**Ciência dos Dados – Projeto 5 (2a etapa)**

Felipe Viacava e Felipe Giorgi

Insper – Instituto de Ensino e Pesquisa

São Paulo, 2015

**a)** Consulte em livros como se calculam os estimadores de *β0*, *β1* e *β2* a partir dos dados. Não é necessário demonstrar expressões.

Para se calcular os estimadores de *β0*, *β1* e *β2* utiliza-se o método dos mínimos quadrados. Este método descobre qual é o plano médio que mais se aproxima dos dados ao zerar a derivada da equação do plano para encontrar o mínimo local.

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

Teste de hipóteses 1 (*β1)*:

*H0*: *β1* = 0

*H1*: *β1* ≠ 0

A rejeição da hipótese nula para o teste de hipóteses 1 acontece caso *β1* estimadoseja distante o suficiente de 0 para ser considerado diferente de 0.

Teste de hipóteses 2 (*β2)*:

*H0*: *β2* = 0

*H1*: *β2* ≠ 0

A rejeição da hipótese nula para o teste de hipóteses 2 acontece caso *β2* estimado seja distante o suficiente de 0 para ser considerado diferente de 0.

**c)** Qual será a interpretação das estimativas dos coeficientes que serão estimados no seu problema? Aqui, faça a interpretação em termos do problema ainda que a estimativa não tenha sido calculada.

Caso *β1* = 0, entende-se que X1 não afeta a variável de interesse (Y). Ou seja: se *β1* = 0, inferimos que os novos casos anuais de câncer no cólon e no reto não são afetados pelo insumo calórico diário por pessoa. Caso a hipótese nula não se confirme (*β1* ≠ 0), entende-se que X1 afeta a variável de interesse. Em outras palavras, o insumo calórico diário por pessoa influencia a incidência de novos casos de câncer no cólon e reto.

Caso *β2* = 0, entende-se que X2 não afeta a variável de interesse (Y). Ou seja: se *β2* = 0, inferimos que os novos casos anuais de câncer no cólon e no reto não são afetados pelo gasto total em saúde por pessoa. Caso a hipótese nula não se confirme (*β2* ≠ 0), entende-se que X2 afeta a variável de interesse. Em outras palavras, o gasto total em saúde por pessoa influencia a influencia a incidência de novos casos de câncer no cólon e no reto.

**d)** 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, ou seja:

Os erros são independentes entre si, ou seja:

O modelo é linear nos parâmetros

Homocedasticidade:

, para qualquer