

Estatística



Sumário

Introdução.....	2
Por que By Casseb?	2
Como funciona a metodologia didática By Casseb?	2
Conceitos.....	3
O que é estatística?	3
População.....	3
Amostra.....	3
Estatística Descritiva	3
Estatística Inferencial	3
Variáveis Qualitativas	4
Variáveis nominais	4
Variáveis ordinais	4
Variáveis Quantitativas	4
Variáveis discretas	4
Variáveis contínuas	4
Medidas de Tendência Central	4
Média (\bar{x})	4
Moda (M_o)	4
Mediana (M_d).....	4
Medidas de Variação	5
Amplitude total (AT)	5
Desvios (d_i).....	5
Variância (σ^2 ou s^2).....	5
Desvio padrão	5
Coeficiente de Variação.....	5
Exemplo com Conjuntos Não Agrupados	6
Conjunto original.....	6
Primeira etapa será ordenar o conjunto:	6
1 - Ordenando o conjunto	6
2 – Média	6
3 – Moda	6
4 – Mediana	6
5 - Calculando Amplitude Total	7

6 – Desvios	7
7 – Variância	7
8 – Desvio Padrão	8
8 – Coeficiente de Variação	8
Exemplo com Conjuntos Agrupados	9
1 – Ordenando o conjunto	9
2 – Quantidade de classes	10
3 – Amplitude das classes	10
4 – Distribuição das frequências	11
5 – Representante da classe	11
6 – Moda	11
7 – Média	12
8 – Frequência cumulativa ou acumulativa	12
9 – Mediana	12
10 – Desvio padrão	13
11 – Coeficiente de Variação	14

Introdução

Estatística é uma matéria muito importante quando se trata do levantamento e análise de dados, sendo possível tomar decisões baseadas na variação salarial de um grupo ou quantidade de alunos em uma escola.

Nesta apostila vamos tratar sobre os principais pontos e exemplos de cálculos estatísticos para um aprendizado e aplicação deste ensinamento.

Por que By Casseb?

By Casseb é todo material criado diretamente por Felipe Casseb, fundador da empresa by Casseb, de forma totalmente independente, para fins não-comerciais baseado em seus estudos e conhecimentos na área aplicada.

Como funciona a metodologia didática By Casseb?

Neste documento você terá acesso ao passo a passo relacionado a linguagens, frameworks e boas práticas, sendo declarado quais são os pré-requisitos necessários para o entendimento de cada ferramenta. Não há uma ordem específica para o uso do mesmo, você poderá estudar somente um capítulo, contanto que já atenda aos pré-requisitos do mesmo, sem os pré-requisitos devidamente estudados, não posso garantir um completo entendimento do conteúdo.

Todo capítulo será dividido no máximo em 4 partes:

1. Pré-requisitos – O que você precisa estudar ou comprar antes de iniciar o estudo para aplicar o conteúdo.
2. Implantação – Passo a passo do que precisa ser instalado e configurado para aplicar o conteúdo.
3. Utilização – Explicações sobre o que é, para que serve e como usar cada parte da ferramenta.
4. Práticas – Exemplos práticos utilizando a ferramenta para solução.

Lembrando que não é necessário estudar todas as 4 partes, caso você já tenha domínio da ferramenta e só não sabe implantá-la, pode somente estudar a parte 2 e desenvolver.

Nem sempre 1 capítulo terá as 4 partes, alguns capítulos são somente implantações objetivando capítulos posteriores.

Conceitos

O que é estatística?

Estatística é a parte da ciência responsável pela coleta, organização e interpretação de dados experimentais e pela extrapolação dos resultados da amostra para a população.

População

Conjunto qualquer relacionado a todos os elementos que deseja aplicar a estatística, ou seja, se surge a necessidade de analisar estatisticamente os funcionários de uma empresa, uma pesquisa com todos os funcionários seria um levantamento da população.

Amostra

Conjunto que representa um todo, ou seja, se for necessário analisar um país, é levantado as informações somente de uma parcela representativa do mesmo, e baseado nela é avaliado o todo.

Estatística Descritiva

É aplicado envolvendo a representação dos dados utilizando ferramentas como gráficos, cálculos de moda, mediana e variância.

Estatística Inferencial

Leva em conta a probabilidade, analisando um grupo pequeno visando um grupo maior de dados.

