

Métodos Matemáticos

Regressão linear e correlação

Você sabia que seu material didático é interativo e multimídia? Isso significa que você pode interagir com o conteúdo de diversas formas, a qualquer hora e lugar. Na versão impressa, porém, alguns conteúdos interativos ficam desabilitados. Por essa razão, fique atento: sempre que possível, opte pela versão digital. Bons estudos!

Nesta webaula conheceremos sobre o modelo de regressão linear. Inicialmente, veremos a origem desse modelo e na sequência os principais objetivos. Em relação a representação, aprenderemos a equação linear simples.

Regressão linear

Origem

Em um estudo para trabalhar a relação entre a altura dos pais e dos filhos (X_i e Y_i). A fim de verificar como a altura do pai influenciava a altura do filho, Galton deu origem, no século XIX, ao que conhecemos hoje como teoria de regressão. O termo regressão veio pelo fato de que no experimento de Galton as medidas estudadas tendiam a regredir à média.

Objetivos

De modo geral, os modelos de regressão têm como objetivos:

| | |
|---|---|
| Predição | ▼ |
| Consiste em prever valores de Y por meio de valores de X não presentes entre os dados originais. Isto é, trabalhamos com possíveis valores de X e obtemos valores de Y, desde que boa parte da variabilidade de Y seja, de fato, explicada por X. | |
| Seleção de variáveis | ▼ |
| Consiste em selecionar as variáveis que têm algum impacto significativo em relação à variável de interesse (variável resposta). | |
| Estimação de parâmetros | ▼ |
| Encontrar valores que são utilizados para inferir resultados sobre os parâmetros da população, por exemplo, cálculo de doses letais, tempo até a cura de uma doença, porcentagem de falha de uma máquina industrial, etc. | |
| Inferência | ▼ |
| Após encontrar os valores dos parâmetros, o processo de inferência serve para discutir/inferir resultados sobre uma dada população e estabelecer critérios como intervalo de confiança sobre tais resultados. | |

Nomenclaturas e o Modelo de Regressão Linear Simples

No que tange aos modelos de regressão, algumas nomenclaturas são importantes. Por exemplo, as variáveis **Xs** são chamadas **variáveis independentes ou explanatórias**, enquanto que a variável **Y** é chamada de **variável dependente ou resposta**.

Com esses nomes vamos estabelecer o modelo de regressão linear simples. Suponha que a relação verdadeira entre X e Y pode ser escrita por uma equação de reta. Então, o valor esperado de Y para cada valor de X é dado por:

$$E(Y|X) = \beta_0 + \beta_1 X$$

sendo que os parâmetros da equação da reta β_0 e β_1 são constantes desconhecidas. Quando $X=0$, β_0 , representa o ponto onde a reta corta o eixo dos Ys. Esse parâmetro é chamado de intercepto. Por outro lado, β_1 é chamado coeficiente de regressão ou declive da reta e ter interpretação descrita como “a cada aumento de 1 unidade em X, temos que aumenta β_1 unidades”.

Agora, dados $(X_1, Y_1), \dots, (X_n, Y_n)$, se for admitido que Y é uma função linear de X, o modelo de regressão linear simples é dado por:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, 2, \dots, n$$

sendo β_0 e β_1 , os parâmetros do modelo e ϵ_i é o erro aleatório do modelo.

Com as informações apresentadas, esperamos que você tenha uma noção sobre o funcionamento do modelo de regressão linear que poderá ser muito utilizado no futuro em sua profissão.

Para visualizar o vídeo, acesse seu material digital.