【参考答案】

1. 对数据集进行描述性统计（3分）

无缺失值（2分）



管理费用和利润存在少量异常值，考虑对结果有影响，选择删除处理。（3分）



连续变量的描述性统计，管理费用和利润右偏，考虑对数据形态进行转换，或删除异常值后再查看是否有好转（2分）



量纲不一致，但考虑解释性问题，暂不考虑做标准化处理（2分）

地区为汉字，且为多分类，考虑做哑变量处理（3分）



（2）异常值处理（4分）

删除异常值后，所有数值型变量基本呈正态分布（做数据形态转换也可酌情给分）



哑变量处理（4分）



 (3) 相关系数矩阵（5分）



高度相关：

[研发支出与管理费用]的相关系数值为0.9257，绝对值大于0.8，表明研发支出与管理费用之间存在高度相关关系;

[研发支出与营销费用]的相关系数值为0.9858，绝对值大于0.8，表明研发支出与营销费用之间存在高度相关关系;

[研发支出与利润]的相关系数值为0.9805，绝对值大于0.8，表明研发支出与利润之间存在高度相关关系;

[管理费用与营销费用]的相关系数值为0.8961，绝对值大于0.8，表明管理费用与营销费用之间存在高度相关关系;

[管理费用与利润]的相关系数值为0.9088，绝对值大于0.8，表明管理费用与利润之间存在高度相关关系;

[营销费用与利润]的相关系数值为0.9666，绝对值大于0.8，表明营销费用与利润之间存在高度相关关系;

自变量之间关系紧密，可能存在较严重的多重共线性

（4）回归过程，选择混合逐步方法（3分）

结果如下：





结果分析（8分）

模型R方为[0.994]，R方>0.8表明模型拟合程度较好；模型F检验P值为[0.0000e+00]，P值<0.05表明模型通过F检验，自变量（至少一个）和因变量间的线性关系是显著的，模型效果较好； ['营销费用', '地区transfer\_0', '研发支出', '地区transfer\_1']这4个自变量的t检验对应的P值是显著的，对因变量有显著影响；模型DW检验值为[1.8914]，在1.7-2.3之间，说明自变量和残差之间没有自相关性，模型构建良好；残差基本满足均值为零、方差齐性和正态性；但是，有两个及以上变量的VIF值大于10，表明该数据存在多重共线性问题，需要重新调整模型

将共线性的变量删除，先删除一个营销费用，重新回归，结果如下（4分）



从上表可知:模型R方为[0.9939]，R方>0.8表明模型拟合程度较好；模型F检验P值为[0.0000e+00]，P值<0.05表明模型通过F检验，自变量（至少一个）和因变量间的线性关系是显著的，模型效果较好 ['研发支出', '地区transfer\_0', '地区transfer\_1']这3个自变量的t检验对应的P值是显著的，对因变量有显著影响；模型DW检验值为[1.8652]，在1.7-2.3之间，说明自变量和残差之间没有自相关性，模型构建良好；残差基本满足均值为零、方差齐性和正态性；变量的VIF值均小于10，表明该数据不存在严重的多重共线性问题

(5) 写出回归方程（2分）

 ['利润'] = 46330.8542 + 0.9589\*研发支出 + 19566.7709\*地区transfer\_0 + 6451.3918\*地区transfer\_1

解释每个自变量对因变量的影响作用（2分）

在存在多重共线性的情况下，很难区分营销费用和管理费用各自对利润的具体影响。

研发支出：系数为0.9589，这意味着在其他条件不变的情况下，研发支出每增加1单位（元），企业的利润将平均增加0.9589单位（百元）。这表明研发支出对企业的利润有正向影响，即增加研发投入有助于提高企业利润。

地区transfer\_0：系数为19566.7709，这表示当企业位于某个特定地区（深圳）时，与基准地区相比，企业的利润将平均增加19566.7709单位。这说明深圳对企业利润有显著的正向影响。

地区transfer\_1：系数为6451.3918，这表示当企业位于另一个特定地区（北京）时，与基准地区相比，企业的利润将增加6451.3918单位。这也表明北京对企业利润有正向影响，但相较于transfer\_0地区，影响程度较小。

业务建议（3分）

增加研发投入：鉴于研发支出对利润的正向影响，企业可以考虑适当增加研发投入，开发新产品、改进技术或优化管理，以提高企业竞争力，从而提升利润。

优化地区布局：企业可以考虑在深圳和北京地区扩大业务规模，利用这些地区对企业利润的正向影响。同时，分析这些地区成功的原因，将其经验复制到其他地区，以提高整体利润。

加强市场调研：了解不同地区消费者的需求和偏好，针对性地调整产品和服务，以更好地满足市场，提高销售额和利润。

提高生产效率：在增加研发投入的同时，关注生产环节的优化，降低成本，提高生产效率，从而提高利润。