a distribuição de Y conhecendo X, porém não nos focaremos nisso no curso.

Ainda assim, vejamos um exemplo a fim de esclarecimentos.

Exemplo 1

Augusto é o gerente de uma revendedora de carros. Toda semana ele tem cinco carros para venda. Se ele vender até dois carros, não ganha qualquer adicional ao seu salário; porém, se conseguir vender três ou mais carros, ganha um prêmio de R$ 500,00 por cada carro vendido. Suponha que as chances de venda dos diversos carros são independentes e que a probabilidade de cada carro ser vendido é 0,6. Determine a função massa de probabilidade do prêmio semanal ganho por Augusto.

Solução: Temos que $X \sim \text{Bin}(5, 0{,}6)$, e se Y é o prêmio semanal ganho por Augusto, temos que

$$Y = \begin{cases} 0, & \text{se } X \leq 2 \\ 500X, & \text{se } X \geq 3. \end{cases}$$

Portanto, temos que:

$$\begin{aligned} \mathbb{P}(Y = 0) &= \mathbb{P}(X \leq 2) = 0{,}3174 \\ \mathbb{P}(Y = 1.500) &= \mathbb{P}(X = 3) = 0{,}3456 \\ \mathbb{P}(Y = 2.000) &= \mathbb{P}(X = 4) = 0{,}2592 \\ \mathbb{P}(Y = 2.500) &= \mathbb{P}(X = 5) = 0{,}0778. \end{aligned}$$

Esperança e variância de função de v.a.

Teorema: Se X é v.a. discreta, então $Y = g(X)$ também é discreta e vale que

$$\mathbb{E}[g(X)] = \sum_{x \in \Omega_X} g(x)\ \mathbb{P}(X = x)$$

$$\mathrm{Var}(g(X)) = \sum_{x \in \Omega_X} (g(x) - \mathbb{E}[g(X)])^2\ \mathbb{P}(X = x) = \mathbb{E}[g(X)^2] - \mathbb{E}[g(X)]^2$$

Teorema: Se X é v.a. contínua com f.d.p. f_X e g é tal que $Y = g(X)$ também o seja, vale que

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx$$

$$\mathrm{Var}(g(X)) = \int_{-\infty}^{\infty} (g(x) - \mathbb{E}[g(X)])^2 f_X(x)\ dx = \mathbb{E}[g(X)^2] - \mathbb{E}[g(X)]^2$$

Exemplo 1 (continuação)

Se quiséssemos saber apenas o prêmio médio semanal ganho por Augusto e o seu respectivo desvio padrão, bastaria notarmos que:

$$\mathbb{E}[Y] = \sum_{k=0}^{2} 0\mathbb{P}(X = k) + \sum_{k=3}^{5} 500k\mathbb{P}(X = k) = \text{R\$ } 1.231{,}30$$

$$\mathbb{E}[Y^2] = \sum_{k=0}^{2} 0^2\mathbb{P}(X = k) + \sum_{k=3}^{5} (500k)^2\mathbb{P}(X = k) = (\text{R\$})^2 \ 2.300.650$$

$$\text{Var}(Y) = 2.300.650 - (1.321{,}30)^2 = (\text{R\$})^2 \ 784.550{,}31$$

$$\text{DP}(Y) = \sqrt{\text{Var}(Y)} = \text{R\$ } 885{,}75$$

Propriedades da média e da variância

- Se c é uma constante, então $\mathbb{E}[c] = c$ e $\text{Var}(c) = 0$
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$, para $a, b \in \mathbb{R}$
- Se $\mathbb{E}[X]$ é finita, então $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$, para $a, b \in \mathbb{R}$
- $\text{DP}(aX + b) = |a|\text{DP}(X)$
- Se $\mu = \mathbb{E}[X]$ e $\sigma^2 = \text{Var}(X)$ são finitas, então $Z = \dfrac{X - \mu}{\sigma}$ tem média zero e variância 1.

Atenção: Se X é normal então $\dfrac{X - \mu}{\sigma}$ também o é, porém isso não vale para qualquer distribuição!

Exemplo 2

*Q*4-2015/02. Uma empresa transporta carga e o custo $C(X)$ do transporte, em u.m., depende do peso X, em toneladas, da carga a ser transportada, conforme a tabela abaixo:

Valores de X	$X < 4$	$4 < X < 5$	$5 < X < 6$	$X > 6$
$C(X)$	400	600	800	1.500

Se o peso da carga a ser transportada for uma variável aleatória X com distribuição $N(5, 0{,}25)$, determine:

a) O valor esperado do custo do transporte.

b) O desvio padrão do custo do transporte

Solução:

a) Temos que:

$$\begin{aligned}\mathbb{P}(X < 4) &= \Phi(-2) = 0{,}0228\\ \mathbb{P}(4 < X < 5) &= \Phi(0) - \Phi(-2) = 0{,}4772\\ \mathbb{P}(5 < X < 6) &= \Phi(2) - \Phi(0) = 0{,}4772\\ \mathbb{P}(X > 6) &= 0{,}0228,\end{aligned}$$

e portanto,

$$\begin{aligned}\mathbb{E}[C(X)] = 400 \times 0{,}0228 + 600 \times 0{,}4772\\ + 800 \times 0{,}4772 + 1.500 \times 0{,}0228 = 711{,}40 \text{ u.m.}\end{aligned}$$

b) Analogamente, calculamos

$$\begin{aligned}\mathrm{Var}(C(X)) = 400^2 \times 0{,}0228 + 600^2 \times 0{,}4772\\ + 800^2 \times 0{,}4772 + 1.500^2 \times 0{,}0228 - (711{,}40)^2 = 26.058 \text{ u.m.}^2,\end{aligned}$$

de modo que $\mathrm{DP}(C(X)) = \sqrt{\mathrm{Var}(C(X))} = 161{,}43$ u.m.

Variáveis aleatórias Bidimensionais

Em diversas ocasiões é interessante estudar a relação entre múltiplas variáveis aleatórias de um mesmo experimento. Por exemplo:

- *Medicina*: Para avaliar a efetividade de um tratamento, pode-se tomar várias medidas por paciente; a pressão arterial, a frequência cardíaca e as taxas de colesterol podem ser mais informativas do que qualquer uma dessas medidas separadamente.
- *Finanças:* Para estudar a evolução do preço das ações de uma empresa, coletamos uma série de medidas ao longo do tempo, e as estumados conjuntamente. Essa série de medidas quando considerada conjuntamente pode ser usada para deduzir a tendência do preço das ações da empresa.

Distribuições conjuntas

Sejam X e Y v.a.s definidas no mesmo espaço amostral. O par (X, Y) é chamado *vetor aleatório bidimensional*.

O vetor aleatório bidimensional (X, Y) é chamado *discreto* se X e Y são v.a.s discretas.

A *função massa de probabilidade conjunta* do v.a. (X, Y) discreto é a função $p_{X,Y}(x, y)$ definida por

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) \text{ para todo } (x, y) \in \Omega_X \times \Omega_Y.$$

Note que a f.m.p. conjunta $p_{X,Y}$ satisfaz

$$\begin{cases} p_{X,Y}(x, y) \geq 0 \text{ para todo } (x, y) \in \Omega_X \times \Omega_Y \\ \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} p_{X,Y}(x, y) = 1. \end{cases}$$

Distribuições marginais

Considere um v.a. bidimensional (X, Y) discreto. A *função massa de probabilidade marginal* de X é a função $p_X(x)$ definida por

$$p_X(x) = \sum_{y \in \Omega_Y} \mathbb{P}(X = x, Y = y) \text{ para todo } x \in \Omega_X.$$

De modo similar, a *função massa de probabilidade marginal* de Y é a função $p_Y(y)$ definida por

$$p_Y(y) = \sum_{x \in \Omega_X} \mathbb{P}(X = x, Y = y) \text{ para todo } y \in \Omega_Y.$$

Exemplo 1

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x,y) = \begin{cases} 1/2, \text{ se } (x,y) = (0,0) \\ 0, \text{ se } (x,y) = (0,1) \\ 1/4, \text{ se } (x,y) = (1,0) \\ 1/4, \text{ se } (x,y) = (1,1) \end{cases}$$

Marginal de X: $p_X(0) = 1/2 = p_X(1) = 1/2$.

Marginal de Y : $p_Y(0) = 3/4$ e $p_Y(1) = 1/4$.

Logo, $X \sim \text{Ber}(1/2)$ e $Y \sim \text{Ber}(1/4)$.

Variáveis aleatórias independentes

Lembre que duas v.a.s discretas X e Y são *independentes* se

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y),$$

para todo $x \in \Omega_X$ e $y \in \Omega_Y$.

Em outras palavras, as v.a.s aleatórias X e Y são independentes se a f.m.p. conjunta $p_{X,Y}$ é o produto das maginais p_X e p_Y:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y),$$

para todo $x \in \Omega_X$ e $y \in \Omega_Y$.

De volta ao exemplo 1

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x,y) = \begin{cases} 1/2, \text{ se } (x,y) = (0,0) \\ 0, \text{ se } (x,y) = (0,1) \\ 1/4, \text{ se } (x,y) = (1,0) \\ 1/4, \text{ se } (x,y) = (1,1) \end{cases}$$

Vimos que a marginal de X é: $p_X(0) = 1/2 = p_X(1) = 1/2$.

Vimos também que a marginal de Y é: $p_Y(0) = 3/4$ e $p_Y(1) = 1/4$.

As v.a.s X e Y são independentes? Não, já que

$$p_{X,Y}(0,1) = 0 \neq \frac{1}{2}\frac{1}{4} = p_X(0)p_Y(1).$$

Distribuição condicional

Seja (X, Y) um v.a. com f.m.p. conjunta $p_{X,Y}(x, y)$. A *função massa de probabilidade condicional* de X dado $Y = y$, com $y \in \Omega_Y$ tal que $p_Y(y) > 0$, é a função $p_{X|Y}(x|y)$ definida por

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \text{ para todo } x \in \Omega_X.$$

Note que a f.m.p. condicional $p_{X|Y}$ satisfaz

$$\begin{cases} p_{X|Y}(x|y) \geq 0 \text{ para todo } (x, y) \in \Omega_X \times \Omega_Y \\ \sum_{x \in \Omega_X} p_{X|Y}(x|y) = 1 \text{ para cada } y \in \Omega_Y \end{cases}$$

Ainda o exemplo 1

Considere o v.a. (X, Y) discreto com f.m.p. conjunta

$$p_{X,Y}(x,y) = \begin{cases} 1/2, \text{ se } (x,y) = (0,0) \\ 0, \text{ se } (x,y) = (0,1) \\ 1/4, \text{ se } (x,y) = (1,0) \\ 1/4, \text{ se } (x,y) = (1,1) \end{cases}$$

A f.m.p condicional de X dado $Y = 0$:

$$p_{X|Y}(x|0) = \frac{p_{X,Y}(x,0)}{p_Y(0)} = \begin{cases} 2/3, \text{ se } x = 0 \\ 1/3, \text{ se } x = 1 \end{cases}$$

A f.m.p. condicional de X dado $Y = 1$:

$$p_{X|Y}(x|1) = \frac{p_{X,Y}(x,1)}{p_Y(1)} = \begin{cases} 0, \text{ se } x = 0 \\ 1, \text{ se } x = 1 \end{cases}$$

Exemplo 2

*Q*4-2018/02. Foi feito o loteamento de uma área rural em terrenos retangulares. Para cada terreno, seu comprimento e sua largura, ambos em decâmetros, podem ser iguais a 1 dam, 2 dam ou 3 dam. Denote por X o comprimento e Y a largura de um terreno sorteado ao acaso. A tabela a seguir fornece (apenas parcialmente) a distribuição conjunta de X e Y, variáveis aleatórias supostas independentes:

	Y			
X	1	2	3	$p_X(x)$
1	0,35		0,14	
2				0,20
3	0,05			
$p_Y(y)$		0,30		

Considere as v.a.s $V = 2X + 2Y =$ perímetro do terreno e $W = XY =$ área do terreno (em dam^2) e calcule:

a) o valor de cada probabilidade, conjunta ou marginal, omitido na tabela acima.

b) a probabilidade condicional de que a área seja igual a 4 dam^2, dado que o perímetro é igual a 8 dam.

Solução:

a) Denote $\mathbb{P}(X=1)=a$ e $\mathbb{P}(Y=1)=b$. Em seguida, use a independência para concluir que $a=0.7$ e $b=0.5$. Com essas informações, podemos deduzir que

$$\begin{cases} \mathbb{P}(X=3)=1-\mathbb{P}(X=1)-\mathbb{P}(X=2)=1-0.7-0.2=0.1 \\ \mathbb{P}(Y=3)=1-\mathbb{P}(Y=1)-P(Y=2)=1-0.5-0.3=0.2 \\ \mathbb{P}(X=1,Y=2)=(0.7)(0.3)=0.21 \\ \mathbb{P}(X=2,Y=1)=(0.2)(0.5)=0.10 \\ \mathbb{P}(X=2,Y=3)=(0.2)(0.2)=0.04 \\ \mathbb{P}(X=3,Y=2)=(0.1)(0.3)=0.03 \\ \mathbb{P}(X=3,Y=3)=(0.1)(0.2)=0.02 \end{cases}$$

Continuação da solução:

b) Queremos calcular $\mathbb{P}(W = 4|V = 8)$. Ora,

$$\mathbb{P}(W = 4|V = 8) = \frac{P(V = 8, W = 4)}{P(V = 8)}$$
$$= \frac{\mathbb{P}(2X + 2Y = 8, XY = 4)}{\mathbb{P}(2X + 2Y = 8)}.$$

Agora, $\mathbb{P}(2X + 2Y = 8)$ pode ser reescrita como

$$\mathbb{P}(X = 1, Y = 3) + \mathbb{P}(X = 2, Y = 2) + \mathbb{P}(X = 3, Y = 1) = 0.25,$$

e $\mathbb{P}(2X + 2Y = 8, XY = 4) = \mathbb{P}(X = 2, Y = 2) = 0.06$. Logo,

$$\mathbb{P}(W = 4|V = 8) = \frac{0.06}{0.25} = 0.24.$$

Exemplo 3

Para cada uma das $N \sim \text{Poi}(\lambda)$ bolas de uma urna atribuímos independentemente uma cor, vermelha ou azul. A cor vermelha é atribuída com probabilidade p e a cor azul com probabilidade $q = 1 - p$. Seja X o número de bolas vermelhas e Y o número de bolas azuis, de forma que $X + Y = N$. Encontre a f.m.p. conjunta de X e Y.

Solução: Note que para $x, y \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(X = x, Y = y) &= \mathbb{P}(N = x + y)\mathbb{P}(X = x, Y = y | N = x + y) \\
&= \mathbb{P}(N = x + y)\mathbb{P}(X = x | N = x + y) \\
&= \frac{e^{-\lambda}\lambda^{x+y}}{(x+y)!}\frac{(x+y)!}{x!y!}p^x q^y \\
&= \frac{e^{-p\lambda}(p\lambda)^x}{x!}\frac{e^{-q\lambda}(q\lambda)^y}{y!}.
\end{aligned}$$

Em particular, $X \sim \text{Poi}(p\lambda)$ e $Y \sim \text{Poi}(q\lambda)$. Além disso, X e Y são independentes!

Esperança e Variância Condicionais

Seja (X, Y) v.a. bidimensional discreto. Definimos a *esperança condicional* de Y dado $X = x \in \Omega_X$ como

$$\begin{aligned}\mathbb{E}(Y|X=x) &= \sum_{y \in \Omega_Y} y p_{Y|X}(y|x) \\ &= \sum_{y \in \Omega_Y} y \mathbb{P}(Y = y | X = x).\end{aligned}$$

Definimos a *Variância condicional* de Y dado $X = x \in \Omega_X$ como

$$\begin{aligned}\mathrm{Var}(Y|X=x) &= \mathbb{E}(Y^2|X=x) - (\mathbb{E}(Y|X=x))^2. \\ &= \sum_{y \in \Omega_Y} y^2 p_{Y|X}(y|x) - \left(\sum_{y \in \Omega_Y} y p_{Y|X}(y|x) \right)^2.\end{aligned}$$

Exemplo 4

Sejam $X, Y \sim \text{Poi}(\lambda)$ independentes e defina a v.a. $T = X + Y$. Determine $\mathbb{E}(X|T = t)$ e $\text{Var}(X|T = t)$.

Solução: Note que $X|T = t \sim \text{Bin}(t, 1/2)$ de forma que

$$\mathbb{E}(X|T = t) = \frac{t}{2} \text{ e } \text{Var}(X|T = t) = \frac{t}{4}.$$

Covariância e Correlação

A *Covariância* entre v.a.s X e Y é definida por

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) \\ &= \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y) \text{ (Linearidade da esperança).}\end{aligned}$$

As seguintes propriedades são imediatas da definição:

1. $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$.
2. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$
3. $\mathrm{Cov}(X,Z)=0$, se Z é v.a. constante com probabilidade 1.
4. $\mathrm{Cov}(aX,Y)=a\mathrm{Cov}(X,Y)$, para $a\in\mathbb{R}$.

5. Considere v.a.s X e Y. Para quaisquer números reais a, b, c e d,

$$\begin{aligned}\mathrm{Cov}\left(aX + bY, cX + dY\right) = ac\mathrm{Var}(X) + bd\mathrm{Var}(Y) \\ + (ad + bc)\mathrm{Cov}(X, Y).\end{aligned}$$

6. $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$.

