Medidas resumo

Tiago M. Magalhães

Departamento de Estatística - ICE-UFJF

Juiz de Fora, 02 de abril de 2024

Roteiro

1. Motivação
2. Medidas de posição
3. Medidas de dispersão
4. Outras medidas
5. Bibliografia

Roteiro

Exemplo

Márcio é aluno da disciplina de Estatística Avançada I, com prova marcada para às 8h, sem a possibilidade de entrada após este horário.

Às 7h42, ele chega ao Terminal Butantã e os dois ônibus das linhas que lhe serve, 8012 e 8022, sairão às 7h44.

Exemplo

Márcio é aluno da disciplina de Estatística Avançada I, com prova marcada para às 8h, sem a possibilidade de entrada após este horário.

Às 7h42, ele chega ao Terminal Butantã e os dois ônibus das linhas que lhe serve, 8012 e 8022, sairão às 7h44.

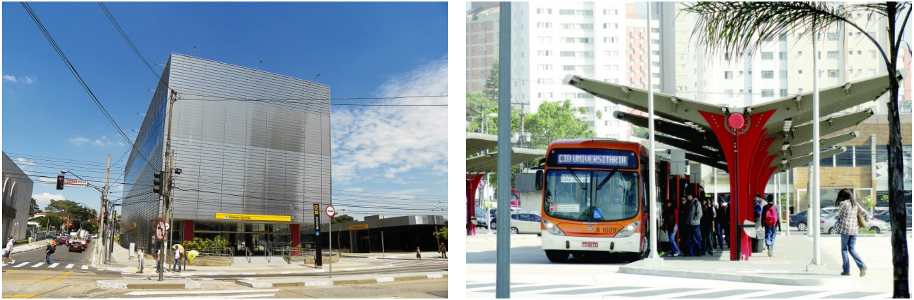

Figura 1: Terminal Butantã, São Paulo/SP (Créditos: `Google imagens`).

Exemplo

Desesperado, ele pergunta para o fiscal quanto tempo cada um deles leva até chegar ao Departamento de Estatística. O fiscal responde: "nos últimos 10 dias, neste mesmo horário, eu observei, em minutos:"

Exemplo

Desesperado, ele pergunta para o fiscal quanto tempo cada um deles leva até chegar ao Departamento de Estatística. O fiscal responde: "nos últimos 10 dias, neste mesmo horário, eu observei, em minutos:"

Tabela 1: Tempo (min) do Terminal Butantã ao Dept. Estatística - USP.

8012	13	15	12	10	19	17	10	14	15	15
8022	16	16	15	16	16	14	15	15	14	15

Exemplo

Desesperado, ele pergunta para o fiscal quanto tempo cada um deles leva até chegar ao Departamento de Estatística. O fiscal responde: "nos últimos 10 dias, neste mesmo horário, eu observei, em minutos:"

Tabela 1: Tempo (min) do Terminal Butantã ao Dept. Estatística - USP.

8012	13	15	12	10	19	17	10	14	15	15
8022	16	16	15	16	16	14	15	15	14	15

Exemplo

Pergunta

Qual dos ônibus Márcio deveria pegar?

Roteiro

Medidas de posição

Média ($\bar{X}$)

A média aritmética, ou simplesmente média, é definida como:

Medidas de posição

Média ($\bar{X}$)

A média aritmética, ou simplesmente média, é definida como:

$$\bar{X} = \frac{\text{Soma de todos os valores}}{\text{Número de observações}}.$$

Medidas de posição

Média ($\bar{X}$)

A média aritmética, ou simplesmente média, é definida como:

$$\bar{X} = \frac{\text{Soma de todos os valores}}{\text{Número de observações}}.$$

Medidas de posição

Para o exemplo, dados na Tabela 1:

$$\bar{X}_{8012} = \frac{13 + 15 + 12 + 10 + 19 + 17 + 10 + 14 + 15 + 15}{10}$$
$$= \frac{140}{10} = 14 \text{ minutos.}$$

Medidas de posição

Para o exemplo, dados na Tabela 1:

$$\begin{aligned}\bar{X}_{8012} &= \frac{13+15+12+10+19+17+10+14+15+15}{10} \\ &= \frac{140}{10} = 14 \text{ minutos.}\end{aligned}$$

Interpretação: Nós podemos interpretar que, se todos os ônibus da linha 8012 levassem exatamente o mesmo tempo para chegar ao Departamento de Estatística, esse tempo seria de 14 minutos.

