



Regressão Simples

Parte III: Coeficiente de determinação,
regressão na origem e método de máxima
verossimilhança

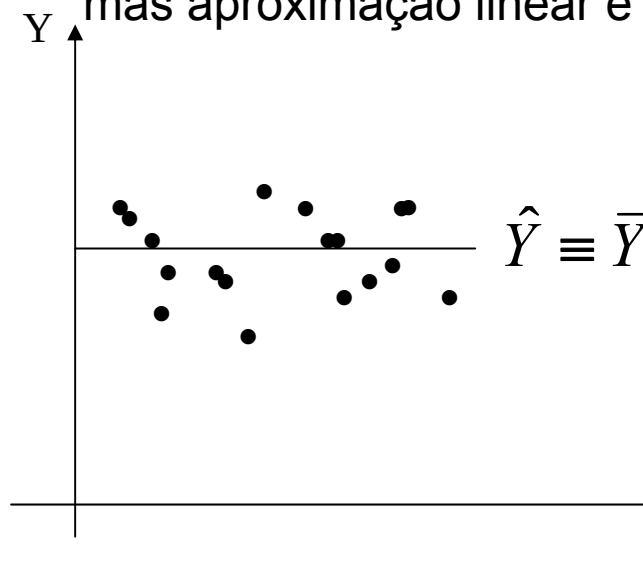
Coeficiente de determinação

- Proporção da variabilidade explicada pelo regressor.

$$R^2 = \frac{\text{Variação explicada}}{\text{Variação total}}$$

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSR}{SST}$$
$$0 \leq R^2 \leq 1$$

R^2 pode ser relativamente grande
mas aproximação linear é pobre



Coeficiente de determinação

- Na regressão simples: a correlação linear é a raiz quadrada do R^2

$$r = \pm\sqrt{R^2}$$

$$-1 \leq r \leq 1$$

- Embora o R^2 não decresça, não podemos dizer que, se mais variáveis preditoras são adicionadas no modelo, o modelo novo é melhor do que o antigo.
- A menos que, se a soma de quadrados de resíduos do modelo novo seja reduzida para o mesmo da soma de quadrados de resíduos do modelo antigo, o novo modelo terá uma média quadrados de resíduos maior que a média de quadrados de resíduos do modelo antigo. Então o modelo novo é pior.
- A magnitude de R^2 também depende da range de variabilidade de x . Um grande valor de R^2 pode ter sido resultado de uma range não realístico de x e um valor pequeno de R^2 pode ter sido resultado de uma insuficiente de um range pequeno de x .

Outras métricas

$$R^2 = 1 - \left(\frac{n-1}{n-p-1} \right) \frac{SSR}{SST}$$

$$0 \leq R^2 \leq 1$$

R² ajustado

*Quanto maior melhor
o modelo*

$$MS_{Res} = \frac{SSR}{n-p-1}$$

Quanto menor melhor o modelo

$$PRESS = \frac{\sum_{i=1}^n (y_i - \hat{y}_i^{-1})^2}{n}$$

Quanto menor melhor o modelo

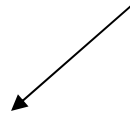
A predição para a média i-ésima resposta considerando o modelo de regressão sem o i-ésimo par de observação (x,y).

