

Entrega 2

Ciência dos Dados

Leonardo Lamounier Grotti

Patrick Serrano Wiegerinck

Prof. Maria Kelly Venezuela

2016

São Paulo

Com a minimização da soma dos quadrados dos resíduos, encontra-se e , que por sua vez trarão a menor diferença entre a previsão de e o realmente observado.

🡪 Equação da reta da previsão

🡪 Equação da reta realmente observada

A minimização é feita ao deixar em relação a e e, então, igualar a 0

Substituindo na equação 1 e dividindo por 2n:

Sendo a média amostral de e a média amostral de .

Substituindo esse resultado na equação 2, temos:

Isolando o , chega-se a segunda resposta:

Pode-se assumir que em regressões lineares os erros modelos, , são representados por distribuições normais e independentes com igual a 0 e variância igual a (), ou seja, a variância é constante e, portanto, existe homoscedasticidade. Além disso, assume-se que não existe correlação entre os erros .

Para verificar isso basta analisar a curva de probabilidade cumulativa dos resíduos e a da distribuição normal. A semelhança das curvas definirá se o erro é ou não uma distribuição normal. Outros métodos de verificação são: construção de um intervalo de confiança para a média, com o objetivo de verificar a suposição da média, e verificar graficamente se isso se confirma.

Normalmente, uma das hipóteses em análise de regressão é avaliar a significância da regressão, ou seja, os testes de hipóteses verificam a qualidade da regressão para a variável resposta. (No nosso caso Expectativa de vida).

A hipótese nula é: e ela diz que não há relação entre x (variável explicativa) e y (variável resposta), por outro lado a hipótese alternativa é beta1 diferente de 0 e nesse caso há relação entre x e y.

Concluindo, caso a hipótese nula seja rejeitada, podemos concluir que há relação entre a variável explicativa e a variável resposta.

Sim, nesse caso estaremos fazendo uma regressão linear múltipla. Para isso acontecer temos que analisar no mínimo 3 variáveis sendo uma a variável resposta e duas ou mais as variáveis explicativas.

Para a equação devemos acrescentar mais termos de acordado com a quantidade de variáveis estudadas ficando assim:

No caso do teste de hipóteses devem ser feitos um teste para cada variável explicativa e eles se comportam exatamente da mesma maneira que na regressão linear simples.

As suposições do modelo continuam iguais, pois, como já foi definido anteriormente, pode-se sempre assumir as suposições do item 2 para regressões lineares.