Teorema 1. Se as v.a.s X e Y são independentes, então elas são *não correlacionadas*, i.e., $\mathrm{Cov}(X, Y) = 0$.

Do Teorema 1, segue mais uma propriedade:

7. Se X e Y são v.a.s independentes

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y).$$

Exemplo 1

A recíproca do Teorema 1 não é verdadeira. De fato, seja $Z \sim \mathcal{N}(0, 1)$, defina as v.a.s $X = Z$ e $Y = Z^2$ e note que

$$\mathrm{Cov}(X, Y) = \mathbb{E}(Z^3) - \mathbb{E}(Z)\mathbb{E}(Z^2) = 0.$$

A *Correlação* entre v.a.s X e Y é definida por

$$\operatorname{Cor}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\operatorname{DP}(X)\operatorname{DP}(Y)}.$$

Teorema 2. Para v.a.s X e Y, temos $-1 \leq \operatorname{Cor}(X, Y) \leq 1$.

Se $\operatorname{Cor}(X, Y) > 0$, dizemos que X e Y são *positivamente correlacionados*. Se $\operatorname{Cor}(X, Y) < 0$, dizemos que X e Y são *negativamente correlacionados*.

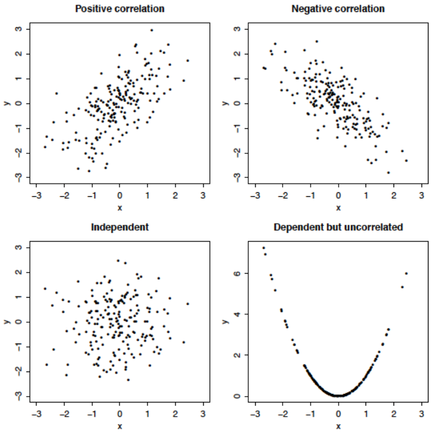


Figura 18: (Blitzstein, J.K. e Hwang, J. Introduction to probability, p.302.) Realizações da distribuição conjunta de (X, Y) sob várias estruturas dependência. Superior esquerdo: X e Y são positivamente correlacionadas. Superior direito: X e Y são negativamente correlacionadas. Inferior esquerdo: X e Y são independentes. Inferior direito: Y é uma função determinística de X, mas X e Y são não correlacionados.

Teorema 3. Se X e Y são v.a.s tais que $Y = aX + b$ com números reais a e b, onde $a \neq 0$, então

$$\text{Cor}(X, Y) = \begin{cases} 1, & \text{se } a > 0 \\ -1, & \text{se } a < 0 \end{cases}$$

De volta ao Exemplo 3

Para cada uma das $N \sim \text{Poi}(\lambda)$ bolas de uma urna atribuímos independentemente uma cor, vermelha ou azul. A cor vermelha é atribuída com probabilidade p e a cor azul com probabilidade $q = 1 - p$. Seja X o número de bolas vermelhas e Y o número de bolas azuis, de forma que $X + Y = N$. Encontre a correlação entre as v.a.s X e N.

Solução: Vimos no exemplo 3 que as v.a.s X e Y são independentes tais que $X \sim \text{Poi}(p\lambda)$ e $Y \sim \text{Poi}(q\lambda)$, onde $q = 1 - p$. Logo, temos que

$$\text{Cov}(N, X) = \text{Cov}(X + Y, X) = \text{Var}(X) = p\lambda,$$

e

$$\text{Cor}(N, X) = \frac{\text{Cov}(N, X)}{\text{DP}(N)\text{DP}(X)} = \frac{p\lambda}{\sqrt{\lambda}\sqrt{p\lambda}} = \sqrt{p}.$$

Exemplo 4

Q4-2017/02. PEGAR ARQUIVO .TEX E COLAR AQUI!

Variáveis aleatórias multidimensionais

Vetores aleatórios multidimensionais são úteis para estudar várias características numéricas do mesmo experimento aleatório (*generalização das variáveis bidimensionais*).

Um vetor $(X_1, X_2, \ldots, X_n)$ cujas componentes são v.a.s definidas no mesmo espaço amostral é chamado *vetor aleatório n-dimensional.*

A *função de distribuição conjunta* de um vetor aleatório $(X_1, \ldots, X_n)$ é uma função $F_{X_1,\ldots,X_n} : \mathbb{R}^n \to \mathbb{R}$ definida, $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$, como

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right)$$

Vetores aleatórios discretos

Um vetor aleatório $(X_1, \ldots, X_n)$ cujas componentes são v.a. discretas é dito vetor aleatório discreto

Dado um vetor discreto $(X_1, \ldots, X_n)$, definimos a função *massa de probabilidade conjunta* $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ por

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := \mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n\right),$$

para todo $(x_1, \ldots, x_n) \in \Omega_{X_1} \times \ldots \times \Omega_{X_n}$.

As seguintes propriedades valem:

- $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$ para todos $(x_1, \ldots, x_n) \in \mathbb{R}^n$
- $\sum_{(x_1,\ldots,x_n)\in\mathbb{R}^n} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 1$

A massa de probabilidade *marginal* de X_i é obtida a partir da *conjunta*:

$$p_{X_i}(x_i) := \mathbb{P}(X_i = x_i) = \sum_{k \neq i} \sum_{x_k \in \Omega_{X_k}} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

Vetores aleatórios contínuos

Um vetor aleatório $(X_1, \ldots, X_n)$ cujas componentes são v.a. contínuas é dito vetor aleatório contínuo.

Nesse caso existe uma função $f_{X_1,\ldots,X_n} : \mathbb{R}^n \to \mathbb{R}$ chamada *densidade conjunta* tal que para todos $A \subset \mathbb{R}^n$

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = \int \underset{A}{\cdots} \int f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n.$$

As seguintes propriedades valem:

- $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$ para todos $(x_1, \ldots, x_n) \in \mathbb{R}^n$
- $\int_{\mathbb{R}^n} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1$

A densidade *marginal* de X_i é obtida da *conjunta*

$$f_{X_i}(x_i) = \int \cdots \int f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \prod_{\substack{j=1 \\ j \neq i}}^{n} dx_j$$

Exemplo: vetores aleatórios contínuos*

Seja (X, Y, Z) um vetor aleatório contínuo cuja densidade conjunta é

$$f_{X,Y,Z}(x,y,z) = \begin{cases} 6xy^2z & \text{se } 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ 0 \leq z \leq \sqrt{2} \\ 0 & \text{caso contrário} \end{cases}$$

Verifique que $f_{X,Y,Z}$ é uma densidade e calcule a marginal de X

Temos que $f_{X,Y,Z} \geq 0$ para todos x, y, z e

$$\int\int\int f_{X,Y,Z}(x,y,z)dxdydz = \int_0^{\sqrt{2}}\int_0^1\int_0^1 6xy^2zdxdydz = 1$$

$$f_X(x) = \int\int f_{X,Y,Z}dydz = \begin{cases} \int_0^{\sqrt{2}}\int_0^1 6xy^2zdydz = 2x & \text{se } 0 \leq x \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

*Do livro *Probabilidade e Variáveis Aleatórias*, Marcos N. Magalhães

Independência

Seja $(X_1, \ldots, X_n)$ um vetor aleatório. As variáveis $X_1, \ldots, X_n$ são *independentes* se, para todos $(x_1, \ldots, x_n) \in \mathbb{R}^n$

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

F_{X_i} é a função de distribuição da variável aleatória X_i

As variáveis $X_1, \ldots, X_n$ de um vetor aleatório discreto são independentes se, para todos $(x_1, \ldots, x_n) \in \Omega_{X_1} \times \ldots \Omega_{X_n}$,

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i).$$

As variáveis $X_1, \ldots, X_n$ de um vetor aleatório contínuo são independentes se, para todos $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Exemplo: independência*

Considere o vetor aleatório (X_1, X_2) cuja distribuição é:

		X_1		
		1	2	3
	2	1/12	1/6	1/12
X_2	3	1/6	0	1/6
	4	0	1/3	0

$\mathbb{P}(X_1 = 2) = 1/2$ e $\mathbb{P}(X_2 = 4) = 1/3$

$\mathbb{P}(X_1 = 2, X_2 = 4) = 1/3 \neq 1/2 \cdot 1/3$

$\implies X_1$ e X_2 não são independentes

Considere agora o vetor aleatório (Y_1, Y_2) cuja distribuição é:

		Y_1		
		1	2	3
	2	1/12	1/6	1/12
Y_2	3	1/12	1/6	1/12
	4	1/12	1/6	1/12

São Y_1 e Y_2 independentes?

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Soma de variáveis aleatórias

Dadas n variáveis aleatórias $X_1, \ldots, X_n$ definidas no mesmo espaço de amostral, podemos definir uma nova variável aleatória como

$$S_n := \sum_{i=1}^{n} X_i.$$

Somas de variáveis aleatórias aparecem em muitos contextos:

- ao fazer a média de varias observações do mesmo experimento
- ao quantificar efeitos cumulativos

Uma questão importante é determinar a distribuição da variável soma S_n a partir da distribuição de cada X_i; em geral isso não é fácil

Exemplo: $X_1 \sim Bin(n_1, p_1)$ e $X_2 \sim Bin(n_2, p_2)$ com $p_1 \neq p_2$, $X_1 + X_2 \overset{?}{\sim}$

Soma de variáveis aleatórias

Apesar que a distribuição de soma de v.a. pode ser complicada de achar, a relação entre valor esperado e variância é mais simples.

Dadas n variáveis aleatórias $X_1, \ldots, X_n$ e constantes reais $a_1, \ldots, a_n$

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i)$$

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i<j} a_i a_j Cov(X_i, X_j)$$

Se as variáveis $X_1, \ldots, X_n$ são independentes ($\implies Cov(X_i, X_j) = 0$)

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

Exemplo: esperança e variância para soma*

Uma classe tem 300 estudantes e cada estudante tem probabilidade de 1/3 de obter a máxima nota A, independentemente dos outros. Seja X o número de estudantes que recebem nota A. Achar $\mathbb{E}(X)$ e $Var(X)$

Podemos representar $X = \sum_{i=1}^{300} X_i$ com $X_i \sim Ber(1/3)$

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{300} X_i\right) = \sum_{i=1}^{300} \mathbb{E}(X_i) = \frac{300}{3} = 100$$

$$Var(X) = Var\left(\sum_{i=1}^{300} X_i\right) = \sum_{i=1}^{300} Var(X_i) = 300\left(\frac{1}{3}\frac{2}{3}\right) = \frac{200}{3}$$

*Exemplo tirado do livro: *Introduction to probability*, D. Bertsekas & J. Tsitsiklis

Distribuição da soma de variáveis independentes

Vimos como calcular esperança e variância de soma de variáveis aleatórias. Em alguns caso é possível conhecer a distribuição toda (que é uma informação bem mais ampla)

Distribuições de soma de variáveis independentes

$$
\begin{aligned}
\mathrm{X}_i \sim Ber(p) \quad &\Longrightarrow \quad \sum_{i=1}^{n} X_i \sim Bin(n, p) \\
\mathrm{X}_i \sim Bin(m_i, p) \quad &\Longrightarrow \quad \sum_{i=1}^{n} X_i \sim Bin\left(\sum_{i=1}^{n} m_i, p\right) \\
\mathrm{X}_i \sim Poi(\lambda_i) \quad &\Longrightarrow \quad \sum_{i=1}^{n} X_i \sim Poi\left(\sum_{i=1}^{n} \lambda_i\right) \\
\mathrm{X}_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad &\Longrightarrow \quad \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)
\end{aligned}
$$

Exemplo:

Seja $X \sim \mathcal{N}(0, 1/2)$ e $Y \sim \mathcal{N}(0, 1/2)$, com X e Y independentes. Calcule $\mathbb{P}(X + Y \leq 0,5)$.

Sabemos que a soma de variáveis normais independentes é uma normal cujos parâmetros são a soma dos parâmetros das variáveis somadas.

Então $X + Y \sim \mathcal{N}(0, 1)$ e temos

$$\mathbb{P}(X + Y \leq 0,5) = \Phi(0,5) = 0,69$$

Lei dos Grandes Números

Sejam $X_1, X_2, \ldots$ v.a. independentes e identicamente distribuídas (i.i.d.) com média μ e variância σ^2 e $S_n := \sum_{i=1}^n X_i$. Sabemos que

$$\mathbb{E}(S_n) = n\mu \qquad\qquad Var(S_n) = n\sigma^2$$

Observação: ao aumentar de n, o valor médio da distribuição de S_n cresse assim como a variância (*distribuição fica mais e mais espalhada*)

Por esse motivo consideramos uma nova variável $M_n := \frac{S_n}{n}$. Temos

$$\mathbb{E}(M_n) = \mu \qquad\qquad Var(M_n) = \frac{\sigma^2}{n}.$$

Observação: o valor médio não depende de n e a variância diminui com n. Ao aumentar de n a distribuição de M_n se concentra ao redor de μ.

Lei dos Grandes Números: para todo $\epsilon > 0$, no limite $n \to \infty$

$$\mathbb{P}(|M_n - \mu| \geq \epsilon) \to 0$$

Exemplo: Lei dos grandes números*

Queremos estimar a taxa de aprovação do professor Fulano. Por isso perguntamos a n estudantes escolhidos aleatoriamente e denotamos por X_i cada resposta: $X_i = 1$ se o estudante aprova o professor e $X_i = 0$ caso contrário.

Modelamos $X_i \sim Ber(p)$ onde p é a verdadeira taxa de aprovação (*desconhecida!*)

Denotamos $M_n = \frac{X_1 + \ldots + X_n}{n}$ a média das respostas. Temos $\mathbb{E}(M_n) = p$ e $Var(M_n) = \frac{p(1-p)}{n}$

A lei dos grandes número nos garante que quando n cresse a média das respostas M_n fica perto da verdadeira taxa de aprovação p com alta probabilidade.

*Exemplo tirado do livro: *Introduction to probability*, D. Bertsekas & J. Tsitsiklis

Teorema Central do Limite (TCL)

Sejam $X_1, X_2, \ldots$ v.a. independentes e identicamente distribuídas (i.i.d.) com média μ e variância σ^2 e $S_n := \sum_{i=1}^n X_i$

Consideramos uma nova variável aleatória

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

Temos que $\mathbb{E}(Z_n) = 0$ e $Var(Z_n) = 1$. Isso significa que contrariamente a $\frac{S_n}{n}$, a v.a. Z_n tem uma distribuição que não se concentra ao redor do valor médio (a variância não vai para zero com n.)

O TCL permite conhecer a distribuição limite de Z_n (quando $n \to \infty$)

Teorema Central do Limite: A distribuição de Z_n se aproxima de uma normal padrão $Z \sim \mathcal{N}(0,1)$, ou seja para todos $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \to \mathbb{P}\left(Z \leq x\right) = \Phi(x)$$

Exemplo: Teorema Central do Limite *

$X_1, X_2, \dots$ i.i.d com média μ e variância σ^2, $S_n = \sum_{i=1}^n X_i$

$$\mathbb{P}(S_n \leq c) = \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{c - n\mu}{\sqrt{n}\sigma}\right) \underset{TCL}{\approx} \mathbb{P}\left(Z \leq \frac{c - n\mu}{\sqrt{n}\sigma}\right) = \Phi\left(\frac{c - n\mu}{\sqrt{n}\sigma}\right)$$

onde a aproximação melhora a medida que n cresce.

Exemplo: Nós carregamos um avião com 100 caixas cujos pesos são modelados por v.a. uniforme entre 5 e 50 kg. Qual é a probabilidade que o peso total excede 3000 kg?

X_i: peso da i-ésima caixa ; $\mu = 27,5$ e $\sigma^2 = (50-5)^2/12 = 168,75$

$$\mathbb{P}(S_{100} \leq 3000) = \mathbb{P}\left(\frac{S_{100} - 100(27,5)}{10\sqrt{168,75}} \leq \frac{300 - 100(27,5)}{10\sqrt{168,75}}\right) \approx \Phi(1,92) = 0,97$$

Então $\mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \leq 3000) = 0,03$

*Exemplo tirado do livro: *Introduction to probability*, D. Bertsekas & J. Tsitsiklis

Aproximação da Binomial pela Normal (De Moivre-Laplace)

Nos vimos que a soma de n v.a. de Bernoulli X_i de parâmetro p e independentes é uma Binomial de parâmetros n e p. Ou seja $S_n = X_1 + \ldots + X_n \sim Bin(n,p)$.

Podemos usar o TCL para estimar a probabilidade $\mathbb{P}(l \leq S_n \leq m)$

$$\mathbb{P}(l \leq S_n \leq m) = \mathbb{P}\left(\frac{l - np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{m - np}{\sqrt{np(1-p)}}\right)$$

$$\underset{TCL}{\approx} \Phi\left(\frac{m - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{l - np}{\sqrt{np(1-p)}}\right)$$

Quanto n tem que ser grande para que a aproximação seja "boa" depende do p. Por isso recomenda-se usar essa aproximação somente se

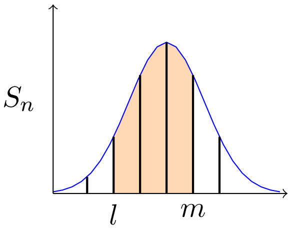
$$np(1-p) \geq 3.$$

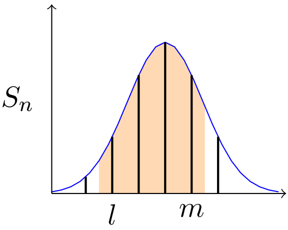
Correção de continuidade

Na aproximação $\mathbb{P}(l \leq S_n \leq m) \approx \Phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{l-np}{\sqrt{np(1-p)}}\right)$ se consideramos o caso $l = m$ a expressão a direita retorna o valor zero apesar da expressão a esquerda poder ser diferente de zero.

