

## 第七部分 统计学

### 考点1 统计与统计数据

- 1、统计一词包含三种含义：统计工作、统计数据、统计学
- 2、统计数据的计量尺度：定类尺度(将人口分为男女两类)、定序尺度(将学生成绩分成优、良、中、及格和不及格五类)、定距尺度(某物长 10 米，重 100 千克等)、定比尺度(相对数或平均数，可以进行加、减、乘、除等数学运算)
- 3、统计数据的类型：分类数据、顺序数据、数值型数据(分类数据和顺序数据是定性数据或品质数据，数值型数据是定量数据或数量数据)
- 4、统计中把说明现象某种特征的概念称为变量，变量的具体表现为变量值，统计数据就是统计变量的具体表现。可分为：分类变量、顺序变量、数值型变量
- 5、数值型变量分为：离散变量和连续变量
- 6、统计指标按其所反映的内容或其数值表现形式，可分为：总量指标、相对指标和平均指标
- 7、总量指标按其所反映的时间状况不同可分为：时期指标和时点指标。
- 8、统计数据的来源：直接来源，称之为第一手或直接的统计数据；间接来源，称第二手或间接的统计数据。
- 9、统计数据的直接来源主要有两个渠道：专门组织的调查、科学实验
- 10、实际中常用的统计调查方式主要有：普查、抽样调查、统计报表。
- 11、普查：为某一特定目的而专门组织的一次性全面调查。普查的特点：通常是一次性的或周期性的；一般需要规定统一的标准调查时间；数据一般比较准确，规范化程度也较高；使用范围比较狭窄
- 12、抽样调查的特点：经济性、时效性强、适应面广、准确性高，它是实际中应用最广泛的一种调查方式和方法
- 13、抽样调查的集中类型：（1）**简单随机抽样**（也叫纯随机抽样）就是从总体中不加任何分组、划类、排队等，完全随机地抽取调查单位。特点是：每个样本单位被抽中的概率相等，样本的每个单位完全独立，彼此间无一定的关联性和排斥性。简单随机抽样是其它各种抽样形式的基础。通常只是在总体单位之间差异程度较小和数目较少时，才采用这种方法。（2）**等距抽样**（也叫机械抽样或系统抽样）就是将总体各单位按一定标志或次序排列成为图形或一览表式（也就是通常所说的排队），然后按相等的距离或间隔抽取样本单位。特点是：抽出的单位在总体中是均匀分布的，且抽取的样本可少于纯随机抽样。等距抽样既可以用同调查项目相关的标志排队，

也可以用同调查项目无关的标志排队。等距抽样是实际工作中应用较多的方法，目前我国城乡居民收支等调查，都是采用这种方式。（3）**类型抽样**（也叫分层抽样）就是将总体单位按其属性特征分成若干类型或层，然后在类型或层中随机抽取样本单位。特点是：由于通过划类分层，增大了各类型中单位间的共同性，容易抽出具有代表性的调查样本。该方法适用于总体情况复杂，各单位之间差异较大，单位较多的情况。（4）**整群抽样**就是从总体中成群成组地抽取调查单位，而不是一个一个地抽取调查样本。特点是：调查单位比较集中，调查工作的组织和进行比较方便。但调查单位在总体中的分布不均匀，准确性要差些。因此，在群间差异性不大或者不适宜单个地抽选调查样本的情况下，可采用这种方式。

14、统计推断是在概率论的基础上依据样本的有关数据和信息，对未知总体的质量特性参数，做出合理的判断和估计。在质量活动和管理实践中，人们关心的是特定产品的质量水平，如产品质量特性的平均值、不合格品率等。这些都需要从总体中抽取样本，通过对样本观察值分析来估计和推断，即根据样本来推断总体分布的未知参数。称为参数估计。参数估计有两种基本形式：点估计和区间估计。

15、点估计是指用样本的统计量去估计总体相应未知参数称为点估计，通常用无偏性和有效性作为评价点估计优良性的标准。

16、区间估计的指用样本确定两个统计量，构筑一个置信水平为  $1 - \alpha$  的区间，对总体未知参数给出估计，称为区间估计。点估计仅仅给出未知参数的一个具体估计值，没有给出估计的精度，而区间估计却体现了估计的精度。所谓置信水平  $1 - \alpha$ ，是指所构造的置信区间覆盖未知参数的概率为  $1 - \alpha$ 。由于置信区间是由选用样本的统计量构筑的，它是会随着样本的变化而变化的，它有时覆盖未知参数，有时却没有覆盖未知参数。但是，用此法构筑的置信区间，在 100 次中大约有  $100(1 - \alpha)$  个区间覆盖未知参数

