Licenciatura em Ciência de Dados Análise Exploratória de Dados



Enunciado – Caso Prático 23

O estudo "Saúde Psicológica e Bem-estar" foi promovido pelo Ministério da Educação, em 2022, após o período de pandemia Covid-19. Um dos principais objetivos é o diagnóstico e conhecimento do panorama da saúde psicológica e bemestar das crianças e adolescentes em idade escolar (do pré-escolar ao 12.º ano), em diferentes regiões de Portugal. Os dados disponibilizados referem-se a um excerto das respostas dos alunos (ou dos seus educadores) ao questionário aplicado, relativamente a algumas das perguntas, enquadradas nos objetivos definidos e alguma informação de caracterização das crianças e jovens.

Com a realização do caso prático pretende-se a preparação e organização do conjunto de dados e correspondente análise descritiva, em função do contexto e objetivos definidos. Com recurso ao Excel, Jamovi e R deverão ser realizadas as seguintes tarefas:

I - Excel

- 1. Aplicar o filtro a todas as variáveis e apresentar a primeira linha fixa;
- 2. Atribuir de código de não resposta a uma variável com valores omissos;
- 3. Atribuir texto às categorias de uma variável que tenha códigos;
- 4. Corrigir possíveis erros;
- 5. Criar regra de validação para duas variáveis: uma quantitativa e outra qualitativa;
- 6. Construir uma tabela de frequências absolutas com funções de contagem e cálculo de percentagens;
- 7. Construir uma tabela de frequências dinâmica;
- 8. Construir uma tabela de cruzamentos dinâmica.

II - Jamovi

Construir de forma automática e com recurso ao módulo R:

- 2 tabelas com medidas descritivas (com 1 e com 2 variáveis)
- 1 gráfico com cruzamento de variáveis

ISCLE SOCIOLOGIA E POLÍTICAS PÚBLICAS

Licenciatura em Ciência de Dados Análise Exploratória de Dados

III - R

- 1. Importar o ficheiro de Excel;
- 2. Mudar o nome de duas variáveis;
- 3. Identificar os valores omissos de uma variável;
- 4. Imputar os valores omissos de uma variável com a mediana dos seus valores;
- 5. Analisar descritivamente os dados (uni e bivariada), incluindo:
 - a. tabelas de frequências com valores absolutos e percentagens
 - b. medidas descritivas de variáveis quantitativas
 - c. gráficos (incluindo gráficos de distribuição)
- 6. Construir, com recurso ao R Markdown, um relatório com a interpretação dos resultados, estruturado de acordo com os objetivos do estudo.

Na realização do trabalho, a preparação e tratamento da base de dados através do Excel deve preceder o carregamento do ficheiro .xlsx no R Studio. Desta forma, não é recomendado que se comece a parte do R ou Jamovi, sem antes se concluir a parte em Excel.

No final, cada grupo de trabalho deverá entregar 4 documentos/ficheiros (e-mail do docente):

- Documento Word, onde deve constar a descrição dos diferentes passos realizados pelo grupo para a execução das tarefas solicitadas no enunciado, Excel e Jamovi (com inclusão das imagens construídas nesta interface) não sendo necessária a análise de nenhuma da informação extraída;
- 2. O ficheiro de Excel com a base de dados transformada e as novas folhas com as tarefas realizadas;
- 3. O ficheiro em R markdown (.Rmd) com todos os comandos incorporados, mas não visíveis. Neste documento deve constar a informação analítica e conclusiva sobre o estudo da base de dados. Este relatório não deve conter detalhes técnicos, de forma que a leitura seja possível de ser feita por qualquer pessoa sem conhecimento na matéria.
- 4. O ficheiro em html com o relatório final.