

## §14.1　获取数据的基本途径及相关概念

学习目标　1.了解普查的意义和抽查的概念，理解抽查的必要性和重要性.2.了解总体、样本、样本容量的概念，会对一些实际问题进行合理的抽样调查．



知识点一　获取数据的途径

|  |  |  |
| --- | --- | --- |
|  | 概念 | 数据名称 |
| 直接获取 | 通过社会调查或观察、试验等途径获取数据 | 直接数据或一手数据 |
| 间接获取 | 借助各种媒介，包括报纸杂志、统计报表和年鉴、广播、电视或互联网等获取数据 | 间接数据或二手数据 |

知识点二　普查和抽查

|  |  |  |
| --- | --- | --- |
| 调查方法  概念、特点 | 普查 | 抽查 |
| 定义 | 为了掌握调查对象的整体情况，对全体调查对象进行研究的一种调查方式 | 从全体调查对象中按照一定的方法抽取一部分对象作为代表进行调查分析，并以此推断全体调查对象的状况的调查方式 |
| 优点 | ①所取得的资料更加全面、系统；  ②调查特定时段的社会经济现象总体的信息 | ①迅速、及时；  ②节约人力、物力、财力，对个体信息的了解更详细 |
| 缺点 | 耗费大量的人力、物力、财力、时间长、任务重 | 获取的信息不够全面、系统，其结果具有不确定性 |

知识点三　总体和样本

1．总体：一般地，在获取数据时，把所考察对象(某一项指标的数据)的全体叫作总体．

2．个体：把组成总体的每一个考察对象叫作个体．

3．样本：从总体中所抽取的一部分个体叫作总体的一个样本．

4．样本容量：样本中个体的数目叫作样本容量．



1．获取数据的方式一般有两种：普查和抽查．(　√　)

2．总体中包含的个体数称为样本容量．(　×　)

3．要想准确知道全班同学的平均年龄，应调查每个同学，这种调查方式是普查．(　√　)

4．要了解一批节能灯的使用寿命，可以采用普查的方式．(　×　)



一、总体、样本等概念的辨析

例1　 (多选)2022年4月某市模考共有70 000多名学生参加，教科室为了了解本校3 390名考生的数学成绩，从中抽取300名考生的数学成绩进行统计分析，在这个问题中有以下说法，其中正确的是(　　)

A．3 390名考生是总体的一个样本

B．3 390名考生的数学成绩是总体

C．样本容量是300

D．70 000多名考生的数学成绩是总体

答案　BC

解析　总体是3 390名考生的数学成绩，样本是抽取的300名考生的数学成绩，样本容量是300.

反思感悟　解决此类问题要明确概念的实质，应注意两个问题

(1)调查对象是什么，如本例调查对象是“每个考生的数学成绩”，不是“每个考生”．

(2)样本容量是样本中个体的数目，无单位．

跟踪训练1　2020年2月20日，欧盟特别峰会在布鲁塞尔举行，主要讨论2021年至2027年长期预算，有27个国家代表参加，最终因各方分歧太大，未达成共识．会后某记者从每个国家与会人员中采访了两名成员，调查得到各成员国在预算总量、主要政策领域分配额、欧盟收入来源以及激励机制等多方面都存在分歧．在这个问题中样本容量是(　　)

A．2 B．27

C．54 D．不确定

答案　C

解析　样本容量是27×2＝54.

二、获取数据的途径

例2　为了研究近年我国高等教育发展状况，小明需要获取近年来我国大学生入学人数的相关数据，他获取这些数据的途径最好是(　　)

A.通过调查获取数据

B．通过试验获取数据

C．通过观察获取数据

D．通过查询获取数据

答案　D

解析　因为近年来我国大学生入学人数的相关数据有所存储，所以小明获取这些数据的途径最好是通过查询获取数据．

反思感悟　获取数据的基本途径

|  |  |  |
| --- | --- | --- |
| 获取数据的基本途径 | 适用类型 | 注意问题 |
| 通过调查获取数据 | 对于有限总体问题，一般通过抽样调查或普查的方法获取数据 | 要充分有效地利用背景信息选择或创建更好的抽样方法，并有效避免抽样过程中的人为错误 |
| 通过试验获取数据 | 没有现存的数据可以查询 | 严格控制试验环境，通过精心的设计安排试验，以提高数据质量 |
| 通过观察获取数据 | 自然现象 | 借助专业测量设备通过长久的持续观察获取数据 |
| 通过查询获取数据 | 众多专家研究过，其收集的数据有所存储 | 必须根据问题背景知识“清洗”数据，去伪存真 |