Para resolver esse problema de aproximação, a seguinte formula, chamada *correção de continuidade*, foi proposta por de Moivre-Laplace

$$\mathbb{P}(l \leq S_n \leq m) \approx \Phi\left(\frac{m + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{l - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$





Exemplo: correção de continuidade *

$S_n \sim Bin(n = 36, p = 0,5)$. Temos $\mathbb{P}(S_n \leq 21) = \sum_{k=0}^{21} \binom{36}{k} (0,5)^{36} = 0,8785$

A aproximação usando o TCL nos da

$$\mathbb{P}(S_n \leq 21) \approx \Phi\left(\frac{21 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{21 - 18}{3}\right) = 0,8413$$

Usando a correção de continuidade na aproximação acima temos

$$\mathbb{P}(S_n \leq 21) \approx \Phi\left(\frac{21,5 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{21,5 - 18}{3}\right) = 0,879$$

Também sabemos que $\mathbb{P}(S_n = 19) = 0,1251$

$$\mathbb{P}(S_n = 19) \approx \begin{cases} \Phi\left(\frac{19,5-18}{3}\right) - \Phi\left(\frac{18,5-18}{3}\right) = 0,124 & TCL\ +\ correção \\ 0 & TCL \end{cases}$$

*Exemplo tirado do livro: *Introduction to probability*, D. Bertsekas & J. Tsitsiklis

Estatística

Estatística

Definição: Estatística é um conjunto de conceitos e métodos para coletar, organizar, analisar e interpretar dados

Estatística descritiva: se preocupa com a organização e apresentação dos dados observados (*tabelas, gráficos, medidas descritivas como média e variância...*)

Inferência estatística: se preocupa de como dar informação sobre um universo (*população*) a partir de um conjunto de dados observados (*amostra*)

Frase resumidora da inferência:
"Não é preciso beber toda a garrafa para saber se o vinho é bom"

Estatística Descritiva

Conceitos básicos

- *População (N):* conjuntos de todos os elementos sob investigação
- *Amostra (n):* subconjuntos finitos da população
- *Parâmetro:* característica numérica de uma da população

Exemplo: *Pesquisa eleitoral no estado do Rio de Janeiro*

população	$\rightarrow$	todos os eleitores
amostra	$\rightarrow$	1000 eleitores entrevistados
parâmetro	$\rightarrow$	idade média da população

Tipologia das variáveis

Dada uma população ou amostra, podemos estar interessados em várias caracteristícas dos elementos constituintes; essas características são chamadas *variáveis*

Quantitativa: assume um valor numérico

- *discreta:* assume un número finito ou enumerável de valores

  Ex: *número de alunos na sala, ...*

- *contínua:* assume um número não-enumerável de valores

  Ex: *tempo, ...*

Qualitativa: é classificada em categorias (não é um valor numérico)

- *nominal:* assume categorias não-ordenadas

  Ex: *sexo, etnia, ...*

- *ordinal:* assume categorias ordenadas

  Ex: *classe social, grau de instrução, ...*

Distribuições de frequências: tabelas e gráficos

Os dados brutos podem não ser práticos para responder a questões de interesse. Por isso é útil resumi-los; as tabelas de distribucões de frequência é uma forma de fazer isso

- *Frequência absoluta:* número de vezes que cada valor é observado
- *Frequência relativa:* numero de vezes que cada valor é observado dividido pelo tamanho da amostra
- *Frequência acumulada:* soma das frequências absolutas dos valores inferiores ou iguais ao valor dado

Exemplo: frequências para variáveis quantitativas

Pesquisa realizada em 20 domicílios do RJ, com o objetivo de contabilizar o número de filhos por família

dados observados: 0, 1, 0, 1, 2, 1, 0, 0, 1, 0, 1, 4, 2, 1, 3, 1, 2, 1, 1, 1

n^o filhos	Frequência absoluta	Freq. relativa	Freq. acumulada
0	5	5/20	5
1	10	10/20	15
2	3	3/20	18
3	1	1/20	19
4	1	1/20	20

Exemplo: frequências para variáveis qualitativas

Observação: Como as variáveis são qualitativas, só podemos contar quantas observações tem um determinado atributo (categoria)

Numa pesquisa sobre os jogadores de vôlei de praia no Rio em 2013 foram entrevistados 3730 jogadores, dos quais 2032 eram do sexo feminino e 1698 eram do sexo masculino. Podemos resumir os dados na tabela seguinte

Categorias	Freq. absoluta	Freq. relativa
Mulheres	2032	54%
Homens	1698	46%

Gráficos para variáveis qualitativas/quantitativas

Utilizar recursos visuais na criação de gráficos permite analisar características dos dados de maneira rápida

Com base nas tabelas de frequências podem ser construidos gráficos da distribuição de frequência

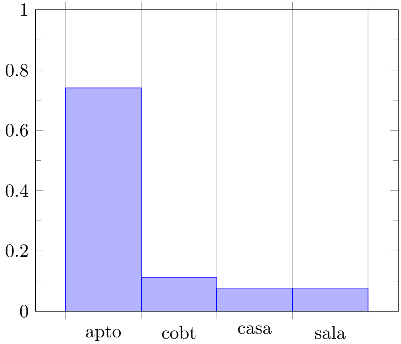
Gráfico de barra: as categorias/valores são representadas por retângulos dispostos ao longo de um eixo e as frequências correspondentes a cada categoria/valor são as alturas dos retângulos

Gráfico de setores (de pizza): as categorias são representadas como setores circulares cujo tamanho é proportional a frequência de cada categoria

<u>Observação:</u> O gráfico de barra, no qual as categorias estão naturalmente ordenadas, é normalmente mais apropriado para variáveis quantitativas, e o gráfico de setores para variáveis qualitativas

Exemplo: gráfico de barra para variáveis qualitativas *

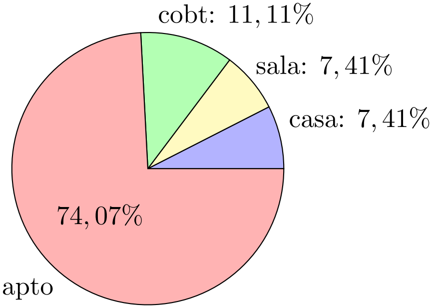
Categorias	Frequência Absoluta	Frequência relativa
Apartamento	20	74,07%
Cobertura	3	11,11%
Casa	2	7,41%
Sala	2	7,41%



*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Exemplo: gráfico de setores para variaveis qualitativas

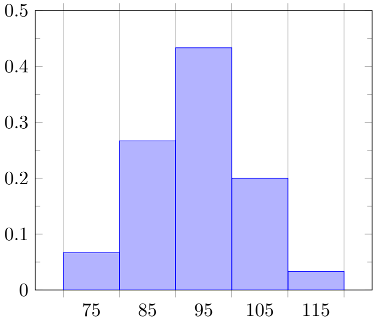
Categorias	Frequência Absoluta	Frequência relativa
Apartamento	20	74,07%
Cobertura	3	11,11%
Casa	2	7,41%
Sala	2	7,41%



Exemplo: gráfico de barra para variáveis quantitativas*

Uma companhia produz cabos náuticos de um determinado tipo, que podem ser usados em barcos a vela. A medição da carga de ruptura para 30 espécimes de cabos desse tipo é apresentado na tabela abaixo

Carga (Kg)	F. absoluta	F. relativa
70 – 80	2	6,67%
80 – 90	8	26,67%
90 – 100	13	43,33%
100 – 110	6	20 %
110 – 120	1	3,33 %



*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Outros graficos para variáveis quantitativas

O *histograma* é parecido com o gráfico de barras. Em um eixo estão representado os valores da variável, e no outro as frequências desses valores. Pode ser vertical (se a frequência está no eixo vertical) ou horizontal (se a frequência está no eixo horizontal)

Observação: Essa representação é boa para visualizar facilmente onde se situam as maiores incidências da uma variável

O *diagrama ramo-folha* é uma gráfico onde os valores assumidos pela variável são divididos em dezenas (ramos) e as unidades são tratadas como folhas e ordenadas crescentemente em cada ramo

Exemplo: diagrama ramo-folha*

Consideramos novamente o exemplo dos cabos náuticos e dessa vez a medição da carga (em Kg) de ruptura de 30 cabos é explicita

83	96	73	102	93	94	99	85	91	118
93	103	87	95	102	84	100	95	90	81
102	98	94	89	91	78	85	83	105	96

O diagrama ramo-folha correspondente a esses dados é

7	38
8	13345579
9	011334455668 9
10	022235
11	8

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Medidas para dados amostrais quantitativos

Dada uma coleção de valores de uma variável quantitativa é útil definir algumas formas de resumir esses dados para poder comparar mais facilmente com outros. Uma maneira de fazer isso é atravez das *medidas de centralidades*

- *média aritmética:* dados os valores $x_1, \ldots, x_n$ definimos

$$\overline{x} := \frac{\sum_{i=1}^{n} x_i}{n}$$

- *mediana:* dados os valores $x_1, \ldots, x_n$, sejam $x_{(1)}, x_{(2)} \ldots, x_{(n)}$ os mesmos valores ordenados. Definimos a mediana $Q2$ como

$$Q2 := \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ é impar (valor na posição central)} \\ \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{se } n \text{ é par (médias dos valor nas posições centrais)} \end{cases}$$

- *moda:* é aquele valor que ocorre com mais frequência

Exemplo: medidas de centralidades

Consideramos mais uma vez o exemplo dos cabos náuticos e dessa vez a medição da carga (em Kg) de ruptura de 30 cabos é explicita

83	96	73	102	93	94	99	85	91	118
93	103	87	95	102	84	100	95	90	81
102	98	94	89	91	78	85	83	105	96

Temos $n = 30$ e $\sum_{i=1}^{30} x_i = 83 + 96 + \ldots + 105 + 96 = 2785$, então

$$\overline{x} = \frac{2785}{30} = 92,83$$

Sendo $n = 30$ par, a mediana é a média aritmética das observações $x_{(15)}$ e $x_{(16)}$ que são 93 e 94, respectivamente. Assim,

$$Q2 = \frac{93 + 94}{2} = 93,5$$

Observação: $\overline{x}$ e $Q2$ são relativamente próximos entre si. Isso se deve ao fato de que os dados distribuem-se de forma aproximadamente simétrica em torno do valor central, conforme é mostrado pelo diagrama ramo-folha

Medidas de dispersão

As medidas de centralidades são insuficientes para caracterizar um conjunto de dados (e.g., conjuntos diferente de dados podem ter a mesma média!)

Para ter uma caracterização melhor outras medidas, chamada *medidas de dispersão* são introduzidas. Essas são indicadores do grau de espalhamento dos valores em torno da média

Dados os valores observados $x_1, \ldots, x_n$ definimos

- *variância amostral:*

$$s^2 := \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

- *desvio padrão amostral:* $s := \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
- *coeficiente de variação amostral:* $cv := \frac{s}{\bar{x}}$

Exemplo: medidas de dispersão*

Continuando com os dados de carga de ruptura

83	96	73	102	93	94	99	85	91	118
93	103	87	95	102	84	100	95	90	81
102	98	94	89	91	78	85	83	105	96

$n = 30$, $\sum_{i=1}^{30} x_i = 2785$, $\sum_{i=1}^{30} x_i^2 = 261037$, $\overline{x} = 92,83$

Variância amostral: $s^2 = \frac{\sum_i x_i^2 - n\overline{x}^2}{n-1} = 86,08$

Desvio padrão amostral: $s = \sqrt{86,08} = 9,3$

Coeficiente de variação amostral: $cv = \frac{s}{\overline{x}} = \frac{9,3}{92,83} = 0,10 = 10\%$

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Quartis

Outras medidas úteis relacionadas a um conjunto de dados são os *quartis*. Eles são valores que dividem os dados em quatro grupos, cada grupo contendo 1/4 do tamanho total do conjunto

Dado os valores $x_1, \ldots, x_n$, sejam $x_{(1)}, x_{(2)} \ldots, x_{(n)}$ os mesmos valores ordenados em ordem crescente. Definimos

- *primeiro quartil Q1:* o valor que tem $0,25$ de dados abaixo dele
- *segundo quartil Q2:* o valor que tem $0,50$ de dados abaixo dele
- *terceiro quartil Q3:* o valor que tem $0,75$ de dados abaixo dele
- *distância interquartil DIQ:* é definida como $DIQ = Q3 - Q1$

Como calcular os quartis

Para o cálculo dos quartis é necessário determinar primeiro a posição que eles ocupam quando os dados são dispostos em ordem crescente

Mediana $Q2$: sua posição é intermediária entre a posição 1 e n

- se n for impar, $Q2$ é o valor na posição $\frac{n+1}{2}$
- se n for par, não tem nenhum valor na posição $\frac{n+1}{2}$ (não é um inteiro). Então $Q2$ é calculado como a média dos valores nas posições $\frac{n}{2}$ e $\frac{n}{2}+1$

Primeiro quartil $Q1$: sua posição é intermediária entre a posição 1 e a posição da mediana $Q2$

- se $\frac{1+\frac{n+1}{2}}{2} = \frac{n+3}{4}$ for inteiro, $Q1$ é o valor nessa posição
- se $\frac{n+3}{4}$ não for inteiro, $Q1$ tem que ser calculado por interpolação entre os valores nas posições vizinhas a $\frac{n+3}{4}$

Exemplo: (*interpolação*) seja $n = 8 \implies (n+3)/4 = 2,75$. Sejam $x_{(2)}$ e $x_{(3)}$ os valores nas posições vizinhas. Temos

$$Q1 = 0,25\, x_{(2)} + 0,75\, x_{(3)}$$

Como calcular os quartis

Terceiro quartil Q3: a sua posição é intermediária entre a posição n e a posição da mediana $Q2$

- se $\frac{n+\frac{n+1}{2}}{2} = \frac{3n+1}{4}$ for inteiro, $Q3$ é o valor nessa posição
- se $\frac{3n+1}{4}$ não for inteiro, $Q3$ tem que ser calculado por interpolação entre os valores cujas posições são vizinhas a $\frac{3n+1}{4}$

Exemplo: Considere os seguinte dados: $8, 2, 4, 10, 1, 7$

Mediana: $(n+1)/2 = 3,5 \implies Q2 = 0.5x_{(3)} + 0.5x_{(4)} = 5,5$

$Q1$: temos $(n+3)/4 = 2,25 \implies Q1 = 0.75x_{(2)} + 0.25x_{(3)} = 2,5$

$Q3$: temos $(3n+1)/4 = 4,75 \implies Q3 = 0.25x_{(4)} + 0.75x_{(5)} = 9,5$

Discrepâncias em variáveis quantitativas

Num conjunto de dados podem existir valores que foram coletados em condições "anormais"

Esses valores podem afetar o resultado da analise estatística, sobretudo quando esses valores são muito maior ou menor dos demais

Esse valores são chamados *discrepância ou outliers*

Dada a possível presença de valores discrepantes nas observações é importante ter uma maneira de poder detectá-los

Valores discrepantes

Existem varias maneiras de identificar valores discrepantes:

1) considerando os dados $x_1, \ldots, x_n$, chamamos discrepantes todos os valores que caem fora do intervalo

$$[\bar{x} - 3s; \bar{x} + 3s]$$

Observação: Esse método não é muito bom porque a média $\bar{x}$ e o desvio padrão s são bastante sensíveis a valores grandes da amostra

2) Um método menos sensível a valores grandes consiste em usar os quartis. Definimos

$$Cerca\ Inferior := Q1 - \frac{3}{2}DIQ \qquad Cerca\ Superior := Q3 + \frac{3}{2}DIQ$$

Chamamos outliers todos os valores fora do intervalo $[CI; CS]$

Observação: ambos os métodos funcionam bem quando a distribuição de valores é simétrica com respeito à medida de centralidade usada (média ou mediana)

Exemplo: valores discrepantes

Consideremos, mais uma vez, os dados de carga de ruptura

83	96	73	102	93	94	99	85	91	118	
93	103	87	95	102	84	100	95	90	81	$\overline{x} = 92,83$
102	98	94	89	91	78	85	83	105	96	s=9,3

Usando o critério 1) para identificação de valores discrepantes, encontramos:

$$\overline{x} - 3s = 64,9 \qquad \overline{x} + 3s = 120,7 \implies [64,9; 120,7]$$

Todos os 30 valores estão dentro do intervalo, não há valores discrepantes

Usando o critério 2) de identificação de valores discrepantes, temos:

$$Q1 = 85,5 \quad Q3 = 98,75 \quad DIQ = 13,25 \quad CI = Q1 - \frac{3}{2}DIQ = 45,75$$

$$CS = Q3 + \frac{3}{2}DIQ = 138,5 \implies [45,75; 138,5]$$

Todos os 30 valores estão entre as duas cercas, não há valores discrepantes

Observação: O que mudaria se o valor 105 fosse trocado para 140?

