Laboratório de Física Geral

Profª Patricia Teles

Sala 3006-A

Email: athenafma@gmail.com

<u>Aula2</u>: Parâmetros de Posição, Dispersão e Correlação. Atividade 1

Parâmetros de Dispersão

Amplitude: Diferença entre os valores máximo e mínimo de uma coleção de dados {x1, x2, . . . , xN}

$$A = x_{\max} - x_{\min}$$

Desvio médio: Média dos módulos dos desvios, em relação à média

$$\overline{|\delta x|} = \frac{1}{N}\sum_{i=1}^{N}|\delta x_i| = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}| = \frac{|x_1 - \bar{x}| + \ldots + |x_N - \bar{x}|}{N}$$

Variância: Média dos quadrados dos desvios (δxi)

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(\delta x_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + \ldots + (x_N - \bar{x})^2}{N}$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2 = \overline{x^2} - \bar{x}^2$$

(expressão simplificada para cálculo da variância)

Parâmetros de Dispersão

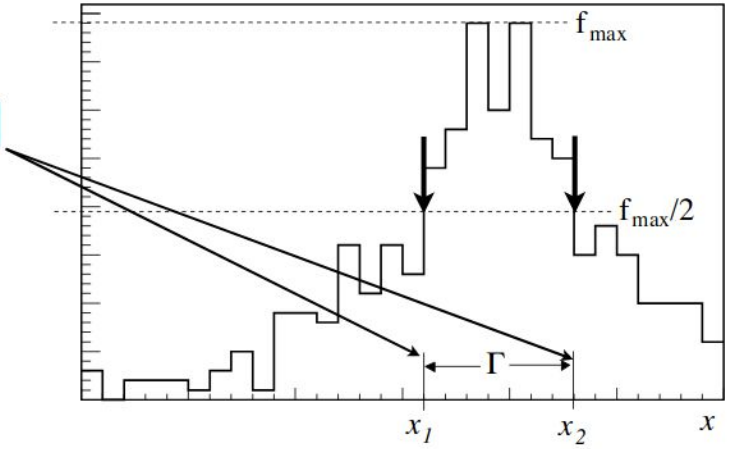
Desvio padrão: Raiz quadrada da variância, ou média quadrática dos desvios

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\delta x_i)^2} = \sqrt{\frac{(x_1-\bar{x})^2 + \ldots + (x_N-\bar{x})^2}{N}} \quad \Rightarrow \quad \sigma_x = \sqrt{\overline{x^2} - \bar{x}^2}$$

Largura a meia altura: Comprimento do intervalo limitado pelos valores (x1,x2) correspondentes à metade da frequência máxima

