

1. A tabela abaixo representa os salários dos funcionários de uma empresa de Informática:

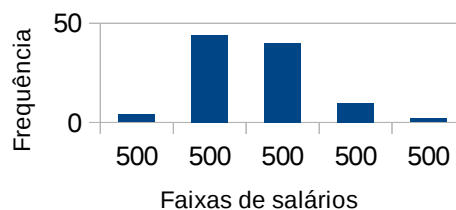
| Classe | Salário (R\$) |            | Número de funcionários |     |          |           |              |          |
|--------|---------------|------------|------------------------|-----|----------|-----------|--------------|----------|
|        | Limite inf    | Limite sup |                        |     |          |           |              |          |
| 1      | 500           | 1000       | 4                      | 4   | 750,00   | 3.000,00  | 656.100,00   | 2624400  |
| 2      | 1000          | 1500       | 44                     | 48  | 1.250,00 | 55.000,00 | 96.100,00    | 4228400  |
| 3      | 1500          | 2000       | 40                     | 88  | 1.750,00 | 70.000,00 | 36.100,00    | 1444000  |
| 4      | 2000          | 2500       | 10                     | 98  | 2.250,00 | 22.500,00 | 476.100,00   | 4761000  |
| 5      | 2500          | 3000       | 2                      | 100 | 2.750,00 | 5.500,00  | 1.416.100,00 | 2832200  |
|        |               |            | <b>100</b>             |     |          |           | 156.000,00   | 15890000 |

|                     |                |               |     |
|---------------------|----------------|---------------|-----|
| média               | R\$ 1.560,00   |               | 0,5 |
| mediana             | R\$ 1.525,00   |               | 0,5 |
| moda                | R\$ 1.454,55   |               | 0,5 |
| variância           | R\$ 158.900,00 | R\$ 26.805,00 | 0,5 |
| desvio padrão       | R\$ 398,62     | R\$ 163,72    | 0,5 |
| Zona de normalidade | R\$ 1.161,38   | R\$ 1.396,28  | 0,5 |
|                     | R\$ 1.958,62   | R\$ 1.723,72  | 0,5 |

- a) Qual é o valor da **média** dos salários dos funcionários?
- a) R\$ 1.560,00      b) R\$ 1.161,38      c) R\$ 1.454,55      d) R\$ 1.958,62      e) R\$ 1.525,00
- b) Qual é o valor da **mediana** dos salários dos funcionários?
- a) R\$ 1.560,00      b) R\$ 1.161,38      c) R\$ 1.454,55      d) R\$ 1.958,62      e) R\$ 1.525,00
- c) Qual é o valor da **moda** dos salários dos funcionários?
- a) R\$ 1.560,00      b) R\$ 1.161,38      c) R\$ 1.454,55      d) R\$ 1.958,62      e) R\$ 1.525,00
- d) Qual é o valor do **variância** dos salários dos funcionários?
- a) R\$ 158.900,00      b) R\$ 156.215,00      c) R\$ 165.677,00      d) R\$ 126.125,00      e) R\$ 159.546,00
- e) Qual é o valor do **desvio padrão da média** dos salários dos funcionários?
- a) R\$ 344,15      b) R\$ 371,04      c) R\$ 398,62      d) R\$ 468,24      e) R\$ 376,00
- f) Quais são os valores (**superior e inferior**) da **primeira Zona de Normalidade**?
- a) R\$ 1.560,00      b) R\$ 1.161,38      c) R\$ 1.454,55      d) R\$ 1.958,62      e) R\$ 1.525,00

g) Fazer histograma.

Histograma



2. O número de pacotes enviados e recebidos, denotada por  $X_S$  e  $X_R$  respectivamente, por diferentes estações em uma rede local foram medidos. Os números observados são mostrados na segunda e terceira linha da Tabela abaixo. Qual das variáveis tem uma maior variação? Faça um esboço do gráfico de dispersão dos dados. Pode ser visto a partir desta figura, se há uma correlação significativa entre as duas variáveis? Determine o coeficiente de correlação sabendo que o desvio padrão de  $X_S$  e  $X_R$  são respectivamente,  $\sigma_S=1.692$  e  $\sigma_R=1.341$