Outras métricas

Método Akaike

$$AIC = -2 \ln L(\hat{\theta}) + 2p$$

Quanto menor melhor o modelo

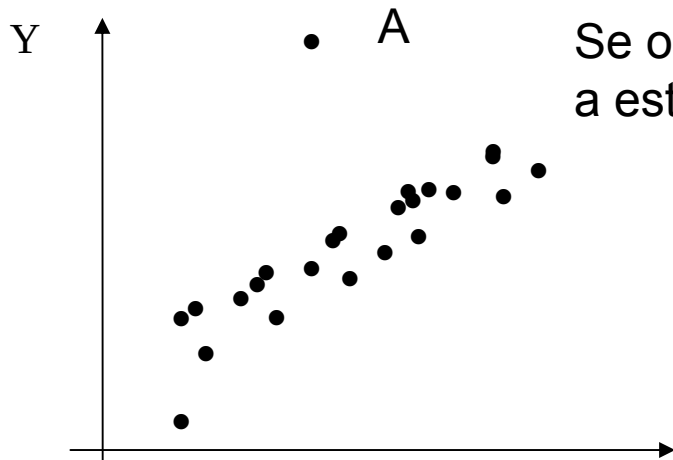


Valor da
verossimilhança no
estimativa dos
parâmetros

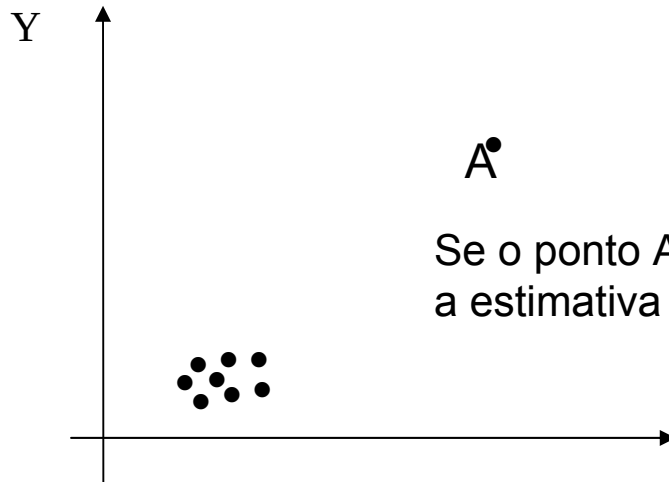
Considerações

- Deve-se ter cuidado para não fazer extrapolações no range da variável preditora.
- Se numa regressão indicar forte relacionamento entre duas variáveis, não implica que as variáveis estão relacionadas em um sentido casual. Por exemplo, seja y = numero de deficiente mental por 10000 habitantes e x o número de emissoras de rádio com licença no período de 1924 a 1937. O p-value do teste t é 0,0 e $R^2=0,9842$.
- Quando o valor da variável x é desconhecida, pode-se usar uma estimativa para x . Contudo a acurácia da estimativa de y dependerá da acurácia da estimativa usada para x .

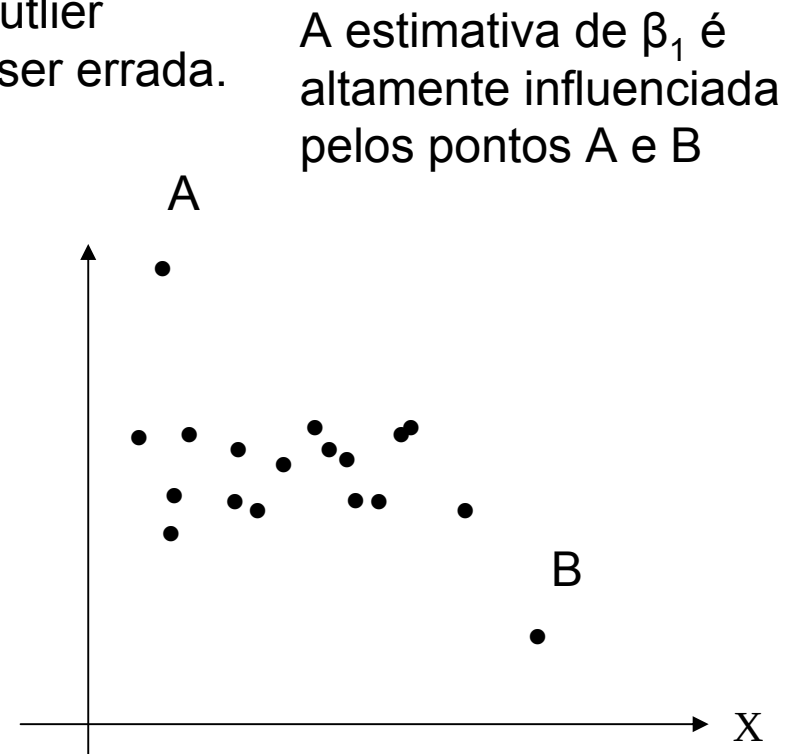
Considerações



Se o ponto A é um outlier
a estimativa β_1 pode ser errada.



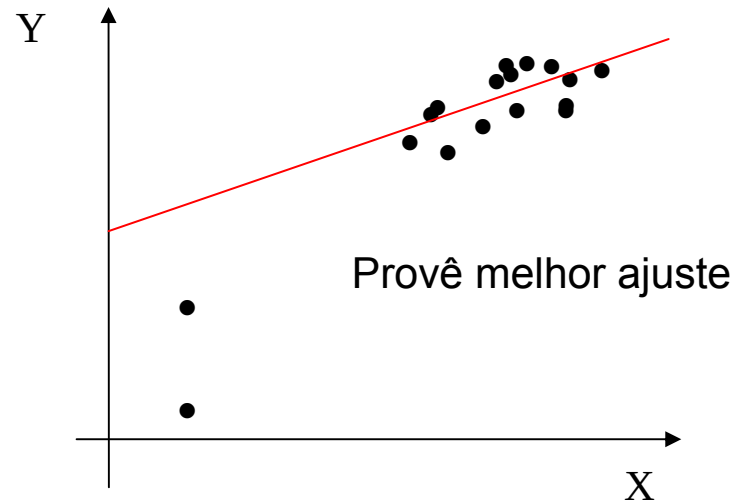
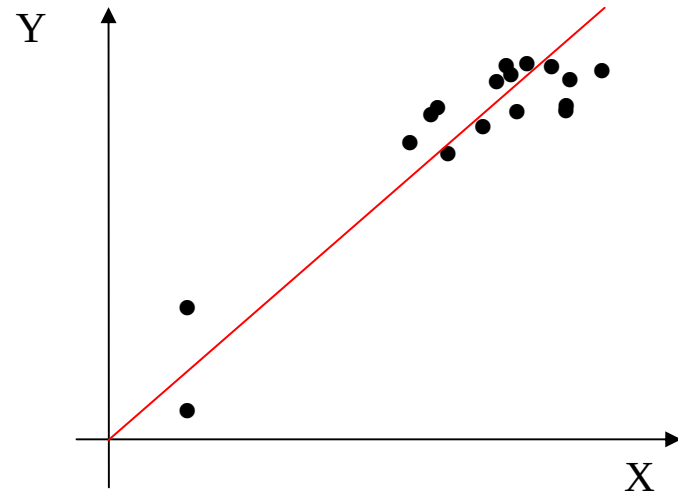
Se o ponto A for removido,
a estimativa de β_1 seria zero



