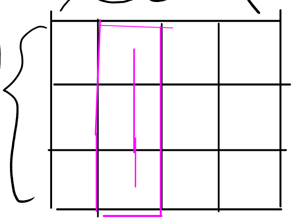
Data Science — 10/10/2023 (teória)

Estrutura Retangular

Variáveis

Entidades da mesma categoria



independente

dependente → Conseguimos explicar ou prever usando outra(s) variável(is) do mesmo dataset

$$\boxed{y = B_0 + B_1 x + \varepsilon} \rightarrow \text{Real}$$

estimador

$$\boxed{\hat{y} = \hat{B}_0 + \hat{B}_1 x + \varepsilon} \rightarrow \text{modelo}$$

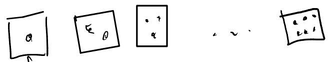
Após calcularmos os Betas

temos que validar o modelo

- plotar o gráfico e ver se o modelo difere dos datapoints
- desvio padrão
- 95% confidence interval (Lola)

Teste de hipótese

Exemplo:



1/6

- Assumimos que não existe bias
- Recolhemos evidência
- Confrontamos os acontecimentos com a análise

$P(\neg ⚅) = 5/6$ — não ser 6

$P(\neg ⚅ \text{ after 4 throws}) = \left(\frac{5}{6}\right)^4 \approx 48\%$

$P(\neg ⚅ \text{ after 16 throws}) = \left(\frac{5}{6}\right)^{16} \approx 5\%$ → Very unlikely

Não provamos necessariamente que existe cheating mas muito provável! Como sair sempre 6 por exemplo.

Exemplo:

TA: Teste acurate

Teste covid: 99% resultado correto

$\boxed{p(\text{acurate}) = 0{,}99}$ $\boxed{P(\neg\text{acurate}) = 0.01}$

$100 \to$ pessoas

$\frac{100}{100} \to$ testes

$$P\left(TA\ \tfrac{100}{100}\right) = (0{,}99)^{100} \approx 36\%$$

$$P\left(T.A.\ {}^{99}/_{100}\right) = \left[(0{,}99)^{99} \times (0{,}01)\right] \times \binom{100}{1} \approx 37\%$$

$\leftarrow$ 1 pessoa teve teste errado

$$\boxed{\binom{n}{k} = \frac{n!}{k!\,(n-k)!}}$$

$$P\left(T.A.\ {}^{98}/_{100}\right) = \binom{100}{2} \times (0{,}99)^{98} \times (0{,}01)^2 \approx 18{,}5\%$$

$$P\left(T.A\ \tfrac{97}{100}\right) = \binom{100}{3} \times (0{,}99)^{97} \times (0{,}01)^3 \simeq 6\%$$

$$P\left(T.A\ \tfrac{96}{100}\right) = \binom{100}{4} \times (0{,}99)^{96} \times (0{,}01)^4 \simeq 1{,}5\%$$

$$P\left(T.A\ \tfrac{95}{100}\right) = \binom{100}{5} \times (0{,}99)^{95} \times (0{,}01)^5 \simeq 0{,}3\%$$

Hipotese nula $\to$ o que significa fazer um teste de hipotese?