**形成性考核一**

一、选择题

1.以下关于统计含义的描述不正确的是（）。

C.统计研究的是确定性关系

2.以产品的等级来衡量某种产品质量的好坏，则产品等级作为一种变量属于（）。

C.分类变量

3.在下列分组中，按照取值大小分组的是（）。

C.家庭按照年消费水平分组

4.按照连续变量的定义，（）一定不属于连续变量。

A.分类变量

5.下列变量中，（）属于连续变量。

B.雇员的年收入

6.下列变量中，（）属于数值变量且属于离散变量。

B.产量

7.假设某地区有800家工业企业，要研究这些企业的产品生产情况，个体是（）。

D.每一件工业产品

8.下列属于时间序列数据的是（）。

A.股票的每日收盘价

9.假设你是一家公司的数据分析员，下列数据来源中属于一手数据的是（）。

D.在公司内部通过调查得到的数据

10.工厂对生产的一批零件进行检查，通常采用（）。

D.随机抽样调查

11.（）描述了变量在不同观测值或不同观测值区间上出现的频数或频率。

A.分布

12.（）是满足特定目标的所有观测对象或要素的集合。

D.总体

13.在制作统计分布表时，将各组的频数与个体总数的比值称为（）。

B.频率

14.按照获取数据的途径不同，统计数据分为（）。

A.观测数据

C.实验数据

15.下列调查方法中，属于概率抽样的是（）。

A.简单随机抽样

B.分层抽样

D.等距抽样

16.下列属于数据分析方法的是（）。

A.回归分析

B.时间序列分析

C.指数分析

D.对比分析

17.数据审核关注的主要问题包括（）。

A.异常值

B.缺失值

C.逻辑错误

D.重复值

二、综合应用题

18.一家小微企业共有20位员工，员工的年龄数据如下。

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 39 | 47 | 31 | 33 | 42 | 36 | 38 | 36 | 44 | 47 |
| 48 | 42 | 29 | 42 | 39 | 39 | 46 | 40 | 34 | 40 |

请结合数据和所学知识回答下列问题：

(1)员工年龄的变量类型是（）。

A.数值变量

(2)在制作员工年龄统计分布表时，第一步是（）。

B.根据变量的观测值进行分组

(3)在根据变量观测值进行分组时，第一步是（）。

C.找到最大值和最小值

(4)在Excel中，制作统计分布表用到的主要功能是（）。

A.数据透视表

(5)请使用Excel制作员工年龄变量的频率分布表（设置组距是4）。根据结果可知，占比最多的年龄组是（）。

C.37-40岁

(6)请使用Excel制作员工年龄变量的频率分布表（设置组距是4）。根据结果可知，占比最少的年龄组是（）。

A.29-32岁

(7)请使用Excel制作员工年龄变量的频率分布表（设置组距是4）。根据结果可知，年龄小于等于40岁的占比是（）。

B.0.6

(8)请使用Excel制作员工年龄变量的频率分布表（设置组距是4）。根据结果可知，年龄小于33岁的占比是（）。

B.0.1

**形成性考核二**

一、选择题

1.在柱形图中，使用（）代表每个类别对应的数值（频数、频率或其他数值结果）。

C.高度

2.在饼图中，使用圆内扇形（）表示数值大小。

A.面积

3.散点图是描述（）变量之间关系的图形。

B.两个

4.现在有关于两支球队的几项关键数据，为了横向对比两支球队的表现，适合采取的数据可视化方法是（）。

D.雷达图

5.在热力图中，用来反映变量取值变化规律的特征是（）。

A.颜色

6.当数据中存在异常值时，不适合用于度量变量分布特征的指标是（）。

C.均值

7.已知总体容量为N，抽取了一个样本量是n的样本，则在样本方差的计算公式中，分母是（）。

A.n-1

8.变异系数的计算结果是一个（）。

A.相对数

9.对于右偏分布而言，在箱线图中，中位数更靠近（）的位置。

B.下四分位数

10.对于对称分布而言，直方图的形状呈现（）特征。

C.对称

11.度量事件发生可能性的测量指标是（）。

C.概率

12.在正态分布中，共有（）参数。

A.2

13.假设z服从标准正态分布，请利用Excel或查表方式计算，P(Z<0.3)，计算结果是（）。

D.0.62

14.假设z服从标准正态分布，请利用Excel或查表方式计算，P(Z<?)=0.38，求解得到?的结果是（）。

B.-0.3

15.点估计方法背后的理论依据是（）。

A.大数定律

16.在假设检验中，显著性水平实质上代表的是（）。

C.犯第一类错误的概率

二、综合应用题

17.为了解学习者在线学习情况，课题研究组从某校本科生和专科生中分别各随机抽取了1000人，调查每日在线学习时长(分钟)。根据调查数据计算度量指标，结果如下表所示：

