Variáveis aleatórias

Prof. Eduardo Vargas Ferreira

Curso de Especialização em
Data Science & Big Data
Universidade Federal do Paraná

26 de março de 2018

Tipos de Variáveis

Qualitativa

- Nominal: não existe ordenação nas possíveis categorias de resposta (ex: sexo, estado civil);
- Ordinal: existe uma certa ordem nas possíveis categorias de resposta (ex: escolaridade);

Quantitativa

- Discreta: os possíveis valores formam um conjunto finito ou enumerável de números, são variáveis de contagem (ex: número de filhos);
- Contínua: os possíveis valores pertencem à um intervalo, aberto ou fechado, dos números reais (ex: peso de um indivíduo).

Variável aleatória

- A variável aleatória X atribui um número a cada resultado:

$$X : \Omega \to \mathbb{R}$$

- A Função de probabilidade de X é dada por:

$$p(a) = P(X = a)$$

- Já a função de probabilidade acumulada é dada por:

$$F(a) = P(X \leq a)$$

Exemplo

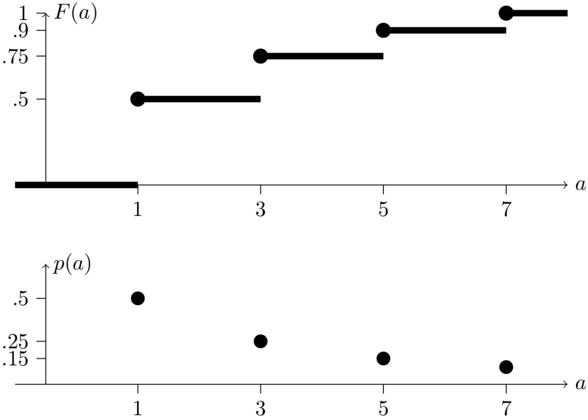
- Suponha que X seja uma variável aleatória de acordo com a tabela abaixo:

Valor de X:	1	3	5	7
$p(a)$	0,5	0,25	0,15	0,10
$F(a)$	0,5	0,75	0,90	1

- $F(3) = 0,75,\ F(5) = 0,90,\ F(6) = 0,9,\ F(0) = 0,\ F(8) = 1.$

- **Propriedades de** $F(a)$**:**
  - Não decrescente;
  - Quando $a \to -\infty$, F é 0;
  - Quando $a \to \infty$, F é 1.

Exemplo



Variável aleatória contínua

- Valores contínuos significa, p. ex.,:

$$[0, 1], [a, b], [0, \infty), (-\infty, \infty).$$

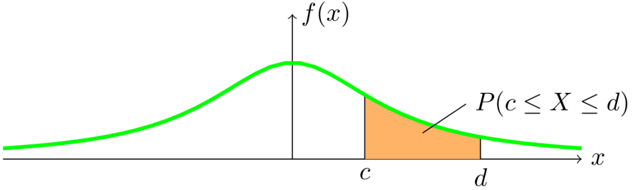
- Função densidade de probabilidade (fdp):

$$f(x) \geq 0; \;\; P(c \leq X \leq d) = \int_c^d f(x)dx.$$

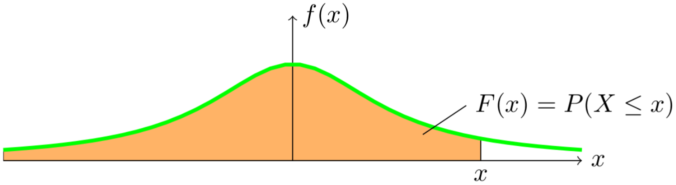
- Função de distribuição acumulada (fda):

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Variável aleatória contínua



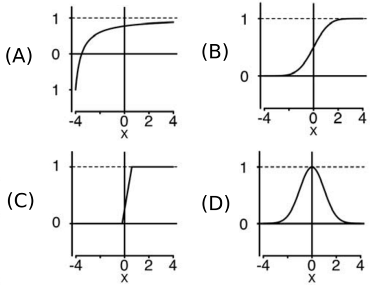
pdf and probability



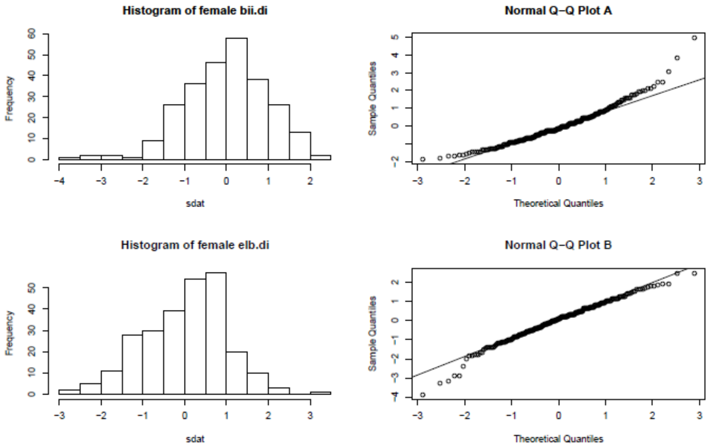
Propriedades da fda:

- $F(x) = P(X \leq x)$;
- $0 \leq F(x) \leq 1$;
- Não decrescente;
- $\lim_{x \to -\infty} F(x) = 0$;
- $\lim_{x \to \infty} F(x) = 1$;
- $P(c < X \leq d) = F(d) - F(c)$;
- $F'(x) = f(x)$.

