

## ESTATÍSTICA DESCRIPTIVA, INFERÊNCIA ESTATÍSTICA E REGRESSÃO LINEAR

### Trabalho de Grupo

O Trabalho é componente obrigatória da Unidade Curricular e tem um peso de 40% da nota final e tem dois objetivos gerais: o primeiro consiste numa análise exploratória de dados amostrais – Estatística Descritiva – e o segundo na estimação de parâmetros e na previsão com base no modelo de Regressão Linear Simples – Estatística Indutiva.

O tratamento de dados deve ser realizado com auxílio do software R.

A estrutura do relatório deve respeitar as orientações que se seguem e não deve exceder 25 páginas A4 (pode ter anexos: imagens e outputs do R).

O trabalho deverá ser defendido oralmente e deverá ser submetido em pasta compactada, no link de submissão do separador Avaliação do Moodle. Esta pasta deve conter:

- O relatório em formato editável (grupoXXX.doc);
- O relatório em formato PDF (grupoXXX.pdf);
- A base de dados (grupoXXX.txt ou grupoXXX.xls);
- O ficheiro Script em R (grupoXXX.R)

A estrutura do relatório deve respeitar as orientações que se seguem

### Estrutura:

#### 1. Introdução

- Contextualização;
- Indicação dos objetivos;

#### 2. Estatística Descritiva

- Caracterização da população e da amostra;
- Representações gráficas;
- Identificação de “outliers”;
- Comparações entre grupos;
- Representação de diagramas de dispersão;
- Determinação dos coeficientes de regressão e da equação da reta associada;
- Interpretação dos coeficientes de correlação e de determinação;
- Análise do ajustamento do modelo;
- Contextualização da análise dos resultados;

#### 3. Estatística Indutiva:

- Identificação de distribuições teóricas e aproximações;
- Aplicação de Testes Paramétricos e Não Paramétricos;

- Estimação de médias populacionais e Intervalos de Confiança;
- Comparação de médias de grupos;
- Representação de diagramas de dispersão;
- Determinação dos coeficientes de regressão e da equação da reta associada;
- Significância dos coeficientes de correlação e de determinação;
- Análise do ajustamento do modelo;
- Previsão de valores “in the box” e “out of the box”.

#### 4. Conclusão

- Concretização dos objetivos propostos;
- Dificuldades encontradas;
- Passos Futuros

### Tema 1: Macroeconomia

Recolheram-se dados relativos a 12 países da Europa. Registou-se a Classificação, o Rendimento Disponível das Famílias e o Consumo Privado, em milhões de euros, o número de habitantes em milhares e o desempenho face às metas ambientais. A classificação varia entre CN, para países localizados no Centro ou Norte Europeu e SE, para aqueles que ocupam as regiões Sul e Este. E o desempenho varia numa escala de 1 a 5 (1 – o desempenho é muito bom; 4 – o desempenho é muito fraco).

### Tema 2: Marketing

Num departamento de marketing de uma empresa nacional recolheram-se dados relativos a 20 registos mensais de faturação de empresas selecionadas aleatoriamente. Registou-se a localização da filial, o número de funcionários, o valor da faturação nesse mês e a respetiva despesa em publicidade, em milhares de euros (m€) e a avaliação geral do gerente em relação à poupança de energia (1 – poupança é grande; 4 – a poupança é fraca). A filial foi classificada por N, para representar uma empresa que serve a região Norte ou S para a região Sul do país.

### Tema 3: Indústria Automóvel

Recolheram-se dados relativos a 21 testes executados em automóveis. Identificou-se o condutor que executou o teste com as letras A ou B, o peso de cada viatura testada, em kg, a distância que cada automóvel percorreu com 10 litros de gasóleo, o consumo médio de litros aos 100 km e a avaliação de desempenho ecológico que varia entre 1 e 4 (1 – o desempenho é muito bom; 4 – o desempenho é muito fraco).

