Estatística Descritiva

Medidas de tendência central

Uma distribuição pode ser convenientemente descrita por meio de algumas medidas numéricas obtidas sobre ela. Estas medidas são chamadas de medidas de tendência central, pois tendem a descrever o centro da distribuição. A primeira é a

Média aritmética:

Para dados isolados, a média aritmética indicada por $\bar{x}$. Sua fórmula é

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

O símbolo $\sum_{i=1}^{n} x_i$ lê-se como somatório de todos os x_i quando i varia entre 1 e n. Nada mais é do que a soma de todos os números. Ele usa a letra grega $\sum$ sigma que corresponde ao nosso "S" e indica soma. n é o tamanho da amostra.

Se os dados estiverem agrupados em classes, ainda pode ser usado o processo acima, mas este vai dar um trabalhão se os dados tiverem que ser recuperados como eram antes da formação das classes. Mas se tal recuperação for impraticável, a fórmula a ser usada é

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i.n_i}{\sum_{i=1}^{n} n_i}$$

Aqui, x_i representa o ponto médio de cada classe $(superior + inferior \div 2)$ e este valor deve ser multiplicado por n_i que é a quantidade de elementos naquela classe

Mediana:

A mediana de um conjunto de valores é o valor que pode ser do próprio conjunto de valores, ou ao contrário ser um valor teórico que tem antes e depois de si exatamente a metade do conjunto de valores. Normalmente é usado um valor do próprio conjunto.

Vale notar que é pré-requisito para obter a mediana ter os dados ordenados segundo o critério a examinar. Para dados isolados, se o número de elementos considerados for ímpar, a obtenção do elemento central é imediata. Se o número for par a mediana deve ser calculada pela média aritmética dos dois elementos centrais. Neste caso o valor encontrado não fará parte da amostra original. O símbolo da mediana é M_d.

Para dados agrupados em classes, deve-se criar uma nova coluna na qual as quantidades de elementos por classe vão sendo acumulados, como em

X cm	n_i	
140-150	8	8
150-160	7	15
160-170	21	36 ←
170-180	13	49
180-190	5	54
190-200	3	57
	57	=

O lugar onde a mediana deve aparecer, neste caso, é $57 \div 2 = 28,5$ e vê-se que está na terceira classe. agora a fórmula é

$$M_d = l_{M_d} + \left(\frac{\frac{\sum n_i}{2} - n_{aM_d}}{n_{M_d}} \right).h_{M_d}$$

sendo

l_{M_d} limite inferior da classe mediana (no nosso exemplo, 160)

n_{aM_d} n_i acumulado anterior ao l_{M_d} (no caso, 15)

n_{M_d} n_i da própria classe mediana (no caso 21)

h_{M_d} tamanho da classe mediana (no caso 10)

Portanto, para o exemplo a mediana é $M_d = 160 + \left(\frac{\frac{57}{2} - 15}{21} \right).10 = 166.42cm$

Moda:

A moda, representada por M_o de um conjunto de dados é o valor que aparece mais vezes, ou seja aquele ao qual esteja associado a frequência absoluta mais alta.

Para dados não agrupados, não há fórmula, basta olhar os dados e ver qual o mais frequente.

Para dados agrupados em classes a fórmula, devida a Czuber, é

$$M_o = l_{M_o} + \left(\frac{n_{aM_o} - n_{M_o}}{n_{aM_o} + n_{pM_o} - 2n_{M_o}} \right).h_{M_o}$$

onde

l_{M_o} limite inferior da classe de maior frequência absoluta (no nosso caso, 160)

n_{aM_o} n_i anterior ao maior n_i (no caso, 7)

n_{M_o} n_i maior n_i da tabela (no caso 21)

n_{pM_o} n_i posterior ao maior n_i (no caso, 13)

h_{M_o} tamanho da classe modal (no caso 10)

e fica no exemplo $M_o = 160 + \left(\frac{7-21}{7+13-2.21} \right).10 = 166.36cm$

Medidas de Variabilidade

Descreve-se aqui algumas medidas que indicam as oscilações de uma variável, daí o nome. Como vimos acima a média aritmética indica onde as medidas tendem a se agrupar, mas a pergunta agora é quanto elas variam em torno àquele valor.

Desvio padrão:

É uma quantidade que mede a amplitude da variação das medidas em torno da média aritmética. É indicado por $s(x)$, que se lê *s de x*.

O processo de calcular o desvio padrão passa pelas seguintes etapas

1. Subtrair de cada valor a média aritmética do conjunto
2. Elevar o item anterior ao quadrado (objetivo é desprezar o sentido da subtração, isto é não importa se a medida foi maior ou menor do que a média)
3. Somar os quadrados obtidos anteriormente
4. Dividir esta soma pelo número de parcelas somadas
5. Extrair a raiz quadrada positiva do resultado

O desvio padrão indica qual o valor médio em que a medida varia em relação a média da população.

Para dados agrupados, o desvio padrão tem a seguinte fórmula

$$s(x) = +\sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2.n_i}{\sum_{i=1}^{n} .n_i}}$$

onde x_i vão ser os pontos médios de cada classe e as diferenças DEPOIS de elevadas ao quadrado é que deverão ser multiplicadas pelos respectivos n_i.

Variância:

Indicado por $s^2(x)$, que se lê *s-quadrado de x*, é exatamente o quadrado do desvio padrão. Seu processo de cálculo é igual ao do desvio padrão deixando-se de fazer a última operação.

Para os dados agrupados a variância segue a fórmula vista acima para o desvio padrão de dados agrupados, mas sem extrair a última raiz.

Exemplos

Considere a seguinte distribuição isolada: $X_i = \{3.1, 7.4, 1.3, 4.7, 1, 6.8, 7.2, 4.4, 3.9, 4.4, 4.4\}$
1. Calcule a média aritmética: 4.4
2. Calcule a mediana: 4.4
3. Calcule a moda: 4.4
4. Calcule o desvio padrão: 2
5. Calcule a variancia: 4.2

Considere a seguinte distribuição agrupada:

L_{inf}	L_{sup}	n_i
27	36	6
36	45	30
45	54	29
54	63	55
63	72	34
72	81	68
81	90	50
90	99	35
99	108	31
108	117	4

6. Calcule a média: 72.5
7. Calcule a mediana: 74.3
8. Calcule a moda: 77.9
9. Calcule o desvio padrão: 19.5
10. Calcule a variância: 381.4

☞ Para você fazer

Considere a seguinte distribuição isolada: $X_i = \{9.9, 1.5, 6, 6, 6.3, 8.3, 9.8, 6, 8.9, 5, 1.5, 6.9\}$
1. Calcule a média aritmética
2. Calcule a mediana
3. Calcule a moda (se não houver: -1 e se empate: o menor)
4. Calcule o desvio padrão
5. Calcule a variancia

Use apenas 1 decimal, e escreva a resposta no local adequado

Considere a seguinte distribuição agrupada:

L_{inf}	L_{sup}	n_i
2.1	2.9	15
2.9	3.7	57
3.7	4.5	41
4.5	5.3	52
5.3	6.1	93
6.1	6.9	55
6.9	7.7	103
7.7	8.5	43
8.5	9.3	42

6. Calcule a média
7. Calcule a mediana
8. Calcule a moda
9. Calcule o desvio padrão
10. Calcule a variância

Use apenas 1 decimal, e escreva a resposta no local adequado.

Responda aqui:

1	2	3	4	5
6	7	8	9	10