

## Mestrado em Ciência de Dados – 2024/25

### Modelação Bayesiana

#### Enunciado do Trabalho

O relatório final deve ter a seguinte estrutura (secções):

1. **Introdução:** Deverá definir de um modo claro o objetivo da análise, o contexto do problema e sua relevância substantiva. É esperada a utilização de conhecimento prévio sobre os construtos (e instrumentos) em análise (para o estabelecimento das distribuições *a priori*).
2. **Dados:** Caracterização sumária da base de dados (não mais do que dois parágrafos); a sua dimensão e as variáveis manifestas (indicadores) e respetiva escala de resposta utilizada na análise e na caracterização.
3. **Identificação do modelo de medida** (i.e., Análise Fatorial Confirmatória) utilizando as variáveis manifestas e as respetivas variáveis latentes (i.e., dimensões). Identificando as distribuições *a priori* e respetivos hiperparâmetros, número de cadeias, a extensão das cadeias, *burn-in* e *thin*. Indique o estimador de sampling utilizado.
4. **Identificação do modelo estrutural** (i.e., Full SEM) partindo do modelo de medida anteriormente definido e estabelecendo as trajetórias diretas entre as variáveis latentes. Identificando as distribuições *a priori* e respetivos hiperparâmetros dos novos parâmetros que adicionar (i.e., coeficientes de regressão). Identificando as distribuições *a priori* e respetivos hiperparâmetros, número de cadeias, extensão das cadeias, *burn-in* e *thin*. Indique o estimador de sampling utilizado.
5. **Conclusão:** Apresente um breve resumo dos resultados principais e uma breve discussão das principais implicações do estudo realizado.
6. **Bibliografia**
7. **Anexos**

Para além desta estrutura, devem respeitar-se as seguintes indicações:

1. O relatório deve ser um documento de análise e de síntese interpretativa. Toda a informação de natureza estritamente técnica deve ser remetida para anexos, que deverão estar claramente

organizados, titulados e paginados. O relatório deve ter no máximo 15 páginas (excetuando os anexos e a folha de rosto que contém o título, nome e número dos alunos);

2. O número de secções e conteúdos devem manter-se de modo a permitir a avaliação relativa dos trabalhos, mas a designação das secções pode naturalmente mudar.

A realização do trabalho processa-se em quatro fases:

### **Fase 1: Aprovação do objetivo da análise e base de dados**

1. É preciso garantir que a base de dados contém informação adequada ao objetivo desejado (e.g., imaginando que tem uma necessidade a responder; escala das variáveis).
2. No email identifiquem claramente:
  - Os elementos do grupo (máximo 4);
  - Objetivo da análise: definir um problema num contexto de tomada de decisão em que o modelo de equações estruturais proposto possa dar resposta;
  - Base de dados e artigo (link e DOI);
  - Variáveis a utilizar: manifestas (indicadores) e latentes (dimensões) mensuradas pelos instrumentos.
3. Concluída esta fase, o conteúdo das Secções 1 e 2 do relatório final deve estar claramente definido.

### **Fase 2: Análise CFA**

Esta fase permite garantir que o trabalho vai bem encaminhado e que o modelo de medida ajustado permite a adequada mensuração das variáveis latentes. Deverá enviar o script de *R* e a base de dados final a utilizar no trabalho. O script pode ter comentários que ajudem a suportar a decisão e que posteriormente serão discutidos no relatório. Concluída esta fase, as análises correspondentes à Secção 3 do relatório final devem estar concluídas.

### **Fase 3: Análise *full SEM***

Esta fase deverá confirmar se o modelo teórico a ser testado faz sentido do ponto de vista teórico (i.e., trajetórias a serem testadas). Deverão enviar o script de *R* e a base de dados final a utilizar no trabalho. O script pode ter comentários que ajudem a suportar a decisão e que posteriormente serão discutidos no relatório. Concluída esta fase, as análises correspondentes à Secção 3 do relatório final devem estar concluídas.

**Fase 4: Envio do relatório final**

1. Apenas as Fases 1 e 4 são obrigatórias.
2. Apenas o relatório final será avaliado.
3. O ficheiro zip, cujo nome deve incluir o nome dos alunos, deverá conter três ficheiros: a base de dados original disponível num repositório (indicar link para o dataset no relatório), o script de *R* que permite replicar e obter todos os resultados constantes no relatório final (e compatível com a base de dados usada) e o pdf do relatório final. Deverá ainda enviar o artigo em pdf, ou pelo menos indicar o DOI no relatório.

NOTA: Eventualmente um grupo pode sugerir um trabalho de replicar um estudo que não seja em SEM, mas o mesmo terá de ser igualmente discutido com o docente para ver a viabilidade e equidade do esforço.

José G. Dias, 19 fevereiro de 2025