

第二十三章 统计与数据科学

关于推断统计的说法，正确的有（ ）。

- A.推断统计的内容包括参数估计和假设检验
- B.推断统计是研究数据收集的统计学方法
- C.推断统计是研究数据整理和描述的统计学方法
- D.推断统计的内容包括用图表描述数据的一般性特征
- E.推断统计是研究如何利用样本数据来推断总体特征的统计学方法

【答案】 AE

【解析】 统计学有两个分支：描述统计和推断统计。描述统计是研究数据收集、整理和描述的统计学方法。推断统计是研究如何利用样本数据来推断总体特征的统计学方法，其内容包括参数估计和假设检验两大类。

关于普查的说法，正确的有（ ）。

- A.普查规范化程度较高
- B.普查数据一般比较标准
- C.普查适用范围较宽
- D.普查具有周期性
- E.普查通常需要间隔较长的时间

【答案】 ABDE

【解析】 普查的适用范围比较窄，只能调查一些最基本及特定的现象。

下列我国重大国情国力普查中，每 10 年进行两次的是（ ）。

- A.经济普查
- B.土地调查
- C.人口普查
- D.农业普查

【答案】 A

【解析】 经济普查每 10 年进行 2 次；农业普查和人口普查每 10 年进行 1 次。

【考点】 第二十三章第四节 统计调查

根据楼房所在的位置，楼层，小区环境，地段等信息来预测房价属于（ ）。

- A.分类分析
- B.回归分析
- C.降维分析
- D.有监督学习
- E.无监督学习

【答案】 BD

【解析】 常见的数据挖掘方法有监督学习，无监督学习，半监督学习。

监督学习包括分类和回归，无监督学习包括聚类和降维。回归分析是通过特征变量确定观测单位因变量的取值，因变量是定量变量。本题属于有监督学习，属于回归。

常用的无监督方法包括（ ）。

- A.聚类分析

- B.降维分析
- C.回归分析
- D.分类分析
- E.半监督分析

【答案】AB

【解析】无监督学习有聚类分析和降维分析。

下列数据挖掘方法中，属于监督学习的有（ ）。

- A.聚类分析
- B.因子分析
- C.支持向量机
- D.逻辑斯特回归
- E.主成分分析

【答案】CD

【解析】监督学习中有两大类典型任务：分类和回归。常用的分类方法有逻辑斯特回归、决策树、随机森林和支持向量机等。常用的回归方法有线性回归、非线性回归和分位数回归等。

下列属于聚类分析的是（ ）。

- A.客户细分
- B.文本归类
- C.结构分组
- D.行为跟踪
- E.因子分析

【答案】ABCD

【解析】下列属于聚类分析的是客户细分、文本归类、结构分组、行为跟踪。

常用的分类技术不包括（ ）。

- A.决策树分类法
- B.罗斯特回归分类法
- C.随机森林
- D.k-均值算法
- E.支持向量机

【答案】ABCE

【解析】分类技术是一种监督学习，常用的分类技术包括决策树分类法、罗斯特回归分类法、支持向量机、随机森林等。

第二十四章 描述统计

在测度数据集中趋势时，与均值、中位数相比，众数的缺点是（ ）。

- A.不适用于分类变量
- B.易受极端值影响
- C.不适用于顺序变量
- D.缺乏稳定性

【答案】D

【解析】众数的缺点是没有充分利用数据的全部信息，缺乏稳定性，而且可能不唯一。

关于众数的说法，正确的有（ ）。

- A.在定量数据中，可能出现多众数和无众数的情况
- B.众数是算数平均数的一种
- C.众数是指一组数据中出现次数最多的变量值
- D.众数适用于描述分类数据的集中趋势
- E.众数适用于描述顺序数据的集中趋势

【答案】 ACDE

【解析】在定量数据中，可能出现多众数和无众数的情况，所以众数不适用于定量变量。众数是位置平均数，B选项错误。众数适用于分类数据和顺序数据。

第二十五章 抽样调查

下列抽样调查方法中，属于概率抽样的有（ ）。

- A.系统抽样
- B.判断抽样
- C.简单随机抽样
- D.分层抽样
- E.方便抽样

【答案】 ACD

【解析】概率抽样具有以下特点：按一定的概率以随机原则抽取样本；总体中每个单元被抽中的概率是已知的，或者是可以计算出来的；当采用样本对总体参数进行估计时，要考虑到每个样本单元被抽中的概率。由此可分析出，属于概率抽样的有：系统抽样、简单随机抽样、分层抽样。

关于抽样调查的说法，正确的有（ ）。

- A.抽样调查只抽取部分样本进行调查
- B.抽样调查时效性强
- C.抽样调查耗费大量人力、物力、财力
- D.抽样调查根据样本调查结果推断总体数量特征
- E.抽样调查工作量小