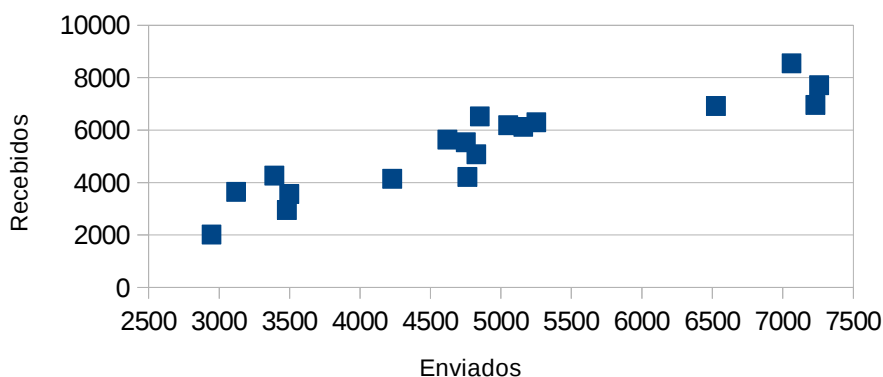
|                   | 1       | 2       | 3       | 4       | 5      | 6      | 7      | 8      | 9     | 10     | 11     | 12     | 13     | 14      | 15      | 16      | 17      | 18      | TOTAL      |
|-------------------|---------|---------|---------|---------|--------|--------|--------|--------|-------|--------|--------|--------|--------|---------|---------|---------|---------|---------|------------|
| pacotes enviados  | 7718    | 6958    | 8551    | 6924    | 6298   | 6120   | 6184   | 6527   | 5081  | 4216   | 5532   | 5638   | 4147   | 3562    | 2955    | 4261    | 3644    | 2020    | 96.336     |
| pacotes recebidos | 7258    | 7232    | 7062    | 6526    | 5251   | 5158   | 5051   | 4850   | 4825  | 4762   | 4750   | 4620   | 4229   | 3497    | 3480    | 3392    | 3120    | 2946    | 88.009     |
|                   | 2366    | 1606    | 3199    | 1572    | 946    | 768    | 832    | 1175   | -271  | -1136  | 180    | 286    | -1205  | -1790   | -2397   | -1091   | -1708   | -3332   |            |
|                   | 2369    | 2343    | 2173    | 1637    | 362    | 269    | 162    | -39    | -64   | -127   | -139   | -269   | -660   | -1392   | -1409   | -1497   | -1769   | -1943   |            |
|                   | 5604134 | 3762233 | 6950183 | 2572753 | 342084 | 206293 | 134460 | -46282 | 17449 | 144714 | -25090 | -77045 | 795769 | 2492376 | 3378305 | 1633651 | 3022116 | 6475372 | 37.383.476 |

**Médias** 5352 4889  
**Dp's** 1692 1341  
**Coeficiente de variação** 0,32 0,27 **1,0**  
**Covariância** 2076860  
**Coeficiente de correção** 0,92 **1,0**

$X_S$

**1,0**

Gráfico de Dispersão (Recebidos X emitidos)



3. O tempo de CPU e disco I/O's de sete programas são apresentados na Tabela 1. Qual deles tem maior variabilidade (coeficiente de variação). Faça um gráfico de dispersão e determine ainda o coeficiente de correlação entre o tempo de CPU ( $x_i$ ) e tempo de disco I/O's ( $y_i$ ). Considere as seguintes informações:

$$\sum_{i=1}^{n=7} x_i = 56 \quad \sum_{i=1}^{n=7} y_i = 3.269 \quad \sum_{i=1}^{n=7} x_i y_i = 49.577 \quad \sum_{i=1}^{n=7} x_i^2 = 630 \quad \sum_{i=1}^{n=7} y_i^2 = 7.950.713$$

Tabela - 1

| Nome do Programa | Função          | Tempo de CPU (x) | Time I/O's (y) |
|------------------|-----------------|------------------|----------------|
| TKB              | Linker          | 14               | 2735           |
| MAC              | Assembler       | 13               | 253            |
| COBOL            | Compiler        | 8                | 27             |
| BASIC            | Compiler        | 6                | 27             |
| Pascal           | Compiler        | 6                | 12             |
| EDT              | Editor de Texto | 4                | 91             |
| SOS              | Editor de Texto | 1                | 33             |

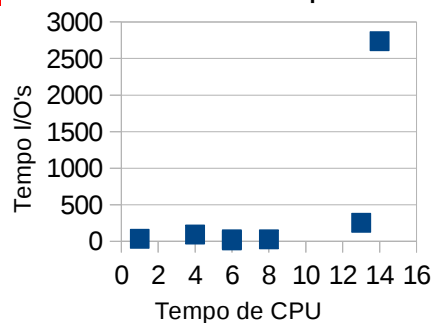
|     |         |       |
|-----|---------|-------|
| 196 | 7480225 | 38290 |
| 169 | 64009   | 3289  |
| 64  | 729     | 216   |
| 36  | 729     | 162   |
| 36  | 144     | 72    |
| 16  | 8281    | 364   |
| 1   | 1089    | 33    |
| 518 | 7555206 | 42426 |

|                                |               |             |
|--------------------------------|---------------|-------------|
| <b>Médias</b>                  | <b>7,4</b>    | <b>454</b>  |
| <b>Dp</b>                      | <b>4,3</b>    | <b>934</b>  |
| <b>Coeficiente de Variação</b> | <b>0,58</b>   | <b>2,06</b> |
| <b>Covariância</b>             | <b>2688</b>   | <b>2688</b> |
| <b>Correlação</b>              | <b>0,6632</b> | <b>1,0</b>  |

