



圣才考研网

www.100exam.com

【圣才考研】——考研专业课辅导中国第一品牌

## 国内外经典教材辅导系列·统计类

适用教材：

①《统计学》（第3版）（袁卫、庞浩、曾五一、贾俊平编著，高等教育出版社）

②《统计学》（贾俊平编著，清华大学出版社）

# 袁卫《统计学》（第3版）

## 笔记和习题（含考研真题）详解

主编：圣才考研网

www.100exam.com

赠

140元大礼包



YZLI0890075925

100元网授班 + 20元真题模考 + 20元圣才学习卡

详情登录：圣才考研网（www.100exam.com）首页的【购书大礼包专区】，

刮开本书所贴防伪标的密码享受购书大礼包增值服务。

特别推荐：圣才考研专业课辅导班【保录班、面授班、网授班等】

圣才考研网

www.100exam.com

面授班网授班 考研考博真题  
购书享受大礼包增值服务

密码

中国石化出版社

HTTP://WWW.SINOPEC-PRESS.COM

教·育·出·版·中·心

## 国内外经典教材辅导系列·统计类

适用教材:

- ①《统计学》(第3版)(袁卫、庞浩、曾五一、贾俊平编著,高等教育出版社)
- ②《统计学》(贾俊平编著,清华大学出版社)

# 袁卫《统计学》(第3版) 笔记和习题(含考研真题)详解

主编: 壹才考研网

www.100exam.com

YZLI



YZLI0890076925

中国石化出版社

## 内 容 提 要

统计类国内外经典教材辅导系列是一套全面解析统计类国内外经典教材的辅导资料。袁卫的《统计学》、贾俊平的《统计学》等是国内最受欢迎的统计学经典教材。本书基本遵循袁卫的《统计学》(第3版)的章目编排,共分10章,每章由三部分组成:第一部分为复习笔记,总结本章的重难点内容;第二部分是课后习题详解,对袁卫的《统计学》(第3版)的所有习题都进行了详细的分析和解答;第三部分是考研真题与典型习题详解,精选部分名校历年考研真题和相关习题,并提供了详细的参考答案。

圣才考研网(www.100exam.com)提供全国所有高校各个专业的考研考博辅导班(保过班、面授班、网授班等)、国内外经典教材名师讲堂(详细介绍参见本书书前彩页)。购书享受大礼包增值服务【100元网授班+20元真题模考+20元圣才学习卡】。本书特别适用于各大院校学习该教材的师生以及在硕士和博士研究生入学考试和复试中参加统计学科目考试的考生。

### 图书在版编目(CIP)数据

袁卫《统计学》(第3版)笔记和习题(含考研真题)详解/圣才学习网主编. —北京:中国石化出版社, 2011.1

(国内外经典教材辅导系列)  
ISBN 978-7-5114-0749-8

I. ①袁… II. ①圣… III. ①统计学-高等学校-教学参考资料 IV. ①C8

中国版本图书馆 CIP 数据核字(2011)第 004647 号

未经本社书面授权,本书任何部分不得被复制、抄袭,或者以任何形式或任何方式传播。版权所有,侵权必究。

### 中国石化出版社出版发行

地址:北京市东城区安定门外大街58号

邮编:100011 电话:(010)84271850

读者服务部电话:(010)84289974

http://www.sinopec-press.com

E-mail:press@sinopec.com.cn

北京东运印刷有限公司印刷

全国各地新华书店经销

\*

787×1092毫米16开本16.5印张4彩插394千字

2011年1月第1版 2011年1月第1次印刷

定价:33.00元

# 言 刊

## 《国内外经典教材辅导系列·统计类》

本系列教材辅导系列由国内著名专家学者精心编写，旨在帮助广大考生系统地掌握国内外经典教材的精髓，提高学习效率。本系列教材辅导系列涵盖了统计学领域的多个重要分支，包括基础理论、应用统计、多元统计等。本系列教材辅导系列由国内著名专家学者精心编写，旨在帮助广大考生系统地掌握国内外经典教材的精髓，提高学习效率。

