Projeto 5

Ana Capriles

2º Etapa – Parte Teórica

A técnica estatística da regressão consiste em quantificar e estudar o efeito que a alteração de um determinado número de variáveis – as quais correspondem a variáveis explicativas na terminologia desta técnica – causa sobre algum fenômeno, o qual é a variável resposta (por ser o fenômeno estudado, isto é, a resposta ao problema). Neste particular projeto, será analisado o efeito que a variação do consumo de CO2 per capta e a idade para a qual se dá o primeiro casamento das mulheres possuem sobre o valor do Índice de Desenvolvimento Humano (IDH).

Para uma dada relação entre variáveis, pode-se afirmar que a correlação explicita a força da correlação linear e que a regressão explicita a sua forma.

A seguinte equação representa um modelo de regressão linear:

(1)

Na qual os parâmetros significam o seguinte:

: valor da variável resposta

: valor da variável explicativa X1, para o i-ésimo elemento

: valor da variável explicativa X2, para o i-ésimo elemento

, e : parâmetros que determinam o ajuste linear

: valor esperado de Y para um dado valor de X1 e X2.

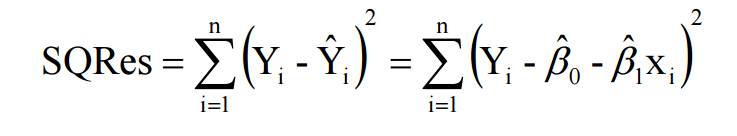
: erro estocástico (aleatório)

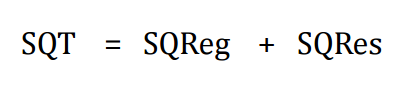
n corresponde ao tamanho da amostra (da base de dados)

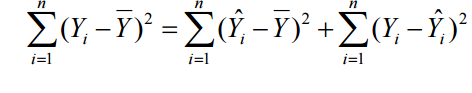
i varia de 1 até n

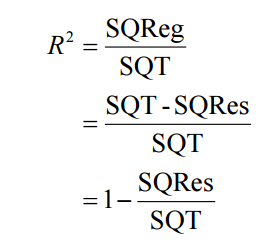
**Como calcular os estimadores de , e a partir da base de dados?**

Para estimá-los, é necessário minimizar o resíduo que é dado pela diferença entre o valor verdadeiro de 𝑦 e seu valor estimado 𝑦. O método utilizado na estimação desses parâmetros é o método dos mínimos quadrados, o qual requer que consideremos a soma dos resíduos quadrados, denotado por SQRes.









O valor de R varia de 0 até 1.

**Como ficam os testes de hipóteses na regressão múltipla e o que a rejeição ou não da particular hipótese nula 𝐻0 significa nesse caso?**

Na regressão, uma das hipóteses em análise avalia a significância da regressão. Isto é, a hipótese nula é a afirmação de que não existe relação entre as variáveis, ou seja,

(2)

Isto é, o parâmetro que ajusta a relação linear vale zero. Para realizar esse teste de hipóteses é necessário atribuir uma distribuição aos erros estocásticos.

A hipótese alternativa afirma que o valor do estimador é diferente de zero, ou seja, que há relação linear entre as variáveis.

**Qual é a interpretação das estimativas dos coeficientes que serão estimados no problema?**

O intercepto é o valor previsto (esperado ou médio) para a variável resposta quando a variável explicativa vale zero. Quando não fizer sentido zerar a variável explicativa, o valor, por si só, não será muito interessante. De maneira geral, a cada variação ∆𝑥 na variável explicativa 𝑥, o estimador é a variação prevista (esperada ou média) na variável resposta.

**Quais as suposições feitas sobre os erros em termos de: distribuição, valor esperado e variância e, ainda responda, como a adequação dessas suposições pode ser checada na prática?**

Os erros têm distribuição normal com média e variância constante, são independentes entre si (ou seja, a sua correlação vale zero. O modelo é linear nos parâmetros e existe homocedasticidade, o que significa que a variância é a mesma para todos os valores da amostra e igual ao quadrado do desvio padrão.