

**UNIVERSIDADE FEDERAL FLUMINENSE**

**Programa de Mestrado e Doutorado em Engenharia de Produção**

**Disciplina: Estatística Multivariada – 2016/2**

**Lista: Multiple Linear Regression**

**Professor: Valdecy Pereira, D. Sc.**

**email: [valdecy.pereira@gmail.com](mailto:valdecy.pereira@gmail.com)**

1) O Data Set MLR-01 é um questionário feito por uma fabricante de carros. O questionário levantou a percepção de atributos que homens (GENDER = 1) e mulheres (GENDER = 2) consideram mais importantes em um carro. No total 93 pessoas responderam o questionário. Os atributos foram medidos em uma escala Likert de 5 pontos (1 = Nenhuma Importância; 2 = Pouca Importância; 3 = Alguma Importância; 4 = Muita Importância; 5 = Extremamente Importante) e são listados a seguir:

VARIÁVEL	DESCRIÇÃO
P01	Qualidade dos Acessórios
P02	Facilidade de Dirigir
P03	Conforto
P04	Visibilidade
P05	Segurança na Estrada
P06	Alta Velocidade
P07	Poder de Aceleração
P08	Absorção de Impacto
P09	Transmissão
P10	Design
P11	Pintura
P12	Ausência de Barulho
P13	Ar Condicionado
P14	Economia de Combustível
P15	Economia nos Reparos
P16	Valor Estimado do Carro Usado
P17	Espaço
P18	Atendimento da Garantia
P19	Disponibilidade das Peças
P20	Qualidade da Assistência Técnica
P21	Preço das Peças
P22	Preço dos Reparos

Considerando a variável GENDER como a variável dependente, faça:

- A amostra satisfaz a Regra de Evans ou a de Doane?
- Realize o MLR com todas as variáveis. Analise de forma completa esse modelo.

- c) Faça a regressão com os seguintes preditores: CONFORT, DESIGN, SAVING ON FUEL, SAVING ON REPAIRS e PRICE OF PARTS. Analise de forma completa esse modelo.
- d) Compare os dois modelos.

2) O Data Set MLR-02 apresenta um histórico de vendas de 2902 casas em um determinado estado americano. O preço de venda final costuma ser bastante diferente do anunciado, podendo variar para mais ou para menos (variando entre 35% e 50%) dependendo da habilidade do corretor de imóveis e do poder de negociação do cliente. Uma corretora de imóveis deseja que você desenvolva um modelo que permita aumentar o seu faturamento, escolhendo um conjunto de casas com determinadas características que façam com que se possa efetuar a venda sempre a um preço maior do que o anunciado. A corretora também pediu que você escreva um protocolo de abordagem que um corretor deverá seguir quando estiver visitando uma casa (com essas características) com um cliente. As informações contidas do banco de dados são listadas a seguir e representam apenas as vendas que foram fechadas acima do preço anunciado:

VARIÁVEL	DESCRIÇÃO
LOTAREA	Área Total do Lote em $\text{ft}^2$
OVERALLQUAL	Qualidade Geral da Casa. 1 = Muito Pobre ; 2 = Pobre ; 3 = Ruim ; 4 = Abaixo da Média ; 5 = Média ; 6 = Acima da Média ; 7 = Boa ; 8 = Muito Boa ; 9 = Excelente e 10 = Perfeita.
EXTERQUAL	Avaliação da Qualidade do Exterior da Casa. 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.
BSMTQUAL	Avaliação da Qualidade do Porão (se houver). 0 = Não Há Porão ; 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.
BSMTTOTAL	Área Total do Porão em $\text{ft}^2$
HEATING	Tipo de Aquecimento. 1 = Gás ; 2 = Água e 3 = Gasóleo.
HEATINGQUAL	Avaliação da Qual. do Aquecimento (se houver). 0 = Não Há Aquecimento ; 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.
CENTRALAIR	Ar Condicionado Central. 0 = Não Possui ; 1 = Possui.
FIRSTFLOOR	Área Total do Primeiro Andar em $\text{ft}^2$
SECONDFLOOR	Área Total do Segundo Andar em $\text{ft}^2$
BEDROOM	Número Total de Quartos.
KITCHEN	Número Total de Cozinhas.
KITCHENQUAL	Avaliação da Qual. da Cozinha (se houver). 0 = Não Há Cozinha ; 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.
TOTALROOMS	Número Total de Cômodos.
FIREPLACE	Número Total de Lareiras.
FIREPLACEQUAL	Avaliação da Qual. das Lareiras (se houver). 0 = Não Há Lareiras ; 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.
GARAGECARS	Nº de Carros que Cabem na Garagem
GARAGEAREA	Área Total do Lote em $\text{ft}^2$