**主编：圣才考研网 (www.100exam.com)**

**编委：李天燕 邱亚辉 郑炳 肖娟 张宇宁**

**周玉芳 程新慧 黄永民 孙新华 田小文**

**严宽 郑云龙 吴平**

本系列教材辅导系列由国内著名专家学者精心编写，旨在帮助广大考生系统地掌握国内外经典教材的精髓，提高学习效率。本系列教材辅导系列涵盖了统计学领域的多个重要分支，包括基础理论、应用统计、多元统计等。本系列教材辅导系列由国内著名专家学者精心编写，旨在帮助广大考生系统地掌握国内外经典教材的精髓，提高学习效率。

咨询电话：010-82519451, 4006-153-151 (北京)

QQ 咨询：437400084, 284384832

网络咨询：圣才网 www.100exam.com

官方网站：www.100exam.com

圣才网

# 序 言

《类书卷·匠表号辞林卷典登术内国》

统计类国内外经典教材辅导系列是一套全面解析统计类国内外经典教材的辅导资料。我国各大院校一般都把国内外通用的权威教科书作为本科生和研究生学习专业课程的参考教材,这些教材甚至被很多考试(特别是硕士和博士入学考试)和培训项目作为指定参考书。但这些国内外优秀教材的内容一般有一定的广度和深度,课(章)后习题一般没有答案或者答案简单,这给许多读者在学习专业教材时带来了一定的困难。为了帮助读者更好地学习专业课,我们有针对性地编著了一套与国内外教材配套的复习资料,整理了各章的笔记,并对课(章)后的习题进行了详细的解答。

袁卫的《统计学》是一本经典的统计学优秀教材。作为该教材的学习辅导书,本书具有以下几个方面的特点:

1. 整理名校笔记,浓缩内容精华。每章的复习笔记以袁卫的《统计学》为主对本章的重难点内容进行了整理,并参考了国内名校名师讲授该教材的课堂笔记对部分知识点进行了延伸和归纳,因此,本书的复习笔记几乎浓缩了该相关经典教材的知识内容。

2. 解析课后习题,提供详尽答案。本书以该教材为基本依据,参考了该教材的国内外配套资料和其他统计学教材的相关知识对该教材的课(章)后习题进行了详细的分析和解答。可以说,习题的答案在以教材知识基础上得到了进一步的完善。

3. 精选名校考研真题,解决知识难点。考研真题与典型习题的内容涉及基础的数理理论以及实际问题分析等,题型包括单选、多选、判断、简答、计算题,并对每一章的习题作了较为详细的分析和解答。

需要特别说明的是:我们深深感谢袁卫教授和高等教育出版社为我们提供了这样一本优秀的统计学教材。本书的笔记和习题解答参考了国内外教材的配套资料和相关参考书,如有不妥,敬请指正,在此表示感谢。

圣才学习网([www.100xuexi.com](http://www.100xuexi.com))是一家为全国各种学历、职称资格考试和专业课学习提供辅导(面授班、网授班等)、教辅图书、视频资源库、题库系统(在线考试)、电子类资料等全方位教育服务的综合性学习型门户网站,包括圣才考研网、中华统计学习网、中华金融学习网、中华证券学习网、中华经济学习网等50个子网站。

圣才考研网([www.100exam.com](http://www.100exam.com))是圣才学习网旗下的考研考博专业网站,提供全国所有院校各个专业的考研考博辅导班(保过班、面授班、网授班等)、经典教材名师讲堂、考研题库系统(在线考试)、全套资料(历年真题及答案、笔记讲义等)、考研教辅图书等。购书享受大礼包增值服务【100元网授班+20元真题模考+20元圣才学习卡】。

考研咨询:010-62516421, 4006-123-191(免长途费)

QQ咨询:474400084, 564948255

考研辅导:圣才考研网 [www.100exam.com](http://www.100exam.com)

官方总站:圣才学习网 [www.100xuexi.com](http://www.100xuexi.com)

