ANÁLISE DE REGRESSÃO

- **Coeficiente de correlação linear produto momento, segundo Pearson (r)**

$$r = \frac{cov(x,y)}{\sqrt{var(x)\,var(y)}} = \frac{\frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1} \cdot \frac{\Sigma(y_i - \bar{y})^2}{n-1}}}$$

$$r = \frac{SPXY}{\sqrt{SQX.SQY}}$$

SPXY = $\Sigma xy - (\Sigma x\ \Sigma y) / n$; SQX = $\Sigma x^2 - (\Sigma x)^2 / n$; SQY = $\Sigma y^2 - (\Sigma y)2^2 / n$

- r: -1 à +1; r: 0, não há correlação linear entre x e y.
- r^2*100%: fração da variância total de x e y explicada pela relação linear: ajuste da distribuição dos pontos em relação à reta.
- teste usado para verificar se a correlação é ou não significativa

$$t = r\sqrt{\frac{n-2}{1-r^2}},$$ com (n-2)g.l.

- **Coeficiente de correlação não paramétrico, segundo Spearman (r_S)**
- variáveis não possuem distribuição normal
- x_i e y_i ordenados por postos (*rank)*, segundo os seus valores (x^*_i e y^*_i)
- $d_i = x^*_i - y^*_i$; d_i^2 ; Σd_i^2

$$r_s = 1 - \frac{6\Sigma d_i^2}{n^3 - n}$$ n = número de pares de valores x^*_i , y^*_i

- caso ocorram muitos casos com valores de posto empatados:

$$r_s = \frac{\Sigma x'_e + \Sigma y'_e - \Sigma d_i^2}{2\sqrt{\Sigma x'_e \ \ \Sigma y'_e}}$$

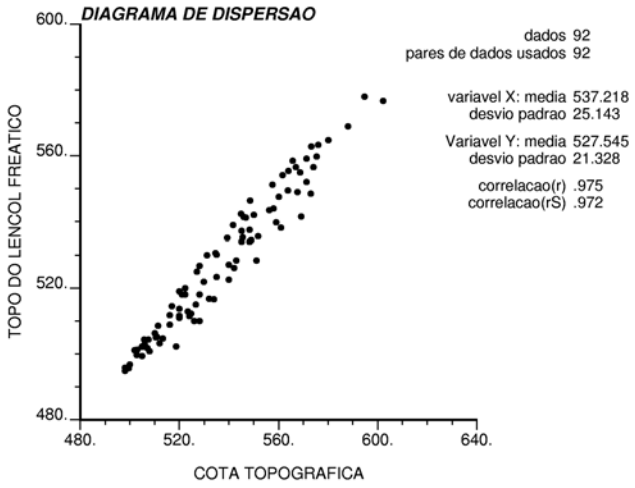
onde

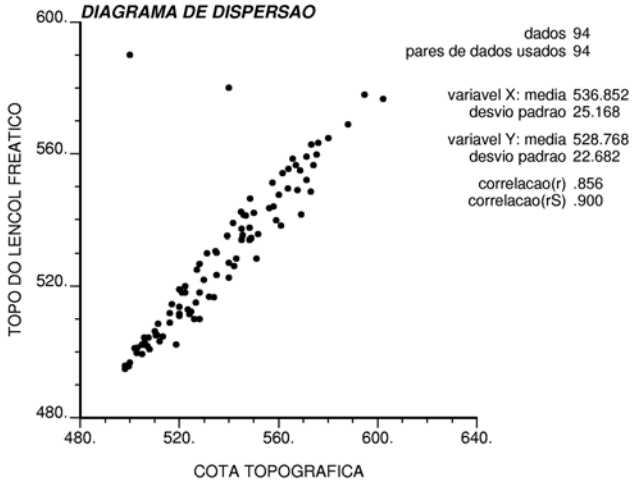
$$\Sigma x'_e = \frac{n^3 - n}{12} - \Sigma T_x \qquad \Sigma y'_e = \frac{n^3 - n}{12} - \Sigma T_y$$

$$T = \frac{t^3 - t}{12}$$

t = número de observações repetidas em um determinado posto.

- **coeficiente de correlação linear é influenciado pela presença de valores anômalos.**
- **grande diferença entre o coeficiente de correlação linear e o coeficiente de correlação por postos reflete tanto uma relação não-linear como presença de pares de valores extremos.**





Regressão linear

- **Verificado pelo valor de r que ocorre uma significante correlação linear entre duas variáveis (x_i, y_i) há necessidade de quantificar tal relação, o que é feito pela análise de regressão.**
- **Equação de uma reta que, disposta num sistema de eixos cartesianos, com valores de y_i (variável dependente) na ordenada e x_i (variável independente) na abcissa, a soma dos quadrados dos desvios verticais dos pontos em relação a ela seja mínima.**

$$y = a + bx,$$

onde y é o valor estimado para um específico valor x_i; b revela a inclinação da reta, ou seja o acréscimo ou decréscimo do valor de y em relação à x; a localiza o ponto de interseção da reta em relação ao sistema de coordenada retangulares.

