问题 - Statistics for Data Science Lecture 4

单项选择题: 以下哪个选项最能描述探索性数据分析（EDA）的作用？

A. 通过EDA，数据科学家可以通过直观的图表来验证假设

B. EDA用于数据清理和预处理

C. EDA用于从数据中提取有用的统计量并进行建模

D. EDA主要用于发现数据中的异常值、趋势和模式

填空题: 在探索性数据分析（EDA）中，首先需要\_\_\_\_\_\_\_\_\_\_整个数据集，了解数据的总体情况。

判断题: 描述性统计只涉及对数据的总结，而不进行任何形式的推断或预测。

计算题: 给定以下钻石价格数据，请计算价格的平均值。  
钻石价格（单位：美元）：[5000, 7000, 6500, 8000, 7500]

简答题: 简要解释DataExplorer包在R中的作用，并举例说明其在EDA中的应用场景。

简答题: 简要描述如何使用图表和数值总结来分析钻石价格数据集，重点关注哪些变量？

# 问题 - Statistics for Data Science Lecture 5

单项选择题: 以下哪个选项描述了推断统计学的基本作用？

A. 从样本数据推导出总体数据的结论

B. 推断统计学只用于数据的图形展示

C. 推断统计学用于计算样本的方差

D. 推断统计学主要用于机器学习模型训练

填空题: 点估计是用于近似总体参数的\_\_\_\_\_\_\_\_\_\_。

判断题: 无偏估计量的偏差为零。

计算题: 已知样本的平均值为10，样本标准差为2，样本量为25，求样本均值的标准误差。

简答题: 简要解释无偏估计量的定义，并举例说明。

简答题: 简要解释如何判断一个估计量是否是好的估计量，涉及哪些统计性质？

# 问题 - Statistics for Data Science Lecture 6

单项选择题: 以下哪个选项正确描述了假设检验的基本作用？

A. 假设检验用于验证数据是否符合正态分布

B. 假设检验用于从样本数据推断总体参数，并检验其有效性

C. 假设检验仅用于计算样本的平均值

D. 假设检验用于推导数据的图表表示

填空题: 在假设检验中，检验的假设通常是关于总体参数的\_\_\_\_\_\_\_\_\_\_。

判断题: 假设检验仅通过样本数据进行推断，并不直接涉及总体数据。

计算题: 假设一个制造公司历史上有70%的灯泡通过可靠性测试，现有新工艺生产的灯泡中，100个样本中有73个通过测试，检验该新工艺是否比旧工艺更可靠。假设显著性水平为0.05。

简答题: 简要解释假设检验的基本流程，并举例说明其应用。

简答题: 简要描述如何使用假设检验来比较两种生产工艺的可靠性。

# 问题 - Statistics for Data Science Lecture 7

单项选择题: 以下哪个选项描述了统计功效的作用？

A. 统计功效用于判断研究的样本大小是否足够

B. 统计功效是指研究拒绝零假设的能力，避免Type II错误

C. 统计功效用于估算样本的标准误差

D. 统计功效主要用于检验研究假设的显著性

填空题: 效应量用于衡量处理效果的\_\_\_\_\_\_\_\_\_\_。

判断题: 样本大小越大，统计功效通常越高。

计算题: 假设一个公交公司生产的刹车垫预计平均寿命为60,000英里，现有新生产工艺生产的刹车垫的样本为50个，已知通过的有73个。计算该工艺的统计功效。假设标准差为5,000英里，显著性水平为0.05。

简答题: 简要解释统计功效的定义，并说明提高统计功效的几种方法。

简答题: 简要描述效应量的定义和常见的效应量指标。

# 问题 - Statistics for Data Science Lecture 8

单项选择题: 以下哪个选项描述了回归分析的作用？

A. 回归分析用于分类数据

B. 回归分析用于预测或描述变量之间的关系

C. 回归分析仅用于描述定性数据

D. 回归分析主要用于计算样本的均值

填空题: 回归分析用于预测或描述\_\_\_\_\_\_\_\_\_\_之间的关系。

判断题: 统计学习和机器学习的区别在于，统计学习强调模型的解释性，机器学习强调应用的规模和预测精度。

计算题: 给定以下数据，使用线性回归预测工资（wage）与年龄（age）的关系。数据如下：  
年龄（age）：[25, 30, 35, 40, 45]  
工资（wage）：[25000, 30000, 35000, 40000, 45000]

