



第四部分 统计



第二十三章 统计与数据科学

【单选题】用1表示“硕士及以上”，用2表示“本科”，用3表示“大专及以下”，这种对变量进行观测的数据是（ ）。

- A. 分类数据
- B. 定量数据
- C. 顺序数据
- D. 数值型数据



第二十三章 统计与数据科学

答案：C

解析：本题考查顺序数据。

数据是对变量进行测量、观测的结果。例如，对企业销售额的测量可以得到其销售金额数据；对员工受教育水平的观测可以得到员工受教育水平的数据。数据根据需要可以是数值、文字或者图像等形式。

选项A错误；分类变量的观测结果称为分类数据，表现为类别，一般用文字来表述，也可用数值代码表示。比如用1表示“男性”，用2表示“女性”。



第二十三章 统计与数据科学

选项C正确；顺序数据是对顺序变量的观测结果，也表现为类别，一般用文字表述，也可用数值代码表示。比如用1表示“硕士及以上”，用2表示“本科”，用3表示“大专及以下”。

选项B错误；数值型数据是对定量变量的观测结果，其取值表现为具体表示大小或多少的数值。对不同类型的数据，可采用不同的统计方法来处理和分析。例如，对分类数据可以计算出各类别的频率，但对其进行加、减、乘或除等数学运算是没有意义的。

选项D错误；而数值型数据则可以进行数学运算，比如计算均值和方差等统计量。



第二十三章 统计与数据科学

【单选题】根据统计数据的收集方式分类，居民消费价格指数（CPI）属于（ ）。

- A. 观测数据
- B. 实验数据
- C. 间接数据
- D. 直接数据



第二十三章 统计与数据科学

答案：A

解析：本题考查观测数据。

选项A正确；观测数据是指在没有对事物施加任何人为控制因素的条件下得到的，几乎所有与社会经济现象有关的统计数据都是观测数据。如GDP、CPI、房价等。直接数据与间接数据是从使用者的角度来区分的，CPI可能是直接数据，也可能是间接数据。



第二十三章 统计与数据科学

选项B错误；通过在实验中控制实验对象以及其所处的实验环境收集到的数据，称为实验数据。比如，一种新产品使用寿命的数据，一种新药疗效的数据。

选项C错误；直接的调查和科学实验，对使用者来说，这是数据的直接来源，称为直接数据或一手数据；

选项D错误；别人的调查或实验的数据，对使用者来说，这是数据的间接来源，称为间接数据或二手数据。



第二十三章 统计与数据科学

【多选题】下列变量中，通常用数值型数据表示的有（ ）。

- A. 注册员工人数
- B. 交通工具
- C. 家庭收入
- D. 居住地区
- E. 年龄



第二十三章 统计与数据科学

答案：ACE

解析：此题考查变量和数据类型。

选项ACE正确；变量是研究对象的属性或特征，它是相对于常数而言的。常数只有一个固定取值，而变量可以有两个或更多个可能的取值。研究对象从不同的角度考察，可以有許多不同的属性或特征，也就有许多变量。比如，研究某个企业，我们可以研究其销售额、注册员工数量、员工受教育水平、所属行业等。当变量的取值是数量时，该变量被称为定量变量或数量变量。例如，企业销售额、注册员工数量等。



第二十三章 统计与数据科学

选项BD错误；当变量的取值表现为类别时则被称为分类变量，比如企业所属行业。当变量的取值表现为类别且具有一定顺序时被称为顺序变量，比如员工受教育水平。分类变量和顺序变量统称为定性变量。