【答案】 ABDE

【解析】选项C错误，抽样调查的优点就是相对节省费用。相比，普查会耗费大量的人力、物力、财力。

为调查我国国有企业在职工延迟退休意愿，分别从东部、中部和西部地区采用简单随机抽样方法各自抽取10个地级行政区域，然后在被抽中企业的所有在职职工发放问卷调查。该项调查中，除了简单随机抽样，还使用到的抽样方法有（ ）。

- A.配额抽样
- B.分层抽样
- C.整群抽样
- D.系统抽样
- E.多阶段抽样

【答案】 BCE

【解析】配额样本，指将总体中的各单元按一定标准划分为若干类型，将样本数额分配到各类型中，从各类型中抽取样本的方法则没有严格限制，一般采用方便抽样的方法抽取样本单元。分层抽样是指先按照某种规则把总体分为不同的层，然后在不同的层内独立、随机地抽取样本，这样所得到的样本称为分层样本。

配额抽样类似于随机抽样中的分层抽样，有两点重要区别：首先，配额抽样的被调查者不是按随机抽样的原则抽选出来的，而分层抽样必须遵守随机抽样的原则；其次，在分层抽样中，用于分类的指标，应联系研究目标来选择，而配额抽样无此要求。系统抽样是指先将总体中的所有单位按一定顺序排列，在规定范围内随机抽取一个初始单元，然后按事先规定的规则抽取其他样本单元。最简单的系统抽样是等距抽样。整群抽样是将总体中所有的基本单位按照一定规则划分为互不重叠的群，抽样时直接抽取群，对抽中的群调查其全部的基本单位，对没有抽中的群则不进行调查。

在大规模抽样调查中，一次抽取到最终样本单位是很难实现的，往往需要经过二个或二个以上阶段才能抽到最终样本单位，这就是多阶段抽样方法。首先从总体中采用随机方法抽取若干个小总体，称为初级单元；再在这些中选的初级单元中随机抽取若干个单位。如果经过二个阶段抽样，抽取到接受调查的最终单位，称为二阶段抽样；

将总体分为若干个一阶单元，如果在每一个一阶单元中，都随机抽取部分二阶单元，由这些二阶单元中的总体基本单元组成的样本，在抽样的方式上，就相当于分层抽样；如果在全部的一阶单元中，只抽取了部分一阶单元，并对抽中的一阶单元中的所有的基本单元都做全面调查，这就是整群抽样。因此，分层抽样实际是第一阶抽样比为 100%时的一种特殊的两阶抽样；而整群抽样实际上是第二阶抽样比为 100%时的一种特殊的两阶抽样，故也称单级整群抽样。

下列属于分层抽样应用条件的有（ ）。

- A.调查对象分布的范围不广阔
- B.层内差异大，层间差异小
- C.抽样框中有足够的辅助信息，能够将总体单位按某种标准划分到各层之中
- D.个体之间的差异不是很大
- E.层内差异小，层间差异大

【答案】CE

【解析】本题考查分层抽样的应用条件。分层抽样的应用条件是：（1）抽样框中有足够的辅助信息，能够将总体单位按某种标准划分到各层之中；（2）同一层内，各单位之间的差异尽可能小；不同层之间各单位的差异尽可能大。

关于样本量的说法。正确的有（ ）。

- A.经费越少样本量越小
- B.总体规模越大样本量要同比例增大
- C.调查误差越小所需的样本量越大
- D.无回答率越高需要抽取的样本量越大
- E.总体方差越小所需的样本量越小

【答案】ACDE

第二十六章 回归分析

关于相关分析和回归分析的说法正确的有（ ）。

- A.相关分析可以从一个变量的变化来推测另一个变量的变化
- B.相关分析和回归分析中都需要确定因变量和自变量

- C.当变量之间存在着相关性时，进行回归分析才有意义
- D.相关分析需要依靠回归分析来表明现象数量相关的具体形式
- E.相关分析与回归分析之间在研究目的和方法上是不同的

【答案】DE

【解析】回归分析可以从一个变量的变化来推测另一个变量的变化，A选项错误；相关分析不需要确定因变量和自变量，B选项错误；只有当变量之间存在着高度相关性时才需要进行回归分析，C选项错误。

关于回归方程决定系数的说法，正确的有（ ）。

- A.决定系数测度回归模型对样本数据的拟合程度
- B.决定系数取值越大，回归模型的拟合效果越差
- C.决定系数等于1，说明回归模型可以解释因变量的所有变化
- D.决定系数取值在[0, 1]之间
- E.如果决定系数等于1，所有观测点都会落在回归线上

【答案】ACDE

【解析】本题考查模型的检验和预测。选项B错误，决定系数越高，模型的拟合效果就越好，即模型解释因变量的能力越强。如果所有观测点都落在回归直线上， $R^2=1$ ，说明回归直线可以解释因变量的所有变化。

第二十七章 时间序列分析

我国2011—2016年期间最终消费支出对国内生产总值增长贡献率的时间顺序如下：

年份	2011	2012	2013	2014	2015	2016
最终消费支出对国内生产总值增长贡献率（%）	61.9	54.9	47.0	48.8	59.7	66.5