BoxPlot para variáveis quantitativas

O gráfico BoxPlot é muito usado para resumir varias informações de um conjuntos de dados

- Todos os possível valores da variável são representado num eixo vertical
- O BoxPlot consiste num retângulo cuja base inferior corresponde ao valor de $Q1$ e cuja base superior corresponde ao valor de $Q3$. A posição da mediana é indicada por um traço horizontal nesse retângulo
- Depois são traçados dois segmentos de reta verticais: o primeiro segmento vai do ponto médio da base inferior até o valor menor na amostra não discrepante. O outro segmento vai do ponto médio da base superior até o maior valor não discrepante
- Finalmente, valores discrepantes são reportados como pontos

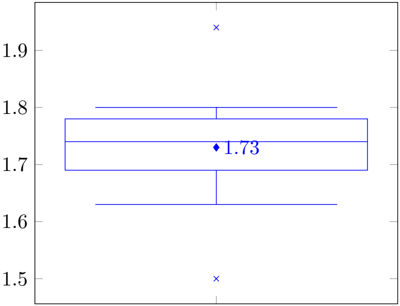
Exemplo: BoxPlot

Os dados abaixo representam as alturas (em metros) para uma amostra de tamanho 10 de alunos da UFRJ

aluno	1	2	3	4	5	6	7	8	9	10
altura	1,68	1,80	1,63	1,76	1,75	1,50	1,79	1,73	1,94	1,72

$\bar{x} = 1,73 \quad s = 0,115 \quad x_{min} = 1,50 \quad x_{max} = 1,94 \quad Q1 = 1,69$

$Q2 = 1,74 \quad Q3 = 1,78 \quad DIQ = 0,135 \quad CI = 1,56 \quad CS = 1,92$



Análise bidimensional: relação entre duas variáveis

Muitas vezes podemos estar interessados em analisar a relação entre duas variáveis (e.g. peso e altura dos alunos)

Consideramos duas variáveis quantitativas, X e Y. Cada dado a partir de uma amostra de tamanho n será representado por um par ordenado (x_i, y_i), onde x_i é a i-ésima observação da variável X e y_i de Y

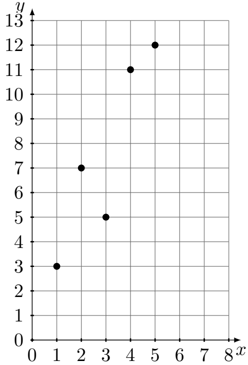
Uma primeira tentativa é tentar descobrir uma possível relação fazendo um gráfico onde no eixo horizontal temos os valores de x e no eixo vertical os valores de y. Esse gráfico é chamado *diagrama de dispersão*

O diagrama de dispersão permite visualizar o tipo de relação entre as variáveis

Diagrama de dispersão

Exemplo: Considere os seguintes dados para *duas variáveis* X e Y

x_i	1	2	3	4	5
y_i	3	7	5	11	12



- Que conclusões preliminares podem ser tiradas do diagrama?
- O que o diagrama sugere sobre a relação entre as duas variáveis? (*relação linear positiva*)

Covariância e correlação entre variáveis quantitativas

Dadas duas variáveis X e Y, seria bom ter uma medida para quantificar o grau de relação entre elas

Por este motivo introduzimos a *covariância amostral* definida como

$$s_{xy} := \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n-1}$$

- A covariância amostral é positiva, se o diagrama de disperção mostra uma tendência crescente
- A covariância amostral é negativa, se o diagrama de disperção mostra uma tendência decrescente
- A covariância amostral é zero ou próxima a zero, se o diagrama de dispersão não mostra uma tendência

Coeficiente de correlação

Além do sinal que serve como indicador da tendência, crescente ou decrescente, muito pouca informação pode ser extraída da covariância. Isso porque o valor de s_{xy} depende fortemente das unidades de X e Y

Por isso introduzimos uma outra medida chamada *coeficiente de correlação* entre X e Y; essa é definida como

$$r_{xy} := \frac{s_{xy}}{s_x s_y}$$

onde s_x (resp. s_y) é o desvio padrão amostral de X (resp. Y)

- O coeficiente de correlação é adimensional
- r_{xy} é uma versão normalizada da covariância e temos $-1 \leq r_{xy} \leq 1$
- O coeficiente de correlação mede um tipo especifico de dependência entre as variáveis, ou seja dependência linear

Observação: se a dependência entre as variáveis não for linear é possível ter um coeficiente baixo apesar de uma dependência forte entre as variáveis

Exemplo: covariância e correlação

Um levantamento obtido, junto aos funcionários de um escritório, busca relacionar as variáveis: tempo de carreira (X) e número de diferentes empregos nos últimos 5 anos (Y)

X	8	9	10	11	12
Y	3	2	2	2	1

Pra calcular a covariância entre X e Y precisamos dos seguintes valores: $\bar{x} = 10$, $\bar{y} = 2$, $\sum_i x_i y_i = 96$. Temos

$$s_{xy} = \frac{96 - 5(10)(2)}{4} = -1$$

Para calcular o coeficiente de correlação precisamos dos valores: $s_x^2 = 2,5$, $s_y^2 = 0,5$. Temos

$$r_{xy} = \frac{-1}{\sqrt{2,5}\sqrt{0,5}} = -0,89$$

Reta de regressão

Quando o diagrama de dispersão sugere uma relação linear entre duas variáveis podemos calcular a reta que "melhor" representa esta relação. Esse procedimento é chamado *regressão linear*

Mensagem principal: Na regressão linear, procuramos estimar a reta que "melhor" representa a relação entre duas variáveis de interesse (X e Y) usando uma coleção de dados (x_i, y_i) com $i = 1, \ldots, n$. Isso significa que estamos interessado em calcular os coeficientes a e b tais que

$$Y \approx a + bX,$$

e a aproximação seja a melhor possível.

Chamamos X de variável independente e Y de variável dependente

Os parâmetros a, b (ou seja a relação linear) são desconhecido e usando os dados (x_i, y_i) nós queremos estimá-los

Regressão linear

Usamos a relação linear (desconhecida) $Y = a + bX$ para calcular

$$\hat{y}_i = a + bx_i$$

Normalmente $\hat{y}_i$ será diferente de y_i, e o erro, ou seja, $y_i - \hat{y}_i$ é chamado i-ésimo *resíduo*

Uma possível escolha para os parâmetros a, b consiste em escolhê-los de forma tal que a soma dos resíduos quadráticos

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (a + bx_i))^2$$

seja miníma. Esse método é chamado, *método dos mínimos quadrados*

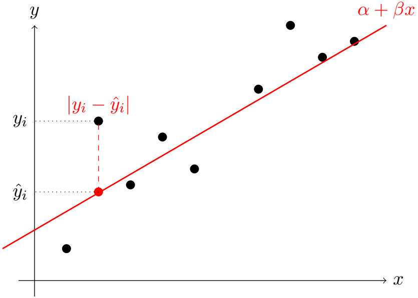
Método dos mínimos quadrados

Minimizando a expressão $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ com respeito aos parâmetros a e b obtemos os seguintes valores para os parâmetros:

$$\beta = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{s_{xy}}{s_x^2}$$

$$\alpha = \overline{y} - \beta\overline{x}$$

Isso significa que escolhendo $a = \alpha$ e $b = \beta$ obtemos a relação linear que minimiza a soma dos resíduos quadrados



Exemplo: reta de regressão

Considere as seguintes observações para duas variáveis X e Y

x_i	2	3	5	1	8
y_i	25	25	20	30	16

Para calcular α e β precisamos das seguintes quantidades:
$\bar{x} = 3,80$; $\bar{y} = 23,2$; $\sum_i x_i^2 = 103$; $\sum_i x_i y_i = 383$ e $n = 5$

$$\beta = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{383 - 5(3,80 \cdot 23,2)}{103 - 5 \cdot (3,80)^2} = -1,88$$

$$\alpha = \bar{y} - \beta\bar{x} = 23,2 - (-1,88)(3,80) = 30,35$$

Então a reta de regressão linear é $Y = -1,88X + 30,35$

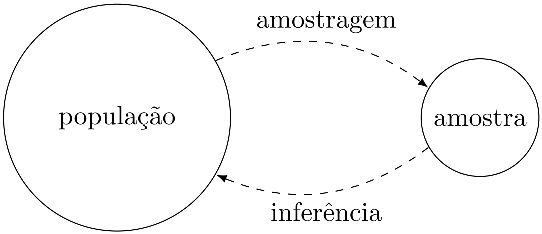
Inferência Estatística

Inferência estatística

A inferência estatística tem como objetivo inferir "alguma coisa" sobre uma população a partir da amostra (sub-conjunto da população)

Ex: *Idade média dos brasileiros, sabendo a idade dos cariocas*

Observação: Caso fosse possível obter a idade de todos brasileiros poderíamos saber a distribuição exata da idade e daí calcular a média. A situação onde é possível observar a população toda é rara (*custos elevados, tempo para a coleta muito longo, etc...*)



Perguntas naturais sobre inferência estatística

Sempre que estudamos uma população a partir de uma amostra existe a possibilidade de cometermos algum tipo de erro

Exemplo: população dos pesos (kg) de 6 pessoas: 64, 75, 68, 80, 72, 70. A média *populacional* é $\mu = 71,5$. Consideramos as duas amostras

$$a)\quad 64, 75, 72, 80 \qquad\qquad b)\quad 75, 80, 72, 70$$

As médias *amostrais* são $72,75$ e $74,25$, claramente diferentes de μ

Algumas perguntas naturais são:

1) Como é que escolhemos uma amostra?
2) É possível medir o erro da estimativa obtida a partir da amostra?

Amostragem

Existem várias formas de selecionar uma amostra de uma população:

Amostragem aleatória simples: sucessivos sorteios aleatórios até completar-se o tamanho da amostra n (*mais utilizado*)

Amostragem sistemática: útil quando a população é ordenada segundo algum critério; aumenta a representatividade da amostra

Exemplo: população $= \{1, 2, \ldots, N\}$ e $n =$ tamanho da amostra. Define-se $b = N/n$ e escolhe-se aleatoriamente um elemento x da população entre o primeiro e o b-ésimo. A amostra é dada por $\{x, x + b, x + 2b, \ldots, x + (n - 1)b\}$

Amostragem estratificada: útil no caso de população heterogênea constituída por subpopulações mais homogêneas. Seleciona-se uma amostra aleatória de cada uma das subpopulações

Exemplo: população = trabalhadores da universidade
população = {professores, secretários, bolsistas, terceirizados}

Como medir o erro de estimação?

Para responder essa pergunta, a Estatística serve-se da Teoria de Probabilidades

A Teoria de Probabilidades é fundamental na Inferência Estatística, pois permite definir um modelo matemático para relacionar as amostras com as populações

estatística descritiva	*população = conjunto de elementos*
inferência estatística	*população = distribuição de uma variável aleatória X*

A distribuição de X é desconhecida e a Inferência Estatística quer utilizar amostras de X para estimar essa distribuição

Conhecimento parcial sobre a população

Para simplificar a análise, nós assumimos que a distribuição da população pertence a uma família de distribuições indexada por um ou mais parâmetros, porém o verdadeiro valor do parâmetro é desconhecido (*conhecimento parcial sobre a população*)

Usamos a notação $f(x|\theta)$, $\theta \in \Theta$ para a família de distribuições, onde θ pode também ser um vetor de mais de um parâmetro

Exemplo: Bernoulli com parâmetro $\theta = p$
Normal com média e variança $(\theta_1 = \mu, \theta_2 = \sigma^2)$

Nesse arcabouço, o objetivo da Inferência Estatística é estimar o verdadeiro parâmetro usando a informação contida na amostra

Definições

População: conjunto de todos os possível resultados de um observação; é modelada por uma variável aleatória X (discreta ou contínua) com uma distribuição $f(x|\theta)$, onde θ é desconhecido

Amostra aleatória (tamanho n): coleção de n variáveis aleatórias $X_1, X_2, \ldots, X_n$ independentes e com a mesma distribuição de X (população)

Intuitivamente, X_i representa a i-ésima observação. Os valores observados correspondentes a amostra $X_1, X_2, \ldots, X_n$ são denotados $x_1, x_2, \ldots, x_n$

Observação: A amostra é um vetor aleatório com distribuição conjunta

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Essa fórmula é útil para calcular probabilidades relacionadas as amostras

Exemplo: população e amostra *

Estamos interessados no tempo de vida de ar condicionados splits

A população consiste de todos os possíveis tempos de vida, que modelamos como uma v.a. exponencial com parâmetro λ desconhecido

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n dessa população. A densidade conjunta da amostra é

$$f(x_1, \ldots, x_n | \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

Usando essa distribuição é possível calcular a probabilidade que o tempo de vida de cada split na amostra seja maior de dois anos:

$$\mathbb{P}(X_1 > 2, \ldots, X_n > 2) = \prod_{i=1}^{n} \mathbb{P}(X_i > 2) = \prod_{i=1}^{n} \int_2^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda 2n}$$

Se o tempo de vida médio $1/\lambda$ é longo essa probabilidade é ≈ 1

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Conceitos básicos da inferência estatística

Parâmetro: medida numérica (em geral desconhecida) que descreve uma caraterística de interesse da população (da distribuição)

É possível resumir a informação contida na amostra em algumas caraterísticas chave da amostra. Isso é o que se chama *estatística*

Estatística: Uma função $T(\boldsymbol{X}) = T(X_1, \ldots, X_n)$ de uma amostra aleatória $X_1, \ldots, X_n$

Estimador: Uma estatística destinada a estimar um parâmetro de uma população

Como os estimadores, e mais em geral as estatísticas, são funções de variáveis aleatórias, elas mesmas são variáveis aleatórias. Podemos então falar de distribuição de um estimador (*distribuição amostral*)

Exemplos de estatísticas

Seja $X_1, \ldots X_n$ uma amostra de tamanho n tirada de uma certa população. Algumas estatísticas importante são:

$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$	Média da amostra
$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$	Variância da amostra
$X_{(1)} = \min(X_1, \ldots, X_n)$	menor valor da amostra
$X_{(n)} = \max(X_1, \ldots, X_n)$	maior valor da amostra

Função de verossimilhança

Uma estatística importante é a *função de verossimilhança*

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n tirada de uma população X com distribuição $f(x|\theta)$. Dada uma observação da amostra $\boldsymbol{x} = (x_1, \ldots, x_n)$ a função de θ definida por

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

é chamada *função de verossimilhança*

Observação: Quando consideramos $f(\boldsymbol{x}|\theta)$ assumimos que θ seja fixo (desconhecido), e $\boldsymbol{x}$ a variável. Quando consideramos $L(\theta|\boldsymbol{x})$ assumimos que $\boldsymbol{x}$ seja fixo, e θ a variável

Exemplo: função de verossimilhança

Seja $X_1, \ldots, X_n$ uma amostra de tamanho $n = 5$ tirada de uma população $X \sim Ber(p)$. Dada uma observação da amostra $\boldsymbol{x} = (x_1, \ldots, x_5)$, onde cada $x_i = 1$ or $x_i = 0$, a função de verossimilhança é

$$L(p|\boldsymbol{x}) = \prod_{i=1}^{5} p^{x_i}(1-p)^{1-x_i}$$

Procedimentos básicos de inferência estatística

Estimação pontual (por ponto): se preocupa com achar "bons" estimadores para o parâmetro de uma população

Estimação intervalar (por intervalo): se preocupa com achar intervalos que contêm o parâmetro da população

Teste de hipótese: se preocupa de como escolher entre duas hipóteses complementares a respeito do parâmetro da população

Em todos os três procedimentos os dados extraídos de uma amostra são usado para inferir algo sobre a população

Estimação pontual

Estimadores pontuais

Estimadores pontuais visam estimar o(s) parâmetro(s) de uma população a partir de uma amostra

Lembrete: dada uma população com distribuição $f(x|\theta)$ o conhecimento do parâmetro θ leva conhecimento sobre a população

Lembrete: um *estimador pontual* é uma estatística, ou seja uma função $T(X_1, \ldots, X_n)$ da amostra $X_1, \ldots, X_n$

1) Como achamos estimadores dos parâmetros?
2) Como podemos avaliar a qualidade da estimativa?