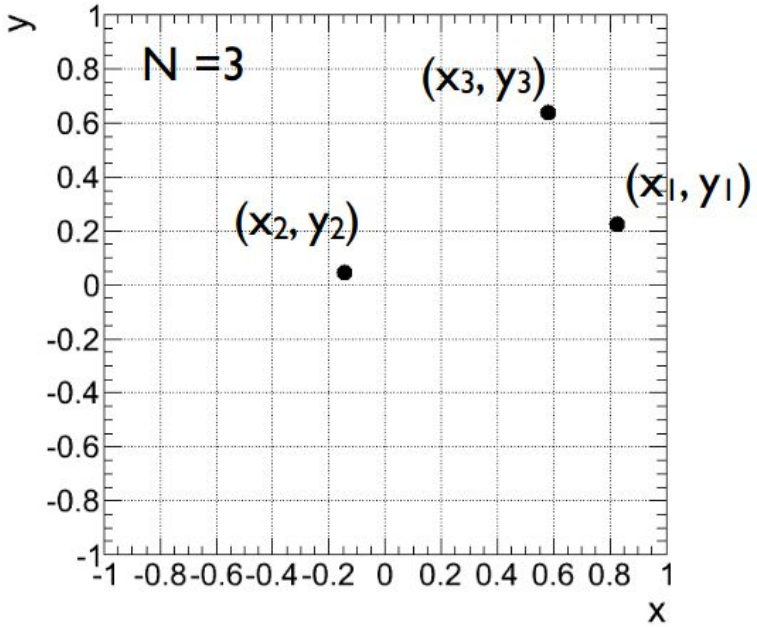
Símbolo: Γ

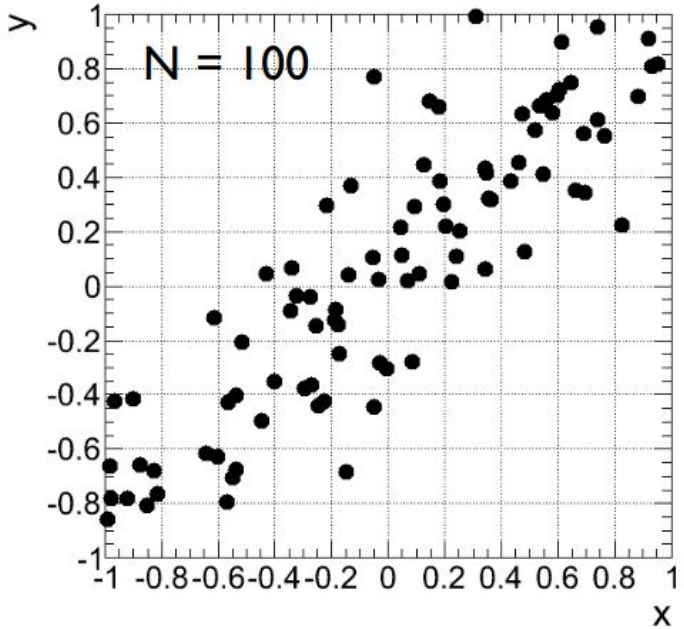
$\Gamma = |x_2 - x_1|$



Representando duas variáveis: x e y

Diagrama de dispersão: Gráfico representando medidas em duas variáveis {(x1, y1), (x2, y2), ..., (xN, yN)}