17、统计报表可分为：全面报表和非全面报表

18、第二手统计数据的主要来源有：公开的出版物、未公开的内部调查等

19、统计数据的误差有：登记性误差和代表性误差

20、统计数据的质量评价标准：精度、准确性、关联性、及时性、一致性、最低成本

**[真题还原]**抽样的作用和意义在于（ ）

[A]杜绝抽样误差，

[B]获得最大容量的样本

[C]使所抽取的样本具有最好的代表性

[D]可以直接测量总体

解析:抽样的特点和作用:①节省人力及费用; ②节省时间, 提高研究的实效性; ③保证研究结果的准确性(从某种意义上讲, 抽样研究有时比总体研究更能避免偶然误差的引入, 从而提高研究的质量)。因此答案选[C]。

**[例题]**参数估计的置信度为  $1-\alpha$  的置信区间表示 ( )

[A]以  $1-\alpha$  的可能性包含了未知总体参数真值的区间

[B]以  $\alpha$  的可能性包含了未知总体参数真值的区间

[C]总体参数取值的变动范围

[D]抽样误差的最大可能范围

解析: 置信区间, 也称置信间距, 是指在某一置信度时, 总体参数所在的区域距离或区域长度。 $1-\alpha$  的置信区间是指总体参数落在该区间之内, 估计正确的概率为  $1-\alpha$ , 因此答案选[A]。

**[例题]**相对于普查而言, 抽样调查的特点不包括 ( )

[A]经济性 [B]时效性强 [C]适应面广 [D]周期性

解析: 本题考查抽样调查的特点, 抽样调查具有以下特点: 经济性、时效性强、适应面广、准确性高。因此答案选[D]。

**[例题]**有一项研究要调查中国人的主观幸福感, 大约要抽取五千名被试, 相对来说比较适宜的抽样方法是 ( )

[A]整群抽样 [B]分层抽样 [C]随机抽样 [D]等距抽样

解析: 分层抽样是指先将总体的单位按某种总特征分为若干次级总体(层), 然后再从每一层内进行单纯随机抽样, 组成一个样本。根据本题特点, 选用分层抽样最佳。因此答案选[B]。

## 考点 2 统计数据的整理和显示

1、品质数据的整理与显示指标: 频数与频数分布、比例、百分比、比率

2、分类数据的图示方法包括: 条形图和圆形图; 条形图有单式和复式等形式 3、顺序数据的整理与显示: 累积频数和累积频率

4、数值型数据的整理与显示: 组距分组的步骤:

(1) 确定分组组数, 经验公式:  $K = 1 + \log_{10} N / \log_{10} 2$

(2) 对原始资料进行排序

(3) 求极差: 最大观察值减去最小观察值

(4) 确定各组组距: 组距 = 极差 / 组数 = 某组的上限值 - 该组的下限值

(5) 确定组限: 组中值 = (上限 + 下限) / 2

(6) 确定各组观察值出现的频数: 组距分组遵循“不重不漏”的原则; 分组时“上组限不在内”以防止重复

(7) 制作频数分布表

5、显示分组数据频数分布特征的图形有：直方图、折线图

6、统计表一般由四个主要部分组成：表头、行标题、列标题和数字资料，必要时可以在统计表的下方加上表外附加

7、统计表的设计要求：科学、实用、简练、美观

8、设计统计表时要注意以下几点：要合理安排统计表的结构；表头一般应包括表号、总标题和表中数据的单位等内容；表中的上下两条横线一般用粗线，中间的其他线要用细线，这样使人看起来清楚、醒目；在使用统计表时，必要时可在表的下方加上注释，特别要注意注明资料的来源，以示对他人劳动成果的尊重，并备读者查阅使用。

### 考点 3 数据特征的测度

1、集中趋势的测度，主要包括位置平均数和数值平均数。

2、位置平均数是指按数据的大小顺序或出现频数的多少，确定的集中趋势的代表值，主要有众数、中位数等。众数是一组数据中出现次数最多的数值，有时众数在一组数中有好几个，简单的说，就是一组数据中占比例最多的那个数。中位数又叫中数是指对一组数进行排序后，正中间的一个数（数字个数为奇数）；或者中间两个数的平均数（数字个数为偶数）。中数是按顺序排列在一起的一组数据中居于中间位置的数，即在这组数据中，有一半的数据比它大，有一半的数据比它小。这个数可能是数据中的某一个，也可能根本不是原有的数。