GARAGEQUAL	Avaliação da Qual. da Garagem (se houver). 0 = Não Há Garagem ; 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.
WOODDECK	Área Total da Varanda de Madeira em $\text{ft}^2$ (se houver).
OPENPORCH	Área Total de Varanda Aberta em $\text{ft}^2$ (se houver).
ENCLOSEDPORCH	Área Total de Varanda Coberta em $\text{ft}^2$ (se houver).
POOLAREA	Área Total da Piscina em $\text{ft}^2$ (se houver).
POOLQUAL	Avaliação da Qual. da Piscina (se houver). 0 = Não Há Piscina ; 1 = Pobre ; 2 = Ruim ; 3 = Mediana ; 4 = Boa e 5 = Excelente.

Considerando a variável SALESPRICE (preço de venda em \$) como a variável dependente, faça:

- A amostra satisfaz a Regra de Evans ou a de Doane?
- Realize o MLR com todas as variáveis. Analise de forma completa esse modelo.
- Qual é a característica de uma casa que tem o potencial de ser vendida a um preço maior do que o anunciado?
- Descreva o seu protocolo de abordagem.

3) O Data Set MLR-03 apresenta dados de 74 restaurantes da franquia *Noodles & Company*. A *Noodles & Company* é um novo modelo de restaurantes casuais de atendimento rápido. A franquidora deseja saber quais são os atributos que impactam diretamente na receita das vendas (SALES\_SQFT em \$/ft<sup>2</sup>). Os seguintes preditores foram levantados:

PREDITORES	DESCRIÇÃO
SQFT	Tamanho do restaurant em $\text{ft}^2$
Seats_Inside	Número de cadeiras dentro do restaurante
Seats_Outside	Número de cadeiras fora do restaurante
Mean_Age	Idade média dos clientes
Sales_Person	Consumo médio por pessoa
Sales_Salad	Venda média de saladas
Sales_Rice	Venda média de arroz como guarnição
Alcohol_%	Percentual médio de vendas de bebidas alcólicas
Water_%	Percentual médio de vendas de água
Soft_Drink_%	Percentual médio de vendas de refrigerantes e sucos
Protein_%	Percentual médio de vendas de pratos com proteínas
House_Dishes_%	Percentual médio de vendas de promoções do dia

Bread_%	Percentual médio de vendas de sanduíches
Deserts_%r	Percentual médio de vendas de sobremesas

Considerando a variável SALES\_SQFT como a variável dependente, faça:

- A amostra satisfaz a Regra de Evans ou a de Doane?
- Realize o MLR com todas as variáveis. Analise de forma completa esse modelo.
- Quais são as características que um restaurante deve ter para que possa maximizar as suas vendas?

4) O Data Set MLR-04 apresenta medidas antropométricas de 250 homens coletadas por um médico. A medição da gordura corporal pode ser feita com precisão através de uma pesagem feita na água, ou estimada através de equações, como por exemplo, a Equação de BROZEK. A Equação de BROZEK necessita como entrada o valor da densidade corporal, uma medida difícil de ser obtida em um consultório tradicional. O médico então o contratou para que você desenvolva um modelo que seja tão bom quanto a Equação de Brozek, mas que tenha como entrada medidas que no máximo necessitem apenas de uma escala e uma fita métrica. As medidas coletadas foram:

VARIÁVEL	DESCRIÇÃO
AGE	Idade (anos)
WEIGHT	Peso em (kg)
HEIGHT	Altura (m)
ADIPOSITY	Adiposidade (kg/m2)
NECK	Circunferência do Pescoço (cm)
CHEST	Circunferência do Peito (cm)
ABDOMEN	Circunferência do Abdomen (cm)
HIP	Circunferência do Quadril (cm)
THIGH	Circunferência da Coxa (cm)
KNEE	Circunferência do Joelho (cm)
ANKLE	Circunferência do Tornozelo (cm)
BICEPS	Circunferência do Bíceps Estendido (cm)
FOREARM	Circunferência do Antebraço (cm)
WRIST	Circunferência do Pulso (cm)

Considerando a variável BROZEK como a variável dependente, faça:

- A amostra satisfaz a Regra de Evans ou a de Doane?
- Realize o MLR com todas as variáveis. Analise de forma completa esse modelo.
- É possível atender a necessidade do médico?
- Existe alguma ressalva a ser feita?