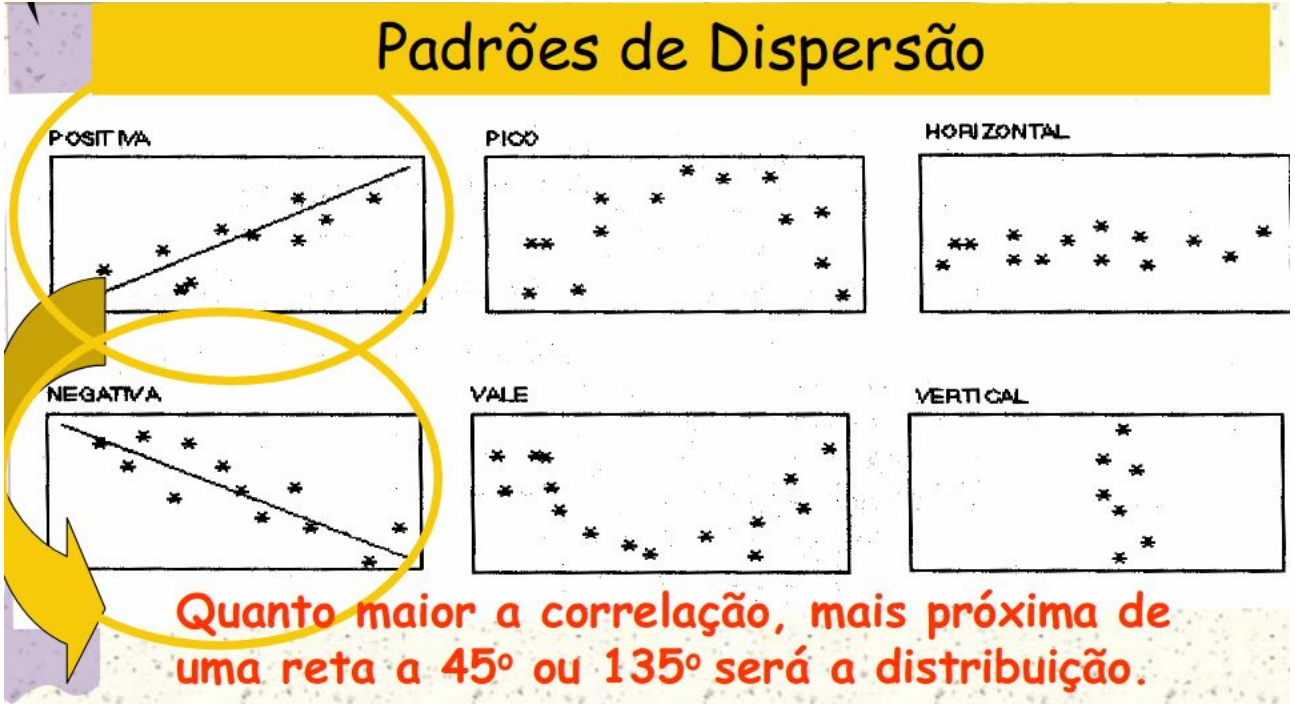


Eixo´x´ : variável que é a iterada por uma modificação no processo (variável independente); geralmente uma possível causa de um problema

Eixo´y´: Variável que pode mudar de acordo com a mudança da variável em´x´ (variável dependente); geralmente um indicador de qualidade ou efeito gerado por uma causa

Analisando Diagramas de Dispersão

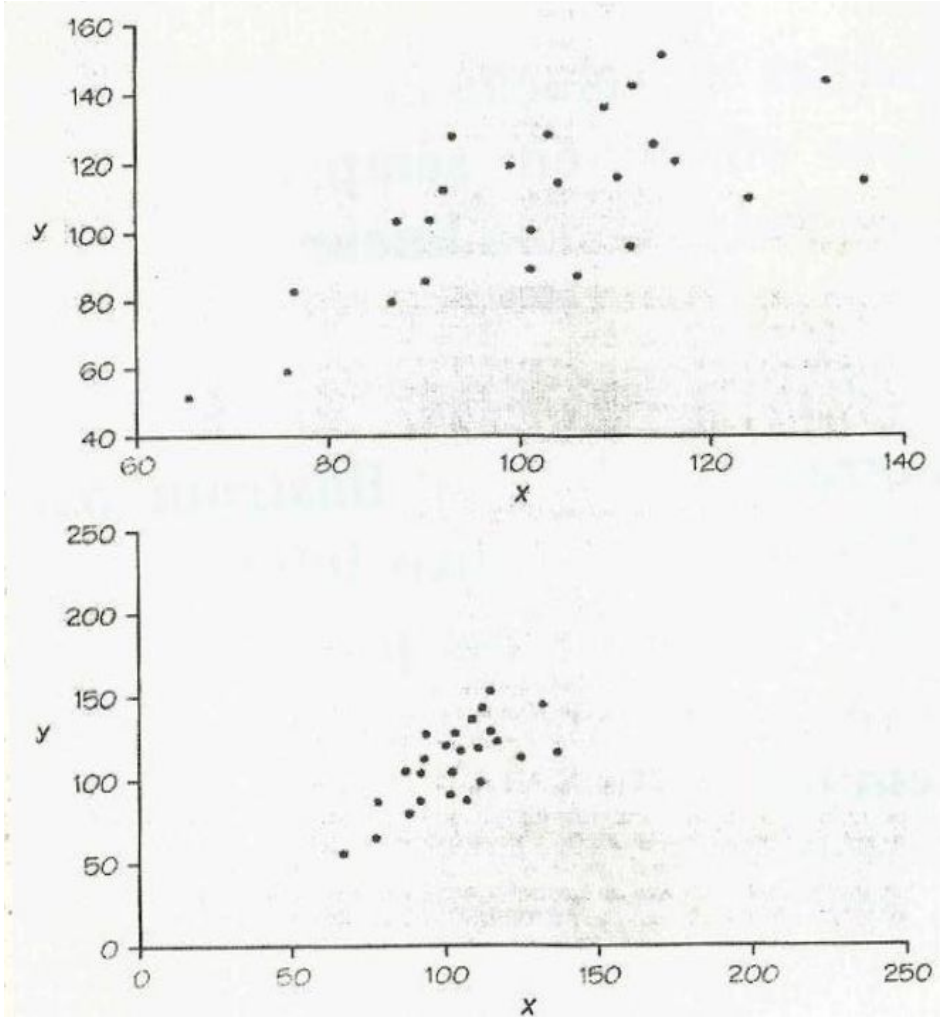
Os aspectos abaixo são relevantes na análise dos Diagramas:
DIREÇÃO (crescente, decrescente)
FORMA (linear, não-linear, aglomerados)
PONTOS DISCREPANTES