第二十四章 描述统计

【单选题】某小学六年级8个班的学生人数由少到多依次为34人、34人、34人、34人、36人、36人、37人、37人，其中位数为（ ）。

A. 34

B. 35

C. 36

D. 37



第二十四章 描述统计

答案：B

解析：此题考查集中趋势测度值，已排序居中的是34人36人，则中位数是 $(34+36) \div 2 = 35$ 人。



第二十四章 描述统计

【单选题】下列指标中，用于描述数据集中趋势，并且易受极端值影响的是（ ）。

- A. 均值
- B. 中位数
- C. 众数
- D. 标准差



第二十四章 描述统计

答案：A

解析：此题考查集中趋势测度值。

选项A正确，BC错误；描述数据集中趋势的包括：均值、中位数和众数，其中均值是易受极端值影响的集中趋势测度值。

选项D错误；标准差是属于描述离散程度的测度标准，易受极端值影响。



第二十四章 描述统计

【单选题】当数据服从对称的钟形分布时，经验法则表明，95%的数据标准分数在（ ）范围内。

- A. $[-1, 1]$
- B. $[-2, 2]$
- C. $[-3, 3]$
- D. $[-4, 4]$



第二十四章 描述统计

答案：B

解析：此题考查分布形态的测度值。

选项A错误；约有68%的标准分数在 $[-1, +1]$ 范围内；

选项B正确；约有95%的标准分数在 $[-2, +2]$ 范围内；

选项C错误；约有99%的标准分数在 $[-3, +3]$ 范围内。

选项D错误；为干扰项。



第二十四章 描述统计

【单选题】关于Pearson相关系数的说法中错误的是
()。

- A. 若 $0 < r \leq 1$ ，表明变量间存在正线性相关
- B. 若 $-1 \leq r < 0$ ，表明变量间存在负线性相关
- C. 若 $r = 0$ ，表明变量间为不相关关系
- D. 若 $r = -1$ ，表明变量间为完全负线性相关



第二十四章 描述统计

答案：C

解析：本题考查Pearson相关系数的取值。

Pearson相关系数的取值范围在+1和-1之间，即 $-1 \leq r \leq 1$ 。

选项A正确；若 $0 < r \leq 1$ ，表明变量X和Y之间存在正线性相关关系；

选项B正确；若 $-1 \leq r < 0$ ，表明变量X和Y之间存在负线性相关关系。

若 $r = 1$ ，表明变量X和Y之间为完全正线性相关；

选项D正确；若 $r = -1$ ，表明变量X和Y之间为完全负线性相关。

可见，当 $|r| = 1$ 时，变量Y的取值完全依赖于X；

选项C错误；当 $r = 0$ 时，说明Y和X之间不存在线性相关关系。



第二十四章 描述统计

【多选题】下列指标中，测度数据分布形态的有（ ）。

- A. 众数
- B. 标准差
- C. 标准分数
- D. 偏态系数
- E. 离散系数



第二十四章 描述统计

答案：CD

解析：本题考查分布形态的测度。

选项CD正确；测度数据的分布形态包括偏态系数和标准分数。

选项A错误；属于集中趋势的测度。

选项BE错误；标准差用于测度数据的离散程度。



第二十五章 抽样调查

【单选题】北京市旅游管理部门要调查了解2023年北京市常住居民出境旅游总消费金额，随机抽取3000位常住居民进行调查，该抽样调查的样本统计量是（ ）。

- A. 随机抽取3000位常住居民出境旅游总消费金额
- B. 所有常住居民出境旅游总消费金额
- C. 被调查的3000位常住居民
- D. 被调查的每一位常住居民出境旅游消费金额



第二十五章 抽样调查

答案：A

解析：抽样调查中涉及的概念有：总体样本、入样单位、样本量、总体参数、样本统计量、抽样框。其中样本统计量是根据样本中各单位的数值计算的，是对总体参数的估计。

选项A正确；题目中，总体为北京市全部常住居民；样本为随机抽取的3000位常住居民；入样单位为3000位常住居民中的每一位；样本量为3000；总体参数为所有常住居民出境旅游总消费金额；样本统计量为随机抽取3000位常住居民出境旅游总消费金额；抽样框为北京市全部常住居民的基本信息（如姓名、电话等）。