Nos veremos dois métodos para achar estimadores

- método dos momentos
- método de máxima verossimilhança

Método dos Momentos

Esse método foi introduzido para Karl Pearson no final do século XIX

$X_1, \ldots, X_n$: amostra de uma população X com dist. $f(x|\theta_1, \theta_2, \ldots, \theta_k)$

$\mu_j(\theta_1, \theta_2, \ldots, \theta_k) = \mathbb{E}(X^j)$: j-ésimo momento da população

Estimadores dos parâmetros $\theta_1, \theta_2, \ldots, \theta_k$ são achados resolvendo um sistema de equações obtido igualando os primeiros k momentos amostrais com os momentos μ_j da população

$$(*) \begin{cases} \frac{1}{n}\sum_{i=1}^n X_i = \mu_1(\theta_1, \theta_2, \ldots, \theta_k) \\ \frac{1}{n}\sum_{i=1}^n X_i^2 = \mu_2(\theta_1, \theta_2, \ldots, \theta_k) \\ \qquad \vdots \qquad\qquad \vdots \\ \frac{1}{n}\sum_{i=1}^n X_i^k = \mu_k(\theta_1, \theta_2, \ldots, \theta_k) \end{cases}$$

Seja $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ a solução do sistema (*). A função $\hat{\theta}_j(X_1, \ldots, X_n)$ nos dá o estimador do parâmetro θ_j da população

Exemplo: método dos momentos*

Seja X uma população com distribuição $\mathcal{N}(\mu, \sigma^2)$ $\quad (\theta_1 = \mu, \theta_2 = \sigma^2)$

Seja $X_1, \dots, X_n$ uma amostra de tamanho n de X

Os primeiros dois momentos da população são $\mu_1 = \mu$ e $\mu_2 = \sigma^2 + \mu^2$

O sistema a resolver é

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mu \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2$$

Resolvendo para μ e σ^2 nós temos

$$\hat{\theta}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Exemplo: método dos momentos*

Seja X uma população com distribuição Binomial(k,p) $\quad(\theta_1 = k, \theta_2 = p)$

Seja $X_1,\ldots,X_n$ uma amostra de tamanho n

Os primeiro e segundo momentos da população são $\mu_1(k,p) = kp$ e $\mu_2(k,p) = kp(1-p) + k^2p^2$

O sistema a resolver é

$$\frac{1}{n}\sum_{i=1}^{n} X_i = kp \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2$$

Resolvendo para k e p nós temos

$$\hat{p} = \frac{n\overline{X}(\overline{X}+1) - \sum_i X_i^2}{n\overline{X}} \qquad \hat{k} = \frac{n\overline{X}^2}{n\overline{X}(\overline{X}+1) - \sum_i X_i^2}$$

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Estimadores de máxima verossimilhança (EMV)

O método EMV é muito usado para obter estimadores

Lembrete: se $X_1, \ldots, X_n$ é uma amostra de uma população com distribuição $f(x|\theta_1, \theta_2, \ldots, \theta_k)$, a função de verossimilhança é

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = L(\theta_1, \theta_2, \ldots, \theta_k|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\theta_1, \theta_2, \ldots, \theta_k)$$

Ideia do EMV: *um bom estimador para um parâmetro é o valor que maximiza a probabilidade da amostra observada*

Para cada realização $\boldsymbol{x} = (x_1, \ldots, x_n)$ da amostra, seja $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ o valor do parâmetro tal que $L(\boldsymbol{\theta}|\boldsymbol{x})$ atinge o valor máximo como função de $\boldsymbol{\theta}$. O *estimador de máxima verossimilhança* para $\boldsymbol{\theta}$ é $\hat{\boldsymbol{\theta}}(X_1, \ldots, X_n)$

Observação: Para achar o EMV é preciso resolver um problema de otimização

Exemplo: máxima verossimilhança*

Seja $X_1, \ldots, X_n$ uma amostra de uma população $X \sim Ber(\theta)$
A função de verossimilhança é

$$L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^y(1-\theta)^{n-y} \qquad y = \sum_{i=1}^{n} x_i$$

Usando o fato que a função log é estritamente crescente, sabemos que o máximo de $L(\theta|\boldsymbol{x})$ para $\theta \in (0,1)$ é o máximo de $\log L(\theta|\boldsymbol{x})$. Então

$$\frac{d}{d\theta}\log(L(\theta|\boldsymbol{x})) = \frac{y}{\theta} - \frac{n-y}{1-\theta} = 0 \implies \hat{\theta}(\boldsymbol{x}) = \frac{y}{n} = \overline{x}$$

Para verificar que é um ponto de máximo para $\theta \in (0,1)$, note que

$$\frac{d^2}{d\theta^2}\log(L(\theta|\boldsymbol{x})) = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2} < 0, \ \forall\theta$$

Então $\overline{x}$ maximiza L em $(0,1)$.

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Exemplo: máxima verossimilhança*

Seja $X_1, \ldots, X_n$ uma amostra de uma população $X \sim \mathcal{N}(\theta, 1)$

Observação: a variância $\sigma^2 = 1$ é conhecida! A função de verossimilhança é

$$L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} (2\pi)^{-1/2} e^{-\frac{(x_i-\theta)^2}{2}} = (2\pi)^{-n/2} e^{-\frac{\sum_i (x_i-\theta)^2}{2}}$$

Temos que $\frac{d}{d\theta} L(\theta|\boldsymbol{x}) = (2\pi)^{-n/2} e^{-\frac{\sum_i (x_i-\theta)^2}{2}} \sum_i (x_i - \theta)$. Então, o ponto estationario é solução da seguinte equação

$$\sum_{i=1}^{n} (x_i - \theta) = 0 \implies \hat{\theta}(x_1, \ldots, x_n) = \overline{x}$$

É possível mostrar que $\frac{d^2}{d\theta^2} L(\theta|\boldsymbol{x})|_{\theta=\overline{x}} < 0$. Então, se σ^2 é conhecida, o estimador EMV da média de uma população normal é $\hat{\theta} = \overline{X}$.

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Exemplo: máxima verossimilhança*

Seja $X_1, \dots, X_n$ uma amostra de uma população $X \sim \mathcal{N}(\theta, \sigma^2)$

Observação: assumimos que a variância σ^2 é desconhecida! A função de verossimilhança é

$$L(\theta, \sigma^2|\boldsymbol{x}) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{-\frac{\sum_i (x_i-\theta)^2}{2\sigma^2}}$$

Primeiro, procuramos os pontos estacionários

$$\frac{\partial}{\partial\theta} L(\theta, \sigma^2|\boldsymbol{x}) = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n} x_i = n\theta$$

$$\frac{\partial}{\partial\sigma^2} L(\theta, \sigma^2|\boldsymbol{x}) = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n} (x_i - \theta)^2 = n\sigma^2$$

As soluções das equações acima são: $\hat{\theta} = \overline{x}$, e $\hat{\sigma}^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 / n$. É possível mostrar que $\hat{\theta}, \hat{\sigma}^2$ maximizam L, então os estimadores EMV para os parâmetros de uma normal são $\hat{\theta} = \overline{X}$ e $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2$.

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

O estimador média amostral

A *média amostral* de uma amostra aleatória $X_1, \dots, X_n$ é definida como

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Sendo uma função de v.a. ela é também uma v.a., é podemos falar da sua distribuição, esperança e variância

- Se $X_1, \dots, X_n$ é uma amostra aleatória de uma população X com distribuição $\mathcal{N}(\mu, \sigma^2)$, é possível provar que

$$S_n = \sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2) \implies \overline{X} = \frac{S_n}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- Se $X_1, \dots, X_n$ é uma amostra de uma população qualquer, temos

$$\text{TCL} \implies \overline{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \text{ para } n \text{ grande}$$

Em geral: $n \geq 30$ é suficiente para usar tal aproximação

Exemplo: média amostral[*]

As especificações de uma característica de qualidade estabelecem um limite máximo de $150,6$ unidades. A medição dessa característica comporta-se como uma v.a. $X \sim \mathcal{N}(150; 2,1)$. Determine a probabilidade de que a média amostral de uma amostra aleatória de tamanho $n = 49$ ultrapasse a especificação limite de 150,6.

Solução: Temos que a média amostral $\overline{X} \sim \mathcal{N}(150; \frac{2,1}{49} = 0,04)$. Então

$$\mathbb{P}\left(\overline{X} > 150,6\right) = \mathbb{P}\left(\frac{\overline{X} - 150}{0,2} > \frac{150,6 - 150}{0,2}\right) = 1 - \Phi(3) = 0,0013$$

<u>Observação:</u> nesse caso não precisamos do TCL!

[*] Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

O estimador proporção amostral

Consideramos uma população cujos elementos podem ser classificados em dois tipos: *sucesso* e *fracasso*. Nesse caso, modelamos a população como uma variável aleatória $X \sim \text{Ber}(p)$

Dada uma amostra aleatória $X_1, \ldots, X_n$, um estimador do parâmetro p é a *proporção amostral* definida por

$$\overline{p} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

O Teorema central do limite, nos garante que se n é "grande", então vale que

$$\frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx \mathcal{N}(0,1)$$

Em geral: essa aproximação vale quando $np(1-p) \geq 3$

Exemplo: proporção amostral*

Uma empresa fabrica diodos usados em placas de circuito impresso. Digamos que tenha sido coletada uma amostra aleatória da linha de produção com $n = 50$ diodos e que exatamente um número Y deles esteja fora das especificações

A cada elemento da amostra corresponde uma v.a. $X_i \sim \text{Ber}(p)$; $X_i = 1$ se o diodo analisado está fora das especificações e $X_i = 0$, caso contrário

Y: número de diodos na amostra que estão fora das especificações $\implies Y \sim \text{Bin}(50; p)$

$$\overline{p} := \frac{Y}{50} \approx \mathcal{N}\left(p, \frac{p(1-p)}{50}\right)$$

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Os estimadores variância/desvio padrão amostral

A *variância amostral* de uma amostra aleatória $X_1, \ldots, X_n$ é definida como

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

É uma estatística da dispersão de uma amostra aleatória

O *desvio padrão amostral* de uma amostra aleatória $X_1, \ldots, X_n$ é definido como $S = \sqrt{S^2}$

Sendo S^2 e S v.a., elas têm uma distribuição. Em, geral não é fácil conhecer a distribuição delas

Porém, se $X_1, \ldots, X_n$ é uma amostra aleatória de uma população $X \sim \mathcal{N}(\mu, \sigma^2)$, é possível provar que

S^2 tem uma distribuição χ^2

Como avaliar a qualidade dos estimadores?

Vimos que existem diferentes métodos para achar estimadores. Por isso é importante introduzir algum critério para avaliá-los

Seja X uma variável aleatória cuja distribuição de probabilidade depende de um parâmetro desconhecido θ, e seja $\hat{\theta} = T(X_1, \ldots, X_n)$ um estimador de θ baseado numa amostra aleatória de X, de tamanho n

Uma das propriedades desejáveis para o estimador $\hat{\theta}$ é que o seu valor esteja o mais "próximo" possível do verdadeiro parâmetro θ

Estimadores não tendenciosos

Lembre que um estimador é uma variável aleatória com uma determinada distribuição. Por isso não está claro o que significa para uma v.a. ser "próxima" de um determinado valor

Possível solução: Uma propriedade desejável para um estimador é que a média da sua distribuição seja igual ao parâmetro sendo estimado (em média, o estimador "acerta")

Definimos o *viés* de um estimador pontual $\hat{\theta}$ de um parâmetro θ como

$$B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

Um estimador pontual $\hat{\theta}$ é dito *não tendencioso* se

$$B(\hat{\theta}) = 0 \iff \mathbb{E}(\hat{\theta}) = \theta$$

Exemplo: estimadores não tendenciosos*

Seja $X_1, \dots, X_n$ uma amostra de uma população $X \sim \mathcal{N}(\mu, \sigma^2)$

Consideramos os estimadores $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$

Temos que $\mathbb{E}(\overline{X}) = \mu \implies \overline{X}$ é não tendencioso para μ Para S^2 temos

$$\begin{aligned}
\mathbb{E}(S^2) &= \frac{1}{n-1}\mathbb{E}\Big(\sum_i (X_i - \overline{X})^2\Big) = \frac{1}{n-1}\mathbb{E}\Big(\sum_i X_i^2 - n\overline{X}^2\Big) \\
&= \frac{1}{n-1}\mathbb{E}\Big(\sum_i (X_i^2 - \overline{X}^2)\Big) = \frac{1}{n-1}\mathbb{E}\Big(\sum_i \left((X_i^2 - \mu^2) - (\overline{X}^2 - \mu^2)\right)\Big) \\
&= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}Var(\overline{X}) = \frac{n}{n-1}\sigma^2 - \frac{1}{n-1}\sigma^2 = \sigma^2
\end{aligned}$$

onde usamos que $Var(\overline{X}) = \frac{\sigma^2}{n}$. Então, S^2 sé não tendencioso para σ^2

Observação: Essa é a motivação para o fator de normalização $1/(n-1)$

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Erro quadrático médio (EQM)

Uma outra métrica para medir se um estimador retorna estimativas que sejam próximas do valor verdadeiro do parâmetro é o *erro quadrático médio*, definido como

$$\text{EQM}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2)$$

Observação: o fato que um estimador seja não tendencioso não garante que o EQM seja pequeno

De fato, para estimadores não tendenciosos, o EQM mede quão espalhadas as estimativas dadas pelo estimador ficam em relação ao valor verdadeiro do parâmetro $\Rightarrow$ Menor o EQM, mais próximas as estimativas ficam do valor verdadeiro do parâmetro, ou seja, melhor é o nosso estimador

Erro quadrático médio (EQM)

O EQM de um estimador pode ser reescrito em função do seu viés:

$$\text{EQM}(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta})$$

Prova: $\mathbb{E}((\hat{\theta} - \theta)^2) = \mathbb{E}((\hat{\theta} - \mathbb{E}(\hat{\theta}))^2) + (\mathbb{E}(\hat{\theta}) - \theta)^2 = Var(\hat{\theta}) + (B(\hat{\theta}))^2$

O EQM incorpora duas componentes: uma mede a precisão (variância) e a outra mede o viés. Estimadores que têm boas propriedades de EQM têm variança e viés pequenos

Observação 1: para estimadores não tendenciosos o EQM é igual à variança do estimador

Observação 2: estimadores não tendenciosos não necessariamente têm menor EQM do que estimadores tendenciosos

Exemplo: erro quadrático médio (EQM)

Seja $X_1, \ldots, X_n$ uma amostra aleatória tirada de uma população $X \sim \text{Ber}(p)$. Consideramos o estimador $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Para calcular o $\text{EQM}(\overline{X})$ podemos calcular $B(\overline{X})$ e $Var(\overline{X})$

$B(\overline{X}) = 0 \implies \overline{X}$ é não tendencioso $\implies \text{EQM}(\overline{X}) = Var(\overline{X})$

$$Var(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}$$

Exemplo: erro quadrático médio*

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população $X \sim \mathcal{N}(\mu, \sigma^2)$. Nós vimos que $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ e $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ são estimadores não tendenciosos para μ e σ^2. Temos

$$\mathrm{EQM}(\overline{X}) = Var(\overline{X}) + \underbrace{Bias(\overline{X})}_{0} = \frac{\sigma^2}{n}$$

$$\mathrm{EQM}(S^2) = Var(S^2) + \underbrace{Bias(S^2)}_{0} = \frac{2\sigma^4}{n-1}$$ pode ser assumido!

Considere o estimador $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{n-1}{n}S^2$ para σ^2. Temos $B(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2$ e $Var(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 Var(S^2) = \frac{n-1}{n^2}2\sigma^4$. Então

$$\mathrm{EQM}(\hat{\sigma}^2) = \frac{n-1}{n^2}2\sigma^4 + \frac{1}{n^2}\sigma^4 = \frac{2n-1}{n^2}\sigma^4$$

Sendo $\frac{2n-1}{n^2} \le \frac{2}{n-1}$ para todo $n > 1$ temos que $\mathrm{EQM}(\hat{\sigma}^2) < \mathrm{EQM}(S^2)$, apesar de ter que S^2 é não tendencioso e $\hat{\sigma}^2$ é tendencioso!

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Erro absoluto da estimação pontual

Mais uma propriedade desejável para um estimador é que ele retorne estimativas que sejam próximas do valor verdadeiro do parâmetro dentro de um certo erro absoluto d fixado, ou seja $|\hat{\theta} - \theta| \leq d$

Sendo $\hat{\theta}$ uma v.a., não é possível garantir que $|\hat{\theta} - \theta| \leq d$ seja sempre válido

Em algumas situações é possível garantir que a condição $|\hat{\theta} - \theta| \leq d$ seja atendida com alta probabilidade, ou seja

$$\mathbb{P}(|\hat{\theta} - \theta| \leq d) \text{ seja perto de } 1$$

Erro absoluto para a média amostral

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população X com média μ (desconhecida) e variância σ^2 (conhecida)

No caso do estimador $\overline{X}$ é possível garantir que $\mathbb{P}(|\overline{X} - \mu| \leq d) \approx 1$:

$$\mathbb{P}(|\overline{X} - \mu| \leq d) = \mathbb{P}\left(\frac{|\overline{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} \leq \frac{d}{\frac{\sigma}{\sqrt{n}}}\right)$$
$$\overset{TCL}{\approx} \mathbb{P}\left(|Z| \leq \frac{d}{\frac{\sigma}{\sqrt{n}}}\right) = 2\Phi\left(\frac{d}{\frac{\sigma}{\sqrt{n}}}\right) - 1,$$

e o valor de Φ pode ser obtido pela tabela da normal padrão

<u>Observação 1:</u> É preciso conhecer o valor de σ!

<u>Observação 2:</u> Ao aumentar o valor de n temos que $\Phi\left(\frac{d}{\frac{\sigma}{\sqrt{n}}}\right) \to 1$

Erro absoluto para a proporção amostral

Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população $X \sim \text{Ber}(p)$

Também para o estimador $\overline{p}$ é possível garantir que $\mathbb{P}(|\overline{p} - p| \leq d) \approx 1$:

$$\mathbb{P}\left(|\overline{p} - p| \leq d\right) = \mathbb{P}\left(\frac{|\overline{p} - p|}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{d}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\overset{TCL}{\approx} \mathbb{P}\left(|Z| \leq \frac{d}{\sqrt{\frac{p(1-p)}{n}}}\right) \geq \mathbb{P}\left(|Z| \leq \frac{d}{\sqrt{\frac{\frac{1}{4}}{n}}}\right) = 2\Phi(2d\sqrt{n}) - 1,$$

onde usamos que $p(1-p) \leq \frac{1}{4}$ para todo $p \in [0, 1]$

Observação: $Var(\overline{p}) = \frac{\sqrt{p(1-p)}}{n}$, mas o parâmetro p é desconhecido!
Para resolver esse problema usamos $p(1-p) \leq \frac{1}{4}$

Exemplo: erro absoluto de estimação*

Foi realizada uma pesquisa de opinião em uma empresa visando determinar o nível médio de satisfação dos empregados. O índice de satisfação de cada empregado varia entre 0 e 100 pontos, e o desvio padrão populacional é de 30 pontos. Se nessa pesquisa foram sorteados 324 empregados ao acaso para uma entrevista, qual a probabilidade de que o índice de satisfação médio seja estimado com erro absoluto menor que três pontos?