(1)本次搜集数据的抽样方法是（）。

A.简单随机抽样

(2)描述学习者每日学习时长分布的数据可视化方法是（）。

C.直方图

(3)下列数据可视化方法中，适合描述本科生每日平均学习时长与专科生每日平均学习时长对比的是（）。

A.柱形图

(4)下列数据可视化方法中，适合用于对比分析本科生与专科生多个相关指标的是（）。

D.雷达图

(5)在Excel中，计算样本方差的函数是（）。

C.VAR.S函数

(6)根据偏度的计算结果可知，本科生组与专科生组的分布形状分别属于（）。

C.轻微右偏分布；严重右偏分布

(7)在95%置信水平下，计算专科生组关于总体均值的置信区间，结果是（）。

C.

(8)有研究显示，该校本科生的平均在线学习时长等于200分钟。现在需要使用假设检验方法基于样本数据验证这一结论是否正确，假设检验的拒绝域在（）。

C.抽样分布的两侧

(9)有研究显示，该校本科生的平均在线学习时长为200分钟。现在需要使用假设检验方法基于样本数据验证这一结论是否正确，适合的检验统计量是（）。

A.Z统计量

(10)在Excel中进行假设检验的计算时，计算临界值会用到的函数是（）。

C.NORM.S.INV函数

(11)在Excel中求解本科生组和专科生组区间估计结果时，用于计算估计误差的函数是（）。

A.CONFIDENCE.NORM函数

(12)在Excel中进行假设检验的计算时，计算P值会用到的函数是（）。

D.NORM.S.DIST函数

**形成性考核三**

一、选择题

1.下列属于结构相对数的是（）。

A.恩格尔系数

2.下列属于比例相对数的是（）。

B.人口性别比

3.通过将同类现象在同一时间、不同空间的指标数值进行对比，来分析不同空间之间的不平衡性的对比分析方法是（）。

B.空间比较分析

4.一家企业2023年计划生产商品1000件，实际完成500件，按照计划完成度相对数的计算公式可得，产量的计划完成相对数是（）。

C.0.5

5.线性相关系数的取值范围是（）。

A.[-1,1]

6.如果两个变量的线性相关系数计算结果是0.7，则说明二者之间可能存在（）。

B.线性正相关

7.下列属于时点时间序列的是（）。

A.年末总资产

8.平稳时间序列只包含（）成分。

A.不规则性

二、综合应用题

9.已知某商场三种商品的销售资料，如下表所示，请根据表中信息回答问题。

(1)拉氏销售量指数属于（）。

A.数量指标指数

(2)帕氏价格指数属于（）。

B.质量指标指数

(3)在表格中，计算【1】位置上数值的方式包括（）。

A.600\*12

B.28700-18000-3500

(4)根据表格数据计算帕氏销售量指数，计算公式正确的是（）。

D.40600/32500

(5)根据平均指数与综合指数的关系，与报告期总值加权的销售量调和平均指数结果一致的是（）。

D.帕氏销售量指数

10.表1给出了一组消费性支出和可支配收入的数据。为了探究两个变量之间的联系，基于Excel工具，使用相关分析、回归分析等方法展开研究，结果如图1所示。

(1)从散点图中，可以得到的初步结论是（）。

A.消费性支出和可支配收入之间的关系几乎呈现线性特征

(2)相关分析结果的含义是（）。

C.消费性支出和可支配收入之间存在高度线性正相关关系

(3)从回归分析的Excel输出结果看，模型拟合优度检验中R平方的值是（）。

B.0.9979

(4)从回归分析的Excel输出结果看，F检验的P值远远小于0，说明模型的显著性检验结论是（）。

B.两个变量之间的线性关系显著

(5)根据回归分析结果，可以写出最终估计得到的一元线性回归方程的表达式是（）。

A.ŷ=54+0.53x

(6)回归系数b的估计值的含义是（）。

D.每增加1元的可支配收入，消费性支出平均增加0.53元

11.现有数据如下表所示，

请结合数据回答下列问题：

(1)现在需要使用数据可视化方法描述2015-2020年年末人口数变化情况，请选择适合的数据可视化方法（）。

B.折线图

(2)现在需要使用数据可视化方法描述年末人口数与最终消费之间的关系，请选择适合的数据可视化方法（）。

A.散点图

(3)计算2015-2020年最终消费序列平均发展水平的公式是（）。

D.2015-2020年最终消费数值的算术平均数

(4)以2015年为基期，年末人口数2017年的环比发展速度等于（）。

B.140011/139232

(5)2015-2020年最终消费的平均发展速度为（）。

A.