圣才学习网编辑部

# 目 录

(305)	.....	.....	.....
(303)	.....	.....	.....
(309)	.....	.....	.....
(317)	.....	.....	.....
(321)	.....	.....	.....
<b>第1章 数据与统计学</b>	.....	.....	(110)
1.1.1 复习笔记	.....	.....	(110)
1.2 课后习题详解	.....	.....	(120)
1.3 考研真题与典型习题详解	.....	.....	(3)
<b>第2章 统计数据的描述</b>	.....	.....	(8)
2.1 复习笔记	.....	.....	(8)
2.2 课后习题详解	.....	.....	(12)
2.3 考研真题与典型习题详解	.....	.....	(27)
<b>第3章 概率、概率分布与抽样分布</b>	.....	.....	(37)
3.1 复习笔记	.....	.....	(37)
3.2 课后习题详解	.....	.....	(41)
3.3 考研真题与典型习题详解	.....	.....	(55)
<b>第4章 参数估计</b>	.....	.....	(66)
4.1 复习笔记	.....	.....	(66)
4.2 课后习题详解	.....	.....	(68)
4.3 考研真题与典型习题详解	.....	.....	(76)
<b>第5章 假设检验</b>	.....	.....	(88)
5.1 复习笔记	.....	.....	(88)
5.2 课后习题详解	.....	.....	(92)
5.3 考研真题与典型习题详解	.....	.....	(99)
<b>第6章 方差分析与试验设计</b>	.....	.....	(112)
6.1 复习笔记	.....	.....	(112)
6.2 课后习题详解	.....	.....	(116)
6.3 考研真题与典型习题详解	.....	.....	(124)
<b>第7章 相关与回归分析</b>	.....	.....	(133)
7.1 复习笔记	.....	.....	(133)
7.2 课后习题详解	.....	.....	(138)
7.3 考研真题与典型习题详解	.....	.....	(151)
<b>第8章 时间序列分析与预测</b>	.....	.....	(167)
8.1 复习笔记	.....	.....	(167)
8.2 课后习题详解	.....	.....	(172)
8.3 考研真题与典型习题详解	.....	.....	(188)

第9章 统计指数 .....	(203)
9.1 复习笔记 .....	(203)
9.2 课后习题详解 .....	(209)
9.3 考研真题与典型习题详解 .....	(217)
第10章 国民经济统计基础知识 .....	(231)
10.1 复习笔记 .....	(231)
10.2 课后习题详解 .....	(241)
10.3 考研真题与典型习题详解 .....	(248)
( 3 ) .....	3.1
( 8 ) .....	3.2
( 8 ) .....	3.1
( 13 ) .....	3.2
( 23 ) .....	3.3
( 37 ) .....	3.2
( 37 ) .....	3.1
( 41 ) .....	3.2
( 22 ) .....	3.3
( 66 ) .....	4.1
( 66 ) .....	4.1
( 68 ) .....	4.2
( 76 ) .....	4.3
( 88 ) .....	4.2
( 88 ) .....	4.1
( 92 ) .....	4.2
( 99 ) .....	4.3
( 111 ) .....	4.2
( 112 ) .....	4.1
( 116 ) .....	4.2
( 124 ) .....	4.3
( 133 ) .....	4.2
( 133 ) .....	4.1
( 138 ) .....	4.2
( 141 ) .....	4.3
( 167 ) .....	4.2
( 167 ) .....	4.1
( 172 ) .....	4.2
( 188 ) .....	4.3



# 第1章 数据与统计学

## 1.1 复习笔记

### 一、统计数据与统计学

#### 1. 统计学的概念

统计学是研究如何搜集数据、整理数据、分析数据，以便从中做出正确推断的认识方法论科学。实际上，它是一门方法论的科学而不是实质性科学。

#### 2. 统计学和统计数据的关系

统计学是由收集、整理、显示和分析统计数据的方法组成的，这些方法来源于对统计数据的研究，目的也在于对统计数据的研究。离开了统计数据，统计方法乃至统计学就失去了其存在的意义。

### 二、统计学的产生和发展

最早的统计是作为国家重要事项的记录，从统计的产生和发展过程来看，可以把统计学划分为三个时期：一是统计学的萌芽时期，主要有国势学派和政治算术学派；二是统计学的近代时期，主要有数理统计学派和社会统计学派；三是统计学的现代期，主要表现为统计学吸收数学营养的程度越来越迅速；统计学向其他学科领域渗透的能力越来越强；统计学的应用日趋广泛和深入，所发挥的功效日益增强。