Solução: Seja $X_1, \ldots, X_{324}$ uma amostra aleatória e μ a média da população. Queremos calcular $\mathbb{P}(|\overline{X} - \mu| \leq 3)$

$$\mathbb{P}(|\overline{X} - \mu| \leq 3) = \mathbb{P}\left(|Z| \leq \frac{3}{30/\sqrt{324}}\right) = \mathbb{P}\left(|Z| \leq 1,8\right) = 2\Phi(1,8) - 1 = 0,9281$$

Isso quer dizer que com $92,8\%$ de chance o erro absoluto na estimação do nível médio de satisfação dos empregados será menor que três pontos

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Dimensionamento da amostra

Como vimos acima, quanto maior é o tamanho da amostra em geral mais precisa é a analise estatística

Qual é o tamanho mínimo da amostra para termos estimativas aceitáveis?

Para poder responder essa pergunta, primeiramente fixamos duas constantes:

d: distância máxima considerada tolerável entre a estimativa e o valor verdadeiro do parâmetro

α: probabilidade que a distância entre estimativa e o valor verdadeiro do parâmetro ultrapasse o limite d

Dimensionando a amostra para média (σ^2 conhecido)

Nesse caso procuramos o menor valor de n tal que

$$\mathbb{P}\left(|\overline{X} - \mu| \leq d\right) = 1 - \alpha$$

Usando que

$$\mathbb{P}\left(|\overline{X} - \mu| \leq d\right) \approx \mathbb{P}\left(|Z| \leq \frac{d}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha$$

obtemos que $z_{1-\frac{\alpha}{2}} = \frac{d}{\frac{\sigma}{\sqrt{n}}}$ o que implica

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2$$

Exemplo: dimensionamento da amostra*

Foi realizada uma pesquisa de opinião em uma empresa visando determinar o nível médio de satisfação dos empregados. O índice de satisfação de cada empregado varia entre 0 e 100 pontos, e o desvio padrão populacional é de 30 pontos. Qual deveria ser o tamanho n da amostra de empregados a serem entrevistados para que o erro absoluto na estimação do índice de satisfação médio estivesse limitado por $1,5$ com uma probabilidade de $92,81\%$?

Solução: Seja $X_1, \ldots, X_n$ uma amostra aleatória e μ a média da população. Queremos achar o menor n tal que $\mathbb{P}(|\overline{X} - \mu| \leq 1,5) = 0,9281$:

$$0,9281 = 1 - 0,0719 = \mathbb{P}(|\overline{X} - \mu| \leq 1,5) \approx \mathbb{P}\left(|Z| \leq \frac{1,5}{30/\sqrt{n}}\right)$$

$$\implies n = \left(\frac{z_{1-\frac{\alpha}{2}}\sigma}{d}\right)^2 = \left(\frac{(1,8)(30)}{1,5}\right)^2 \approx 1,296 \text{ empregados}$$

Observação: repare que σ é conhecido!

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Dimensionando a amostra para estimar a proporção

Quando o objetivo é estimar a proporção populacional, o tamanho mínimo da amostra para garantir um distância máxima d da proporção verdadeira com uma probabilidade pelo menos α é calculada a partir da equação:

$$\mathbb{P}\left(|\overline{p} - p| \leq d\right) = 1 - \alpha$$

Usando a aproximação Normal, como anteriormente, obtemos:

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d}\right)^2 p(1-p)$$

Observação: o tamanho da amostra depende do parâmetro que estamos querendo estimar (p), que de fato não conhecemos!

Usando $p(1-p) \leq 1/4$ temos que o tamanho mínimo é $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{2d}\right)^2$

Se por acaso sabemos que $p \neq 0.5$ é preciso estimar $p(1-p)$ com o valor mais alto que pode assumir no domínio dos possíveis valores de p

Exemplo: dimensionamento da amostra*

Uma empresa fabrica diodos usados em placas de circuito impresso. Cada um desses diodos pode estar fora das especificações padrões com probabilidade p (desconhecida). Suponha que queremos que o erro absoluto da estimativa da fração de diodos fora das especificações não ultrapasse $0,05$ com uma probabilidade de 90%. Qual deveria ser o tamanho amostral se:

a) não tivermos qualquer informação sobre o verdadeiro valor de p?

b) sabemos que o verdadeiro valor de p é inferior a $0,2$?

Solução: Temos $d = 0,05$ e para $1 - \alpha = 0,90$ temos $z_{1-\frac{\alpha}{2}} = 1,64$

a) Como nada sabemos sobre p, na formula $n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{d}\right)^2 p(1-p)$ pegamos

$p = 0,5 \implies n \geq \left(\frac{1{,}64}{(0{,}05)2}\right)^2 = 269$ diodos

b) Sabemos que $p \leq 0,2 \implies p(1-p) \leq 0,2(1-0,2)$. Então,

$n \geq \left(\frac{1{,}64}{(0{,}05)}\right)^2 (0,2)(0,8) = 172,13 \implies 173$ diodos

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Estimação por intervalos

Intervalos de confiança

Na estimação pontual o objetivo é inferir o parâmetro da população a partir de uma amostra; um estimador pontual especifica um único valor.

Na estimativa por intervalo o objetivo é inferir a partir da amostra um *intervalo* de valores para o qual nos temos alguma *confiança* que isso contém o verdadeiro parâmetro

Exemplo: Retira-se uma amostra de 500 brasileiros e calcula-se a média das alturas encontrando-se $1,70$ metro. Logo uma estimação pontual da verdadeira altura média (μ) é dada por $\bar{x} = 1,70\text{m}$. Através do intervalo de confiança é possível encontrar um intervalo, por exemplo $[1,60\text{m}, 1,80\text{m}]$ que, em 95% das vezes, contem o valor verdadeiro μ

A lógica na construção dos intervalos de confiança

1) θ: parâmetro da população (média, variância, etc)

2) $\hat{\theta}$: estimador de θ (é uma variável aleatória!)

3) "Conhecida" a distribuição de probabilidade de $\hat{\theta}$, é possível construir um intervalo $\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2$ e exigir que

$$\mathbb{P}(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha$$

4) $(1-\alpha)$: nível de confiança; normalmente $1 - \alpha = 0.90, 0.95, 0.99$

Seja $X_1, \ldots, X_n$ uma amostra de uma população com parâmetro θ. Um *intervalo de confiança* de θ é dado por um par de funções L e U da amostra (i.e., L e U são estatísticas) tais que $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ para todas possível realizações da amostra $\boldsymbol{x}$.

O intervalo aleatório $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ é chamado *estimador do intervalo*

Exemplo: intervalo de confiança*

X_1, X_2, X_3, X_4: amostra de uma população com distribuição $\mathcal{N}(\theta, 1)$.
Um estimador do intervalo para θ é, por exemplo, $[\overline{\boldsymbol{X}} - 1, \overline{\boldsymbol{X}} + 1]$

Na estimação pontual, nós estimamos que $\theta = \overline{\boldsymbol{X}}$, enquanto agora nós temos uma estimativa menos precisa, ou seja, $\theta \in [\overline{\boldsymbol{X}} - 1, \overline{\boldsymbol{X}} + 1]$

Porem, na estimação pontual temos que $\mathbb{P}(\overline{\boldsymbol{X}} = \theta) = 0$, ou seja não temos um bom controle sobre o possível erro. Por outro lado, é possível calcular a *probabilidade que o intervalo contenha o verdadeiro parâmetro*:

$$\mathbb{P}(\theta \in [\overline{\boldsymbol{X}} - 1, \overline{\boldsymbol{X}} + 1]) = \mathbb{P}(\overline{\boldsymbol{X}} - 1 \leq \theta \leq \overline{\boldsymbol{X}} + 1) = \mathbb{P}(-1 \leq \overline{\boldsymbol{X}} - \theta \leq +1)$$

$$= \mathbb{P}(-2 \leq \frac{\overline{\boldsymbol{X}} - \theta}{\sqrt{1/4}} \leq +2) = \mathbb{P}(-2 \leq Z \leq +2) = 0.95 \qquad Z \sim \mathcal{N}(0, 1)$$

Observação: a motivação para utilizar estimativas por intervalos é ter alguma garantia de capturar o verdadeiro parâmetro

*Do livro *Statistical Inference*, G. Casella & R. L. Berger

Intervalo de confiança para a média populacional (1/3)

Situação 1: A população X tem distribuição $\mathcal{N}(\mu, \sigma^2)$, $\underline{\sigma^2 \text{ é conhecida}}$

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n, e $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média amostral. Sabemos que

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

Então, fixado um nível de confiança $(1-\alpha)$, obtem-se o intervalo de confiança para μ usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \implies \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

Intervalo de confiança para a média populacional (2/3)

Situação 2: A população X têm distribuição $\mathcal{N}(\mu, \sigma^2)$, <u>$\sigma^2$ é desconhecida</u>

Seja $X_1, \ldots, X_n$ uma amostra de tamanho n, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ média amostral e $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ a variância amostral (*repare que precisamos estimar a variância!*). Temos que

$T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ tem distribuição t de Student com $(n-1)$ grau de liberdade

Então, fixado um nível de confiança $(1-\alpha)$, obtem-se o intervalo de confiança para μ usando

$$\mathbb{P}\left(-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{n-1;1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Longrightarrow \left[\overline{X} - t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}; \overline{X} + t_{n-1;1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

$t_{n-1;1-\frac{\alpha}{2}}$ encontra-se na tabela da t de Student com $(n-1)$ graus de liberdade

Distribuição t de Student

Se $X_1, \ldots, X_n$ são i.i.d com distribuição $\mathcal{N}(\mu, \sigma^2)$, a variável aleatória $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ têm uma distribuição que não depende de μ nem σ^2, e apenas depende de um parâmetro que é $(n-1)$

Tal distribuição é chamada t de Student com $(n-1)$ graus de liberdades

Observação: a dependência na distribuição de T do parâmetro n é devida a $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$

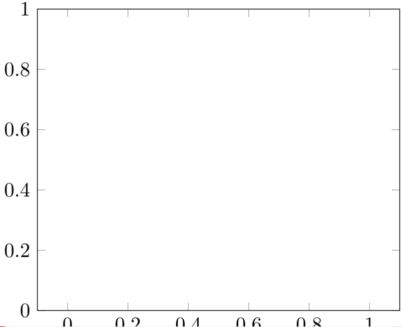
Observação: Se $X_1, \ldots, X_n$ não foram normais, não é mas verdade que a variável T tem distribuição t de Student!

Historia: Foi introduzida in 1908 por William Gosset quando trabalhava em uma cervejaria em Dublin. Ele publicou usando o pseudônimo Student porque não foi permitido publicar com o seu próprio nome. Ele estava interessado na seleção de cevadas de melhor qualidade, e trabalhava com amostras pequenas

Distribuição t de Student

A distribuição t de Student tem as seguintes propriedades:

- depende de um parâmetro chamado graus de liberdade
- como a normal é simétrica e centrada em 0
- quando os graus de liberdade tendem ao infinito a distribuição se aproxima da distribuição normal padrão
- o cálculo de probabilidades pode ser feito por meio de consulta a uma tabela de probabilidade (como no caso da normal padrão)



Exemplo: intervalo de confiança - situação 1

Os seguintes dados foram coletados para uma amostra de uma população normal, i.e., $X \sim \mathcal{N}(?, 4)$: 15, 10, 7, 13, 6, 9, 13, 7, 10

Calcule o intervalo de 90% de confiança para a média populacional

amostra pequena, população normal e σ conhecido $\Longrightarrow$ usamos a Normal

Para $\alpha = 0,10$ temos $z_{1-\alpha/2} = 1,65$. Então,

$$\left[\bar{x} - z_{0,95}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0,95}\frac{\sigma}{\sqrt{n}}\right] = \left[10 - 1,65\left(\frac{2}{3}\right); 10 + 1,65\left(\frac{2}{3}\right)\right]$$
$$= [8,9; 11,1]$$

Exemplo: intervalo de confiança - situação 2

Os seguintes dados foram coletados para uma amostra de uma população normal, i.e., $X \sim \mathcal{N}(?, ?)$: 15, 10, 7, 13, 6, 9, 13, 7, 10.

Qual é o intervalo de 95% de confiança para a média da população?

amostra pequena e σ desconhecido $\Longrightarrow$ usamos a t de Student

Temos $\overline{x} = \frac{1}{n}\sum_{i=1}^{9} x_i = 10$ e $s^2 = \frac{1}{n-1}\sum_{i=1}^{9}(x_i - \overline{x})^2 = \frac{78}{8} = 9,75$ e $s = 3,12$. Usamos a distribuição t de Student com 8 graus de liberdade. Para $\alpha = 0,05$ temos $t_{8;1-\alpha/2} = 2,306$. Então,

$$\left[\bar{x} - t_{8;0,975}\frac{s}{\sqrt{n}}; \bar{x} + t_{8;0,975}\frac{s}{\sqrt{n}}\right] = \left[10 - \frac{3,12}{3}2,306; 10 + \frac{3,12}{3}2,306\right]$$
$$= [7,6; 12,4]$$

Intervalo de confiança para a média populacional (3/3)

Situação 3 (Grande amostra): A população X tem uma distribuição qualquer mas a amostra é grande $(n \geq 30)$ para que o TCL se aplique

Seja $X_1, \ldots, X_n$ uma amostra $(n \geq 30)$ e $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ a média amostral

σ^2 conhecido: usando o TCL podemos dizer que $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1)$;
o intervalo de $(1-\alpha)$ confiança para μ é calculado usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \implies \left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

σ^2 desconhecido: usamos $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ e, nesse caso também,
o TCL nos garante que $Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \approx \mathcal{N}(0,1)$;
o intervalo de $(1-\alpha)$ confiança para μ é calculado usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \implies \left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

Em ambos casos, $z_{1-\frac{\alpha}{2}}$ encontra-se na tabela da distribuição Normal Padrão

Exemplo: intervalo de confiança - situação 3*

Deseja-se estimar a resistência média de um certo tipo de fibra usada na fabricação de um tecido. Uma amostra aleatória de 40 espécimes da fibra tem uma média amostral de $12,4$ bar. Se o desvio padrão σ da população é conhecido e igual a $2,1$ bar, determine um intervalo de confiança de 95% para o verdadeiro valor da média populacional μ

Temos $n = 40, \overline{x} = 12,4, \sigma = 2,1$, $\alpha = 0,05$ e $z_{1-\alpha/2} = 1,96$. Como $n \geq 30$ e σ é conhecido, podemos aplicar o TCL e construir um intervalo de $1-\alpha$ confiança usando $\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$:

$$\left[12,4 - (1,96)\frac{2,1}{\sqrt{40}}; 12,4 + (1,96)\frac{2,1}{\sqrt{40}}\right] = [11,75; 13,05]$$

Observação: A população pode ter qualquer distribuição, não precisa ser normal; isso porque o TCL pode ser aplicado!

*Do livro *Probabilidade e estatística: quantificando a incerteza*, J. Pinheiro et al.

Intervalo de confiança para a proporção populacional

Seja X uma população cujos elementos podem ser classificados em dois tipos: *sucesso* e *insucesso*, ou seja $X \sim Ber(p)$

Dada uma amostra $X_1, \ldots, X_n$ de tamanho n, um estimador do parâmetro p é a proporção amostral $\overline{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. O TCL $\Longrightarrow$

$$\frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx \mathcal{N}(0,1) \quad \text{se } n \text{ é suficientemente grande, i.e., } np(1-p) \geq 3$$

Calculamos o intervalo de $(1-\alpha)$ confiança para p usando

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \Longrightarrow \left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]$$

<u>Observação:</u> O parâmetro p é de fato desconhecido, então não está claro como calcular $\sqrt{\frac{p(1-p)}{n}}$ e obter o intervalo

Dois tipos de intervalos para a proporção populacional

Para achar o intervalo de confiança para a proporção p é preciso calcular $\sqrt{\frac{p(1-p)}{n}}$. Como fazer isso se o parâmetro p é desconhecido?