Quais gráficos representam uma fda?



Teste de normalidade



Exercício

1. Suponha que X pertença ao intervalo $[0, 2]$, com $f(x) = kx^2$.

   (a) Qual o valor de k?

   (b) Calcule a $F(x)$;

   (c) Calcule $P(1 \leq X \leq 2)$.

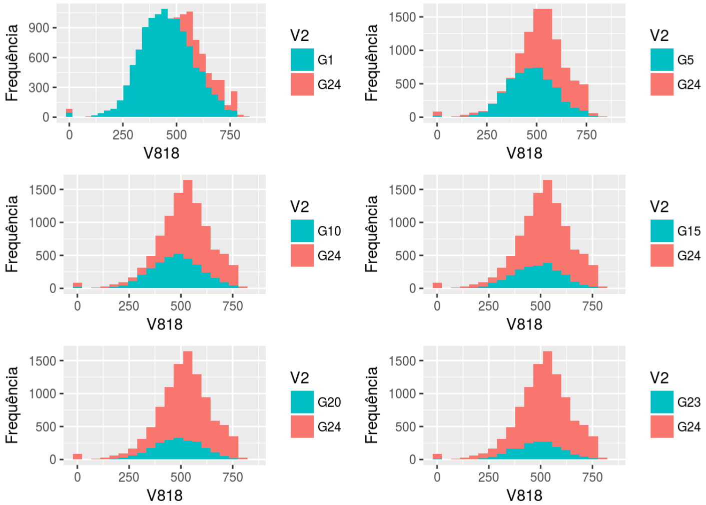
   Resposta:

   (a) $\int_0^2 f(x)dx = \int_0^2 kx^2 dx = k\frac{8}{3} = 1 \Rightarrow k = \frac{3}{8}$.

   (b) $F(x) = \int_0^x ku^2 du = \frac{k}{3}x^3 = \frac{x^3}{8}$, se $x \in [0, 2]$.

   (c) $P(1 \leq X \leq 2) = F(2) - F(1) = 1 - \frac{1}{8} = \frac{7}{8}$.

Esperança de uma variável aleatória



Entendendo o valor esperado

- Suponha que lançamos um dado 5 vezes e obtivemos $(3, 1, 6, 1, 2)$;
- Para encontrar a média, somamos os valores e dividimos por 5:

$$\text{Média} = \frac{3+1+6+1+2}{5} = 2,6$$

- Entretanto, com poucas jogadas, tais resultados não são representativos;
- Agora, suponha que lancemos os dados 6000 vezes. Teríamos resultados próximos do que se encontra abaixo:

Valor de X	1	2	3	4	5	6
Frequência	1000	1000	1000	1000	1000	1000
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

$$\begin{aligned}\text{Média} &= \frac{1000\times 1 + 1000\times 2 + 1000\times 3 + 1000\times 4 + 1000\times 5 + 1000\times 6}{6000} \\ &= 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = 3,5\end{aligned}$$

Esperança de uma variável aleatória

- O valor esperado de X, para o caso discreto, é definido por:

$$E(X) = p(x_1)x_1 + p(x_2)x_2 + \ldots + p(x_n)x_n = \sum_{i=1}^{n} p(x_i)x_i.$$

- Propriedade de $E(X)$:
  - $E(X + Y) = E(X) + E(Y)$;
  - $E(aX + b) = aE(X) + b$;
  - $E[h(X)] = \sum_i h(x_i)p(x_i)$.

Exemplo de classe

- Considere a variável aleatória X com a seguinte distribuição:

X	-2	-1	0	1	2
$P(X = x)$	1/5	1/5	1/5	1/5	1/5

- Calcule $E(X)$.