Medidas de posição

Para o exemplo, dados na Tabela 1:

$$
\begin{aligned}
\bar{X}_{8012} &= \frac{13 + 15 + 12 + 10 + 19 + 17 + 10 + 14 + 15 + 15}{10} \\
&= \frac{140}{10} = 14 \text{ minutos.}
\end{aligned}
$$

Interpretação: Nós podemos interpretar que, se todos os ônibus da linha 8012 levassem exatamente o mesmo tempo para chegar ao Departamento de Estatística, esse tempo seria de 14 minutos.

Medidas de posição

Propriedades da média

1. Se nós adicionarmos a **todos os elementos** de um conjunto **um mesmo valor**, a nova média será a média anterior somada deste valor.

Medidas de posição

Propriedades da média

1. Se nós adicionarmos a **todos os elementos** de um conjunto **um mesmo valor**, a nova média será a média anterior somada deste valor.
2. Se nós multiplicarmos **todos os elementos** de um conjunto por **um mesmo valor**, a nova média será a média anterior multiplicada por este valor.

Medidas de posição

Propriedades da média

1. Se nós adicionarmos a **todos os elementos** de um conjunto **um mesmo valor**, a nova média será a média anterior somada deste valor.
2. Se nós multiplicarmos **todos os elementos** de um conjunto por **um mesmo valor**, a nova média será a média anterior multiplicada por este valor.

Medidas de posição

Por exemplo, suponham que o fiscal percebeu que

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012.

Medidas de posição

Por exemplo, suponham que o fiscal percebeu que

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012.

Então, a nova média seria $14 + 1 = 15$ minutos.

Medidas de posição

Por exemplo, suponham que o fiscal percebeu que

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012.

Então, a nova média seria $14 + 1 = 15$ minutos.

(2) houve erro na marcação do tempo da linha 8012, e os tempos precisariam ser multiplicados por 0,95 (redução de 0,05 nos tempos observados).

Medidas de posição

Por exemplo, suponham que o fiscal percebeu que

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012. Então, a nova média seria $14 + 1 = 15$ minutos.

(2) houve erro na marcação do tempo da linha 8012, e os tempos precisariam ser multiplicados por 0,95 (redução de 0,05 nos tempos observados). Então, a nova média seria $14 \times 0{,}95 = 13{,}3$ minutos.

Medidas de posição

Por exemplo, suponham que o fiscal percebeu que

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012. Então, a nova média seria $14 + 1 = 15$ minutos.

(2) houve erro na marcação do tempo da linha 8012, e os tempos precisariam ser multiplicados por 0,95 (redução de 0,05 nos tempos observados). Então, a nova média seria $14 \times 0{,}95 = 13{,}3$ minutos.

Medidas de posição

Mediana (Md)

É o valor que divide um conjunto de dados exatamente ao meio,

Medidas de posição

Mediana (Md)

É o valor que divide um conjunto de dados exatamente ao meio, isto é, 50% das observações estão à esquerda da mediana e os outros 50% à sua direita.

Medidas de posição

Mediana (Md)

É o valor que divide um conjunto de dados exatamente ao meio, isto é, 50% das observações estão à esquerda da mediana e os outros 50% à sua direita.

Mediana

Para o exemplo, dados na Tabela 1, os tempos **ordenados** para a linha 8012 são:

$$10, 10, 12, 13, 14, \underset{\xleftarrow{50\%}}{} \quad Md \quad \underset{\xrightarrow{50\%}}{}, 15, 15, 15, 17, 19.$$

Mediana

Para o exemplo, dados na Tabela 1, os tempos **ordenados** para a linha 8012 são:

$$10, 10, 12, 13, 14, \overset{50\%}{\longleftarrow} \quad \text{Md} \quad \overset{50\%}{\longrightarrow}, 15, 15, 15, 17, 19.$$

No caso, a mediana é **14,5 minutos**, a média entre 14 e 15.

Mediana

Para o exemplo, dados na Tabela 1, os tempos **ordenados** para a linha 8012 são:

$$10, 10, 12, 13, \underbrace{14,}_{\xleftarrow{50\%}} \quad \text{Md} \quad \underbrace{, 15,}_{\xrightarrow{50\%}} 15, 15, 17, 19.$$

No caso, a mediana é **14,5 minutos**, a média entre 14 e 15.

Medidas de posição

Moda (Mo)

É o valor mais frequente de um conjunto de observações.

Medidas de posição

Moda (Mo)

É o valor mais frequente de um conjunto de observações.

Para o exemplo, dados na Tabela 1, a moda é 15 minutos para linha 8012.

Medidas de posição

Moda (Mo)

É o valor mais frequente de um conjunto de observações.

Para o exemplo, dados na Tabela 1, a moda é 15 minutos para linha 8012.