按时间序列的分类，该时间序列属于（ ）。

- A.平均数时间序列
- B.时点序列
- C.相对数时间序列
- D.时期序列

【答案】C

【解析】本题考查时间序列及其分类。相对数时间序列是由同类相对数指标值按时间先后顺序排列后形成的序列。贡献率为相对数指标，所对应的时间序列为相对数时间序列。

在同一时间序列中，累计增长量与相应时期逐期增长量之间的数量关系是（ ）。

- A.累计增长量等于相应时期逐期增长量的加权平均数
- B.累计增长量等于相应时期逐期增长量之积
- C.累计增长量等于相应时期逐期增长量之和除以逐期增长量个数
- D.累计增长量等于相应时期逐期增长量之和

【答案】D

【解析】本题考查累计增长量与逐期增长量之间的数量关系。累计增长量等于相应时期逐期增长量之和。

关于时间序列的说法，正确的有（ ）。

- A.平均增长量等于累计增长量与逐期增长量之比
- B.同一时间序列中，累计增长量等于相应时期逐期增长量之和

- C.定基增长速度等于相应时期内各环比增长速度的连乘积
- D.定基发展速度等于相应时期内各环比发展速度的连乘积
- E.两个相邻时期定基发展速度的比率等于相应时期的环比发展速度

【答案】 BDE

【解析】 同一时间序列中，累计增长量等于相应时期逐期增长量之和；定基发展速度等于相应时期内各环比发展速度的连乘积；两个相邻时期定基发展速度的比率等于相应时期的环比发展速度。

关于增长量和平均增长量的说法，正确的有（ ）。

- A.逐期增长量是报告期水平与上一期水平之差
- B.平均增长量是时间序列中累计增长量的序时平均数
- C.增长量可分为逐期增长量和累计增长量
- D.增长量是报告期发展水平与基期发展水平之差
- E.累计增长量是报告期水平与某一固定时期（通常是时间序列最初水平）之差

【答案】 ACDE

【解析】 平均增长量是时间序列中逐期增长量的序时平均数，不是累积增长量。

在同一时间序列中，关于增长量表述正确的有（ ）。

- A.增长量是不同时期发展水平之比
- B.累计增长量等于相应时期逐期增长量之和
- C.逐期增长量是报告期水平与某一固定时期水平之差
- D.逐期增长量是报告期水平与上一期水平之差
- E.累计增长量是报告期与上一期水平之差

【答案】 BD

【解析】 本题考查增长量的含义和分类。逐期增长量是报告期水平与上一期水平之差。累计增长量是报告期水平与某一固定时期水平之差。同一时间序列中，累计增长量等于相应时期逐期增长量之和。增长量是报告期发展水平与基期发展水平之差。

已知在一个时间序列中，逐期增长量分别为 21、30、35、41，则累计增长量为（ ）。

- A.80
- B.90
- C.105
- D.127

【答案】 D

【解析】 累计增长量等于相应时期逐期增长量之和，得到 $21+30+35+41=127$ 。

以 2000 年为基期，2008 年和 2009 年我国粮食总产量定基增长速度分别为 14.40%和 14.85%。2009 年对 2008 年的环比发展速度为（ ）。

- A.0.39%
- B.14.63%
- C.100.39%
- D.114.63%

【答案】 C

【解析】 本题考查环比发展速度的计算。定基发展速度=定基增长速度+1，且两个相邻时期定基发展速度的比率等于相应时期的环比发展速度。所以本题中，环比发展速度= $(14.85\%+1) / (14.40\%+1) = 100.39\%$ 。

某企业 2000~2006 年间销售收入的年平均增长速度是 27.6%，这期间相应的年平均发展速度是（ ）。

- A.4.6%
- B.17.6%
- C.127.6%
- D.72.4%

【答案】 C

【解析】 本题考查平均发展速度与平均增长速度之间的关系。平均发展速度=1+平均增长速度=1+27.6%=127.6%。

某公司 A 产品的销售额 2007 年比 2002 年增长 53.5%，2006 年比 2002 年增长 40.2%，则 2007 年比 2006 年增长（ ）。

- A.15.90%
- B.9.50%
- C.13.30%
- D.33.09%

【答案】 B

【解析】 设 2002 年为销售额为 1，2007 年销售额为 153.5%。2006 年销售额为 140.2%，所以 2007 年比 2006 年增长 $(153.5\%/140.2\%) - 1 = 9.5\%$ 。

2011 年-2016 年我国工业生产者出厂价格指数分别为 106.0、98.3、98.1、98.1、98.4、98.6，选取移动间隔 K=3，应用移动平均预测法 2017 年工业生产者出厂价格指数，则预测值为（ ）。

- A.99.0
- B.98.3
- C.96.7
- D.97.2

【答案】 D

【解析】 移动平均法使用时间数列中最近 K 期数据值的平均数作为下一期的预测值。则 2017 年预测值= $(98.1 + 94.8 + 98.6) / 3 = 97.17 \approx 97.2$ 。