简答题: 简要解释统计学习与机器学习的区别，并说明两者的侧重点。

简答题: 简要描述回归分析的应用，并举例说明如何使用回归模型预测工资。

# 问题 - Statistics for Data Science Lecture 9

单项选择题: 以下哪个选项描述了回归分析的基本作用？

A. 回归分析用于预测定性变量之间的关系

B. 回归分析用于描述定量变量之间的关系

C. 回归分析仅用于分类任务

D. 回归分析用于计算样本的标准差

填空题: 在回归分析中，\_\_\_\_\_\_\_\_\_\_回归是指只有一个预测变量的回归模型。

判断题: 多元线性回归模型考虑多个预测变量，而简单线性回归仅考虑一个预测变量。

计算题: 已知以下数据集：  
广告费（TV）：[100, 150, 200, 250, 300]  
销售额（Sales）：[30, 40, 50, 60, 70]  
请建立一个简单线性回归模型，预测广告费用（TV）与销售额（Sales）之间的关系。

简答题: 简要解释简单线性回归与多元线性回归的区别，并说明它们各自的应用场景。

简答题: 简要描述回归模型在广告费用与销售额关系中的应用，并说明如何进行预测。

# 问题 - Statistics for Data Science Lecture 10

单项选择题: 以下哪个选项描述了K折交叉验证的主要优势？

A. K折交叉验证比训练集验证方法计算量大

B. K折交叉验证能够平衡模型的偏差和方差

C. K折交叉验证只适用于简单的模型

D. K折交叉验证不能用于大数据

填空题: 均方误差（MSE）对\_\_\_\_\_\_\_\_\_\_比较敏感，因此在存在异常值时使用均方根误差（RMSE）更为合适。

判断题: 重复K折交叉验证通过多次重复K折交叉验证过程来减少结果的方差。

计算题: 已知回归模型的预测值与实际值分别如下：  
预测值（ŷ）：[100, 150, 200, 250]  
实际值（y）：[120, 145, 190, 260]  
请计算均方误差（MSE）。

简答题: 简要解释均方误差（MSE）和均方根误差（RMSE）的区别，并说明它们的适用场景。

简答题: 简要描述R平方（R²）的定义及其在回归分析中的应用。

# 问题 - Statistics for Data Science Lecture 10

单项选择题: 以下哪个选项描述了描述性统计的作用？

A. 描述性统计用于分析数据的因果关系

B. 描述性统计用于总结和描述数据，通常出现在研究结果部分

C. 描述性统计只用于数据的分布图表展示

D. 描述性统计主要用于对数据进行推断

填空题: 在统计研究中，描述数据集的关键特征时，通常需要回答\_\_\_\_\_\_\_\_\_\_、\_\_\_\_\_\_\_\_\_\_、\_\_\_\_\_\_\_\_\_\_等核心问题。

判断题: 描述性统计不涉及数据推断，只是对样本数据进行总结和展示。

计算题: 给定以下数据集，计算其样本均值。  
样本数据：[10, 20, 30, 40, 50]

简答题: 简要解释描述性统计的作用，并说明其在数据分析中的应用。

简答题: 简要描述在统计研究中如何理解和定义变量，并说明变量定义在数据分析中的重要性。

# 问题 - Statistics for Data Science Lecture 2

单项选择题: 以下哪个选项正确描述了原始数据和二手数据的区别？

A. 原始数据是从他人收集的数据，二手数据是直接收集的数据

B. 原始数据通常用于第一次研究收集，二手数据是历史数据

C. 二手数据通常在实验中收集，原始数据是回顾性数据

D. 二手数据不适用于市场分析，而原始数据可以

填空题: 在数据分析中，\_\_\_\_\_\_\_\_\_\_数据通常是直接收集并且是为特定研究目的收集的数据。

判断题: 定类数据具有数学意义，并且可以进行数学运算。

计算题: 给定以下数据：  
性别（Gender）：['Male', 'Female', 'Female', 'Male']  
请计算该数据集的不同类别数量。

简答题: 简要解释原始数据和二手数据的区别，并举例说明两者的使用场景。

简答题: 简要描述定类数据的定义及其在数据分析中的应用。