A estimativa de β_1 é
altamente influenciada
pelos pontos A e B

Regressão na origem

- É frequente apropriado em análise de dados de processos químicos. For exemplo: a saída de um processo químico é zero quando a temperatura manipulada é zero.
- É comum o uso inapropriado desse modelo quando existem valores de x distantes da origem.



Estimação dos parâmetros pelo método da máxima verossimilhança

Como foi especificado uma distribuição de probabilidades para os erros podemos obter estimadores para β_0 , β_1 e σ^2 pelo MMV.

O método de máxima verossimilhança determina como estimativas de máxima verossimilhança, os valores de β_0 , β_1 e σ^2 os quais produzem o maior valor para a verossimilhança.

Em geral, a densidade de uma observação Y_i para o modelo de regressão com erros normais, utilizando o fato de que $E(Y_i) = \beta_0 + \beta_1 X_i$ e variância σ^2 é dada por :

$$f_i = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma} \right)^2 \right]$$

A função de verossimilhança para n observações $Y_1, Y_2, \dots, Y_n$, é o produto das densidades individuais (é a conjunta). Como a variância σ^2 dos erros é desconhecida, a conjunta é uma função de três parâmetros, β_0 , β_1 e σ^2 :

Estimação dos parâmetros pelo método da máxima verossimilhança

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2} (Y_i - \beta_0 - \beta_1 X_i)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2\right] \end{aligned}$$

Devemos encontrar valores de β_0 , β_1 e σ^2 que maximizam a função de verossimilhança L , calculando-se as derivadas parciais de L com respeito a β_0 , β_1 e σ^2 e igualando cada derivada parcial a zero e resolvendo o sistema de equações obtido. Podemos trabalhar com $\log_e L$ ao invés de L , pois ambos são maximizadas para os mesmos valores de β_0 , β_1 e σ^2 :

$$\log_e L = -\frac{n}{2} \log_e 2\pi - \frac{n}{2} \log_e \sigma^2 - \frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

Estimação dos parâmetros pelo método da máxima verossimilhança

As derivadas parciais do logaritmo da função de verossimilhança, são dadas por:

$$\frac{\partial(\log_e L)}{\partial \beta_0} = \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial(\log_e L)}{\partial \beta_1} = \frac{1}{\sigma^2} \sum X_i (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial(\log_e L)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

Fazendo as derivadas parciais iguais a zero, substituindo β_0 , β_1 e σ^2 pelos estimadores $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\sigma}^2$

Obtemos:

$$\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\sum X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\sum \frac{(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n} = \hat{\sigma}^2$$

Estimação dos parâmetros pelo método da máxima verossimilhança

As duas primeiras equações são idênticas as equações normais encontradas pelo método de mínimos quadrados. O MMV produz um estimador viesado para σ^2 .

Parâmetros	Estimadores
β_0	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
β_1	$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$
σ^2	$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}$

Os estimadores de β_0 e β_1 são os mesmos do método de mínimos quadrados

Estimação dos parâmetros pelo método da máxima verossimilhança

Em geral os estimadores de máxima verossimilhança tem melhores propriedades estatísticas que os estimadores de mínimos quadrados.

1. São não enviesados
2. Variância mínima.
3. São consistentes;
4. São estatísticas suficientes.

Por outro lado: Requerem suposição da distribuição para os erros além das suposições requeridas pelo método de mínimos quadrados.

Estística suficiente

- Seja Q uma população arbitrária cujos elementos podem possuir ou não uma característica w cuja proporção de incidência vale θ .
- Deseja-se **estimar** o valor de θ e para tal dispõe-se de uma amostra $\{w_1, w_2, \dots, w_n\}$ onde w_i vale 1 ou 0 indicando se o elemento i possui ou não a característica w . Parece razoável perceber que o que efetivamente interessa para a estimação de θ é a proporção da incidência da característica na amostra ou seja : que é então dita ser uma estatística suficiente para θ .

$$\hat{g} = \frac{\sum_{i=1}^n w_i}{n}$$

Teste de Hipótese para coeficiente correlação

$$H_0 : \rho = 0$$

$$H_0 : \rho \neq 0$$

Estatística do teste:

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \rightarrow t_{n-2} \quad \text{Se } H_0 \text{ é verdadeira}$$

Região crítica: Rejeita-se se

$$|t_0| > t_{\alpha/2, n-2}$$