

ESTATÍSTICA DESCRITIVA

p-Quantil

Lista de exercícios

1 Medidas Separatrizes

O quantil q de ordem p , ou p -quantil (mediana, quartil, decil, centil, ...) é o valor de uma série ordenada, definido como:

$$q(p) = \begin{cases} x_{(i)} & \text{se } p = p_i = \frac{i - 0,5}{n}, \quad i = 1, \dots, n \\ (1 - f_i)q(p_i) + f_i q(p_{i+1}) & \text{se } p_i < p < p_{i+1} \\ x_{(1)} & \text{se } p < p_1 \\ x_{(n)} & \text{se } p > p_n \end{cases}$$

onde $f_i = \frac{p - p_i}{p_{i+1} - p_i}$.

Para uma distribuição aproximadamente simétrica as seguintes relações são válidas:

- $q(0, 50) - x_{(1)} \simeq x_{(n)} - q(0, 50)$ (dispersão inferior e superior, respectivamente);
- $q(0, 50) - q(0, 25) \simeq q(0, 75) - q(0, 25)$;
- $q(0, 25) - x_{(1)} \simeq x_{(n)} - q(0, 75)$

EXEMPLO

Considere a amostra $X = (15; 5; 3; 8; 10; 2; 7; 11; 12)$. Determine os percentis 10%, 20%, 50% e 75%

- Entrada dos valores
> $x = c(15, 5, 3, 8, 10, 2, 7, 11, 12)$
- Ordenando a série:
> $(x = \text{sort}(x))$
[1] 2 3 5 7 8 10 11 12 15
- Calculando a frequência relativa acumulada:
> $(p.ac = \text{round}(((1:9) - 0.5)/9, 2))$
[1] 0.06 0.17 0.28 0.39 0.50 0.61 0.72 0.83 0.94

a) **Caso 01:** Calculando o percentil 10% ($p = 0,1$). O resultado dá o valor que separa os 10% elementos inferiores dos 90% superiores. Como $p_1 < 0,1 < p_2$

- ```
> (f.1 = (0.1 - p.ac[1]) / (p.ac[2] - p.ac[1]))
[1] 0.3636364
> (q.p = (1 - f.1) * x[1] + f.1 * x[2])
[1] 2.363636
```

Assim, 10% dos valores estão abaixo de 2.4.

- b) **Caso 02:** Calculando o percentil 10% ( $p = 0,2$ ). O resultado dá o valor que separa os 20% elementos inferiores dos 80% superiores. Como  $p_2 < 0,20 < p_3$

```
> (f.2 = (0.2 - p.ac[2])/(p.ac[3] - p.ac[2]))
```

```
[1] 0.2727273
```

```
> (q.p = (1 - f.2) * x[2] + f.2 * x[3])
```

```
[1] 3.545455
```

Assim, 20% dos valores estão abaixo de 3.5.

- c) **Caso 03:** Calculando o percentil 50% ( $p = 0,5$ ). O resultado dá o valor que separa os 50% elementos inferiores dos 50% superiores. Como  $p = p_5$  então o valor coincide com  $x_5$ .

```
> (q.p = x[5])
```

```
[1] 8
```

Assim, 50% dos valores estão abaixo de 8.

- d) **Caso 04:** Calculando o percentil 75% ( $p = 0,75$ ). O resultado dá o valor que separa os 75% elementos inferiores dos 25% superiores. Como  $p_7 < 0,75 < p_8$

```
> (f.4 = (0.75 - p.ac[7])/(p.ac[8] - p.ac[7]))
```

```
[1] 0.2727273
```

```
> (q.p = (1 - f.4) * x[7] + f.4 * x[8])
```

```
[1] 11.27273
```

Assim, 75% dos valores estão abaixo de 11.3.