- **Utilizando o método dos mínimos quadrados, os valores da equação da reta são determinados por:**

$$b = \frac{SPXY}{SQX} \qquad a = \bar{y} - b\bar{x}$$

$$\bar{y} = \frac{\Sigma y_i}{n} \; ; \; \bar{x} = \frac{\Sigma x_i}{n}$$

Eixo maior reduzido

- **desconhecimento de uma variável independente ou sem erro**
- **no lugar de desvios verticais dos pontos em relação à reta, áreas dos triângulos compreendidos entre os pontos e a reta**

$y = a + bx,$

$b = \pm (Sy / Sx)$, **sendo o sinal de "b" o do correspondente r**

$$b = \sqrt{\frac{SQY}{SQX}} = \left[\frac{\Sigma y^2 - (\Sigma y)^2 / n}{\Sigma x^2 - (\Sigma x)^2 / n}\right]^{1/2} \; ; \; a = \bar{y} + b\bar{x}$$

Regressão curvilínea

$$Y^* = a_0 + a_1X + a_2X^2 + a_3X^3 + ...$$

- potências crescentes de x_i, variável independente e coeficientes
- x_i e x_i^2: parábola com um único ponto de inflexão
- com potências crescentes de x_i, curva mais complexa para ajuste
- processo por etapas (*stepwise*)
- O modelo para a regressão polinomial de grau k é

$$Y = \alpha_o + \alpha_1 X_i + \alpha_2 X_i^2 + \ ... \ + \alpha_k X_i^k + \varepsilon$$

- cálculo dos coeficientes de regressão α

$$[X] = \begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 & \cdots & \Sigma x_i^k \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 & \cdots & \Sigma x_i^{k+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \Sigma x_i^k & \Sigma x_i^{k+1} & \Sigma x_i^{k+2} & \cdots & \Sigma x_i^{k+k} \end{bmatrix}$$

$$[Y] = \begin{bmatrix} \Sigma y_i \\ \Sigma y_i x_i \\ \Sigma y_i x_i^2 \\ \vdots \\ \Sigma y_i x_i^k \end{bmatrix} \qquad [\hat{a}] = \begin{bmatrix} \hat{a}_o \\ \hat{a}_1 \\ \vdots \\ \hat{a}_k \end{bmatrix}$$

$$[\hat{a}] = [X]^{-1}[Y]$$

Regressão polinomial

- superfícies contínuas podem ser encontradas, por critérios de regressão polinomial, onde Z_i é a variável dependente em função linear das coordenadas X-Y dos pontos amostrados e irregularmente distribuídos
- o modelo para a representação da superfície pelo método dos polinômios não ortogonais é:

$$z_i(X,Y) = [a_0 + a_1x_i + a_2y_i + a_3x_i^2 + a_4x_iy_i + a_5y_i^2 + \ ... \ ] + e_i(x_i, y_i) \ ,$$

onde $z_i(X,Y)$ é a variável mapeada em função das coordenadas x_i e y_i e $e_i(x_i, y_i)$ representa os resíduos, ou seja, a fonte não-sistemática de variação.

- a representação de uma superfície linear é dada por:

$$z(X,Y) = a_0 + a_1x_i + a_2y_i + e_i$$

- para o cálculo dos coeficientes a_i, dispõe-se os dados num sistema de equações normais :

$$\begin{bmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_iy_i \\ \sum y_i & \sum x_iy_i & \sum y_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum z_i \\ \sum z_ix_i \\ \sum z_iy_i \end{bmatrix}$$

$$[XY]\,[A] = [Z]$$

multiplicando ambos os termos pelo inverso de [XY],

$$[XY]^{-1}[XY][A] = [XY]^{-1}[Z]$$

como $[XY]^{-1}[XY] = [I]$ = matriz de identidade e $[I][A] = [A]$

$$[A] = [XY]^{-1}[Z]$$

para o cálculo do vetor de coeficientes [A], inverter a matriz [XY] e multiplicar esse resultado pelo vetor [Z].

- **A superfície quadrática é representada por:**

$$z_i(X,Y) = b_0 + b_1 x_i + b_2 y_i + b_3 x_i^2 + b_4 x_i y_i + b_5 y_i^2 + e_i,$$

e a determinação dos coeficientes b_0, b_1, b_2, b_3 ,b_4 e b_5 para a superfície de grau 2 torna-se:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} n & \sum x_i & \sum y_i & \sum x_i^2 & \sum x_i y_i & \sum y_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i y_i & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i y_i^2 \\ \sum y_i & \sum x_i y_i & \sum y_i^2 & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum y_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i^4 & \sum x_i^3 y_i & \sum x_i^2 y_i^2 \\ \sum x_i y_i & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i^3 y_i & \sum x_i^2 y_i^2 & \sum x_i y_i^3 \\ \sum y_i^2 & \sum x_i y_i^2 & \sum y_i^3 & \sum x_i^2 y_i^2 & \sum x_i y_i^3 & \sum y_i^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum z_i \\ \sum x_i z_i \\ \sum y_i z_i \\ \sum x_i^2 z_i \\ \sum x_i y_i z_i \\ \sum y_i^2 z_i \end{bmatrix}$$

As superfícies de grau superior a dois seguem o mesmo processo de desenvolvimento polinomial.