Pois é o tempo que mais se repetiu.

Medidas de posição

Moda (Mo)

É o valor mais frequente de um conjunto de observações.

Para o exemplo, dados na Tabela 1, a moda é 15 minutos para linha 8012.

Pois é o tempo que mais se repetiu.

Medidas de posição

Quartis

Os quartis Q1, Q2 e Q3 são os valores que dividem um conjunto de observações em quatro partes.

Medidas de posição

Quartis

Os quartis Q1, Q2 e Q3 são os valores que dividem um conjunto de observações em quatro partes.

- Q1: 25% das observações são menores que ele e 75% maiores.

Medidas de posição

Quartis

Os quartis Q1, Q2 e Q3 são os valores que dividem um conjunto de observações em quatro partes.

- Q1: 25% das observações são menores que ele e 75% maiores.
- Q2: 50% das observações são menores que ele e 50% maiores.

Medidas de posição

Quartis

Os quartis Q1, Q2 e Q3 são os valores que dividem um conjunto de observações em quatro partes.

- Q1: 25% das observações são menores que ele e 75% maiores.
- Q2: 50% das observações são menores que ele e 50% maiores.
- Q3: 75% das observações são menores que ele e 25% maiores.

Medidas de posição

Quartis

Os quartis Q1, Q2 e Q3 são os valores que dividem um conjunto de observações em quatro partes.

- Q1: 25% das observações são menores que ele e 75% maiores.
- Q2: 50% das observações são menores que ele e 50% maiores.
- Q3: 75% das observações são menores que ele e 25% maiores.

Medidas resumo

Com o conhecimento de medidas de posição, o fiscal obteve:

Tabela 2: Medidas de posição para o tempo (min).

Linha	Min	Q1	Md	$\bar{X}$	Moda	Q3	Max
8012	10	12,3	14,5	14,0	15	15	19
8022	14	15,0	15,0	15,2	15 e 16	16	16

Medidas resumo

Com o conhecimento de medidas de posição, o fiscal obteve:

Tabela 2: Medidas de posição para o tempo (min).

Linha	Min	Q1	Md	$\bar{X}$	Moda	Q3	Max
8012	10	12,3	14,5	14,0	15	15	19
8022	14	15,0	15,0	15,2	15 e 16	16	16

Pergunta

Agora, com as informações da Tabela 2, qual dos ônibus Márcio deveria pegar?

Roteiro

Medidas de dispersão

Variância (S^2)

Mede o grau de dispersão dos valores em torno da média.

Medidas de dispersão

Variância (S^2)

Mede o grau de dispersão dos valores em torno da média. Quanto menor a variância, maior a homogeneidade.

Medidas de dispersão

Variância (S^2)

Mede o grau de dispersão dos valores em torno da média. Quanto menor a variância, maior a homogeneidade.

Em outras palavras, a variância serve para medir o quanto a média representa bem ou não os dados.

Medidas de dispersão

Variância (S^2)

Mede o grau de dispersão dos valores em torno da média. Quanto menor a variância, maior a homogeneidade.

Em outras palavras, a variância serve para medir o quanto a média representa bem ou não os dados.

Medidas de dispersão

A variância pode ser calculada da seguinte forma:

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}.$$

Notem que, a unidade de medida da variância é o quadrado da unidade de medida das observações.

Medidas de dispersão

A variância pode ser calculada da seguinte forma:

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}.$$

Notem que, a unidade de medida da variância é o quadrado da unidade de medida das observações.

Medidas de dispersão

Para o exemplo, dados na Tabela 1:

$$S^2_{8012} = \frac{(13^2 + 15^2 + \cdots + 15^2) - 10 \times 14^2}{10 - 1}$$

$$= \frac{2.034 - 1.960}{9} = \frac{74}{9} = 8{,}2 \text{ minutos}^2.$$

Medidas de dispersão

Para o exemplo, dados na Tabela 1:

$$S^2_{8012} = \frac{\left(13^2 + 15^2 + \cdots + 15^2\right) - 10 \times 14^2}{10 - 1}$$

$$= \frac{2.034 - 1.960}{9} = \frac{74}{9} = 8{,}2 \text{ minutos}^2.$$

Medidas de dispersão

Propriedades da variância

1. Se todos os valores de um conjunto de dados são iguais, a variância é zero. Pois não há variabilidade;

Medidas de dispersão

Propriedades da variância

1. Se todos os valores de um conjunto de dados são iguais, a variância é zero. Pois não há variabilidade;
2. Se nós adicionarmos a **todos os elementos** de um conjunto **um mesmo valor**, a variância **não se altera**.