跟踪训练2　“中国天眼”全称为500米口径球面射电望远镜(Five­hundred­meter Aperture Spherical radio Telescope，简称FAST)，是具有我国自主知识产权、世界最大单口径、最灵敏的射电望远镜．建造“中国天眼”的目的是(　　)



A．通过调查获取数据 B．通过试验获取数据

C．通过观察获取数据 D．通过查询获取数据

答案　C

解析　“中国天眼”主要是通过观察获取数据．

三、普查与抽查

例3　(多选)下列调查中，适合采用抽样调查的是(　　)

A．调查黄河的水质情况

B．调查某快餐店中8位店员的生活质量情况

C．调查某药品生产厂家一批药品的质量情况

D．农业科技人员调查今年麦穗的单穗平均质量

答案　ACD

解析　A项，因为无法对所有的黄河水质进行普查，所以只能采取抽样调查的方式；B项，容量小可以普查；C项，对药品的质量检验具有破坏性，所以只能采取抽样调查；D项，总体容量大，用抽样调查．

反思感悟　对普查与抽样调查的理解

(1)普查是一项非常艰巨的工作，它要对所有的对象进行调查，当检验对象很大或检验对象具有破坏性时，采用普查的方法是行不通的，要进行抽样调查．

(2)普查与抽样调查的适用条件是不同的，在具体问题中，用普查还是抽样调查的方式，要根据它们的特点和适用范围进行判断.

跟踪训练3　下列调查方式中合适的是(　　)

A．要了解一批节能灯的使用寿命，采用普查方式

B．调查你所在班级同学的身高，采用抽样调查方式

C．调查沱江某段水域的水质情况，采用抽样调查方式

D．调查全市中学生每天的就寝时间，采用普查方式

答案　C

解析　要了解节能灯的使用寿命，由于调查具有破坏性，所以宜采取抽样调查的方式；要调查所在班级同学的身高， 由于人数较少，宜采用普查的方式；对全市中学生每天的就寝时间的调查不宜采用普查方式．故选C.

四、抽样调查中样本的抽取问题

例4　为了调查2022年中央电视台春节联欢晚会播出的收视率，节目组设置了《春节联欢晚会收视率调查表》，并向社会广泛征求调查意见．某校有两名同学为电视台设计了调查方案：

同学A：我把《春节联欢晚会收视率调查表》放在互联网上，只要上网登录该网址的人就可以看到这张表，他们填表的信息可以很快反馈到我的电脑中．这样，我就可以很快统计出收视率了．

同学B：我在电话号码本上随机地选出一定数量的电话号码，然后逐个给他们打电话，问一下他们是否收看了中央电视台春节联欢晚会，我不出家门就可以统计出中央电视台春节联欢晚会的收视率了．

请问上述两名同学设计的调查方案能够获得比较准确的收视率吗？为什么？

解　 调查的总体是所有可能看电视的人群．同学A的设计方案考虑的人群是：上网而且登录某网址的人群，那些不能上网的人群，或者不登录该网址的人群就被排除在外了．因此同学A设计的方案抽取样本的代表性差，不能较好的估计总体分布．

同学B的设计方案考虑的人群是：在电话号码本上统计接电话的人群，也有一定的片面性，因此同学B设计的方案抽取样本的代表性也差，也不能较好的估计总体分布．

延伸探究　你认为怎样设计调查方案，能得到较有说服力的数据？

解　发放调查问卷或打电话询问都要考虑不同的人群，要根据人口比例分别对农村、城市分成不同的群体调查，对各个不同群体中最好要按年龄、职业等属性再细化．

反思感悟　根据调查问题的特点设计抽样调查的不同方案，应遵循以下原则

(1)要考虑如何合理地获取样本，以确保其典型性、代表性．即抽取的部分个体具有广泛的代表性，能很好地代表总体．

(2)要考虑如何保证调查内容的真实性．

跟踪训练4　下列调查方式中合适的是(　　)