**1,0**

**0,5**

Gráfico de Dispersão



- a) Qual deles tem maior variabilidade (coeficiente de variação).

$$Coefv(X) = \frac{\sigma_x}{\bar{x}} \quad Var(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

VARIÂNCIA X:

$$Var(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{630}{7} - 56^2 = 18,8$$

$$Coefv(X) = \frac{\sigma_x}{\bar{x}} = \frac{\sqrt{18,8}}{56/7} = 0,584$$

VARIÂNCIA Y:

$$Var(Y) = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 = \frac{7.555.206}{7} - \left( \frac{31.780}{7} \right)^2 = 873.199,1$$

$$Coefv(Y) = \frac{\sigma_y}{\bar{y}} = \frac{\sqrt{873.199,1}}{31.780/7} = 2,058$$

- b) Determine ainda o coeficiente de correlação entre o tempo de CPU ( $x_i$ ) e tempo de disco I/O's ( $y_i$ ).

$$Cov(X,Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} = \frac{42.426}{7} - \frac{52}{7} \frac{31.780}{7} = 2688,3 \quad Corr(X,Y) = \frac{Cov(X,Y)}{(\sigma_x \sigma_y)} = \frac{2688,3}{4,3 * 934,5} = 0,66$$

4. Num grupo de 7 pessoas da FATEC foram pesquisados os anos de estudos e a renda familiar (Tabela ao lado). Sabemos que o número médio de anos de estudo  $\bar{x}$  (obtida considerando a soma total de anos de estudos) é de 14 anos e que neste caso os dados tem uma variância de 17,7. Quais alunos estão na zona de normalidade de anos de Estudo? Sabendo que a média da renda familiar é 4 salários mínimos, com desvio padrão de 1,5 salários mínimos, podemos afirmar que existe uma correlação entre os anos de estudos e a renda familiar? Qual é esta correlação?

| Aluno | Anos de estudos (X) | Renda familiar (Y) |      |      |      |
|-------|---------------------|--------------------|------|------|------|
| A     | 13                  | 4                  | -0,6 | 0,0  | 0,0  |
| B     | 16                  | 5                  | 2,4  | 1,0  | 2,4  |
| C     | 22                  | 7                  | 8,4  | 3,0  | 25,3 |
| D     | 9                   | 2                  | -4,6 | -2,0 | 9,1  |
| E     | 11                  | 3                  | -2,6 | -1,0 | 2,6  |
| F     | 14                  | 4                  | 0,4  | 0,0  | 0,0  |
| G     | 10                  | 3                  | -3,6 | -1,0 | 3,6  |
|       | 95                  | 28                 |      |      | 43,0 |

|                               |               |      |
|-------------------------------|---------------|------|
| Médias                        | 13,6          | 4,0  |
| Variância                     | 16,8          | 2,3  |
| Dp's                          | 4,1           | 1,5  |
| Zona de Normal Anos Estudos   | 9,5           | 17,7 |
| Alunos na Zona de Normalidade | A, B, E, F, G | 1,0  |
| Covariância                   | 6,1           |      |
| Correlação                    | 0,99          | 2,0  |

## Formulário

### A fórmula para dados no rol

Variância da variável X:

$$Var(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$Var(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

Desvio padrão da variável X:  $\sigma_x = \sqrt{Var(X)}$

Coefficiente de variação da variável X:  $Coefv(X) = \frac{\sigma_x}{\bar{x}}$

Covariância entre as variáveis X e Y:

$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

Coefficiente de correlação entre as variáveis X e Y:

$$Corr(X, Y) = \frac{Cov(X, Y)}{(\sigma_x \sigma_y)}$$

### A fórmula para dados agrupados

$$Mediana = l_i + \frac{\frac{n}{2} - Fac_{anterior}}{f_i} \cdot h$$

$$Moda = l_i + \frac{\Delta_{anterior}}{\Delta_{anterior} + \Delta_{posterior}} \cdot h$$

- $l_i$  limite inferior da classe i
- $f_i$  frequência da classe i
- $h$  amplitude dos intervalos de cada classe
- $n$  número de dados analisados
- $Fac_{anterior}$  frequência acumulada da classe anterior a classe da mediana
- $\Delta_{anterior}$  variação das frequência da classe da moda e da classe anterior;
- $\Delta_{posterior}$  variação das frequência da classe da moda e da classe posterior;