Intervalo de confiança conservativo: lembrando que $p(1-p) \leq 1/4$ para todo p podemos achar o intervalo de confiança usando esse valor

$$\left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \subset \left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}}\right]$$

Intervalo de confiança não conservativo: podemos usar na expressão da variância o valor estimado $\bar{p}$ e achar o intervalo usando

$$\left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\overline{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}; \overline{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right]$$

Observação: A escolha conservativa sempre levará a intervalo maiores

Exemplo: intervalo de confiança para proporção

Uma loja em Botafogo deseja inferir a proporção de clientes que estão satisfeitos com seu serviço. Para isto, entrevistou 30 clientes e obteve que 20 são satisfeitos e 10 são insatisfeitos

Construa um intervalo conservativo e um não conservativo de 96% confiança para a proporção de clientes satisfeitos

O intervalo não conservativo de 96% de confiança para p é

$$\left[\bar{p} - z_{0,98}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{0,98}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] = \left[\frac{2}{3} - 2,05(\sqrt{\frac{2/9}{30}}); \frac{2}{3} + 2,05(\sqrt{\frac{2/9}{30}})\right] = \left[\frac{2}{3} - 0,18; \frac{2}{3} + 0,18\right] = [0,49; 0,85]$$

O intervalo conservativo de 96% de confiança para p é

$$\left[\bar{p} - z_{0,98}\sqrt{\frac{1}{4n}}; \bar{p} + z_{0,98}\sqrt{\frac{1}{4n}}\right] = \left[\frac{2}{3} - 2,05(\sqrt{\frac{1/4}{30}}); \frac{2}{3} + 2,05(\sqrt{\frac{1/4}{30}})\right] = \left[\frac{2}{3} - 0,19; \frac{2}{3} + 0,19\right] = [0,48; 0,86]$$

Testes de Hipótese

Introdução

Ao fazer um teste de hipótese, o objetivo é formular hipóteses sobre um parâmetro da população, e com base em uma amostra observada, tomar alguma decisão sobre o comportamento de tal parâmetro.

Exemplo: Em 1710, o polímata John Arbuthnot examinou dados sobre nascimentos de pessoas em Londres de 1629 até 1710. Em todos os 82 anos examinados, o número de nascimentos de homens excedia o de mulheres. Ele estava interessado em decidir se esse excesso era estatisticamente significante ou ocorreu simplesmente por acaso. Modelando o sexo de um recém-nascido como uma distribuição de Bernoulli de parâmetro p desconhecido, ele estava interessado em decidir se $p = 1/2$ ou $p \neq 1/2$. Assumindo que $p = 1/2$ (nascimento de homens e mulheres são equiprováveis), a probabilidade do excesso ser observado nos 82 anos seria de $1/2^{82}$, ou seja, 1 em $4,84 \times 10^{24}$. Portanto, ele concluiu que tal diferença de fato existia, e não foi observada por acaso, apesar de não conhecer o seu motivo.

Conceitos básicos

Exemplo: Consideremos uma empresa que produz cabos náuticos de determinado tipo. O fabricante garante que a carga de ruptura média de seu produto é de pelo menos 90kg. Um potencial consumidor, interessado na compra de grande quantidade do produto, decide fazer ensaios de carga de ruptura com 20 espécimes, obtendo uma média amostral de 88,4kg. Podemos concluir que a carga média de ruptura é de fato inferior à apresentada pelo fabricante?

- População: todos os cabos fabricados pela companhia
- Propriedade de interesse: carga de ruptura, em kg, representada pela variável aleatória X
- Distribuição de X: Assumiremos ser normal, ou que o tamanho da amostra é suficientemente grande para garantir a validade do TCL
- Parâmetro: média populacional μ
- Hipótese nula: $\mathcal{H}_0 : \mu \geq 90$
- Hipótese alternativa: $\mathcal{H}_1 : \mu < 90$

Parece razoável considerar a estatística $\overline{X}$, e dependendo do seu valor na amostra, aceitar ou rejeitar a hipótese nula.

Conceitos básicos

Mais geralmente, queremos avaliar a validade (ou não) de uma afirmação sobre uma característica (parâmetro θ da distribuição de probabilidade) da população X, através de uma amostra aleatória $X_1, \ldots, X_n$ dessa população. Alguns passos importantes nesse procedimento são:

1) Formulamos duas hipóteses mutuamente exclusivas sobre o valor correto de θ:
   - A *hipótese nula*, denotada por $\mathcal{H}_0$, representando o *status quo* da natureza, a alternativa mais conservadora sobre θ
   - A *hipótese alternativa*, denotada por $\mathcal{H}_1$, representando a alternativa inovadora sobre θ
2) Elegemos uma *estatística de teste* $T(X_1, \ldots, X_n)$, que resuma toda a "informação relevante" sobre θ
3) Dividimos o conjunto de valores possíveis de T em duas regiões:
   - *Região de aceitação*, A: Se $T(x_1, \ldots, x_n) \in A$, aceitamos $\mathcal{H}_0$
   - *Região de rejeição* ou *região crítica* R: Se $T(x_1, \ldots, x_n) \in R$, rejeitamos $\mathcal{H}_0$

Pergunta: Como escolhemos a estatística de teste, a região de aceitação e a região de rejeição?

Conceitos básicos

Trataremos somente de testes sobre a média ou proporção populacional, portanto nossas estatísticas de teste estarão relacionadas com $\overline{X}$ ou $\overline{p}$.
Foquemo-nos em como determinar as regiões de aceitação e rejeição. Para isso, precisamos falar dos *erros tipo I e II*.

Como a amostra nos fornece somente informações parciais sobre a população, a decisão que tomamos sobre aceitar ou rejeitar $\mathcal{H}_0$ pode estar incorreta. Queremos a "melhor decisão" possível, ou seja, alguma que minimize algum erro. Temos dois tipos de erros possíveis:

- Erro tipo I: Rejeitar $\mathcal{H}_0$ quando $\mathcal{H}_0$ é verdadeira
- Erro tipo II: Não rejeitar $\mathcal{H}_0$ quando $\mathcal{H}_0$ é falsa

Cada um desses erros tem uma certa probabilidade de ocorrência:

- $\alpha = \mathbb{P}(\text{erro tipo I}) = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira})$, dito o *nível de significância do teste*
- $\beta = \mathbb{P}(\text{erro tipo II}) = \mathbb{P}(\text{não rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é falsa})$

Gostaríamos de construir as regiões de aceitação e rejeição de modo que tanto α quanto β fossem pequenos, porém isso é impossível, fixado o tamanho da amostra!

O que é feito é fixar α igual a um valor pequeno (usualmente 0,05 ou 0,01) para assim construir as regiões de aceitação e rejeição. Mas porque essa assimetria?

Assimetria nos erros tipo I e II

Note que não tratamos os erros tipo I e II acima como simétricos. Isso se dá pois, da maneira como formulamos $\mathcal{H}_0$ e $\mathcal{H}_1$, o erro tipo I representa algo mais grave que o erro tipo II, mesmo sendo essa noção subjetiva:

Exemplo: No exemplo dos cabos náuticos, o fabricante afirma que a carga de ruptura é de pelo menos 90kg, porém um consumidor está desconfiado que ela é menor, com base na sua observação, e formulamos as seguintes hipótese:

$$\begin{cases} \mathcal{H}_0 : & \mu \geq 90 \\ \mathcal{H}_1 : & \mu < 90 \end{cases}$$

Note que, do ponto de vista do fabricante, o erro tipo I é mais grave, pois ele representa concluir (erradamente) que a carga de ruptura média é menor do que o por ele afirmado. Porém, do ponto de vista do consumidor, o erro tipo II é mais grave, pois implica em uma crença em um valor maior que o real da carga de ruptura, implicando, possivelmente, em um acidente. Portanto, do ponto de vista do consumidor, seria mais adequado formular o teste como:

$$\begin{cases} \mathcal{H}_0' : & \mu \leq 90 \\ \mathcal{H}_1' : & \mu > 90 \end{cases}$$

As hipóteses sempre serão formuladas de modo que $\mathcal{H}_0$ seja aquela hipótese cuja rejeição equivocada constitui o erro de maior importância.

Hipótese simples vs. composta

Dizemos que uma hipótese, seja $\mathcal{H}_0$ ou $\mathcal{H}_1$ é *simples* se ela corresponde a um único valor, ou seja, é da forma $\theta = \theta_0$.

Hipóteses que contém mais de um valor de θ, em particular da forma $\theta \leq \theta_0$, $\theta < \theta_0$, $\theta \geq \theta_0$ e $\theta > \theta_0$, são ditas *compostas*.

O que isso influencia no cálculo das probabilidades dos erros tipo I e II?

Considere que queremos testar $\mathcal{H}_0 : \theta \geq \theta_0$ *versus* $\mathcal{H}_1 : \theta < \theta_0$. Lembremos que $\alpha = \mathbb{P}(\text{erro tipo I}) = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira})$. Se $\mathcal{H}_0$ é uma hipótese composta, tal probabilidade seria bastante difícil de ser calculada, pois não temos um valor específico para o parâmetro θ.

$\Longrightarrow$ seria conveniente poder considerar $\mathcal{H}_0$ simples, por exemplo, como $\mathcal{H}_0 : \theta = \theta_0$!

Porque isso faz sentido?

Pode-se provar que um procedimento capaz de distinguir eficientemente entre $\mathcal{H}_0' : \theta = \theta_0$ (valor mais desfavorável possível para θ) e $\mathcal{H}_1 : \theta < \theta_0$ também será "bom" para distinguir as hipóteses originais $\mathcal{H}_0 : \theta \geq \theta_0$ e $\mathcal{H}_1 : \theta < \theta_0$.

Mesmo que as hipóteses sejam do tipo $\mathcal{H}_0 : \theta \geq \theta_0$ *vs.* $\mathcal{H}_1 : \theta < \theta_0$ ou $\mathcal{H}_0 : \theta \leq \theta_0$ *vs.* $\mathcal{H}_1 : \theta > \theta_0$, a hipótese nula $\mathcal{H}_0$ sempre será considerada como $\theta = \theta_0$, pois é o valor mais desfavorável para θ e especifica uma única distribuição para a população.

Um exemplo para fixar as ideias

Exemplo: Para concluir o exemplo dos cabos náuticos, o fabricante afirma que a carga de ruptura é de pelo menos 90kg, e um consumidor desconfiado, em uma amostra de 20 cabos, obtém uma carga de ruptura de 88,4kg. Supondo que o desvio padrão populacional é de 10kg, teste as hipóteses $\mathcal{H}_0 : \mu \geq 90$ *versus* $\mathcal{H}_1 : \mu < 90$, ao nível de significância de 5%.

Solução:

- Lembremos que consideraremos $\mathcal{H}_0$ simplesmente como $\mu = 90$
- Parece razoável rejeitarmos $\mathcal{H}_0$ se observarmos um valor suficientemente baixo para $\overline{X}$, digamos x_c. Usaremos o nível de significância para obter tal valor. Lembremos que se $\mu = 90$, então $Z = \dfrac{\overline{X} - 90}{10/\sqrt{20}} \sim N(0,1)$:

$$0,05 = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \mathbb{P}(\overline{X} < x_c \mid \mu = 90) = \mathbb{P}\left(Z < \frac{x_c - 90}{10/\sqrt{20}}\right)$$

$$\Longrightarrow \frac{x_c - 90}{10/\sqrt{20}} = z_{0,05} = -z_{0,95} = -1,64 \Longrightarrow x_c = 86,33\text{kg}.$$

- Portanto, a região de rejeição R é dada por $\{\overline{X} < 86,33\}$, e a região de aceitação A é dada por $\{\overline{X} \geq 86,33\}$.
- Como $\overline{x} = 88,4 \in A$, não temos evidência para rejeitar $\mathcal{H}_0$, ao nível de significância de 5%.

Teste para a média amostral: σ conhecido

Seja μ o parâmetro de interesse de uma população X. Assumiremos que X tenha uma distribuição normal, ou que a amostra seja suficientemente grande para poder aplicar o TCL. Um roteiro para construir um teste de hipótese sobre μ, quando σ é conhecido, é o seguinte:

- Especificar as hipóteses nula e alternativa dentre as três possibilidades a seguir:
  1) $\mathcal{H}_0 : \mu = \mu_0$ *versus* $\mathcal{H}_1 : \mu \neq \mu_0$ (teste *bilateral*)
  2) $\mathcal{H}_0 : \mu \leq \mu_0$ *versus* $\mathcal{H}_1 : \mu > \mu_0$ (teste *unilateral*)
  3) $\mathcal{H}_0 : \mu \geq \mu_0$ *versus* $\mathcal{H}_1 : \mu < \mu_0$ (teste *unilateral*)

  Nos casos 2) e 3) trabalhamos com $\mathcal{H}_0 : \mu = \mu_0$
- Fixar o nível de significância
  $\alpha = \mathbb{P}(\text{erro tipo I}) = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira})$
- Usar $\overline{X}$ como estatística de teste. Note que se $\mu = \mu_0$, então $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ é normal padrão (resp. aproximadamente normal padrão), se X é normal (resp. se X tem distribuição qualquer mas n é grande para valer o TCL)
- Usar o nível de significância para determinar as regiões de rejeição e aceitação
- Coletar os dados, calcular $\overline{x}$ e decidir por aceitar ou rejeitar $\mathcal{H}_0$, se $\overline{x} \in A$ ou $\overline{x} \in R$, respectivamente

Exemplo: σ conhecido

O desempenho médio na estrada para um certo tipo de automóvel SUV movido a diesel é de 15 km/ℓ segundo informações da montadora. Uma revista especializada verificou o desempenho na estrada em 27 desses automóveis e apurou um desempenho médio amostral de 14,3km/ℓ. Admitindo que o desempenho ao longo da estrada siga um modelo Normal com desvio padrão de 3 km/ℓ:

a) Teste ao nível de significância de 5% a afirmação da montadora de que a média de consumo é igual a 15 km/ℓ contra a alternativa de ser diferente de 15 km/ℓ. O que você pode concluir?

Solução:

a) Queremos testar $\mathcal{H}_0 : \mu = 15$ *versus* $\mathcal{H}_1 : \mu \neq 15$. Parece razoável rejeitarmos $\mathcal{H}_0$ se observarmos um valor de $\overline{X}$ que diste suficientemente de 15. Determinemos tal distância mais precisamente usando o nível de significância. Como σ é conhecido, notemos que se $\mu = 15$, então $Z = \dfrac{\overline{X} - 15}{3/\sqrt{27}} \sim N(0,1)$:

$$0,05 = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \mathbb{P}(|\overline{X} - 15| > x_c \mid \mu = 15)$$
$$= \mathbb{P}\left(|Z| > \frac{x_c}{3/\sqrt{27}}\right) \implies \frac{x_c}{3/\sqrt{27}} = 1,96 \implies x_c = 1,13.$$

Dessa forma, rejeitamos $\mathcal{H}_0$ se $|\overline{X} - 15| > 1,13$, ou equivalentemente, se $\overline{X} > 16,13$km/ℓ ou $\overline{X} < 13,87$km/ℓ. Como o valor observado foi $\overline{x} = 14,3$km/ℓ, não rejeitamos a hipótese nula, ao nível de significância de 5%.

Exemplo: σ conhecido (continuação)

O desempenho médio na estrada para um certo tipo de automóvel SUV movido a diesel é de 15 km/ℓ segundo informações da montadora. Uma revista especializada verificou o desempenho na estrada em 27 desses automóveis e apurou um desempenho médio amostral de 14,3km/ℓ. Admitindo que o desempenho ao longo da estrada siga um modelo Normal com desvio padrão de 3 km/ℓ:

b) Determine a probabilidade do erro do tipo II assumindo $\mu = 14$ km/ℓ. O que esta probabilidade significa?

Solução:

b) Lembremos que $\beta = \mathbb{P}(\text{erro tipo II}) = \mathbb{P}(\text{não rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é falsa})$. Como estamos especificando um valor para μ de modo que $\mathcal{H}_0$ seja falsa, temos que:

$$\begin{aligned}\beta(14) &= \mathbb{P}(\text{não rejeitar } \mathcal{H}_0 \mid \mu = 14) = \mathbb{P}(13,87 \leq \overline{X} \leq 16,13 \mid \mu = 14) \\ &= \mathbb{P}\left(\frac{13,87-14}{3/\sqrt{27}} \leq \frac{\overline{X}-14}{3/\sqrt{27}} \leq \frac{16,13-14}{3/\sqrt{27}}\right) = \mathbb{P}(-0,23 \leq Z \leq 3,69) = 0,59.\end{aligned}$$

Assim, se $\mathcal{H}_0$ é falsa com $\mu = 14$, teríamos uma chance de 59% de equivocadamente não a rejeitarmos.

Exemplo: σ conhecido (continuação)

O desempenho médio na estrada para um certo tipo de automóvel SUV movido a diesel é de 15 km/ℓ segundo informações da montadora. Uma revista especializada verificou o desempenho na estrada em 27 desses automóveis e apurou um desempenho médio amostral de 14,3km/ℓ. Admitindo que o desempenho ao longo da estrada siga um modelo Normal com desvio padrão de 3 km/ℓ:

c) Uma concorrente afirma que a montadora em questão não diz a verdade, e que o desempenho médio do veículo é na verdade de 14km/ℓ. Teste agora, ainda ao nível de significância de 5%, qual das duas afirmativas são suportadas pelos dados observados.

Solução:

c) Queremos testar $\mathcal{H}_0 : \mu = 15$ *versus* $\mathcal{H}_1 : \mu = 14$. Parece razoável seguirmos a lógica de um teste unilateral, e rejeitarmos $\mathcal{H}_0$ se observarmos um valor suficientemente baixo para $\overline{X}$. Determinemos tal valor mais precisamente usando o nível de significância. Como σ é conhecido, notemos que se $\mu = 15$, então $Z = \dfrac{\overline{X} - 15}{3/\sqrt{27}} \sim N(0,1)$:

$$0,05 = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \mathbb{P}(\overline{X} < x_c \mid \mu = 15)$$

$$= \mathbb{P}\left(Z < \frac{x_c - 15}{3/\sqrt{27}}\right) \implies \frac{x_c - 15}{3/\sqrt{27}} = -1,65 \implies x_c = 14,04.$$

Dessa forma, rejeitamos $\mathcal{H}_0$ se $\overline{X} < 14,04$. Como o valor observado foi $\overline{x} = 14,3$km/ℓ, não rejeitamos a hipótese nula, ao nível de significância de 5%, e os dados ainda suportam a hipótese do desempenho ser 15km/ℓ.