A．某单位将新购买的准备开业庆典的20箱礼炮全部进行质检

B．某班有40名同学，指定家庭最富有的3人参加“学代会”

C．某服装厂的一批5 000件出口服装，随机抽50件进行抽样调查

D．为了调查最近上映影片的一周内的票房情况，特选周六、周日两天进行调查

答案　C

解析　选项A中，对礼炮的质检带有破坏性，虽然总量不大，但不宜采用普查方式．选项B中，“家庭最富有的”不具有代表性，样本选取错误．选项D中，因调查一周的票房，时间不长，最好采用普查的方式，即使是用抽样调查，周六、周日两天的选取也不具备代表性．



1．下列调查中，适合采用普查方式的是(　　)

A．调查某品牌电视机的市场占有率

B．调查某电视连续剧在全国的观看人数

C．调查某校七年级各班男女同学的比例

D．调查某型号炮弹的射程

答案　C

解析　A项，调查某品牌电视机的市场占有率，适于抽样调查；B项，调查某电视连续剧在全国的观看人数，适于抽样调查；C项，调查某校七年级各班男女同学的比例，适于普查；D项，调查某型号炮弹的射程，适于抽样调查，故选C.

2．(多选)从某年级的500名学生中抽取60名学生进行体重的统计分析，下列说法正确的是(　　)

A．500名学生的体重是总体

B．每个学生是个体

C．抽取的60名学生的体重是一个样本

D．抽取的60名学生的体重是样本容量

答案　AC

解析　由题意可知在此抽样调查中，总体是500名学生的体重，A对；个体是每个学生的体重，B错；样本是抽取的60名学生的体重，C对；样本容量为60，D错．

3．下列调查中属于抽样调查的是(　　)

①每隔5年进行一次人口普查；②调查某商品的质量优劣；③某报社对某个事情进行舆论调查；④高考考生的查体．

A．②③ B．①④ C．③④ D．①②

答案　A

解析　人口普查和高考考生的查体都属于普查，调查某商品的质量优劣和对某个事情进行舆论调查只能是抽样，没必要进行普查．故选A.

4．小明从网上查询到某贫困地区10户居民家庭年收入(单位：万元)如下所示：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 年收入 | 1.2 | 1.3 | 1.8 | 2.0 | 4.6 | 1.7 | 0.9 | 2.1 | 1.0 | 1.6 |

根据以上数据，我们认为有一个数据是不准确的，需要剔除，这个数据是\_\_\_\_\_\_\_\_．

答案　4.6

解析　由于编号为5的数据为4.6，明显高于其他数据，所以这个数据是不准确的．

5．国家统计局、国家残联决定对国家残疾人生活、就业等情况进行调查，某同学设计的调查方案是在国家残联的网站上设立一个调查表，根据网站上的数据进行分析．你认为他的方案\_\_\_\_\_\_\_\_(填“合理”或“不合理”)．

答案　不合理

解析　很多残疾人不具有上网条件，因此获取的数据不具有代表性．



1．知识清单：

(1)获取数据的途径．

(2)普查和抽查．

(3)总体、样本的概念及样本的选取问题．

2．常见误区：

(1)弄错调查对象导致对总体、样本的叙述错误．

(2)使用不恰当的抽取方法导致抽取的样本代表性差．



1．2022年某省实行新高考，考试及录取发生了很大的变化．为了报考理想的大学，小明需要获取近年来我国各大学会计专业录取人数的相关数据，他获取这些数据的途径最好是(　　)

A．通过调查获取数据

B．通过试验获取数据

C．通过观察获取数据

D．通过查询获取数据

答案　D

解析　因为近年来我国各大学会计专业录取人数的相关数据有存储，所以小明获取这些数据的途径最好是通过查询获取数据．

2．为了了解某年级同学每天参加体育锻炼的时间，比较恰当地收集数据的方法是(　　)