Problemas da Análise Gráfica
A análise gráfica da relação entre variáveis é importante, mas os olhos nem sempre são um bom juiz da intensidade de uma relação linear.

Os diagramas ao lado ilustram precisamente os mesmos dados, mas o gráfico inferior é menor em um campo mais amplo (escala diferente).

Nossos olhos podem ser enganados por uma mudança de escalas, ou pela quantidade de espaço em branco em torno do aglomerado dos pontos.
Deve-se, então, utilizar uma medida numérica para suplementar o gráfico.
Coeficiente de Correlação Linear (r)



Parâmetros de correlação

Covariância: média dos produtos dos desvios nas duas variáveis (δxi e δyi)

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N} \delta x_i \delta y_i = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \ldots + (x_N - \bar{x})(y_N - \bar{y})}{N}$$

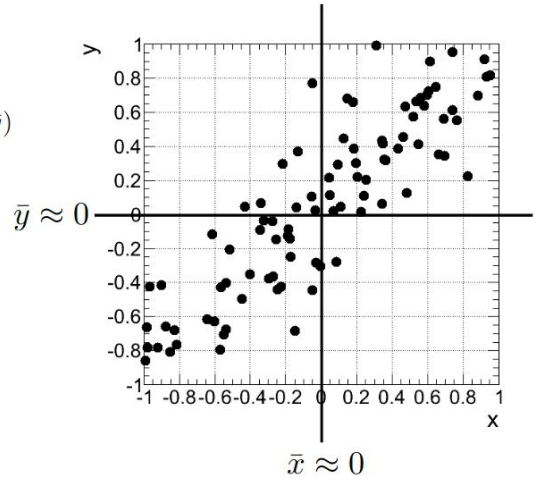
$\sigma_{xy} = \overline{xy} - \bar{x}\bar{y}$ (expressão simplificada para cálculo da covariância; não importa a ordem das variáveis)

$\sigma_{xu} = \sigma_{ux}$

Covariância:

$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

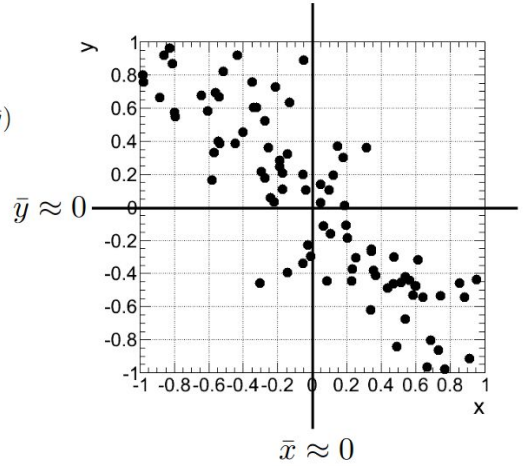
$\sigma_{xy} > 0$



Covariância:

$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

$\sigma_{xy} < 0$



Parâmetros de correlação

Coeficiente de correlação linear de Pearson: covariância entre duas variáveis, dividida por seus desvios padrão