## EXERCÍCIOS

- 1.1. Experimente as funções `mean()`, `var()`, `sd()`, `median()`, `quantile()` nos dados `women` disponível no banco de dados do R. Veja a documentação das funções e suas opções de uso.
- 1.2. Carregue o conjunto de dados `USArrests` com o comando `data(USArrests)`. Examine sua documentação com `help(USArrests)` e construa um script que responda as seguintes perguntas:
  - a) Qual o número médio e mediano de cada um dos crimes?
  - b) Encontre os quartis para cada crime e interprete os resultados.
  - c) Encontre o número máximo e mínimo para cada crime.
  - d) Faça um gráfico adequado para o número de assassinatos (*murder*).
  - e) Faça um diagrama ramo-e-folhas para o número de estupros (*rape*) e discuta o resultado.
- 1.3. Para facilitar um projeto de ampliação da rede de esgoto de uma certa região de uma cidade, as autoridades tomaram uma amostra de tamanho 50 dos 270 quarteirões que compõem a região e foram encontrados os seguintes número de casas por quarteirão:

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 2  | 2  | 3  | 10 | 13 | 14 | 15 | 15 | 16 | 16 |
| 18 | 18 | 20 | 21 | 22 | 22 | 23 | 24 | 25 | 25 |
| 26 | 27 | 29 | 29 | 30 | 32 | 36 | 42 | 44 | 45 |
| 45 | 46 | 48 | 52 | 58 | 59 | 61 | 61 | 61 | 65 |
| 66 | 66 | 68 | 75 | 78 | 80 | 89 | 90 | 92 | 97 |

- a) Identifique o tipo de variável;
- b) Determine as medidas de tendência central (média, mediana e moda) para os dados isolados (sem agrupamento em classes);
- c) Determine as medidas de dispersão (variância, desvio padrão e coeficiente de variação);
- d) Construa um histograma e um gráfico de ramo-e-folhas.

- e) Verifique se a distribuição é simétrica utilizando medidas do  $p$ -quantil.  
 f) Construa um box-plot. O resultado é compatível com sua conclusão no item anterior?
- 1.4. As taxas médias geométricas de incremento anual (por 100 habitantes) dos 30 maiores municípios do Brasil são:

|      |      |      |      |      |
|------|------|------|------|------|
| 3,67 | 1,82 | 3,73 | 4,10 | 4,30 |
| 1,28 | 8,14 | 2,43 | 4,17 | 5,36 |
| 3,96 | 6,54 | 5,84 | 7,35 | 3,63 |
| 2,93 | 2,82 | 8,45 | 5,28 | 5,41 |
| 7,77 | 4,65 | 1,88 | 2,12 | 4,26 |
| 2,78 | 5,54 | 0,90 | 5,09 | 4,07 |

- a) Identifique o tipo de variável;  
 b) Determine as medidas de tendência central (média, mediana e moda) para os dados isolados (sem agrupamento em classes);  
 c) Determine as medidas de dispersão (variância, desvio padrão e coeficiente de variação);  
 d) Construa um histograma e um gráfico de ramo-e-folhas.  
 e) Verifique se a distribuição é simétrica utilizando medidas do  $p$ -quantil.  
 f) Construa um box-plot. O resultado é compatível com sua conclusão no item anterior?
- 1.5. Em uma granja foi observada a distribuição dos frangos em relação ao peso, cujo resultado é dado na tabela a seguir:

| Peso (gramas) |    |       | $n_i$ |
|---------------|----|-------|-------|
| 960           | +— | 980   | 60    |
| 980           | +— | 1.000 | 160   |
| 1.000         | +— | 1.020 | 280   |
| 1.020         | +— | 1.040 | 260   |
| 1.040         | +— | 1.060 | 160   |
| 1.060         | +— | 1.080 | 80    |

- a) Qual a média da distribuição?  
 b) Qual a variância da distribuição?  
 c) Queremos dividir os frangos em quatro categorias:
- os 20% mais leves na categoria D;
  - os 30% seguintes na categoria C;
  - os 30% seguintes na categoria B;
  - os 20% mais pesados na categoria A.
- Quais os limites de peso de classificação?