### 三、统计学的内容

统计学的内容由描述统计和推断统计组成。描述统计是用图形、表格和概括性的数字对数据进行描述的统计方法。推断统计是根据样本信息对总体进行估计、假设检验、预测或其他推断的统计方法。推断统计主要有两种类型，即参数估计和假设检验。

### 四、统计数据的来源

统计数据来源于直接组织的调查、观察和科学试验，称之为第一手数据或直接的数据；或者来源于已有的数据，称之为第二手数据或间接的数据。

### 五、统计数据的质量

1. 抽样误差是由于抽样的随机性引起的样本结果与总体真值之间的误差。这种误差虽然不可避免，但是可以控制。

2. 非抽样误差是相对于抽样误差而言的，是指除了抽样误差之外的，由于其他原因引起的样本观察值与总体真值之间的差异。非抽样误差特别是其中的系统偏差是可以避免的，但如果不注意，这类误差造成的结果对调查质量来说是致命的。

### 六、统计学的基本概念

1. 总体：是指包含所研究的全部个体的集合。

2. 变量：是说明现象某种特征的概念，特点是从一次观察到下一次观察结果会呈现出差别或变化。

3. 样本：是指从总体中抽取的一部分元素的集合。



## 1.2 课后习题详解

### 1. 什么是统计学？怎样理解统计学与统计数据的关系？

答：统计学是一门收集、整理、显示和分析统计数据的科学，其目的是探索数据内在的数量规律性。

统计学是由收集、整理、显示和分析统计数据的方法组成的，这些方法来源于对统计数据的研究，目的也在于对统计数据的研究。离开了统计数据，统计方法乃至统计学就失去了其存在的意义。

### 2. 试举出日常生活或工作中统计数据及其规律性的例子。

答：(1)对人类性别比例的调查，新生儿男女性别比为105:100，如果没有人为的干扰，其规律是婴幼儿时男性略多于女性，中青年时男女人数大致相同，老年时女性又略多于男性。

(2)施肥量与粮食产量之间的数量关系的调查研究，其规律性为：某种粮食作物的产量会随施肥量的增加而增加。当开始增加施肥量时，产量增加较快，以后增加同样的施肥量，粮食产量的增加量逐渐减少。当施肥量增加到一定数量，产量不再增加。这时如果再增加肥料，产量反而会减少。

(3)商品广告费用与销售额的关系的调查，其规律性为：随着广告费用的增加，商品的知名度和销售额会相应增加。

### 3. 简要说明统计数据的来源。

答：统计数据的来源大致分为两种，其中来源于直接组织的调查、观察和科学试验的数据，称为第一手数据或直接的数据；来源于已有的数据，称为第二手数据或间接的数据。

### 4. 获取直接统计数据的渠道主要有哪些？

答：(1)对于社会经济管理和决策而言，主要是通过统计调查的方式获取数据，如客户满意度调查、电视收视率调查、家庭收支情况调查、居民闲暇时间利用调查等。

(2)在自然科学和工程的研究领域，通常是通过科学实验的方法获得研究的统计数据。

### 5. 简要说明抽样误差和非抽样误差。

答：(1)抽样误差是利用样本推断总体时产生的误差。抽样误差对任何一个随机样本来讲都是不可避免的，但它又是可以计量的，并且是可以控制的。在坚持随机原则的条件下，一般来讲，样本量越大，抽样误差就越小。

(2)非抽样误差是由于调查过程中各有关环节工作失误造成的。它包括调查方案中有关规定或解释不明确所导致的填报错误、抄录错误、汇总错误，不完整的抽样框导致的误差，调查中由于被调查者不回答产生的误差等。非抽样误差在普查、抽样调查中都可能发生。从理论上讲，这类误差是可以避免的。

6. 一家大型油漆零售商收到了客户关于油漆罐分量不足的许多抱怨。因此，他们开始检查供货商的集装箱，有问题的将其退回。最近的一个集装箱装的是2440加仑的油漆罐。这家零售商抽查了50罐油漆，每一罐的质量精确到4位小数。装满的油漆罐应为4.536kg。