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad -1 \leq r \leq 1$$

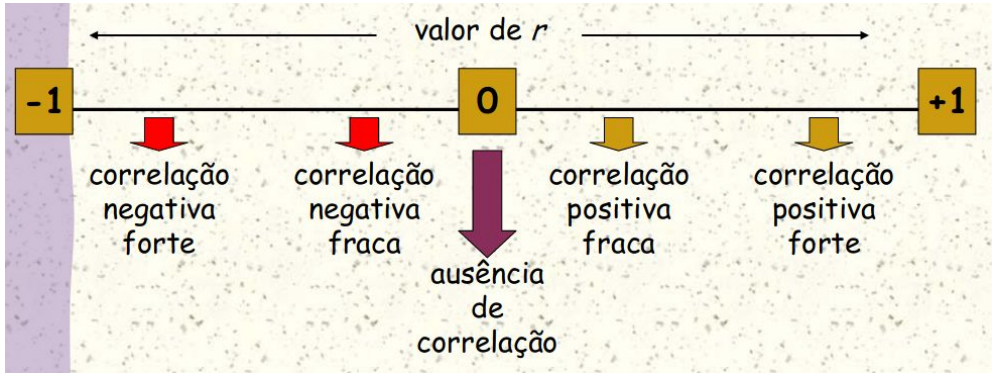
Correlação linear, perfeita e positiva: $r = 1$

Correlação linear, perfeita e negativa: $r = -1$

r = 0 : duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma dependência não linear que deve ser investigada por outros meios

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.
Exemplo: velocidade x consumo combustível

mede o grau de relacionamento linear entre valores emparelhados x e y em uma amostra.
Mede a intensidade e a direção da relação linear entre duas variáveis quantitativas.



Ex.: Alturas e Pesos de Ursos Siberianos

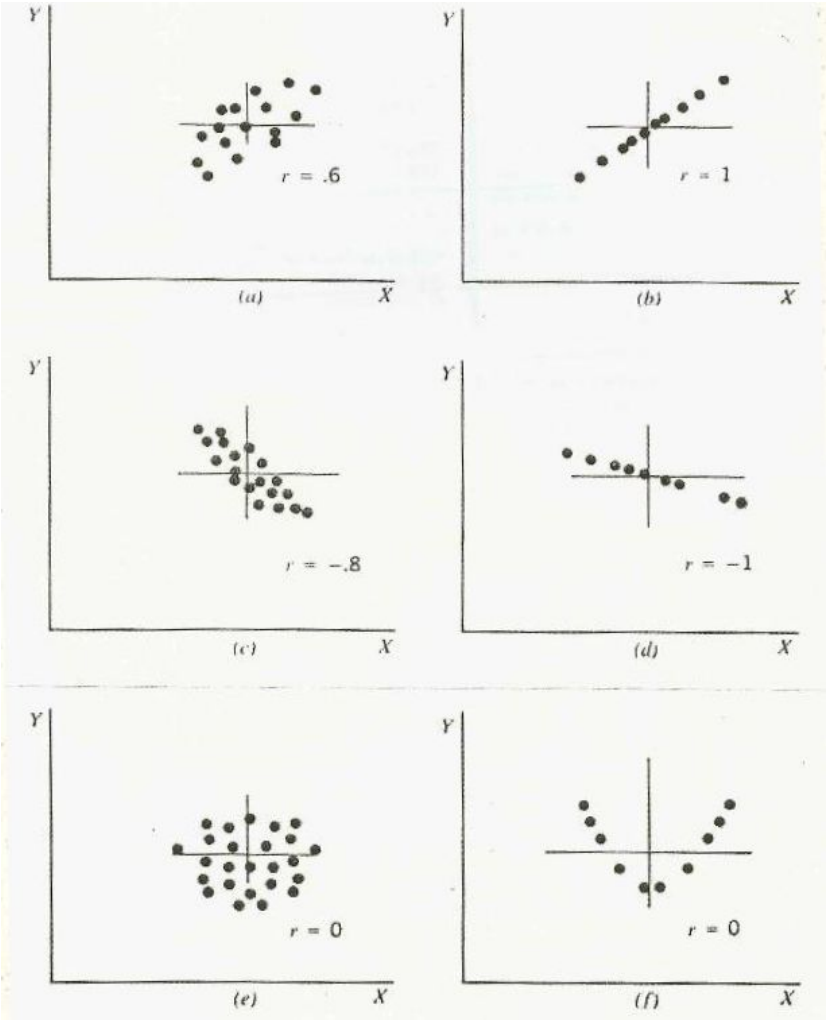
Comprimento (pol.)		Peso (lb.)			
	x	y	x.y	x2	y2
	53,0	80	4.240	2.809,00	6.400
	67,5	344	23.220	4.556,25	118.336
	72,0	416	29.952	5.184,00	173.056
	72,0	348	25.056	5.184,00	121.104
	73,5	262	19.257	5.402,25	68.644
	68,5	360	24.660	4.692,25	129.600
	73,0	332	24.236	5.329,00	110.224
	37,0	34	1.258	1.369,00	1.156
Totais	517	2.176	151.879	34.525,75	728.520