Medidas de dispersão

Propriedades da variância

1. Se todos os valores de um conjunto de dados são iguais, a variância é zero. Pois não há variabilidade;
2. Se nós adicionarmos a **todos os elementos** de um conjunto **um mesmo valor**, a variância **não se altera**.
3. Se nós multiplicarmos **todos os elementos** de um conjunto por **um mesmo valor**, a nova variância será a variância anterior multiplicada pelo **quadrado deste valor**.

Medidas de dispersão

Propriedades da variância

1. Se todos os valores de um conjunto de dados são iguais, a variância é zero. Pois não há variabilidade;
2. Se nós adicionarmos a **todos os elementos** de um conjunto **um mesmo valor**, a variância **não se altera**.
3. Se nós multiplicarmos **todos os elementos** de um conjunto por **um mesmo valor**, a nova variância será a variância anterior multiplicada pelo **quadrado deste valor**.

Medidas de dispersão

Por exemplo, sabendo que variância do tempo da linha 8012 é 8,2 minutos^2, suponham que o fiscal percebeu que:

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012. Então, a nova variância não se alteraria, isto é, ela continuaria sendo 8,2 minutos^2.

Medidas de dispersão

Por exemplo, sabendo que variância do tempo da linha 8012 é 8,2 minutos^2, suponham que o fiscal percebeu que:

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012. Então, a nova variância não se alteraria, isto é, ela continuaria sendo 8,2 minutos^2.

(2) houve erro na marcação do tempo da linha 8012, e os tempos precisariam ser multiplicados por 0,95.

Medidas de dispersão

Por exemplo, sabendo que variância do tempo da linha 8012 é 8,2 minutos^2, suponham que o fiscal percebeu que:

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012. Então, a nova variância não se alteraria, isto é, ela continuaria sendo 8,2 minutos^2.

(2) houve erro na marcação do tempo da linha 8012, e os tempos precisariam ser multiplicados por 0,95. Então, a nova variância seria $8{,}2 \times 0,95^2 = 7{,}4$ minutos^2.

Medidas de dispersão

Por exemplo, sabendo que variância do tempo da linha 8012 é 8,2 minutos^2, suponham que o fiscal percebeu que:

(1) deveria adicionar 1 minuto em cada tempo marcado para a linha 8012. Então, a nova variância não se alteraria, isto é, ela continuaria sendo 8,2 minutos^2.

(2) houve erro na marcação do tempo da linha 8012, e os tempos precisariam ser multiplicados por 0,95. Então, a nova variância seria $8{,}2 \times 0,95^2 = 7{,}4$ minutos^2.

Medidas de dispersão

Desvio padrão (S)

É a raiz quadrada da variância, isto é, $S = \sqrt{S^2}$.

Medidas de dispersão

Desvio padrão (S)

É a raiz quadrada da variância, isto é, $S = \sqrt{S^2}$.

O desvio padrão é uma medida de dispersão na **mesma escala** dos dados.

Medidas de dispersão

Desvio padrão (S)

É a raiz quadrada da variância, isto é, $S = \sqrt{S^2}$.

O desvio padrão é uma medida de dispersão na **mesma escala** dos dados.

Medidas de dispersão

Coeficiente de variação (CV)

É o quociente entre o desvio padrão e a média, em porcentagem,

Medidas de dispersão

Coeficiente de variação (CV)

É o quociente entre o desvio padrão e a média, em porcentagem,

$$\text{CV} = \frac{S}{\bar{X}} \times 100.$$

Medidas de dispersão

Coeficiente de variação (CV)

É o quociente entre o desvio padrão e a média, em porcentagem,

$$\text{CV} = \frac{S}{\bar{X}} \times 100.$$

O CV é uma medida de dispersão **sem escalas**.

Medidas de dispersão

Coeficiente de variação (CV)

É o quociente entre o desvio padrão e a média, em porcentagem,

$$\mathrm{CV} = \frac{S}{\bar{X}} \times 100.$$

O CV é uma medida de dispersão **sem escalas**. O CV pode ser utilizado para comparar conjunto de observações com escalas diferentes.

Medidas de dispersão

Coeficiente de variação (CV)

É o quociente entre o desvio padrão e a média, em porcentagem,

$$\text{CV} = \frac{S}{\bar{X}} \times 100.$$

O CV é uma medida de dispersão **sem escalas**. O CV pode ser utilizado para comparar conjunto de observações com escalas diferentes.

Medidas resumo

Com o conhecimento de medidas de dispersão, o fiscal obteve:

Tabela 3: Medidas de posição e dispersão para o tempo (min).