(6)在使用Excel计算定基发展速度时，会用到绝对引用符号，表示正确的是（）。

A.$

**形成性考核四**

**共享单车骑行数据分析报告**

**一．背景与目的**

1. 行业背景

共享单车作为城市短途出行的重要方式，2023年全国用户规模达3亿人，日均订单量超4000万次。某城市共享单车企业面临车辆调度效率低、用户流失率高等问题。

2. 业务问题

- 高峰时段车辆供需失衡

- 用户单次骑行时长分布异常

- 会员续费率低于行业平均水平（45% vs 60%）

3. 分析目标

- 挖掘用户骑行行为规律

- 识别高价值用户特征

- 优化车辆调度策略

4. 分析意义

预计提升车辆周转率15%，年增收约120万元

**二．分析思路**

分析方法组合：

1. 描述性统计：骑行时长分布、用户类型占比

2. 数据可视化：热力图（时段-区域需求）、折线图（骑行趋势）

3. 假设检验：独立样本t检验（会员与非会员骑行差异）

**三．分析过程**

（一）数据来源及变量说明

数据概况：

- 来源：企业2023年Q2运营数据（脱敏处理）

- 样本量：50,000条骑行记录

- 核心变量：

| 变量名 | 类型 | Excel公式应用示例 |

| 骑行时长 | 数值型 | `=TEXT(结束时间-开始时间,"h:mm")` |

| 用户类型 | 分类型 | `=IF(会员标识=1,"会员","散客")` |

| 起始区域 | 文本型 | `=VLOOKUP(站点ID,区域对照表,2,FALSE)` |

（二）数据预处理

1. 异常值处理：

- 删除骑行时长<1分钟或>4小时的记录

- 公式：`=FILTER(A2:E50000,(E2:E50000>0.0167)\*(E2:E50000<0.1667))`

2. 数据分箱：

- 将骑行时长划分为5个区间

- 公式：`=FLOOR(骑行时长\*24,0.5)`（单位：小时）

（三）数据分析

1. 描述性统计分析

（1）数据分布特征

| 指标 | 计算公式 | 结果 |

| 平均骑行时长 | `=AVERAGE(D2:D50000)` | 18.6分钟 |

| 会员占比 | `=COUNTIF(B2:B50000,1)/COUNTA(B2:B50000)` | 43.7% |

（2）可视化分析

① 时段需求热力图（Excel操作）：

- 步骤：插入→条件格式→色阶

- 结论：早8-9点、晚18-19点形成双高峰

② 用户类型对比柱状图（Excel操作）：

- 步骤：数据透视表→插入簇状柱形图

- 结论：会员平均骑行次数是散客的2.3倍

2. 假设检验分析

检验命题：会员与非会员骑行时长存在显著差异

- Excel操作：数据分析工具库→t-检验: 双样本异方差

- 结果：t=4.32，p=0.00017<0.05

- 结论：拒绝原假设，两类用户骑行行为差异显著

**四．结论与建议**

核心结论

1. 会员用户贡献65%的订单量但仅占43.7%的用户基数

2. 工作日早高峰车辆缺口达23%

运营建议

1. 动态调度策略：

- 依据热力图数据，在晚高峰前向商务区预调度车辆

- 计算公式：`=FORECAST.ETS(未来时段需求量,历史数据,季节性周期)`

2. 会员体系优化：

- 对月骑行超20次的用户自动升级为VIP会员

- 流失预警公式：`=IF(最后骑行日期>TODAY()-30,"活跃","流失")`

**五．附录**

1. Excel关键公式说明

| 功能 | 公式示例 |

|---|---|

| 时长转换 | `=TEXT(时间差,"h""小时""m""分""")` |

| 区域统计 | `=COUNTIFS(起始区域列,"A区",用户类型列,"会员")` |

2. 数据抽样方法

采用系统抽样法，每100条记录抽取1条，保证数据代表性