$$r = \frac{n\sum(x_i \cdot y_i) - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2} \cdot \sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \therefore$$

$$r = \frac{8(151.879) - (516{,}5)(2.176)}{\sqrt{8(34.525{,}75) - (516{,}5)^2}\sqrt{8(728.520) - (2.176)^2}} =$$

$$= \frac{91.128}{\sqrt{9433{,}75 \cdot 1.093.184}} = 0{,}897$$

Correlação positiva forte



Para curiosidade*

Interpretação geométrica [editar | editar código-fonte]

As duas séries de valores $X(x_1, \ldots, x_n)$ e $Y(y_1, \ldots, y_n)$ podem ser consideradas como vetores em um espaço de n dimensões. $X(x_1 - \bar{x}, \ldots, x_n - \bar{x})$ e $Y(y_1 - \bar{y}, \ldots, y_n - \bar{y})$.

O cosseno do ângulo α entre estes vetores é dado pela fórmula (produto escalar normado): $\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\| \cos\theta$

$$\cos(\alpha) = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Portanto $\cos(\alpha) = \rho$

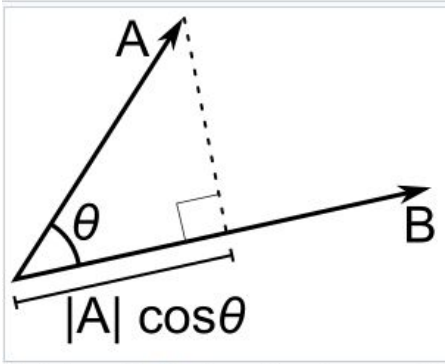
O coeficiente de correlação não é outro senão o cosseno do ângulo α entre os dois vetores!

Se $\rho = 1$, o ângulo α = 0, os dois vetores são colineares (paralelos).

Se $\rho = 0$, o ângulo α = 90°, os dois vetores são ortogonais.

Se $\rho = -1$, o ângulo α = 180°, os dois vetores são colineares com sentidos opostos.

Mais geralmente : $\alpha = \arccos(\rho)$, (arccos é a inversa da função cosseno).



Produto escalar de vetores. Percebe-se que ||**A**||•cos(θ) é a projeção escalar de **A** em **B**.

*não será usada nesse curso

Ferramentas para análise de dados

É possível utilizar pacotes como Excel™ e similares para representação, análise e visualização de dados.

Como veremos, há outras ferramentas com mais recursos, incluindo as linguagens de programação Python e R, e suas bibliotecas e extensões.

Podemos listar como vantagens em utilizar tais linguagens de programação como ferramenta para análise de dados:

- São grátis (possuem licença livre).

- São intuitivas e de rápida aprendizagem.

- Integram uma vasta biblioteca para computação científica e análise de dados.

- São "portáteis": o código escrito nestas linguagens pode ser processado em qualquer sistema ou plataforma.

Estatistica Descritiva com Python
http://felipegalvao.com.br/blog/2016/03/31/estatistica-descritiva-com-python/

Curso Int. Pyhon UERJ: link