## Tema 4: Indústria de Confeção

Recolheram-se dados relativos a 23 peças de pele que foram cortadas automaticamente pela mesma máquina. Registou-se o tipo de molde que foi cortado, A ou B, o comprimento do desperdício, em centímetros, e o custo associado, em euros. Também se registaram a espessura da pele, em mm e a qualidade da peça que varia entre 1 – qualidade muito boa e 4 – qualidade muito fraca.

## Tema 5: Setor imobiliário

Numa imobiliária, recolheram-se dados relativos a 25 apartamentos. Registou-se o meio envolvente, a idade do imóvel, em anos, o preço de venda e o lucro absoluto em milhares de euros, e a eficiência energética. O meio envolvente pode assumir o valor U, para apartamentos localizados em meio Urbano ou R para apartamentos localizados em meio Rural. A eficiência energética assume valores entre 1 para pouco eficiente e 3 para muito eficiente.

## Tema 6: Stand automóvel

Um negociante do ramo automóvel recolheu dados relativos aos lucros de 17 vendas de veículos usados com garantia de 1 ano. Registou o género do condutor habitual (atual proprietário), o número de avarias durante o período de garantia, o Lucro da venda e o valor da despesa de reparações, em euros. O género toma o valor M para representar o sexo feminino e H o masculino e o perfil do cliente foi codificado de 1 a 3 onde 1 representa um cliente muito frequente e 3 para um cliente pouco frequente.

## Tema 7: Empresa Construção Civil

Recolheram-se dados relativos a uma amostra de projetos de uma empresa de construção civil. Utilizou-se para indicador a coluna “Codigo”. Para cada projeto registou-se a região da filial de intervenção (Filial), o número de habitações do projeto (n\_habit), a solução apresentada pela empresa (Solucao), o valor inicialmente previsto (Inicial) e o valor final da obra (Final).

## Tema 8: Instituto Nacional de Estatística

Para se compararem duas regiões de Portugal face ao custo da habitação, recolheram-se dados relativos a uma amostra de Municípios. Utilizou-se para indicador a coluna “Mun”. Para cada Município registou-se a região (NUTII), o valor de referência da Aquisição (VRefAq), o valor de referência do arrendamento (VRefAr), o número de núcleos precários (n. Nucleos) existentes nesse município e a Equipa responsável pelo levantamento (Equipa).

## Tema 9: Fundos UE

Uma empresa que gere fundos da União Europeia organizou uma amostra de candidaturas. Utilizou-se para indicador a coluna “ID\_Processo”. Para cada candidatura registou-se a região onde se localiza a filial (Filial), o número de habitações da candidatura (N\_Fogos), o tipo de cliente Beneficiário (Beneficiário), o tipo de solução encontrada (Solucao), o valor do Investimento (investimento) e a equipa de trabalho responsável pela candidatura (Equipa\_Trab).

## Tema 10: Fundos Comunitários

Uma empresa que gere fundos comunitários organizou uma amostra de candidaturas. Utilizou-se para indicador a coluna “candidat”. Para cada candidatura registou-se o número de habitações da candidatura (N\_habit), o tipo de cliente (Cliente), o tipo de projeto (Projeto), o valor do empréstimo concedido (emprestimo) e a equipa de trabalho responsável pela candidatura (Equipa\_Trab).

## Tema 11: Famílias prioritárias

Para se compararem duas regiões de Portugal face ao custo da habitação, recolheram-se dados relativos a uma amostra de Municípios. Utilizou-se para indicador a coluna “Mun”. Para cada Município registou-se a região (NUTII), o valor de referência da Aquisição (VRefAq), o valor de referência do arrendamento (VRefAr), o número de famílias que foram sinalizadas pelos serviços sociais (n. Familias) a residir nesse município e a prioridade que uma empresa estipulou para as intervenções a desenvolver nesse município (Prioridade).

## Tema 12: Agência Bancária

Um banco organizou, em tabela, dados de uma amostra processos de empréstimo dispersos pelo país. Utilizou-se para indicador a coluna “ID\_Processo”. Para cada processo registou-se a região onde se enquadra a Filial Bancária (Filial), o número de fogos atribuídos a esse cliente (N\_Fogos), a equipa de trabalho responsável pelo empréstimo (Equipa\_Trab), o valor do investimento pedido e aprovado, em euros.