A．查阅资料 B．问卷调查

C．做试验 D．以上均不对

答案　B

解析　首先明确收集数据的几种方法：查资料、做试验和做调查；接下来根据人数的多少以及调查的目的进行选择即可．问卷调查能达到目的，比较适合．

3．(多选)以下获取的数据是通过查询获取的是(　　)

A．某领导想了解*A*市的大气环境质量，向当地有关部门咨询该市的PM2.5的浓度

B．张三利用互联网了解到，2019年某市居民平均寿命达到82.2岁

C．某中学为了了解学生对课堂禁用手机的认同度，进行了问卷调查

D．从某公司员工年度报告中获知信息

答案　ABD

解析　A，B，D都是通过查询获取的数据，C是通过调查获取的数据．

4．下面问题可以用普查的方式进行调查的是(　　)

A．检验一批钢材的抗拉强度

B．检验海水中微生物的含量

C．检验10件产品的质量

D．检验一批汽车的使用寿命

答案　C

解析　A不能用普查的方式调查，因为这种试验具有破坏性；B用普查的方式无法完成；C可以用普查的方式进行调查；D中试验具有破坏性，且需要耗费大量的时间，在实际生产中无法应用．

5．下列调查方案中，抽样方法合适、样本具有代表性的是(　　)

A．用一本书第1页的字数估计全书的字数

B．为调查某校学生对航天科技知识的了解程度，上学期间，在该校门口，每隔2分钟随机调查一位学生

C．在省内选取一所城市中学，一所农村中学，向每个学生发一张卡片，上面印有一些名人的名字，要求每个学生只能在一个名字下面画“√”，以了解全省中学生最崇拜的人物是谁

D．为了调查我国小学生的健康状况，共抽取了100名小学生进行调查

答案　B

解析　A中，样本缺少代表性(第1页的字数一般较少)；B中，抽样保证了随机性原则，样本具有代表性；C中，城市中学与农村中学的规模往往不同，学生崇拜的人物也未必在所列的名单之中，这些都会影响数据的代表性；D中，总体数量很大，而样本容量太少，不足以体现总体特征．

6．为了了解高一年级学生的视力情况，特别是近视率问题，抽测了其中100名同学的视力情况．在这个过程中，100名同学的视力情况(数据)是\_\_\_\_\_\_\_\_．

答案　总体的一个样本

解析　100名同学的视力情况(数据)是从总体中抽取的一部分个体所组成的集合，所以是总体的一个样本．

7．为了分析高三年级的8个班400名学生第一次高考模拟考试的数学成绩，决定在8个班中每班随机抽取12份试卷进行分析，这个问题中样本容量是\_\_\_\_\_\_\_\_．

答案　96

解析　本题中，400名学生第一次高考模拟考试的数学成绩是总体，从8个班中每班抽取的12名学生的数学成绩是样本，400是总体个数，96是样本容量．

8．下列试验适合用抽样调查方法获取数据的序号是\_\_\_\_\_\_\_\_．

①考察一片草皮的平均高度；

②检查某食品单位职工的身体状况；

③考察参加某次考试的3万名考生的数学答题情况；

④检验一个人的血液中白细胞的含量是否正常．

答案　①③④

解析　①该问题用普查的方法很难实现，适合用抽样调查的方法获取数据；

②体检，必须了解每个职工的身体状况，不适合用抽样调查的方法获取数据；

③3万名考生的答题情况用普查的方法获取数据不合适，适合用抽样调查的方法获取数据；

④该问题只能用抽样调查的方法获取数据．

9．为了准备高考听力考试，某校于高考前新进了45套听力设备．现需要检查这批听力设备的质量，是全部检查还是抽取部分检查？谈谈你的想法和理由．

解　必须全部检查(采用普查)．因为高考是一件非常严肃、责任重大的事件，高考讲究公平竞争，要求十分严格，所配设备必须全部合格，且这批设备数量较少，全部检查是可行的，这样可确保万无一失．

10．为了创建“和谐平安”校园，某校决定在开学前将学校的电灯电路使用情况进行检查，以便排除安全隐患，获得路灯的相关数据应该用什么方法？为什么？

解　由于一个学校的电灯电路数目不算大，属于有限总体问题，所以应该通过调查获取数据，并且对创建“和谐平安”校园来说，必须排除任一潜在或已存在的安全隐患，故必须用普查的方式．



