

ANÁLISE DE REGRESSÃO LOGÍSTICA

- 1) Determine o valor da desigualdade e a probabilidade estimada para os seguintes casos:
 - a) $Z=0,5$
 - b) $Z=1$
 - c) $Z=0,1$
 - d) $Z=2$
- 2) Discorra sobre as diferenças da Regressão Linear e Regressão Logística.
- 3) Na base de dados `binary.csv`, estão os dados de 400 alunos que foram ou não aprovados em um curso. As variáveis constantes no arquivo são: `admit` (mostra se o aluno foi ou não aprovado no curso) `gre` e `gpa` são notas obtidas em duas provas e `rank` é uma variável qualitativa que vai de 1 a 4 que mostra o nível de concorrência do curso. Aplique a técnica de análise de regressão logística com o método *stepwise*. Analise as principais estatísticas referentes à qualidade de ajuste do modelo.
- 4) Considere o arquivo `Propensao Doença`, que indica a incidência ou não de uma doença de acordo com seis tipos de proteína encontradas no corpo de um paciente

Faça o processamento da análise de regressão logística *stepwise*. Faça uma análise completa das principais informações obtidas.

- 5) Considere o arquivo `prcancer` referente a uma amostra de 53 homens. O tratamento e prognóstico de câncer depende de quanto a doença se espalhou. Uma das regiões em que o câncer pode se espalhar refere-se aos nódulos linfáticos. Se os mesmos forem atingidos, o prognóstico é geralmente mais pobre do que em caso negativo. Por isso é desejável estabelecer o quanto antes se os nódulos são cancerosos. Para certos tipos de câncer, cirurgia exploratória é feita só para determinar se os nódulos são cancerosos, uma vez que isso determinará qual o tratamento necessário. Se for possível prever se os nódulos são afetados ou não com base nos dados sem a realização de cirurgia, considerável desconforto e gasto poderão ser evitados. Os dados referem-se a 53 homens com câncer de próstata. Para cada paciente, temos: idade (**age**), serum acid phosphatase (**acid** – um valor de laboratório que é elevado se o tumor se espalhou para certas regiões), o estágio da doença (**stage** – uma indicação do avanço da doença), o grau do tumor (**grade** – uma indicação da agressividade) e os resultados do raio x (**xray**), assim como se o câncer se espalhou para os nódulos da região linfática na fase da cirurgia. O problema é prever se os nódulos são positivos para câncer com base nos valores das variáveis que podem ser medidas sem cirurgia (variável dependente: **node**). As variáveis `xray`, `stage` e `grade` são categóricas, codificadas como 0 e 1. O valor 1 sempre indica a pior situação (raio x positivo, estágio avançado e agressividade). Aplique a técnica de regressão logística com o método *stepwise*. Interprete os resultados obtidos e responda às questões:

- a) probabilidade de um homem com 66 anos, **acid** = 48 e 0 para as demais variáveis ter os nódulos linfáticos atingidos;
- b) probabilidade de um homem com 60 anos, **acid** = 62, valor 1 para **xray** e 0 para as demais variáveis ter os nódulos linfáticos atingidos;
- c) probabilidade de um homem com 60 anos, **acid** = 62, valor 1 para **xray** e **grade** e 0 para **stage** ter os nódulos linfáticos atingidos;
- d) compare os resultados de b) e c) com o coeficiente estimado para a variável **grade** no modelo logístico.