3、数值平均数是指是指根据全部数据计算出来的平均数。主要有：算术平均数、几何平均数。

4、算术平均数，又称均值，是统计学中最基本、最常用的一种平均指标，分为简单算术平均数、加权算术平均数，它主要适用于数值型数据。①算术平均数是一个良好的集中量数，具有反应灵敏、确定严密、简明易懂、计算简单、适合进一步演算和较小受抽样变化的影响等优点。②算术平均数易受极端数据的影响，这是因为平均数反应灵敏，每个数据的或大或小的变化都会影响到最终结果。

5、几何平均数是指  $n$  个观察值连乘积的  $n$  次方根，几何平均数  $\leq$  算术平均数。

6、离散程度的测度主要包括：极差、标准差和方差、离散系数。标准差（Standard Deviation），也称均方差(mean square error)，是各数据偏离平均数的距离的平均数。标准差是方差的算术平方根。标准差能反映一个数据集的离散程度。平均数相同的，标准差未必相同。方差是各个数据与平均数之差的平方和的平均数

**【例题】**一个有 10 个数据的样本，它们中的每一个分别与 20 相减后所得的差相加是 100，那么这组数据的均值是（ ）

[A]20 [B]10 [C]30 [D]50

解析：按照平均数的性质，一组数据中，每一个数都加上或减去一个常数 C，则所得的平均数为原来的平均数加上或减去常数 C。题中，每一个数减去 20 后，所得的平均数为  $100/10=10$ ，则原来的平均数为  $10+20=30$ 。因此答案选[C]。

**【例题】**下列易受极端数据影响的统计量是（ ）

[A]算术平均数 [B]中数 [C]众数 [D]四分差

解析：算术平均数易受极端值的影响。由于平均数反映灵敏，因此当数据分布呈偏态时，受极值的影响，平均数就不能恰当地描述分布的真实情况。因此答案选[A]。

**【例题】**已知某小学一年级学生的平均体重为 26 千克，体重的标准差是 3.2 千克，平均身高 110 厘米，标准差为 6.0 厘米，问体重与身高的离散程度哪个大（ ）？

[A]体重离散程度大

[B]身高离散程度大

[C]离散程度一样

[D]无法比较

解析：对于同一被试团体，两种不同属性离散程度的比较应当选用变异系数。 $CV = (\text{标准差}/\text{均值}) \times 100\%$ 。所以， $CV(\text{体重}) = (3.2 \div 26) = 0.123$ ， $CV(\text{身高}) = (6 \div 110) = 0.055$ ，由此可见，体重的离散程度要大于身高的离散程度。因此答案选[A]。

**【例题】**关于众数在分布曲线上对应的位置，如下说法正确的是（ ）

[A]总是对径分布曲线的最高点

[B]总是对应分布曲线的最低点

[C]位置不确定

[D]以上说法都不正确。

解析：众数又称范数、密集数、通常数等，是集中量数的一种，常用符号  $M_0$  表示，是指在次数分布中出现次数最多的那个数的数值。因此答案选[A]。

## 考点 4 时间序列

1、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

2、依据指标值的时间特点，绝对数时间序列又可分为：时期序列和时点序列

3、增长量：时间序列中的报告期水平与基准水平之差，用于描述现象在观察期内增长的绝对数量。按照采用基期的不同，增长量有逐期增长量、累积增长量之分

4、逐期增长量是报告期水平与前一时期水平之差，表示本期比前一时期增长的绝对数量；累计增长量是报告期水平与某一固定时期水平之差，说明报告期与某一固定时期相比增长的绝对数量。整个观察期内各逐期增长量之和等于最末期的累积增长量

5、发展速度：报告期发展水平与基期发展水平之比，用于描述现象在观察期内的发展变化程度。分为定基发展速度和环比发展速度

6、增长速度：也称增长率，是增长量与基期水平之比，用于描述现象的相对增长程度。按照采用基期的不同，增长速度可分为定基增长速度与环比增长速度，两者之间没有直接的换算关系，在由环比增长速度推算定基增长速度时，可先将各环比增长速度加 1 后连乘，再将结果减 1，即得定基增长速度

