

统计学原理-1

单选题

问题 1:

二分类变量的总体比例通过（）可以转化为总体均值

选项:标准化变换

选项:0-1 变换

选项:位移变换

选项:中心化变换

答案: 0-1 变换

问题 2:

假设在一个不透明的袋中有 10 个球，其中白色的球有 5 个，红色的球有 2 个，绿色的球有 3 个。现在随机抽取一个球，请问被抽到的球是白色的概率是多少？

（）

选项:0.5

选项:0.3

选项:0.4

选项:0.2

答案: 0.5

问题 3:

现有一组数据如下，{1,4,5,5,10,10,7,5}，该组数据的众数是（）。

选项:5

选项:4

选项:10

选项:7

答案: 5

问题 4:

众数是指（）。

选项:出现次数最少的频数

选项:出现次数最少的变量值

选项:出现次数最多的变量值

选项:出现次数最多的频数

答案: 出现次数最多的变量值

问题 5:

下列属于时点时间序列的是（）。

选项:年末总资产

选项:第一季度营业收入

选项:年净资产收益率

选项:人均 GDP

答案: 年末总资产

问题 6:

为了横向对比 2019—2022 年公司三项主要的财务指标数据，适合采用的数据可视化方法是（）。

选项:散点图

选项:热力图

选项:文字云

选项:雷达图

答案: 雷达图

问题 7:

（）是指根据现象之间的客观联系，将两个有联系的统计指标进行对比，反映研究对象的变化或差异。

选项:回归分析

选项:对比分析

选项:相关分析

选项:时间序列分析

答案: 对比分析

问题 8:

按照连续变量的定义，（）一定不属于连续变量。

选项:分类变量

选项:数值变量

选项:实数变量

选项:可积变量

答案: 分类变量

问题 9:

（）刻画了变量在不同观测值或不同观测区间上的全部信息。

选项:分布

选项:分布特征

选项:总体

选项:个体

答案: 分布

问题 10:

下列选项中，属于分类变量的是（）。

选项:一个家庭资产的类型

选项:一个人每天通勤所需时长

选项:单位时间内光顾店铺的顾客人数

选项:一家公司的年销售量

答案: 一个家庭资产的类型

多选题

问题 1:
按照时间序列数据的变化趋势,时间序列可以分为 ()。

选项:平稳时间序列

选项:绝对数时间序列

选项:相对数时间序列

选项:非平稳时间序列

答案: 平稳时间序列

答案: 非平稳时间序列

问题 2:
下列属于集中趋势特征度量指标的是 ()。

选项:四分位差

选项:中位数

选项:众数

选项:变异系数

答案: 中位数

答案: 众数

问题 3:
在下列常见随机变量的概率分布中,属于离散随机变量分布的是 ()。

选项:泊松分布

选项:二项分布

选项:正态分布

选项:指数分布

答案: 泊松分布

答案: 二项分布

问题 4:
按照数据与时间的关系不同,统计数据分为 ()。

选项:时间序列数据

选项:横截面数据

选项:面板数据

选项:观测数据

答案: 时间序列数据

答案: 横截面数据

答案: 面板数据

问题 5:
应用数据可视化方法的注意事项包括 ()。

选项:注意数据可视化的受众

选项:图形设计要直观

选项:图形的色彩风格统一

选项:简洁地展示关键信息

答案: 注意数据可视化的受众

答案: 图形设计要直观

答案: 图形的色彩风格统一

答案: 简洁地展示关键信息

判断题

问题 1:
总体和个体的概念是成对出现的,需要相互定义。()

选项:√

选项:×

答案:√

问题 2:
季节性是指时间序列在一段时期内出现重复的周期性波动特征。()

选项:√

选项:×

答案:√

问题 3:
无论样本量 n 是奇数还是偶数,中位数的计算公式相同。()

选项:√

选项:×

答案:×

问题 4:
相关表是指根据两个变量的对应数值编成的统计表。()

选项:√

选项:×

答案:√

问题 5:
在大样本情况下,样本比例的抽样分布与正态分布无关。()

选项:v

选项:x

答案: x

问答题

问题 1:

在一些统计调查的场景中使用抽样调查是非常必要的,请列举 2 个适合使用抽样调查的场景。

参考 1: 常见的场景举例:1.当一项调查具有破坏性时,需要使用抽样调查,比如在工厂调查灯泡的使用寿命。2.当一项调查需要耗费大量的人力、财力和物力时,无法对总体进行全面调查,需要使用抽样调查。3.当对调查要求不高,只是进行初步试验性的调查时,适合使用抽样调查。4.现场调查的误差对调查精度影响更大时,适合采用抽样调查,从而将更多资源配置给现场调查。(本题答案不唯一,言之有理,即可给分)

问题 2:

简述假设检验的基本步骤。

参考 1: 假设检验的基本步骤包括以下 7 步:第 1 步,说明要检验的问题,提出原假设和备选假设。第 2 步,设定显著性水平 α (风险水平)。第 3 步,根据检验问题,确定适合的检验统计量及抽样分布。第 4 步,根据给定的显著性水平和抽样分布,确定临界值和拒绝域。第 5 步,利用样本数据,计算检验统计量的值并计算 P 值。第 6 步,做出

统计决策。一是临界值法,即利用检验统计量的值与临界值做比较,来决定是否拒绝原假设;二是 P 值法,即利用 P 值与显著性水平做比较,来判断是否拒绝原假设。第 7 步,给出关于实际问题的结论。

综合题

为了了解某产品的使用寿命情况,收集了同批次 100 个产品的使用寿命(小时)数据。现在使用描述性统计分析方法进行研究,相关结果如下图所示:请回答下列问题:

问题 1:

员工年龄的变量类型是()。

选项:数值变量

选项:分类变量

选项:离散变量

选项:属性变量

答案: 数值变量

问题 1:

在根据变量观测值进行分组时,第一步是()。

选项:计算组距

选项:确定组限

选项:找到最大值和最小值

选项:确定组数

答案: 找到最大值和最小值

问题 1:

从数据结果看,产品寿命分布呈现()。

选项:左偏扁平分布

选项:右偏扁平分布

选项:左偏尖峰分布

选项:右偏尖峰分布

答案: 右偏尖峰分布

问题 1:

图中使用的数据可视化方法是()。

选项:散点图

选项:饼图

选项:气泡图

选项:直方图

答案: 直方图

问题 1:

在频率分布表中,根据结果可知,占比最多的寿命时长组是()。

选项:701-710 小时

选项:671-680 小时

选项:691-700 小时

选项:681-690 小时

答案: 691-700 小时

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:请结合数据结果回答下列问题:

问题 2:

在 Excel 中,计算均值的函数是 ()。

选项:AVERAGE 函数

选项:MEDIAN 函数

选项:VAR.S 函数

选项:MODE.SNGL 函数

答案: AVERAGE 函数

问题 2:

根据表格中的指标结果,本科生组和专科生组的极差分别是 ()。

选项:270; 300

选项:300; 270

选项:150; 210

选项:210; 150

答案: 270; 300

问题 2:

在 95%置信水平下,计算本科生组关于总体均值的置信区间,估计误差的计算结果是 ()。

选项:1.96*92.466

选项:1.96*2.924

选项:1.96*92.466/1000

选项:1.96*150/1000

答案: 1.96*2.924

问题 2:

在表格中呈现的指标中,“样本均值的标准差”与“标准差”之间的关系是 ()。

选项:样本均值标准差等于标准差除以 n

选项:样本均值标准差等于标准差除以根号 n

选项:标准差等于样本均值标准差乘以 n

选项:样本均值标准差等于标准差开方

答案: 样本均值标准差等于标准差除以根号 n

问题 2:

在 Excel 中进行假设检验的计算时,计算 P 值会用到的函数是()。

选项:CONFIDENCE.NORM 函数

选项:CONFIDENCE.T 函数

选项:NORM.S.INV 函数

选项:NORM.S.DIST 函数

答案: NORM.S.DIST 函数

统计学原理-2

单选题

问题 1:

统计学的基本理论建立在抽样基础上,具体指的是 ()。

选项:简单随机抽样

选项:分层抽样

选项:典型抽样

选项:重点抽样

答案: 简单随机抽样

问题 2:

假设在一个不透明的袋中有 4 个质地均匀的小球。现在随机抽取 1 个球,任意一个小球被抽中的概率是 ()

选项:0.25

选项:0.5

选项:0.4

选项:0.6

答案: 0.25

问题 3:

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 ()。

选项:众数

选项:中位数

选项:均值

选项:四分位数

答案: 均值

问题 4:

现有一组数据如下, {1,2,5,7,10,13,15},该组数据的极差是 ()。

选项:6

选项:14

选项:7

选项:10

答案: 14

问题 5:

()是时间序列中对应某个时期(或时点)的指标数值,反映社会经济现象在某个时期(或时点)实际达到的规模和水平。

- 选项:发展水平
- 选项:发展速度
- 选项:增长速度
- 选项:增长量

答案: 发展水平

问题 6:

为了描述 2010-2020 年我国居民恩格尔系数的变化趋势,适合采用的数据可视化方法是 ()。

- 选项:饼图
- 选项:直方图
- 选项:散点图
- 选项:折线图

答案: 折线图

问题 7:

如果想要了解一家公司 2022 年较 2021 年的经营变化情况,适合使用的对比分析方法是 ()。

- 选项:动态对比分析
- 选项:空间比较分析
- 选项:强度对比分析
- 选项:比例对比分析

答案: 动态对比分析

问题 8:

下列选项中,属于分类变量的是 ()。

- 选项:企业年销售量
- 选项:年龄
- 选项:产品品种
- 选项:家庭收入水平

答案: 产品品种

问题 9:

()是按照随机原则选择若干个体的一种调查方式。

- 选项:配额抽样调查
- 选项:判断抽样调查
- 选项:概率抽样调查
- 选项:普查

答案: 概率抽样调查

问题 10:

2010-2021 年所有直辖市的 GDP 数据属于 ()。

- 选项:实验数据
- 选项:时间序列数据
- 选项:面板数据
- 选项:横截面数据

答案: 面板数据

多选题

问题 1:

在时间序列指标分析中,下列属于水平分析指标的是 ()。

- 选项:平均发展水平
- 选项:平均增长速度
- 选项:增长量
- 选项:平均增长量

答案: 平均发展水平

答案: 增长量

答案: 平均增长量

问题 2:

下列属于离散程度特征度量指标的是 ()。

- 选项:标准差
- 选项:中位数
- 选项:峰度
- 选项:极差

答案: 标准差

答案: 极差

问题 3:

区间估计和点估计的理论核心是 ()。

- 选项:中心极限定理
- 选项:大数定理
- 选项:柯西中值定理
- 选项:拉格朗日定理

答案: 中心极限定理

答案: 大数定理

问题 4:

下列调查方法中,属于概率抽样的是 ()。

选项:简单随机抽样

选项:分层抽样

选项:方便抽样

选项:等距抽样

答案: 简单随机抽样

答案: 分层抽样

答案: 等距抽样

问题 5:

常用于数值变量的数据可视化方法包括 ()。

选项:直方图

选项:南丁格尔玫瑰图

选项:散点图

选项:气泡图

答案: 直方图

答案: 散点图

答案: 气泡图

判断题

问题 1:

在概率抽样中,一个个体是否将被抽中是不确定的。()

选项:√

选项:×

答案:√

问题 2:

简单移动平均法是指将使用时间序列中最近 k 期的观测值计算

的平均值作为下一期预测值的一种预测方法。()

选项:√

选项:×

答案:√

问题 3:

对于右偏分布而言,直方图的特点是右侧有一个长的拖尾。()

选项:√

选项:×

答案:√

问题 4:

线性相关关系能体现变量之间的因果关系。()

选项:√

选项:×

答案:×

问题 5:

在假设检验中,可以使用 P 值来判断检验的结果。()

选项:√

选项:×

答案:√

问答题

问题 1:

简述数据分析的基本流程。

参考 1: (1)确定实际业务问题。
(2)从实际业务问题中抽象出

统计问题。(3)根据统计问题确定变量、收集数据。(4)对收集到的数据进行数据整理。

(5)选择适合的数据分析方法研究数据。(6)分析统计结果,得出统计结论。(7)结合实际问题,给出实际结论和决策建议。

问题 2:

简述随机试验需要满足的三个条件。

参考 1: 随机试验需要满足的三个条件分别是: 1)试验是可重复的; 2)试验的所有可能结果是已知的; 3)试验前,一次具体试验的结果无法确知。

综合题

现有 2018-2022 年国内生产总值数据如图 1 所示。请结合图中数据回答下列问题:

问题 1:

图中用于描述 2018-2022 年国内生产总值的数据可视化方法是 ()。

选项:箱线图

选项:柱形图

选项:直方图

选项:饼图

答案: 柱形图

问题 1:

图中用于描述 2018-2022 年国内生产总值增长速度的数据可视化方法是 ()。

选项:箱线图

选项:2021 年

选项:1.96*92.466

选项:柱形图

选项:2022 年

选项:1.96*2.924

选项:折线图

答案: 2020 年

选项:1.96*92.466/1000

选项:饼图

选项:1.96*150/1000

答案: 折线图

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: 请结合数据结果回答下列问题:

答案: 1.96*2.924

问题 1:

国内生产总值数据的极差等于 ()。

选项:1210207-919281

选项:1149237-919281

选项:1210207-986515

选项:1149237-919281

答案: 1210207-919281

问题 2:

在 Excel 中, 计算均值的函数是 ()。

选项:AVERAGE 函数

选项:MEDIAN 函数

选项:VAR.S 函数

选项:MODE.SNGL 函数

答案: AVERAGE 函数

问题 2:

在表格中呈现的指标中, “样本均值的标准差”与“标准差”之间的关系是 ()。

选项:样本均值标准差等于标准差除以 n

选项:样本均值标准差等于标准差除以根号 n

选项:标准差等于样本均值标准差乘以 n

选项:样本均值标准差等于标准差开方

答案: 样本均值标准差等于标准差除以根号 n

问题 1:

国内生产总值数据的中位数等于 ()。

选项:2020 年的数值

选项:2020 年的数值与 2021 年的数值的平均值

选项:2019 年的数值与 2020 年的数值的平均值

选项:2021 年的数值

答案: 2020 年的数值

问题 2:

根据表格中的指标结果, 本科生组和专科生组的极差分别是 ()。

选项:270; 300

选项:300; 270

选项:150; 210

选项:210; 150

答案: 270; 300

问题 2:

在 Excel 中进行假设检验的计算时, 计算 P 值会用到的函数是()。

选项:CONFIDENCE.NORM 函数

选项:CONFIDENCE.T 函数

选项:NORM.S.INV 函数

选项:NORM.S.DIST 函数

答案: NORM.S.DIST 函数

问题 1:

从增长速度的数据结果可知, 2018-2022 年国内生产总值增长最慢的是 ()。

选项:2019 年

选项:2020 年

问题 2:

在 95%置信水平下, 计算本科生组关于总体均值的置信区间, 估计误差的计算结果是 ()。

统计学原理-3

单选题

问题 1:

下列说法中，表述正确的是（）。

选项:置信水平越大，估计的可靠性就越大

选项:置信水平越大，估计的可靠性就越小

选项:置信水平越小，估计的可靠性就越大

选项:置信水平的大小与估计的可靠性无关

答案: 置信水平越大，估计的可靠性就越大

问题 2:

度量事件发生可能性的指标是（）。

选项:随机变量

选项:随机试验

选项:概率

选项:事件

答案: 概率

问题 3:

变异系数的计算结果是一个（）。

选项:相对数

选项:绝对数

选项:平均数

选项:正数

答案: 相对数

问题 4:

在下列分布特征中，反映分布形状特征的是（）。

选项:偏度

选项:均值

选项:众数

选项:中位数

答案: 偏度

问题 5:

平均增长速度的计算方式是（）。

选项:定基增长速度-1

选项:平均发展速度-1

选项:平均发展速度+1

选项:各期环比增长速度求平均数

答案: 平均发展速度-1

问题 6:

通过颜色冷暖来反映变量数值大小的一类图形是（）。

选项:直方图

选项:热力图

选项:文字云

选项:南丁格尔玫瑰图

答案: 热力图

问题 7:

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（）。

选项:0.33

选项:0.25

选项:0.5

选项:0.75

答案: 0.5

问题 8:

下列属于时间序列数据的是（）。

选项:股票的每日收盘价

选项:2021 年不同地区的 GDP

选项:全国 70 个城市的新房价格

选项:10 家上市公司 2022 年末的净资产数据

答案: 股票的每日收盘价

问题 9:

下列变量中，属于连续变量的是（）。

选项:人体的身高体重

选项:合格品数量

选项:图书馆藏书的数量

选项:一批电脑的开机不正常台数

答案: 人体的身高体重

问题 10:

下列变量中，（）属于数值变量且属于连续变量。

选项:店铺每月的营业额

选项:店铺商品每月的销售量

选项:F 分布

选项:单位时间内浏览店铺的顾客数

选项:几何分布

选项:店铺经营的商品种类

选项:二项分布

答案: 店铺每月的营业额

选项:T 分布

多选题

问题 1:

在时间序列指标分析中,下列属于速度分析指标的是 ()。

答案: 几何分布

答案: 二项分布

选项:发展速度

问题 4:

下列调查方法中,属于概率抽样的是 ()。

选项:增长速度

选项:简单随机抽样

选项:发展水平

选项:整群抽样

选项:增长量

选项:方便抽样

选项:判断抽样

答案: 发展速度

答案: 简单随机抽样

答案: 增长速度

答案: 整群抽样

问题 2:

下列属于集中趋势特征度量指标的是 ()。

问题 5:

根据数据可视化结果,可以从图中可以得到的结论包括 ()。

选项:方差

选项:均值

选项:标准差

选项:中位数

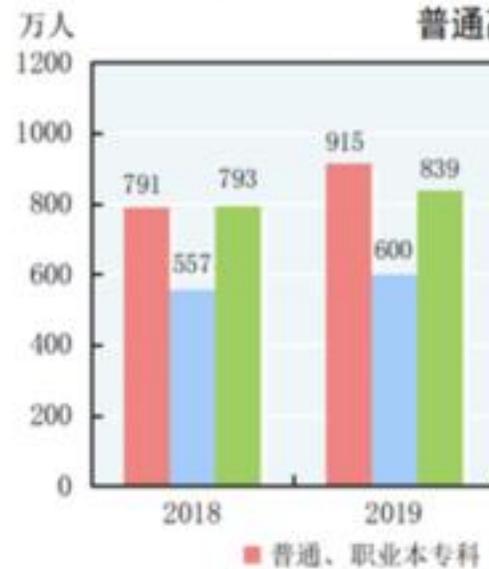
答案: 均值

答案: 中位数

问题 3:

在下列常见随机变量的概率分布中,属于离散随机变量分布的是 ()。

图21 2018-2022年普通



选项:普通、职业本专科的招生人数逐年递增

选项:普通高中的招生人数逐年递增

选项:中等职业教育的招生人数逐渐递增

选项:相比 2021 年,中等职业教育的招生人数在 2022 年出现回落

答案: 普通、职业本专科的招生人数逐年递增

答案: 普通高中的招生人数逐年递增

答案: 相比 2021 年,中等职业教育的招生人数在 2022 年出现回落

判断题

问题 1:

普查和抽样调查的划分是以最后取得的数据是否完全来划分的。（）

选项:√

选项:×

答案:×

问题 2:

时间序列反映了某个事物或某种现象随时间变化的发展规律。（）

选项:√

选项:×

答案:√

问题 3:

箱线图和分布特征度量指标之间没有联系。（）

选项:√

选项:×

答案:×

问题 4:

在散点图中,如果散点在一条曲线周围波动,则两个变量的相关关系是非线性相关。（）

选项:√

选项:×

答案:√

问题 5:

样本量 n 影响总体均值的区间估计结果。（）

选项:√

选项:×

答案:√

问答题

问题 1:

列举三个二手数据的获取途径。

参考 1: (1) 国家统计局和政府部门公开发布的官方数据。(2) 市场调查公司或数据公司的专题数据。(3) 微观调查数据库。(4) 行业数据。(答案合理,即可酌情给分)

问题 2:

简述饼图和柱状图的异同点。

参考 1: 相同点: 饼图和柱状图都是描述数据分布的一种分布图, 都能反映变量的总体结构, 二者都适用于描述分类变量的分布。不同点: 饼图可以反映分类变量或分组变量数目极少时的分布, 适合于变量值极少情况; 柱状图在反映分类变量分布时能够容纳的变量值个数比饼图多一些; 相比饼图, 柱状图更能显现不同类别的频率差异。

综合题

现有 2018-2022 年国内生产总值数据如图 1 所示。请结合图中数据回答下列问题:

图1 2018-2022年国内生产总值



问题 1:

图中用于描述 2018-2022 年国内生产总值的数据可视化方法是（）。

选项:箱线图

选项:柱形图

选项:直方图

选项:饼图

答案: 柱形图

问题 1:

图中用于描述 2018-2022 年国内生产总值增长速度的数据可视化方法是（）。

选项:箱线图

选项:柱形图

选项:折线图

选项:饼图

答案: 折线图

问题 1:

国内生产总值数据的极差等于（）。

选项:1210207-919281

选项:1149237-919281

选项:1210207-986515

选项:1149237-919281

答案: 1210207-919281

问题 1:

国内生产总值数据的中位数等于 ()。

选项:2020 年的数值

选项:2020 年的数值与 2021 年的数值的平均值

选项:2019 年的数值与 2020 年的数值的平均值

选项:2021 年的数值

答案: 2020 年的数值

问题 1:

从增长速度的数据结果可知, 2018-2022 年国内生产总值增长最慢的是 ()。

选项:2019 年

选项:2020 年

选项:2021 年

选项:2022 年

答案: 2020 年

为了解某大型社区居民的收入水平, 研究人员从全社区中随机抽取了 300 名居民, 调查收入 (元)。相关描述性度量指标的计算结果如下图所示: 请结合数据结果回答下列问题:

均值	3000
标准误差	18.2574186
众数	2400
中位数	2500
方差	100000
标准差	316.227766

问题 2:

案例中使用的抽样调查方法是 ()。

选项:简单随机抽样

选项:整群抽样

选项:分层抽样

选项:等距抽样

答案: 简单随机抽样

问题 2:

从数据结果看, 被调查居民的收入呈现 ()。

选项:左偏扁平分布

选项:右偏扁平分布

选项:左偏尖峰分布

选项:右偏尖峰分布

答案: 右偏尖峰分布

问题 2:

在 95%置信水平下, 计算关于居民平均收入水平的置信区间, 估计误差 (保留两位小数) 的结果是 ()。

选项:1.96*18.26

选项:1.96*18.26/300

选项:1.96*316.23

选项:1.96*316.23/300

答案: 1.96*18.26

问题 2:

有报告显示该社区居民的平均收入水平等于 2900 元。现在需要利用假设检验方法验证这一结论是否正确, 适合的检验统计量是 ()。

选项:Z 统计量

选项:t 统计量

选项:F 统计量

选项:卡方统计量

答案: Z 统计量

问题 2:

有报告显示该社区居民的平均收入水平等于 2900 元。现在需要利用假设检验方法验证这一结论是否正确, 假设检验的拒绝域在 ()。

选项:抽样分布的左侧

选项:抽样分布的右侧

选项:抽样分布的两侧

选项:抽样分布的中间

答案: 抽样分布的两侧

单选题

问题 1:
在点估计中,常常用于估计总体比例的估计量是 ()。

- 选项:样本标准差
- 选项:样本方差
- 选项:样本比例
- 选项:样本众数

答案: 样本比例

问题 2:
在总体方差的计算公式中,需要首先计算的是 ()。

- 选项:总体均值
- 选项:样本均值
- 选项:总体比率
- 选项:总体总量

答案: 总体均值

问题 3:
将变量各观测值按从小到大顺序排列,处于中间位置的数值是 ()。

- 选项:最小值
- 选项:最大值
- 选项:众数
- 选项:中位数

答案: 中位数

问题 4:
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 ()。

- 选项: $n-1$
- 选项: N
- 选项: $n+1$
- 选项: $N-1$

答案: $n-1$

问题 5:
一般将不同时期的发展水平加以平均得到的平均数称为 ()。

- 选项:平均发展水平
- 选项:平均发展速度
- 选项:平均增长速度
- 选项:增长量

答案: 平均发展水平

问题 6:
一家公司核心部门有三个, 为了描述核心部门员工的性别分布, 适合采用的数据可视化方法是 ()。

- 选项:饼图
- 选项:簇状柱形图
- 选项:雷达图
- 选项:南丁格尔玫瑰图

答案: 簇状柱形图

问题 7:
与拉氏指数计算结果一致的是 ()。

选项:基期总值加权的算术平均指数

选项:报告期总值加权的算术平均指数

选项:报告期总值加权的调和平均指数

选项:几何平均指数

答案: 基期总值加权的算术平均指数

问题 8:
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 ()。

- 选项:时间变量
- 选项:数值变量
- 选项:分类变量
- 选项:连续变量

答案: 分类变量

问题 9:
下列分组中按观测值类别分组的是 ()。

- 选项:产成品按品种分组
- 选项:企业按年生产产量分组
- 选项:家庭年收入金额分组
- 选项:人口按年龄分组

答案: 产成品按品种分组

问题 10:
下列变量中, () 属于连续变量。

- 选项:运动员获得的奖牌数
- 选项:雇员的年收入

选项:患者的常住地

选项:指数分布

答案: 交通通信价格环比下降
1.2%

选项:单位时间内浏览店铺的
顾客数

选项:几何分布

答案: 食品烟酒类价格环比上涨
0.6%

答案: 雇员的年收入

选项:正态分布

答案: 居住价格持平

选项:泊松分布

多选题

答案: 指数分布

答案: 正态分布

问题 1:

平稳时间序列预测方法包括
()。

选项:指数平滑法

问题 4:

下列属于统计应用场景的是()。

选项:移动平均法

选项:金融时间序列分析

选项:线性趋势模型法

选项:财务分析

选项:指数趋势模型法

选项:人工智能

选项:生物医药

答案: 指数平滑法

答案: 金融时间序列分析

答案: 移动平均法

答案: 财务分析

答案: 人工智能

答案: 生物医药

问题 2:

下列说法正确的是()。

选项:平均数不受异常值影响

选项:中位数不受异常值影响

选项:中位数和众数更加稳健

选项:当数据呈现偏态分布时,
中位数和众数的代表性更好

答案: 中位数不受异常值影响

答案: 中位数和众数更加稳健

答案: 当数据呈现偏态分布时,
中位数和众数的代表性更好

问题 3:

在下列常见随机变量的概率分
布中,属于连续随机变量分布的
是()。

判断题

问题 1:

在统计数据表中,既可能有分类
变量也可能有数值变量。()

选项:√

选项:×

答案:√

问题 2:

在时间序列中,每个数据点的观
测值只能是在一个具体时间点
上的测量值。()

选项:√

选项:×

答案:×

问题 3:

数据可视化结果和数据分布特
征之间并无联系。()

选项:√

选项:×

答案:×

问题 4:

线性相关系数会受异常值的影
响。()

选项:√

选项:x

答案:√

问题 5:

样本均值的抽样分布是卡方分布。()

选项:√

选项:x

答案:x

问答题

问题 1:

简述总体和个体的区别与联系。

参考 1: 总体和个体是统计学中最基础的核心概念。区别: 总体是满足特定目标的所有观测对象或要素的集合。个体是总体中的每一个观测对象或要素。联系: 总体是由许多同类个体构成的集合, 个体是构成总体的元素。根据研究问题的不同, 总体和个体是相互定义的相对概念。

问题 2:

简述假设检验的基本步骤。

参考 1: 假设检验的基本步骤包括以下 7 步: 第 1 步, 说明要检验的问题, 提出原假设和备选假设。第 2 步, 设定显著性水平 α (风险水平)。第 3 步, 根据检验问题, 确定适合的检验统计量及抽样分布。第 4 步, 根据给定的显著性水平和抽样分布, 确定临界值和拒绝域。第 5 步, 利用样本数据, 计算检验统计量的值并计算 P 值。第 6 步, 做出

统计决策。一是临界值法, 即利用检验统计量的值与临界值做比较, 来决定是否拒绝原假设; 二是 P 值法, 即利用 P 值与显著性水平做比较, 来判断是否拒绝原假设。第 7 步, 给出关于实际问题的结论。

综合题

为了了解某产品的使用寿命情况, 收集了同批次 100 个产品的使用寿命(小时)数据。现在使用描述性统计分析方法进行研究, 相关结果如下图所示: 请回答下列问题:

问题 1:

员工年龄的变量类型是()。

选项:数值变量

选项:分类变量

选项:离散变量

选项:属性变量

答案: 数值变量

问题 1:

在根据变量观测值进行分组时, 第一步是()。

选项:计算组距

选项:确定组限

选项:找到最大值和最小值

选项:确定组数

答案: 找到最大值和最小值

问题 1:

从数据结果看, 产品寿命分布呈现()。

选项:左偏扁平分布

选项:右偏扁平分布

选项:左偏尖峰分布

选项:右偏尖峰分布

答案: 右偏尖峰分布

问题 1:

图中使用的数据可视化方法是()。

选项:散点图

选项:饼图

选项:气泡图

选项:直方图

答案: 直方图

问题 1:

在频率分布表中, 根据结果可知, 占比最多的寿命时长组是()。

选项:701-710 小时

选项:671-680 小时

选项:691-700 小时

选项:681-690 小时

答案: 691-700 小时

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:

问题 2:
每日在线学习时长变量属于()。

- 选项:离散变量
- 选项:连续变量
- 选项:分类变量
- 选项:属性变量

答案: 连续变量

问题 2:
在 Excel 中, 计算均值的函数是 ()。

- 选项:MEDIAN 函数
- 选项:AVERAGE 函数
- 选项:VAR.S 函数
- 选项:MODE.SNGL 函数

答案: AVERAGE 函数

问题 2:
仅观察方差和标准差的计算结果可知, 本科生组的数据离散程度 () 专科生组的数据离散程度。

- 选项:低于
- 选项:等于
- 选项:高于
- 选项:不确定

答案: 低于

问题 2:
根据指标计算结果可知, 两组数据的变异系数分别为 ()。

- 选项:本科生组: 92.466/150;
专科生组: 111.243/120

选项:本科生组: 92.466/30; 专科生组: 111.243/30

选项:本科生组: 92.466/150;
专科生组: 111.243/90

选项:本科生组: 8550/30; 专科生组: 12375/30

答案: 本科生组: 92.466/150;
专科生组: 111.243/120

问题 2:
在 95%置信水平下, 计算本科生组关于总体均值的置信区间, 结果是 ()。

统计学原理-5

单选题

问题 1:
当一个随机变量服从正态分布时, 变量值取值个数为 ()。

- 选项:100 个
- 选项:无限多个
- 选项:500 个
- 选项:1000 个

答案: 无限多个

问题 2:
点估计方法背后的理论依据是 ()。

- 选项:大数定律
- 选项:中心极限定理
- 选项:中值定理
- 选项:拉格朗日定理

答案: 大数定律

问题 3:
对于右偏分布而言, 在箱线图中, 中位数更靠近 () 的位置。

- 选项:上四分位数
- 选项:下四分位数
- 选项:中间
- 选项:最大值

答案: 下四分位数

问题 4:
反映分布相对于均值对称轴线的偏离方向和程度的指标是 ()。

- 选项:众数
- 选项:偏度
- 选项:峰度
- 选项:中位数

答案: 偏度

问题 5:
平稳时间序列是在有限范围内围绕均值 () 波动的时间序列。

- 选项:不规则性
- 选项:趋势性
- 选项:周期性
- 选项:季节性

答案: 不规则性

问题 6:
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)。

- 选项:颜色

选项:面积

选项:高度

选项:宽度

答案: 高度

问题 7:

以个体指数为基础通过平均形式编制的总指数称为是 ()。

选项:数量指标指数

选项:综合指数

选项:平均指数

选项:个体指数

答案: 平均指数

问题 8:

按随机原则进行抽样的抽样称为 ()。

选项:判断抽样

选项:方便抽样

选项:随意抽样

选项:概率抽样

答案: 概率抽样

问题 9:

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 ()。

选项:800 家工业企业

选项:每个工业企业

选项:全部工业产品

选项:每一件工业产品

答案: 每一件工业产品

问题 10:

下列变量中, () 属于数值变量且属于离散变量。

选项:资产的类别

选项:产量

选项:销售额

选项:净利润

答案: 产量

多选题

问题 1:

按照时间序列数据观测值的表现形式不同,时间序列可以分为 ()。

选项:绝对数时间序列

选项:相对数时间序列

选项:平均数时间序列

选项:中位数时间序列

答案: 绝对数时间序列

答案: 相对数时间序列

答案: 平均数时间序列

问题 2:

下列说法正确的是 ()。

选项:极差容易受到异常值影响

选项:标准差的大小会受到数据本身数值大小的影响

选项:标准差相同的两组数据的差异程度可能不同

选项:变异系数除以均值等于标准差

答案: 极差容易受到异常值影响

答案: 标准差的大小会受到数据本身数值大小的影响

答案: 标准差相同的两组数据的差异程度可能不同

问题 3:

以下关于正态分布的描述正确的是 ()。

选项:正态分布是对称分布

选项:正态分布的均值等于中位数

选项:正态分布取值不可以为负数

选项:正态分布有两个参数

答案: 正态分布是对称分布

答案: 正态分布的均值等于中位数

答案: 正态分布有两个参数

问题 4:

一手数据的主要来源包括 ()。

选项:统计调查

选项:实验

选项:第三方数据库

选项:人口普查数据

答案: 统计调查

答案: 实验

问题 5:

常用于分类变量的数据可视化方法包括 ()。

选项:饼图

选项:南丁格尔玫瑰图

选项:散点图

选项:柱形图

答案: 饼图

答案: 南丁格尔玫瑰图

答案: 柱形图

判断题

问题 1:

统计调查仅指对样本数据进行的抽样调查。 ()

选项:√

选项:×

答案: ×

问题 2:

在指数平滑法中,平滑参数的数值越大,意味着对临近期的观测值赋予越大的权重。 ()

选项:√

选项:×

答案: √

问题 3:

变异系数消除了数值大小对标准差的影响。 ()

选项:√

选项:×

答案: √

问题 4:

弱线性相关不一定表明变量之间没有关系。 ()

选项:√

选项:×

答案: √

问题 5:

在假设检验中, P 值就等于设定的显著性水平。 ()

选项:√

选项:×

答案: ×

问答题

问题 1:

简述饼图和柱状图的异同点。

参考 1: 相同点: 饼图和柱状图都是描述数据分布的一种分布图, 都能反映变量的总体结构, 二者都适用于描述分类变量的分布。不同点: 饼图可以反映分类变量或分组变量数目极少时的分布, 适合于变量值极少情况; 柱状图在反映分类变量分布时能够容纳的变量值个数比饼图多一些; 相比饼图, 柱状图更能显现不同类别的频率差异。

问题 2:

简述随机试验需要满足的三个条件。

参考 1: 随机试验需要满足的三个条件分别是: 1) 试验是可重复的; 2) 试验的所有可能结果

是已知的; 3) 试验前, 一次具体试验的结果无法确知。

综合题

现有数据如下表所示, 请结合数据回答下列问题:

问题 1:

现在需要使用数据可视化方法展示不同年份的年末人口数, 适合的数据可视化方法是 ()。

选项:箱线图

选项:柱形图

选项:直方图

选项:饼图

答案: 柱形图

问题 1:

年末人口数序列的极差等于 ()。

选项:141178-138326

选项:141178-139232

选项:141008-138326

选项:141008-139232

答案: 141178-138326

问题 1:

最终消费序列的中位数等于 ()。

选项:2017 年的数值

选项:2018 年的数值与 2019 年的数值的平均值

选项:2017 年的数值与 2018 年的数值的平均值

选项:2018 年的数值

答案: 2017 年的数值与 2018 年的数值的平均值

选项: 整群抽样

选项: F 统计量

选项: 分层抽样

选项: 卡方统计量

选项: 等距抽样

答案: Z 统计量

问题 1:

根据变量的分类, 年末人口数属于 ()。

选项: 连续变量

选项: 数值变量

选项: 分类变量

选项: 定性变量

答案: 数值变量

问题 1:

表中展示的年末人口数和最终消费数据 ()。

选项: 均是时点序列

选项: 均是时期序列

选项: 前者是时点序列, 后者是时期序列

选项: 前者是时期序列, 后者是时点序列

答案: 前者是时点序列, 后者是时期序列

为了解某大型社区居民的收入水平, 研究人员从全社区中随机抽取了 300 名居民, 调查收入 (元)。相关描述性度量指标的计算结果如下图所示: 请结合数据结果回答下列问题:

问题 2:

案例中使用的抽样调查方法是 ()。

选项: 简单随机抽样

答案: 简单随机抽样

问题 2:

从数据结果看, 被调查居民的收入呈现 ()。

选项: 左偏扁平分布

选项: 右偏扁平分布

选项: 左偏尖峰分布

选项: 右偏尖峰分布

答案: 右偏尖峰分布

问题 2:

在 95% 置信水平下, 计算关于居民平均收入水平的置信区间, 估计误差 (保留两位小数) 的结果是 ()。

选项: $1.96 * 18.26$

选项: $1.96 * 18.26 / 300$

选项: $1.96 * 316.23$

选项: $1.96 * 316.23 / 300$

答案: $1.96 * 18.26$

问题 2:

有报告显示该社区居民的平均收入水平等于 2900 元。现在需要利用假设检验方法验证这一结论是否正确, 适合的检验统计量是 ()。

选项: Z 统计量

选项: t 统计量

问题 2:

有报告显示该社区居民的平均收入水平等于 2900 元。现在需要利用假设检验方法验证这一结论是否正确, 假设检验的拒绝域在 ()。

选项: 抽样分布的左侧

选项: 抽样分布的右侧

选项: 抽样分布的两侧

选项: 抽样分布的中间

答案: 抽样分布的两侧