11．下列调查所抽取的样本具有代表性的是(　　)

A．利用某地七月份的日平均最高气温值估计该地全年的日平均最高气温

B．在农村调查市民的平均寿命

C．利用一块实验水稻田的产量估计水稻的实际产量

D．为了了解一批洗衣粉的质量情况，从仓库中任意抽取100袋进行检验

答案　D

解析　A项中某地七月份的日平均最高气温值不能代表全年的日平均最高气温；B项中在农村调查得到的平均寿命，不具代表性；C项中利用一块实验水稻田的产量估计水稻的实际产量，不具代表性；D项具有代表性．

12．研究下列问题：

①某城市元旦前后的气温；②某种新型电器元件使用寿命的测定；③电视台想知道某一个节目的收视率；④银行在收进储户现金时想知道有没有假钞．

一般通过试验获取数据的是(　　)

A．①② B．③④

C．② D．④

答案　C

解析　①通过观察获取数据，③④通过调查获取数据，只有②通过试验获取数据．

13．为制定本市七、八、九年级学生校服的生产计划，有关部门准备对180名初中男生的身高做调查，现有三种调查方案：

(1)测量少年体校中180名男子篮球、排球队员的身高；

(2)网上查阅有关我国其他地市180名男生身高的统计资料；

(3)按本市七、八、九年级学生数目的比例分别从三个年级共抽取180名男生调查其身高．

为了达到估计本市初中这三个年级男生身高分布的目的，则上述调查方案不合理的是\_\_\_\_\_\_，合理的是\_\_\_\_\_\_\_\_．(填序号)

答案　(1)(2)　(3)

解析　(1)中，少年体校的男子篮球、排球的运动员的身高一般高于平均水平，因此不能用测量的结果去估计总体的结果，故方案(1)不合理；(2)中，用外地学生的身高也不能准确地反映本地学生身高的实际情况，故方案(2)不合理；(3)中，由于初中三个年级的男生身高是不同的，所以应该用按比例分别抽取的方法从初中三个年级抽取180名男生调查其身高，方案(3)合理．

14．在一次竞选中，规定一个人获胜的条件是：

(1)在竞选中得票最多；

(2)得票数不低于总票数的一半．如果在计票时，周鹏得票数据丢失．

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 候选人 | 赵明 | 钱红 | 孙华 | 李丽 | 周鹏 |
| 得票数 | 300 | 100 | 30 | 60 | *x* |

请问如果周鹏获胜，那么周鹏的得票数*x*的最小值为\_\_\_\_\_\_\_\_．

答案　490

解析　根据条件，如果周鹏获胜，周鹏的得票数*x*不低于总票数的一半，即≥，解得*x*≥490，且*x*∈**N**，即周鹏得票数至少为490票．



15．为了了解某班学生的会考合格率，要从该班70人中选30人进行考察分析，则70人的会考成绩的全体是\_\_\_\_\_\_，样本是\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_．

答案　总体　30人的会考成绩

解析　70人的会考成绩的全体是总体，样本是30人的会考成绩．

16．某年秋季，某著名的全国性连锁服装店进行了一项关于当年秋季服装流行色的民意调查，调查者通过向顾客发放饮料，并让顾客通过挑选饮料瓶的颜色来对自己喜欢的服装颜色“投票”．这次调查结果显示，某大城市服装颜色的众数(大多数人的选择)为红色，而当年全国服装协会发布的秋季服装流行色是咖啡色．这个结果是否意味着该城市的人比其他城市的人较少倾向于选择咖啡色？你认为这两种调查结果的差异是由什么引起的？

解　这个结果意味着该城市光顾这家连锁店的人比其他城市的人较少倾向于选择咖啡色，由于光顾这家连锁店的人是一种比较容易得到的样本，不一定能代表该城市其他人的想法．而该城市的调查结果来自于该城市光顾这家连锁店的人，这个样本也不能很好地代表全国民众的观点，从而带来了调查结果的差异．