Variáveis Qualitativas

Representam categorias de dados como gênero, escolaridade, datas, etc...

Variáveis nominais

Sem ordenação entre si, exemplo (Gênero, cor dos olhos, doente/sadio).

Variáveis ordinais

Com ordenação, exemplo (escolaridade, estágio de doença (avançado, intermediário, leve)).

Variáveis Quantitativas

Representam valores representadas por números.

Variáveis discretas

Adquiridas por contagens, podem somente ser apresentadas por inteiros, Exemplo (número de filhos, número de integrantes).

Variáveis contínuas

Adquiridas por medições, podem ser representadas por números decimais, Exemplo (altura, pressão arterial).

Medidas de Tendência Central

Serve para representar o conjunto de dados

Média (\bar{x})

Média padrão mais utilizada no cotidiano, onde é somado todos os elementos e depois dividido pela quantidade de elementos.

Observação: Uma média não é confiável pois pode ser afetada pelos seus extremos, falhando em precisão.

Moda (M_o)

Elemento ou elementos que mais se repetem na amostra ou população.

Mediana (M_d)

Representa a metade do conjunto de valores, caso a quantidade for ímpar será o elemento do meio, caso par, será a média dos dois elementos do meio.

Medidas de Variação

Amplitude total (AT)

Representa diferença do maior valor com o menor.

Desvios (d_i)

Diferença de cada valor com a média do conjunto.

$$d_i = x_i - \bar{x}$$

Variância (σ^2 ou ς^2)

Utilizado quando o elemento analisado possui duas dimensões, ex: m².

$$\sigma^2 = \frac{\sum d_i^2}{n} \text{ (Populacional)} \quad \varsigma^2 = \frac{\sum d_i^2}{n-1} \text{ (Amostral)}$$

Desvio padrão

Utilizado quando o elemento possui uma dimensão.

$$\sigma = \sqrt{\sigma^2} \text{ (Populacional)} \quad \varsigma = \sqrt{\varsigma^2} \text{ (Amostral)}$$

Coefficiente de Variação

Representa a dispersão dos elementos em relação com a média.

$$cv = \frac{\sigma}{\bar{x}} * 100 \text{ (Populacional)} \quad cv = \frac{\varsigma}{\bar{x}} * 100 \text{ (Amostral)}$$

Exemplo com Conjuntos Não Agrupados

Conjunto original

15	13	20
29	66	49
2	83	66

Primeira etapa será ordenar o conjunto:

1 - Ordenando o conjunto

2	13	15	20	29	49	66	66	83
---	----	----	----	----	----	----	----	----

2 – Média

$$\bar{x} = \frac{2 + 13 + 15 + 20 + 29 + 49 + 66 + 66 + 83}{9}$$

$$\bar{x} = \frac{343}{9}$$

$$\bar{x} = 38,11$$

3 – Moda

A definição de moda é o elemento que mais se repete, neste caso 66 é o número que mais se repete, logo:

$$Mo = 66$$

4 – Mediana

É a metade do conjunto, já que a quantidade de elementos é ímpar, o elemento exatamente do meio é a Mediana, neste caso o 29.

$$Md = 29$$

5 - Calculando Amplitude Total

Uma vez ordenado sabemos que o menor elemento é 2 e o maior elemento é 83. Logo:

$$At = \text{ElementoMaior} - \text{ElementoMenor}$$

$$At = 83 - 2 = 81$$

Ou seja, a amplitude total deste conjunto é 81.

6 – Desvios

Para calcular os desvios devemos aplicar elemento a elemento:

2	$2 - 38,11 = -36,11$
13	$13 - 38,11 = -25,11$
15	$15 - 38,11 = -23,11$
20	$20 - 38,11 = -18,11$
29	$29 - 38,11 = -9,11$
49	$49 - 38,11 = 10,89$
66	$66 - 38,11 = 27,89$
66	$66 - 38,11 = 27,89$
83	$83 - 38,11 = 44,89$

7 – Variância

O cálculo da variância relaciona a soma dos elementos dos desvios elevados ao quadrado dividido pela quantidade de elementos.