Teste para a média amostral: σ desconhecido

Seja X a quantidade de interesse da população, cuja distribuição assumiremos normal. Um roteiro para construir um teste de hipótese sobre μ, quando σ é desconhecido, é o seguinte:

- Especificar as hipóteses nula e alternativa dentre as três possibilidades a seguir:
  1) $\mathcal{H}_0 : \mu = \mu_0$ *versus* $\mathcal{H}_1 : \mu \neq \mu_0$ (teste *bilateral*)
  2) $\mathcal{H}_0 : \mu \leq \mu_0$ *versus* $\mathcal{H}_1 : \mu > \mu_0$ (teste *unilateral*)
  3) $\mathcal{H}_0 : \mu \geq \mu_0$ *versus* $\mathcal{H}_1 : \mu < \mu_0$ (teste *unilateral*)

  Nos casos 2) e 3) trabalhamos com $\mathcal{H}_0 : \mu = \mu_0$
- Fixar o nível de significância
  $\alpha = \mathbb{P}(\text{erro tipo I}) = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira})$
- Usar $\overline{X}$ como estatística de teste. Note que se $\mu = \mu_0$, então $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ tem distribuição t de Student com $(n-1)$ graus de liberdade. Lembre que para isso valer, X precisa ser necessariamente normal!
- Usar o nível de significância para determinar as regiões de rejeição e aceitação
- Coletar os dados, calcular $\overline{x}$ e decidir por aceitar ou rejeitar $\mathcal{H}_0$, se $\overline{x} \in A$ ou $\overline{x} \in R$, respectivamente

Exemplo: σ desconhecido

Exemplo: Na situação dos cabos náuticos, o fabricante afirma que a carga de ruptura é de pelo menos 90kg, e um consumidor desconfiado, em uma amostra de 20 cabos, obtém uma carga de ruptura de 88,4kg. Agora, o desvio padrão populacional é desconhecido, porém o desvio padrão amostral observado foi de 11,0kg. O que podemos dizer agora sobre as hipóteses $\mathcal{H}_0 : \mu \geq 90$ *versus* $\mathcal{H}_1 : \mu < 90$, ainda ao nível de significância de 5%?

Solução:

- Lembremos que consideraremos $\mathcal{H}_0$ simplesmente como $\mu = 90$
- Parece razoável rejeitarmos $\mathcal{H}_0$ se observarmos um valor suficientemente baixo para $\overline{X}$, digamos x_c. Usaremos o nível de significância para obter tal valor. Lembremos que se $\mu = 90$, então $T = \dfrac{\overline{X} - 90}{S/\sqrt{20}} \sim t$ com $n - 1 = 19$ graus de liberdade:

$$0,05 = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \mathbb{P}(\overline{X} < x_c \mid \mu = 90) = \mathbb{P}\left(T < \frac{x_c - 90}{11/\sqrt{20}}\right)$$

$$\Longrightarrow \frac{x_c - 90}{11/\sqrt{20}} = t_{19;0,05} = -t_{19;0,95} = -1,729 \Longrightarrow x_c = 85,75\text{kg}.$$

- Portanto, a região de rejeição R é dada por $\{\overline{X} < 85,75\}$, e a região de aceitação A é dada por $\{\overline{X} \geq 85,75\}$.
- Como $\overline{x} = 88,4 \in A$, não temos evidência para rejeitar $\mathcal{H}_0$, ao nível de significância de 5%.

p-valor de um teste: motivação

A fim de motivarmos o conceito de p-valor de um teste, vejamos um exemplo.

Exemplo: Um produto químico tem seu teor de pureza normalmente distribuído, com média 0,72 e desvio padrão 0,02. A fim de aumentar a pureza, o produto é submetido a um tratamento. Dezesseis unidades do produto são selecionadas de forma aleatória e submetidas a esse tratamento. Em seguida, a pureza de cada unidade é determinada obtendo-se, para elas, uma média amostral de 0,73. Podemos dizer que o tratamento contribuiu para o aumento da pureza? Assuma que o tratamento somente altera a média populacional, mantendo seu desvio padrão inalterado.

Solução: Está claro que queremos testar $\mathcal{H}_0 : \mu = 0,72$ *versus* $\mathcal{H}_1 : \mu > 0,72$. Rejeitaremos então $\mathcal{H}_0$ ao observarmos um valor suficientemente alto para $\overline{X}$. Considerando $\mu = 0,72$, lembremos que $Z = \dfrac{\overline{X} - 0,72}{0,02/\sqrt{16}} \sim N(0,1)$. Testemos, primeiramente, ao nível de significância de 5%:

$$0,05 = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \mathbb{P}(\overline{X} > x_c \mid \mu = 0,72) = \mathbb{P}\left(Z > \frac{x_c - 0,72}{0,02/\sqrt{16}}\right)$$

$$\implies \frac{x_c - 0,72}{0,02/\sqrt{16}} = z_{0,95} = 1,65 \implies x_c = 0,728.$$

Como o valor observado foi $\overline{x} = 0,73 > 0,728$, temos evidências para rejeitar a hipótese nula, ao nível de significância de 5%, e podemos afirmar que o tratamento aumentou a pureza do produto.

p-valor de um teste: motivação (continuação)

Exemplo: Um produto químico tem seu teor de pureza normalmente distribuído, com média 0,72 e desvio padrão 0,02. A fim de aumentar a pureza, o produto é submetido a um tratamento. Dezesseis unidades do produto são selecionadas de forma aleatória e submetidas a esse tratamento. Em seguida, a pureza de cada unidade é determinada obtendo-se, para elas, uma média amostral de 0,73. Podemos dizer que o tratamento contribuiu para o aumento da pureza? Assuma que o tratamento somente altera a média populacional, mantendo seu desvio padrão inalterado.

Solução: Um determinado padrão de qualidade, mais criterioso, exige que tal teste seja feito ao nível de significância de 1%. Qual será a nova conclusão?

Ainda queremos testar $\mathcal{H}_0 : \mu = 0,72$ *versus* $\mathcal{H}_1 : \mu > 0,72$, e rejeitaremos então $\mathcal{H}_0$ ao observarmos um valor suficientemente alto para $\overline{X}$. Considerando $\mu = 0,72$, lembremos que $Z = \dfrac{\overline{X} - 0,72}{0,02/\sqrt{16}} \sim N(0,1)$. Temos que:

$$0,01 = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ é verdadeira}) = \mathbb{P}(\overline{X} > x_c \mid \mu = 0,72) = \mathbb{P}\left(Z > \frac{x_c - 0,72}{0,02/\sqrt{16}}\right)$$

$$\Longrightarrow \frac{x_c - 0,72}{0,02/\sqrt{16}} = z_{0,99} = 2,33 \Longrightarrow x_c = 0,732.$$

Como o valor observado foi $\overline{x} = 0,73 < 0,732$, não temos evidências para rejeitar a hipótese nula, ao nível de significância de 1%, e não podemos afirmar que o tratamento aumentou a pureza do produto.

p-valor de um teste: definição

O exemplo acima nos leva a algumas indagações:

- Quão perto está o valor observado da fronteira entre as regiões de aceitação e rejeição?
- Equivalentemente, quão sensível é a decisão tomada com respeito a alterações no nível de significância?

Isso nos leva à seguinte definição:

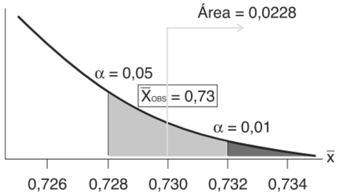
Definição: O *p-valor* ou *nível crítico* de um teste, fixado uma amostra $x_1, \ldots, x_n$ é o menor nível de significância para o qual a hipótese nula ainda será rejeitada. Denotamos o p-valor de um teste por $\tilde{\alpha}$.

Em particular, a hipótese nula não será rejeitada para qualquer nível de significância menor que ou igual a $\tilde{\alpha}$.

p-valor de um teste: intuição e cálculo

Para ganhar intuição, relembremos o exemplo da purificação de um produto químico

- Ao nível de significância de 5%, rejeitamos $\mathcal{H}_0$ se $\overline{X} > 0,728$
- Ao nível de significância de 1%, rejeitamos $\mathcal{H}_0$ se $\overline{X} > 0,732$
- Observamos o valor de $\overline{x} = 0,73$



- O valor de 0,05 corresponde a área à direita do valor de corte $0,728$, no nível de significância de 5%
- Analogamente, o valor de 0,01 corresponde a área à direita do valor de corte $0,732$, no nível de significância de 1%
- Portanto, o menor nível de significância para o qual rejeitaríamos $\mathcal{H}_0$ para esses dados será a probabilidade de observarmos uma média amostral maior que o valor observado $\overline{x}$, sob $\mathcal{H}_0$:

$$\tilde{\alpha} = \mathbb{P}(\overline{X} > 0,73 \mid \mu = 0,72) = 0,0228.$$

- Assim, os dados observados não suportam rejeitar $\mathcal{H}_0$, para qualquer nível de significância abaixo de 0,0228.

p-valor de um teste: intuição e cálculo

A intuição acima nos dá então uma outra interpretação do p-valor de um teste, bem como uma maneira de calculá-lo:

O p-valor de um teste para a média populacional, fixado uma amostra $x_1, \ldots, x_n$, é calculado como a probabilidade de observarmos um valor mais extremo que $\overline{x}$ para $\overline{X}$, onde a distribuição de X é calculada assumindo $\mathcal{H}_0$ como verdadeira.

Portanto, mais geralmente, calculamos o p-valor das seguintes formas:

- Teste unilateral $\begin{cases} \mathcal{H}_0: & \mu \geq \mu_0 \\ \mathcal{H}_1: & \mu < \mu_0 \end{cases} \implies \tilde{\alpha} = \mathbb{P}(\overline{X} < \overline{x} \mid \mu = \mu_0)$
- Teste unilateral $\begin{cases} \mathcal{H}_0: & \mu \leq \mu_0 \\ \mathcal{H}_1: & \mu > \mu_0 \end{cases} \implies \tilde{\alpha} = \mathbb{P}(\overline{X} > \overline{x} \mid \mu = \mu_0)$
- Teste bilateral $\begin{cases} \mathcal{H}_0: & \mu = \mu_0 \\ \mathcal{H}_1: & \mu \neq \mu_0 \end{cases} \implies \tilde{\alpha} = \mathbb{P}(|\overline{X} - \mu_0| > |\overline{x} - \mu_0| \mid \mu = \mu_0)$

Exemplo

Relembremos o exemplo do desempenho de um automóvel na estrada: A montadora afirma que o desempenho é de 15km/ℓ, porém ao testar 27 amostras desses automóveis, observa-se um desempenho médio amostral de 14,3km/ℓ. Lembremos que o desvio padrão populacional é conhecido e igual a 3km/ℓ. Qual o p-valor desse teste?

Solução: Como o teste é bilateral, o p-valor será a probabilidade, sob $\mathcal{H}_0$, de observarmos $\overline{X}$ distando mais que $|14,3 - 15| = 0,7$ de $\mu_0 = 15$. Mais precisamente:

$$\begin{aligned}\tilde{\alpha} &= \mathbb{P}(|\overline{X} - \mu_0| > 0,7 \mid \mu = \mu_0) = \mathbb{P}(|\overline{X} - 15| > 0,7) \\ &= \mathbb{P}\left(|Z| > \frac{0,7}{3/\sqrt{27}}\right) = \mathbb{P}(|Z| > 1,212) = 2[1 - \mathbb{P}(Z < 1,212)] = 0,2256.\end{aligned}$$

Portanto, qualquer nível de significância abaixo de 22,56% implica em não rejeitar $\mathcal{H}_0$.

Teste para proporções

- Considere agora que o atributo de interesse de uma população pode ser modelado por $X \sim Ber(p)$, onde p, a proporção populacional dos elementos que possuem tal atributo, é desconhecida. Queremos testar hipóteses sobre os possíveis valores de p a partir de uma amostra aleatória dessa população
- Seja $X_1, \ldots, X_n$ uma amostra de tamanho n e $\overline{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a proporção amostral (estimador de p)
- TCL $\implies \dfrac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$, se n é suficientemente grande $(np(1-p) \geq 3)$
- Temos aqui também três testes de interesse:

$$\begin{cases} \mathcal{H}_0: & p \geq p_0 \\ \mathcal{H}_1: & p < p_0 \end{cases}, \qquad \begin{cases} \mathcal{H}_0: & p \leq p_0 \\ \mathcal{H}_1: & p > p_0 \end{cases} \quad \text{e} \quad \begin{cases} \mathcal{H}_0: & p = p_0 \\ \mathcal{H}_1: & p \neq p_0 \end{cases}$$

- Analogamente ao teste para a média amostral, no caso unilateral consideraremos a hipótese nula como $\mathcal{H}_0 : p = p_0$. Notemos que nesse caso, a distribuição da proporção amostral torna-se fácil de se trabalhar com, pois teremos que $\overline{p} \approx N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$
- O p-valor é calculado como anteriormente, e valem as mesmas interpretações

Exemplo

Seja p a proporção de produtos que atendem às especificações de qualidade de uma fábrica. Foi selecionada uma amostra aleatória de tamanho 40 da linha de produção, verificando-se que 32 atendem às especificações. Desejamos testar as hipóteses $\mathcal{H}_0 : p \geq 0,85$ contra $\mathcal{H}_1 : p < 0,85$.

a) Qual decisão deve ser tomada ao nível de significância de 5%? Por que?

b) Qual o p-valor desse teste? Interprete seu resultado.

c) Qual a probabilidade de se cometer um erro tipo II no teste obtido no item a) se a proporção populacional de peças que atendem às especificações de qualidade for $p = 0,75$?

Solução:

a) Rejeitamos $\mathcal{H}_0$ ao observarmos um valor suficientemente baixo para $\overline{p}$. Sob $\mathcal{H}_0$, temos que $Z = \dfrac{\overline{p} - 0,85}{\sqrt{\frac{(0,85)(0,15)}{40}}} \approx N(0,1)$, de modo que:

$$0,05 = \mathbb{P}(\overline{p} < p_c \mid p = 0,85) = \mathbb{P}\left(Z < \frac{p_c - 0,85}{\sqrt{\frac{(0,85)(0,15)}{40}}}\right)$$

$$\implies \frac{p_c - 0,85}{\sqrt{\frac{(0,85)(0,15)}{40}}} \approx z_{0,05} = -z_{0,95} = -1,65 \implies p_c \approx 0,757.$$

Como observamos $\overline{p}_{obs} = 0,8 > 0,757$, não temos evidências para rejeitar $\mathcal{H}_0$.

Exemplo

Seja p a proporção de produtos que atendem às especificações de qualidade de uma fábrica. Foi selecionada uma amostra aleatória de tamanho 40 da linha de produção, verificando-se que 32 atendem às especificações. Desejamos testar as hipóteses $\mathcal{H}_0 : p \geq 0,85$ contra $\mathcal{H}_1 : p < 0,85$.

a) Qual decisão deve ser tomada ao nível de significância de 5%? Por que?

b) Qual o p-valor desse teste? Interprete seu resultado.

c) Qual a probabilidade de se cometer um erro tipo II no teste obtido no item a) se a proporção populacional de peças que atendem às especificações de qualidade for $p = 0,75$?

Solução:

b) O p-valor é calculado da seguinte forma:

$$\tilde{\alpha} = \mathbb{P}(\overline{p} < 0,8 \mid p = 0,85) = \mathbb{P}\left(Z < \frac{0,8 - 0,85}{\sqrt{\frac{(0,85)(0,15)}{40}}}\right) = \mathbb{P}(Z < -0,89) \approx 0,1867.$$

Portanto, qualquer nível de significância abaixo de 18,67% nos levará anão rejeitar $\mathcal{H}_0$.

Exemplo

Seja p a proporção de produtos que atendem às especificações de qualidade de uma fábrica. Foi selecionada uma amostra aleatória de tamanho 40 da linha de produção, verificando-se que 32 atendem às especificações. Desejamos testar as hipóteses $\mathcal{H}_0 : p \geq 0,85$ contra $\mathcal{H}_1 : p < 0,85$.

a) Qual decisão deve ser tomada ao nível de significância de 5%? Por que?

b) Qual o p-valor desse teste? Interprete seu resultado.

c) Qual a probabilidade de se cometer um erro tipo II no teste obtido no item a) se a proporção populacional de peças que atendem às especificações de qualidade for $p = 0,75$?

Solução:

c) Como estamos assumindo que $p = 0,75$, note que $Z = \dfrac{\overline{p} - 0,75}{\sqrt{\frac{(0,75)(0,25)}{40}}} \approx N(0,1)$, de modo que:

$$\beta(0,75) = \mathbb{P}(\overline{p} \geq 0,757 \mid p = 0,75) = \mathbb{P}\left(Z \geq \frac{0,757 - 0,75}{\sqrt{\frac{(0,75)(0,25)}{40}}}\right) = \mathbb{P}(Z \geq 0,10) \approx 0,4602.$$

Portanto, se $p = 0,75$, temos uma chance de 46,02% de aceitarmos, erroneamente, a hipótese nula.