要求：

(1)描述总体；

(2)描述研究变量；

(3)描述样本；

#### (4) 描述推断。

- 答：(1) 总体是“最近的一个集装箱装的 2440 加仑的油漆罐”；  
(2) 研究变量是“装满的油漆罐的质量”；  
(3) 样本是“抽查的 50 罐油漆”；  
(4) 根据样本信息对总体进行估计、假设检验，从而推断油漆罐分量是否充足。

7. “可乐战”是描述市场上“可口可乐”与“百事可乐”激烈竞争的一个流行术语。这场战役因影视明星、体育明星的参与以及消费者对品尝试验优先权的抱怨而颇具特色。假定作为百事可乐营销战役的一部分，选择了 1000 名消费者进行匿名性质的品尝试验（即在品尝试验中，两个品牌不做外观标记），请每一名被测试者说出 A 品牌或 B 品牌中哪个口味更好。要求：

- (1) 描述总体；
- (2) 描述研究变量；
- (3) 描述样本；

#### (4) 描述推断。

- 答：(1) 总体是“市场上所有的可口可乐与百事可乐消费者”；  
(2) 研究变量是“消费者偏爱的可乐口味”；  
(3) 样本是“抽取的 1000 名消费者”；  
(4) 推断是①A 品牌口味更好；②B 品牌口味更好。

### 1.3 考研真题与典型习题详解

#### 一、单项选择题

1. 对由实验或调查而得到的数据进行登记、审核、整理、归类，计算出各种反映总体数量特征的综合指标，并加以分析，从中抽出有用的信息，用表格或图形表示出来。这种方法属于( )。[中南财大 2003 研]

- A. 统计观察法 B. 统计描述法 C. 统计推断法 D. 参数估计法

【答案】B

【解析】描述统计学是研究为了反映客观现象总体的数量特征，采用的数据采集方法、数据加工整理方法、数据综合分析方法，计算各项指标反映数据的构成和分布，以及用一定形式的表式和图形把结果显示出来等等。而推断统计学是在概率论的基础上研究由随机样本的数量特征信息来推断总体的数量特征，并做出具有一定可靠程度的估计或检验，包括参数估计和假设检验两个大类。

2. 普查是为了某种特定的目的而( )。[江苏大学 2009 研]

A. 专门组织的一次性全面调查 B. 专门组织的经常性全面调查  
C. 非专门组织的经常性全面调查 D. 非专门组织的一次性全面调查

【答案】A

【解析】普查是专门组织的一次性的全面调查，用来调查属于一定时点上或时期内的现象的总量。它比任何其他调查方式所搜集的资料都更全面、更系统，但普查涉及千家万户，所花费的时间、人力、财力和物力都极其可观，因而只能间隔较长时间进行一次，而两次普查之间的年份以抽样调查方法获得连续的统计数据。

3. 下列不属于描述统计问题的是( )。

- A. 根据样本信息对总体进行的推断
- B. 了解数据分布的特征
- C. 分析感兴趣的总体特征
- D. 利用图、表或其他数据汇总工具分析数据

**【答案】A**

**【解析】**统计学的内容由描述统计和推断统计组成。其中，描述统计研究的是数据收集、处理、汇总、图表描述、概括与分析等统计方法；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方法。根据样本信息对总体进行的推断属于推断统计，所以 A 项属于推断统计的内容。

4. 下列叙述中，采用推断统计方法的是( )。
- A. 用饼图描述某企业职工的学历构成
  - B. 反映大学生统计学成绩的条形图
  - C. 一个城市在 1 月份的平均汽油价格
  - D. 从一个果园中采摘 36 个橘子，利用这 36 个橘子的平均重量估计果园中橘子的平均重量

**【答案】D**

**【解析】**推断统计是根据样本信息对总体进行估计、假设检验、预测或其他推断的统计方法。D 项，根据 36 个橘子的平均重量估计果园中橘子的平均重量属于推断统计方法。ABC 三项属于描述统计的内容。

5. 如果一个样本因人故意操纵而出现偏差，这种误差属于( )。
- A. 抽样误差
  - B. 非抽样误差
  - C. 设计误差
  - D. 实验误差

**【答案】B**

**【解析】**非抽样误差是由于调查过程中各有关环节工作失误造成的。它包括调查方案中有关规定或解释不明确所导致的填报错误、抄录错误、汇总错误，不完整的抽样框导致的误差，人为干扰造成的误差，调查中由于被调查者不回答产生的误差等。

6. 下列说法错误的是( )。

- A. 抽样误差只存在于概率抽样中
- B. 非抽样误差只存在于非概率抽样中
- C. 无论是概率抽样还是非概率抽样都存在非抽样误差
- D. 在全面调查中也存在非抽样误差