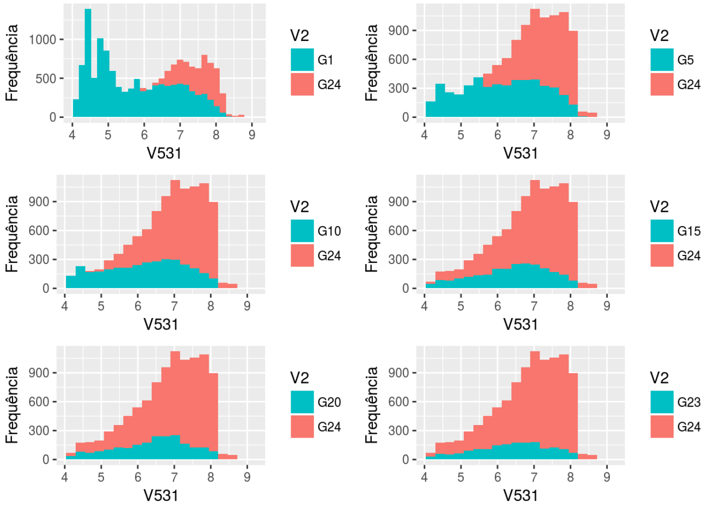
$$E(X) = (-2) \times \frac{1}{5} + (-1) \times \frac{1}{5} + 0 \times \frac{1}{5} + 1 \times \frac{1}{5} + 2 \times \frac{1}{5} = 0$$

- Calcule $E(X^2)$.

X^2	0	1	4
$P(X = x)$	1/5	2/5	2/5

$$\begin{aligned} E(X^2) &= 0 \times \frac{1}{5} + 1 \times \frac{2}{5} + 4 \times \frac{2}{5} = 2 \\ &= \sum_i h(x_i)p(x_i) \end{aligned}$$

Variância e desvio padrão



Variância e desvio padrão

- Seja X uma variável aleatória com média $E(X) = \mu$;
- A variância é a propagação da massa de probabilidade em torno da média;
- Definição como esperança (soma ponderada):

$$Var(X) = E[(X - \mu)^2].$$

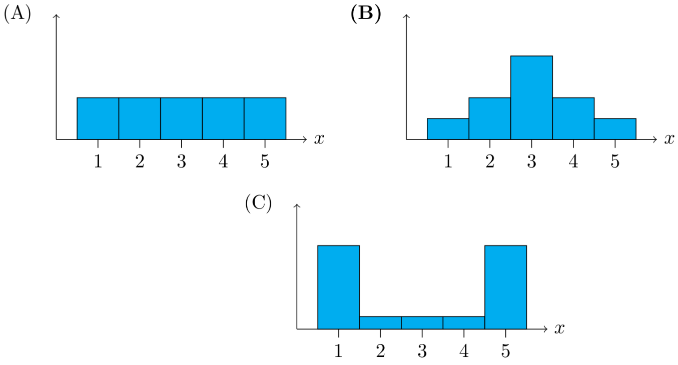
- Cálculo da soma:

$$Var(X) = \sum_{i=1}^{n} p(x_i)(x_i - \mu)^2.$$

- Desvio padrão: $\sigma = \sqrt{Var(X)}$.

Questão conceitual

- Abaixo temos a função de probabilidade para 3 variáveis aleatórias.



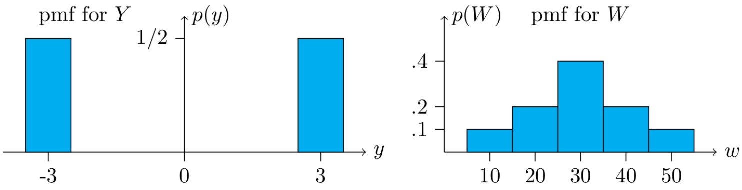
- Ordene as figuras segundo a variância (da maior para a menor).

Resposta

- Todas as 3 variáveis têm o mesmo intervalo de 1-5;
- E todas são simétricas, com média em 3;
- O gráfico (C) tem a maior parte do seu peso nos extremos, por isso tem maior variabilidade;
- O gráfico (B) tem o maior peso no centro, por isso tem a menor variância;
- Onde essa questão tem relação com “outliers”?

Questão conceitual

- Qual das funções de probabilidade apresenta maior variância?



- Faça a tabela de probabilidade para Y e W e calcule a variância.

Resposta

- $E(Y) = 0,5 \times (-3) + 0,5 \times (3) = 0;$
- $E[(Y - \mu)^2] = 0,5 \times (9) + 0,5 \times (9) = 9;$

Y	-3	3
$P(Y = y)$	0,5	0,5
$(Y - \mu)^2$	9	9

- $E(W) = 1 + 4 + 12 + 8 + 5 = 30;$
- $E[(W - \mu)^2] = .1(400) + .2(100) + .4(0) + .2(100) + .1(100) = 120;$

W	10	20	30	40	50
$p(w)$	0,1	0,2	0,4	0,2	0,1
$(W - \mu)^2$	400	100	0	100	400

- **Note que a escala é importante para a variação.**

Referências

- Some of the exercises and figures in this presentation are taken from MIT OpenCourseWare;