-36,11	$-36,11^2 = 1303,93$
-25,11	$-25,11^2 = 630,51$
-23,11	$-23,11^2 = 534,07$
-18,11	$-18,11^2 = 327,97$
-9,11	$-9,11^2 = 82,99$
10,89	$10,89^2 = 118,59$
27,89	$27,89^2 = 777,85$
27,89	$27,89^2 = 777,85$
44,89	$44,89^2 = 2015,11$
Soma	6568,88

Com a soma em mãos basta dividir pela quantidade de elementos (populacional). logo:

$$\sigma^2 = \frac{6568,88}{9} = 729,87 \text{ (Populacional)} \quad \varsigma^2 = \frac{6568,88}{8} = 821,11 \text{ (Amostral)}$$

8 – Desvio Padrão

$$\sigma = \sqrt{729,97^2} = 27,01 \text{ (Populacional)} \quad \varsigma = \sqrt{821,11^2} = 28,65 \text{ (Amostral)}$$

8 – Coeficiente de Variação

$$cv = \frac{27,01}{38,11} * 100 = 70,87\% \text{ (Populacional)}$$

$$cv = \frac{28,65}{38,11} * 100 = 75,17\% \text{ (Amostral)}$$

Exemplo com Conjuntos Agrupados

Quando se trata de uma grande quantidade de dados, podemos agrupar estes dados em classes e analisar as classes separadamente, isto facilita a análise do mesmo.

Abaixo um conjunto de dados que iremos analisar:

206	175	166	198	211	176
196	191	191	208	211	200
199	189	155	206	176	162
177	178	151	166	204	182
188	162	175	180	200	156
156	198	168	203	174	185
173	180	168	151	212	184
196	196	210	199	165	161
153	211	168	211	153	188
165	200	175	177	190	204

1 – Ordenando o conjunto

Primeiro precisamos ordenar este conjunto, você pode fazer isto de diversas formas, uma prática é a que irei apresentar a seguir:

Crie uma tabela com todas as variações dos dois primeiros dígitos ordenados, desta forma:

15
16
17
18
19
20
21

Após isso, preencha esta tabela com todas as variações de cada raio definido, ficando desta forma:

15	1	3	5	6		
16	1	2	5	6	8	
17	3	4	5	6	7	8
18	0	2	4	5	8	9
19	0	1	6	8	9	
20	0	3	4	6	8	
21	0	1	2			

Com esta tabela em mãos basta preencher ordenadamente os conjuntos procurando quantas vezes cada número repetido, chegando neste resultado:

151	151	153	155	155	155
156	156	161	162	162	165
165	166	166	168	168	173
174	175	175	175	176	176
177	177	178	180	180	182
184	185	188	188	189	190
191	191	196	196	196	198
198	199	199	200	200	203
204	204	206	206	208	201
210	211	211	211	211	212

2 – Quantidade de classes

Para localizar a quantidade de classes necessárias para representar este conjunto basta calcular $K = \sqrt{n}$ sendo n a quantidade de números no elemento.

$$n = 60$$

$$K = \sqrt{60}$$

$$K = 7,74$$

Já que o estamos definindo quantidade de classes, a quantidade necessária é 8.

3 – Amplitude das classes

A amplitude das classes é o “tamanho” que cada classe terá, ou seja, se a amplitude da classe for 10, então uma classe possui um máximo de 10 variações. Abaixo formula para seu cálculo:

$$h = \frac{x_f - x_i}{k}$$

$$h = \frac{212 - 151}{5}$$

$$h = 12,2 \sim 13$$

A quantidade de classe correta é 8, mas para ficar mais didático separamos em 5 classes. Desta forma sabemos que vamos usar 5 classes com valores que variam em 13 possibilidades.

4 – Distribuição das frequências

A distribuição é realizada informando a quantidade de cada classe utilizando o padrão abaixo:

i	Classe	Fi
1	151 – 164	11
2	164 – 177	13
3	177 – 190	11
4	190 – 203	12
5	203 – 216	13

Nesta tabela podemos ver que entre 151 e 163 há 11 elementos.