**【答案】B**

**【解析】**抽样误差是由于抽样的随机性引起的样本结果与总体真值之间的误差；非抽样误差是相对抽样误差而言的，是指除抽样误差之外的，由于其他原因引起的样本观察结果与总体真值之间的差异。抽样误差是一种随机误差，只存在于概率抽样中，非抽样误差则不同，无论是概率抽样、非概率抽样，或是在全面性调查中，都有可能产生非抽样误差。

7. 对电视机的平均寿命进行调查，应该采用( )。

- A. 普查
- B. 重点调查
- C. 典型调查
- D. 抽样调查

**【答案】D**

8. 要了解某市工业企业生产设备情况，则统计总体是( )。

- A. 该市全部工业企业
- B. 该市每一个工业企业

- C. 该市工业企业的每一台设备                      D. 该市工业企业的全部生产设备

【答案】D

【解析】总体是人们研究的所有基本单位(通常是人、物体、交易或事件)的总和。要了解某市工业企业生产设备情况,则统计总体应该为该市工业企业的全部生产设备。C项是一个个体。

9. 为了估计全国高中学生的平均身高,从20个城市选取了100所中学进行调查。在该项研究中,样本是( )。

- A. 100所中学                      B. 20个城市

- C. 全国的高中学生                      D. 100所中学的高中学生

【答案】D

【解析】样本是从总体中抽取的一部分元素的集合,构成样本的元素的数目称为样本量。抽样的目的是根据样本的信息推断总体的特征。从20个城市选取了100所中学进行调查,这100所中学的高中学生就构成了一个样本。

### 二、多项选择题

1. 经济普查是( )。[首经贸2006研]

- A. 专门调查                      B. 一次性调查                      C. 全面调查                      D. 非全面调查

【答案】ABC

【解析】普查是专门组织的一种全面调查,它主要用以搜集某些不能或不宜用定期报表搜集的统计资料。我国的周期性普查制度包括人口普查、农业普查和经济普查。

2. 搜集数据的组织方式有( )。

- A. 普查                      B. 抽样调查                      C. 重点调查                      D. 统计报表制度

E. 系统抽样

【答案】ABCDE

3. 推断统计的两种主要类型是( )。

- A. 参数估计                      B. 预测                      C. 决策                      D. 假设检验

E. 描述统计

【答案】AD

【解析】推断统计是根据样本信息对总体进行估计、假设检验、预测或其他推断的统计方法,推断统计的两种主要类型是参数估计和假设检验。

### 三、判断题

1. 统计学是一门以大量现象的数量方面为其研究对象的认识方法论科学。[首经贸2006研]

【答案】✓

【解析】统计学是一门收集、整理、显示和分析统计数据的科学,其目的是探索数据内在的数量规律性。统计学的研究对象是客观现象总体数量的数量特征和数量关系。

2. 第二手数据可以通过抽样调查获得。( )

【答案】×

【解析】抽样调查获得的数据属于第一手数据;第二手数据是来源于已知的数据,不需要调查即可利用的数据。

3. 我国的人口普查和经济普查都是每10年进行一次。( )

【答案】×

【解析】我国的人口普查是每10年进行一次，在逢0的年份实施；经济普查是每5年进行一次，分别是在逢3、逢8的年份实施。

4. 对某家公司进行审计，该公司年度内的所有发票是55400张，审计人员从中随机抽查了100张发票进行审查，发现有2张发票有差错。则总体是100张发票，样本是2张发票。

【答案】×

【解析】总体是指包含所研究的全部个体的集合。样本是指从总体中抽取的一部分元素的集合。该题中总体是55400张发票，样本是100张发票。

#### 四、简答题

1. 描述统计学和推断统计学的关系是什么？[西安交大2006研]

