【参考答案】

1. 描述性统计：

该数据集共1000条记录，该数据除菜系和是否促销为离散型变量外，其他数据均为连续型数据。具体数据情况如下：（描述性统计3分）

缺失值：月收入缺失值11个，考虑缺失值较少，可以采取删除的方式解决。（缺失值及处理3分，删除和均值填充等方式均可）







数据形态：通过偏度系数和直方图可知，所有变量的偏度都不明显，基本满足正态分布（数据形态分析3分）

因此先不做boxcox变换

  

量纲分析：由箱线图和数据单位可知，数据存在量纲差异，但考虑回归的解释性要求，暂不对数据进行标准化处理（量纲分析1分，理由和方法选择合理即可）



异常值：由箱线图可以看出，月收入存在极少量异常值，查看数据可知符合实际情况，故不做异常值处理。（异常值分析3分，理由和方法选择合理即可）

标签编码：将菜系（Cuisine\_Type）和促销活动（Promotions）进行标签编码处理（2分）

原假设：菜系（Cuisine\_Type）和餐厅的月收入（Monthly\_Revenue）没有显著关联；

备择假设：菜系（Cuisine\_Type）和餐厅的月收入（Monthly\_Revenue）有显著关联。

单因素方差分析结果：



P值为0.93，大于0.05，不拒绝原假设。即菜系（Cuisine\_Type）和餐厅的月收入（Monthly\_Revenue）没有显著关联。

（原假设备择假设合理4分，p值2分，结果分析合理并且结论正确2分）

原假设：促销活动（Promotions）和餐厅的月收入（Monthly\_Revenue）没有显著关联；

备择假设：促销活动（Promotions）和餐厅的月收入（Monthly\_Revenue）有显著关联。

单因素方差分析结果：



P值为1.02e-19，小于0.05，拒绝原假设。即促销活动（Promotions）和餐厅的月收入（Monthly\_Revenue）有显著关联。

（原假设备择假设合理4分，p值2分，结果分析合理并且结论正确2分）

1. 相关系数矩阵：（2分）



判定规则：相关系数的绝对值大于0.3且小于0.8时，则认为变量间存在中度相关关系；

相关系数的绝对值大于0.8时，则认为变量间存在高度相关关系。在上述判定规则基础上，由相关系数矩阵可以看出，自变量间无高度相关，说明多重共线性问题可能不严重，还需进一步事后VIF检验（判定规则明确且多重共线性分析合理3分）

变量选择：根据上述第2步，不选择菜系作为自变量，因为它与因变量的关系不显著，只选择促销活动、和上述连续变量作为自变量建立线性回归模型（变量选择合理3分）

建模：采用混合逐步法，得到餐厅月收入预测的线性回归模型结果如下：（方法选择恰当3分）



结果显著性分析：（分析恰当6分，应包含R方（最好调整R方）、F检验、t检验、残差分析、回归方程和VIF检验6个方面的结果分析，缺一个方面扣1分）

1、模型R方为[0.6924]，0.5 ≤ R方 ≤ 0.8表明模型拟合程度一般，并不代表模型效果很好，需要结合具体业务判断模型效果

2、模型F检验P值为[4.3307e-250]，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

3、['Number\_of\_Customers', 'Marketing\_Spend', 'Menu\_Price', 'Promotions\_transfer']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

4、该模型线性回归方程为['Monthly\_Revenue'] = -1.3903 + 2.8389\*Number\_of\_Customers + 4.8399\*Marketing\_Spend + 2.0309\*Menu\_Price + 18.0209\*Promotions\_transfer

5、模型DW检验值为[1.4224]，明显偏离2，说明自变量和残差之间具有正自相关性，可以考虑增加新的自变量或选择其他的回归方法建立模型

6、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

方程解释：（2分）

线性回归方程为['Monthly\_Revenue'] = -1.3903 + 2.8389\*Number\_of\_Customers + 4.8399\*Marketing\_Spend + 2.0309\*Menu\_Price + 18.0209\*Promotions\_transfer

Number\_of\_Customers每增加一个单位，月收入平均增加2.8389个单位

Marketing\_Spend每增加一个单位，月收入平均增加4.8399个单位

Menu\_Price每增加一个单位，月收入平均增加2.0309个单位

搞促销与不搞促销相比，月收入平均增加18.0209个单位

建议：（8分，4个变量均有建议且分析恰当给8分，每少一项建议扣2分）

['Number\_of\_Customers', 'Marketing\_Spend', 'Menu\_Price', 'Promotions\_transfer']这4个自变量的t检验对应的P值是显著的，对因变量月收入（Monthly\_Revenue）均有显著正向影响。

表示这4个变量取值值越大，餐厅月收入越高，尤其是促销活动可以显著提升餐厅的月收入，可考虑多搞一些促销活动