5 – Representante da classe

É a média de cada classe, ficando desta forma:

i	Classe	Fi	xi
1	151 – 164	11	157,5
2	164 – 177	13	170,5
3	177 – 190	11	183,5
4	190 – 203	12	196,5
5	203 – 216	13	209,5
Σ		60	

6 – Moda

É a classe, ou classes, que possuem mais elementos, neste exemplo é a classe 2 e 5 que possuem 13 elementos. Deve ser respondido pela sua respectiva representante:

$Mo = 170,5 \text{ e } 209,5$ (Bimodal)

7 – Média

É necessário calcular a soma da multiplicação entre f_i e x_i dividido pela soma de f_i , abaixo a tabela com estes dados:

i	Classe	F_i	x_i	$x_i * f_i$
1	151 – 164	11	157,5	1732,50
2	164 – 177	13	170,5	2216,50
3	177 – 190	11	183,5	2018,50
4	190 – 203	12	196,5	2358,00
5	203 – 216	13	209,5	2723,50
Σ		60		11049,00

Executando este cálculo:

$$\bar{x} = \frac{\sum x_i * f_i}{\sum f_i}$$

$$\bar{x} = \frac{11049}{60}$$

$$\bar{x} = 184,15$$

8 – Frequência cumulativa ou acumulativa

É a frequência que vai se acumulando, para preenche-la basta ir somando todas as f_i de elemento a elemento, como mostro na tabela abaixo:

i	Classe	F_i	x_i	$x_i * f_i$	f_c
1	151 – 164	11	157,5	1732,50	11
2	164 – 177	13	170,5	2216,50	24
3	177 – 190	11	183,5	2018,50	35
4	190 – 203	12	196,5	2358,00	47
5	203 – 216	13	209,5	2723,50	60
Σ		60		11049,00	

9 – Mediana

Nesta etapa iremos localizar a metade deste conjunto, para isso primeiro devemos localizar a classe mediana, para isso calculamos:

$$\frac{\sum f_i}{2} = \frac{60}{2} = 30$$

Este 30 mostra que a classe mediana é a 3, pois analisando a f_c vimos que 30 está entre a classe 2 ($f_c = 24$) e classe 3 ($f_c = 35$), portanto usaremos a classe 3.

Com esta informação poderemos usar a seguinte fórmula:

$$Md = l + \frac{\left[\frac{\sum fi}{2} - fc^{(ant)} \right] * h}{fi}$$

Sendo l o menor elemento da classe (no nosso exemplo é 177), $fc^{(ant)}$ é o fc da classe anterior (no nosso exemplo é 24) e h sendo a amplitude da classe (no nosso caso 13).

$$Md = l + \frac{\left[\frac{\sum fi}{2} - fc^{(ant)} \right] * h}{fi}$$

$$Md = 177 + \frac{\left[\frac{60}{2} - 24 \right] * 13}{11}$$

$$Md = 177 + 7,09$$

$$Md = 184,09$$

10 – Desvio padrão

Deve ser calculado a subtração da representando com a média elevado ao quadrado vezes a frequência de cada classe, ficando desta forma:

i	Classe	Fi	xi	xi*fi	fc	$(xi - \bar{x})^2 * fi$
1	151 - 164	11	157,5	1732,50	11	7812,45
2	164 - 177	13	170,5	2216,50	24	2602,89
3	177 - 190	11	183,5	2018,50	35	4,65
4	190 - 203	12	196,5	2358,00	47	1830,27
5	203 - 216	13	209,5	2723,50	60	8354,09
\sum		60		11049,00		20604,35

Utilizando este resultado na seguinte fórmula:

$$\sigma = \sqrt{\frac{[\sum (xi - \bar{x})^2 * fi]}{\sum fi}} \text{ (Populacional)} \quad \varsigma = \sqrt{\frac{[\sum (xi - \bar{x})^2 * fi]}{\sum fi}} \text{ (Amostral)}$$

$$\sigma = \sqrt{\frac{[20604,35]}{60}} \text{ (Populacional)} \quad \varsigma = \sqrt{\frac{20604,35}{59}} \text{ (Amostral)}$$

$$\sigma = 18,53 \text{ (Populacional)} \quad \varsigma = 18,69 \text{ (Amostral)}$$

11 – Coeficiente de Variação

$$cv = \frac{\sigma}{\bar{x}} * 100 = \frac{18,53}{184,15} * 100 = 10,06\% \text{ (Populacional)}$$

$$cv = \frac{s}{\bar{x}} * 100 = \frac{18,69}{184,15} * 100 = 10,15\% \text{ (Amostral)}$$