Linha	Min	Q1	Md	$\bar{X}$	Moda	Q3	Max	S	CV
8012	10	12,3	14,5	14,0	15	15	19	2,9	20,5
8022	14	15,0	15,0	15,2	15 e 16	16	16	0,8	5,2

Medidas resumo

Com o conhecimento de medidas de dispersão, o fiscal obteve:

Tabela 3: Medidas de posição e dispersão para o tempo (min).

Linha	Min	Q1	Md	$\bar{X}$	Moda	Q3	Max	S	CV
8012	10	12,3	14,5	14,0	15	15	19	2,9	20,5
8022	14	15,0	15,0	15,2	15 e 16	16	16	0,8	5,2

Pergunta

Finalmente, com as informações da Tabela 3, qual dos ônibus Márcio deveria pegar?

Roteiro

Média móvel

As médias móveis são úteis para suavizar uma serie de observações e observar uma possível tendência. Seja $X_1, X_2, \ldots, X_n$ um conjunto de observações. A i-ésima média móvel, $\bar{X}_i$, de k observações é dada por:

Média móvel

As médias móveis são úteis para suavizar uma serie de observações e observar uma possível tendência. Seja $X_1, X_2, \ldots, X_n$ um conjunto de observações. A i-ésima média móvel, $\bar{X}_i$, de k observações é dada por:

$$\bar{X}_i = \frac{X_i + X_{i+1} + \cdots + X_{i+k-1}}{k}.$$

Média móvel

As médias móveis são úteis para suavizar uma serie de observações e observar uma possível tendência. Seja $X_1, X_2, \ldots, X_n$ um conjunto de observações. A i-ésima média móvel, $\bar{X}_i$, de k observações é dada por:

$$\bar{X}_i = \frac{X_i + X_{i+1} + \cdots + X_{i+k-1}}{k}.$$

Covariância

A covariância é uma forma de mediar a variabilidade conjunta entre dois conjuntos de dados. Sejam $X_1, X_2, \ldots, X_n$ e $Y_1, Y_2, \ldots, Y_n$.

Covariância

A covariância é uma forma de mediar a variabilidade conjunta entre dois conjuntos de dados. Sejam $X_1, X_2, \ldots, X_n$ e $Y_1, Y_2, \ldots, Y_n$. A covariância amostral é definida da seguinte forma:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}.$$

Covariância

A covariância é uma forma de mediar a variabilidade conjunta entre dois conjuntos de dados. Sejam $X_1, X_2, \ldots, X_n$ e $Y_1, Y_2, \ldots, Y_n$. A covariância amostral é definida da seguinte forma:

$$\mathrm{Cov}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}.$$

Correlação

A correlação mede a relação linear entre duas variáveis, sendo ela definida da seguinte forma:

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Correlação

A correlação mede a relação linear entre duas variáveis, sendo ela definida da seguinte forma:

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Com $r(X, Y)$ estando entre -1 e 1.

Correlação

A correlação mede a relação linear entre duas variáveis, sendo ela definida da seguinte forma:

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Com $r(X, Y)$ estando entre -1 e 1.

Autocorrelação

A autocorrelação mede a relação entre as observações dentro de um mesmo conjunto. É uma medida de influência entre as variáveis. Seja X_t um conjunto de observações e k a distância (*lag*) entre duas observações,

Autocorrelação

A autocorrelação mede a relação entre as observações dentro de um mesmo conjunto. É uma medida de influência entre as variáveis. Seja X_t um conjunto de observações e k a distância (*lag*) entre duas observações, a autocorrelação entre X_t e X_{t+k} é dada por:

$$r(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)}.$$

Autocorrelação

A autocorrelação mede a relação entre as observações dentro de um mesmo conjunto. É uma medida de influência entre as variáveis. Seja X_t um conjunto de observações e k a distância (*lag*) entre duas observações, a autocorrelação entre X_t e X_{t+k} é dada por:

$$r(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)}.$$

Roteiro

Bibliografia

Anderson, D. R., D. J. Sweeney, T. A. Williams, J. D. Camm, and J. J. Cochran (2019). *Estatística Aplicada à Administração e Economia* (8th ed.). São Paulo: CENGAGE Learning.

Doane, D. P. and L. E. Seward (2014). *Estatística Aplicada à Administração e Economia* (4th ed.). Porto Alegre: McGraw-Hill.

Martins, G. A. and O. Domingues (2019). *Estatística Geral e Aplicada* (6th ed.). São Paulo: Atlas.

Obrigado!

✉ tiago.magalhaes@ufjf.br

🖱 ufjf.br/tiago_magalhaes

🌍 Departamento de Estatística, Sala 319