7、平均发展速度：各个时期环比发展速度的平均数，用于描述现象在整个观察期内平均发展变化的程度

8、平均增长速度(平均增长率)：用于描述现象在整个观察期内平均增长变化的程度，它通常用平均发展速度减 1 来求得

9、增长 1%绝对值：表示每增长一个百分点而增加的绝对数量。等于逐期增长量/环比增长速度。

## 考点 5 统计指数

1、广义地讲，任何两个数值对比形成的相对数都可以称为指数；狭义地讲，指数是用于测定多个项目在不同场合下综合变动的一种特殊相对数。

2、指数可以分为以下几种主要类型：(1)按所反映的内容不同，可以分为数量指数和质量指数。①数量指数是反映物量变动水平的，如产品产量指数、商品销售指数等；②质量指数是反映事物内含数量的变动水平的，如价格指数、产品成本指数等。(2)按计入指数的项目多少不同，可分为个体指数和综合指数。①个体指数是反映某一个项目或变量变动的相对数，如一种商品的价格或销售量的相对变动水平；②综合指数是反映多个项目或变量综合变动的相对数，如多种商品的价格或销售量的综合变动水平。(3)按计算形式不同，可分为简单指数和加权指数。①简单指数又称不加权指数，它把计入指数的各个项目的重要性视为相同；②加权指数则对计入指数的项目依据重要程度赋予不同的权数，而后再进行计算。目前普遍应用的是加权指数。



3、基期加权综合指数：又称拉氏指数，它是把作为权数的各变量值固定在基期的一种计算方法。拉氏数量指数在实际中用得较多，而质量指数用的较少。数量指数：反应物量变动。如：产品产量指数和销售量指数。质量指数：反映质量变动。如：价格指数和产品成本指数。

4、报告期加权综合指数：又称帕氏指数，它是把作为权数的各变量值固定在报告期的一种计算方法。帕氏质量指数在实际中用得较多，而数量指数用得较少

5、指数与各因素指数之间构成的数量关系式称为指数体系。实际中比较常用的是基期权数加权的数量指数和报告期权数加权的质量指数体系。

6、几种常用的价格指数：零售价格指数、消费价格指数、股票价格指数

## 考点6 国民经济主要统计指标

1、国内生产总值是一国或一个地区在一定时期内生产的社会最终产品的价值总和，也就是社会总产品价值扣除中间投入价值后的余额。

2、国内生产总值三种表现形态，即价值形态、收入形态和产品形态

3、生产法国内生产总值是从生产过程中价值形成的角度计算国内生产总值的方法。目前中国国内生产总值的统计以生产法为主。

4、中间投入都是一次性的使用和消耗，不包含固定资产折旧。

5、增加值是总产出扣除中间投入后的余额，是在生产经营过程中追加的价值，又叫附加值

6、农业增加值是指各种类型的农业生产单位和农户在一定时期从事农业生产经营活动所创造的最终产品的价值。这里的农业是种植业，林业，畜牧业和渔业的总称。

7、目前中国农业增加值采用生产法进行核算，即用总产出扣除中间投入得到的增加值。

8、工业增加值是考核工业生产成果的代表性总量指标，是国内生产总值的重要组成部分。目前中国主要采用“生产法”计算工业增加值，即从工业生产过程中产品和劳务价值形成的角度入手，剔除生产环节中间投入的价值，从而得到新增价值的方法。

9、第三产业具体包括：交通运输、仓储和邮政业、信息传输、计算机服务和软件业、批发和零售业、住宿和餐饮业、金融业、房地产业、租赁和商务服务业，科学研究、技术服务和地质勘查业，水利、环境和公共设施

管理业，居民服务和其他服务业，教育、卫生、社会保障和社会福利业，文化、体育和娱乐业，公共管理和社会组织，国际组织等。

9、第三产业的核算分行业分别进行，一般采用收入法，即首先计算增加值的四个构成项目：劳动者报酬、固定资产折旧、生产税净额和营业盈余，然后将四项构成相加得到增加值。

10、反映社会消费总需求的指标主要有两个：社会消费品零售总额和最终消费。其中，社会消费品零售总额的使用比最终消费更为普遍。

11、固定资产投资统计的范围包括：基本建设投资，更新改造投资，房地产开发投资，国有其他固定资产投资，城镇集体固定资产投资，农村集体固定资产投资，城镇和工矿区私人建房投资，农村个人固定资产投资。按经济类型可分为国有、集体、个体、联营、股份制、外商、港澳台商、其他等。