选项B错误；为总体参数；选项C错误；为样本量；选项D错误；为干扰项。



第二十五章 抽样调查

【单选题】以下情况中不属于计量误差的是（ ）。

- A. 调查员在调查中有意无意地诱导被调查者
- B. 被调查者不愿意告诉实情而拒绝回答
- C. 调查人员有意作弊
- D. 受访者提供虚假信息



第二十五章 抽样调查

答案：B

解析：此题考查抽样调查中的误差。

选项ACD属于；计量误差是指由于调查所获得的数据与其真值之间不一致造成的误差。这种误差可能是由调查人员，问卷设计，受访者等原因造成的。

选项B不属于计量误差；属于无回答误差。



第二十五章 抽样调查

【多选题】抽样调查的一般步骤包括（ ）。

- A. 分析研究背景
- B. 确定调查问题
- C. 调查方案设计
- D. 实施调查过程
- E. 数据处理分析



第二十五章 抽样调查

答案：BCDE

解析：本题考查抽样调查的一般步骤。

选项BCDE正确；一个完整的抽样调查过程需要以下几个步骤：确定调查问题；调查方案设计；实施调查过程；数据处理分析；撰写调查报告。

选项A错误；为干扰项。



第二十五章 抽样调查

【多选题】下列属于样本量影响因素的有（ ）。

- A. 调查的精度
- B. 总体的离散程度
- C. 总体的规模
- D. 无回答情况
- E. 系统录入误差



第二十五章 抽样调查

答案：ABCD

解析：本题考查样本量的影响因素。样本量的影响因素包括：调查的精度、总体的离散程度、总体的规模、无回答情况、经费的制约。此外。调查的限定时间，实施调查的人力资源也是影响样本量的客观因素。

综上所述本题选择ABCD；E错误为干扰项



第二十六章 回归分析

- 【单选题】关于一元线性回归模型，下列表述错误的是（ ）。
- A. 只涉及一个自变量的回归模型称为一元线性回归模型
 - B. 因变量Y是自变量X的线性函数加上误差项
 - C. $\beta_0 + \beta_1 X$ 反映了由于自变量X的变化而引起的因变量Y的线性变化
 - D. 误差项 ε 是个随机变量，表示除线性关系之外的随机因素对Y的影响，它是能由X和Y的线性关系所解释的Y的变异性



第二十六章 回归分析

答案：D

解析：只涉及一个自变量的一元线性回归模型表示为 $\beta_0 + \beta_1 X + \varepsilon$ ，因变量Y是自变量X的线性函数（ $\beta_0 + \beta_1 X$ ）加上误差项 ε ； $\beta_0 + \beta_1 X$ 反映了由于自变量X的变化而引起的因变量Y的线性变化。误差项 ε 是个随机变量，表示除线性关系之外的随机因素对Y的影响，是不能由X和Y的线性关系所解释的Y的变异性，D项错误。



第二十七章 时间序列分析

【单选题】某企业本年度前四个月的生产情况如下表：

月份	1月	2月	3月	4月
产值（万元）	44	50	58	64

该企业1月至4月的平均产值为（ ）万元

A. 44

B. 54

C. 59

D. 66



第二十七章 时间序列分析

答案：B

解析：此题考查平均发展水平。题目属于时期序列。采用算术平均值计算。 $(44+50+58+64) \div 4=54$ 。



第二十七章 时间序列分析

【多选题】下表中能源生产总量是（ ）时间序列。

我国2010——2013年能源生产总量

年份	2010	2011	2012	2013
能源生产总量 (万吨标准煤)	109126	106988	120900	138369

- A. 相对数
- B. 时期
- C. 绝对数
- D. 平均数
- E. 时点



第二十七章 时间序列分析

答案：BC

解析：本题考查时间序列的分类。时间序列按照其构成要素中统计指标值的表现形式，分为绝对数时间序列、相对数时间序列和平均数时间序列三种类型。依据指标值的时间特点，绝对数时间序列又分为时期序列和时点序列。题干中的能源生产总量是绝对数指标，也是时期指标，所以选BC。