答：描述统计学是用图形、表格和概括性的数字对数据进行描述的统计方法。推断统计学是根据样本信息对总体进行估计、假设检验、预测或其他推断的统计方法。描述统计是整个统计学的基础和统计研究工作的第一步。推断统计是现代统计学的核心和统计研究工作的关键环节。如果没有描述统计收集可靠的数据并提供有效的样本信息，即使很高的统计学家和很科学的推断方法也难以得出准确的结论。因而，推断统计对描述统计又有很强的依赖性。

2. 分别简述统计总体与样本的含义及其特点。[中南财大2006研]

答：(1) 总体

##### ① 总体的含义

总体是指人们研究的所有基本单位(通常是人、物体、交易或事件)的总和。每一个总和都包括了研究总体的所有单位。

##### ② 总体的特点

a. 同质性。构成总体的各个单位必须具有某一方面的共性，这个共性是我们确定总体范围的标准。

b. 大量性。总体是由许多单位所组成的，而不是只有个别单位；

c. 差异性。总体单位之间，除了必须在某一方面有共性之外，在其他方面必然存在差异。

##### (2) 样本

##### ① 样本的含义

从总体中抽取出来，作为代表这一总体的部分单位组成的集合体称为样本。

##### ② 样本的特点

a. 样本的单位必须取自全及总体内部，不许总体外部单位参加；

b. 从一个全及总体可以抽取许多个样本；

c. 样本的代表性。抽取样本不是目的，而是手段，是用来推断总体的，因此存在样本的代表性问题；

d. 样本的客观性。从全部总体中抽取样本，必须排除主观因素的影响。

3. 常用的统计调查方式主要有哪些？[北京林业大学2005研]

答：常用的统计调查方式主要有：

(1) 普查，是专门组织的一种全面调查，它主要用以搜集某些不能或不宜用定期报表搜



集的统计资料。对国情国力的调查一般采用普查，如人口普查、工业普查和农业普查等。

(2)统计报表制度，是依照国家有关法规，自上而下地统一布置，以一定的原始记录为依据，按照统一的表式，统一的指标项目，统一的报送时间和报送程序，自下而上逐级地定期提供统计数据的一种调查方式。

(3)抽样调查，是一种非全面调查，包括概率抽样和非概率抽样。

(4)重点调查，是指在调查对象中，只选择一部分重点单位进行的非全面调查。

(5)典型调查，是一种专门组织的非全面调查。它是根据调查的目的，在对所研究的对象进行初步分析的基础上，有意识地选取若干具有代表性的单位进行调查和研究，借以认识事物发展变化的规律。

#### 4. 简述普查和抽样调查的特点。

答：(1)普查是为某一特定目的而专门组织的一次性全面调查。它具有如下特点：

①普查通常是周期性的。由于普查涉及面广、调查单位多，需要耗费大量的人力、物力、财力和时间，通常需要间隔较长的时间进行一次。

②普查一般需要规定统一的标准调查时间，以免遗漏，保证普查结果的准确性。

③普查的数据一般比较准确，规范化程度也比较高，因此它可以为抽样调查或其他调查提供基本依据。

④普查的使用范围比较狭窄，只能调查一些最基本、特定的现象。

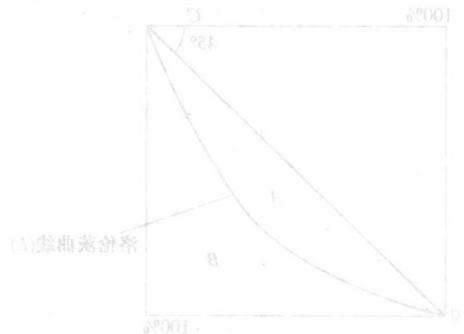
(2)抽样调查则是从调查对象的总体中随机抽取一部分单位作为样本进行调查，根据样本调查结果来推断总体数量特征的一种数据收集方法。它具有以下几个特点：

①经济性。这是抽样调查的一个最显著的优点。由于调查的样本单位通常是总体单位中很小的一部分，调查的工作量小，因而可以节省大量的人力、物力、财力和时间。

②时效性强。由于工作量小，调查的准备时间、调查时间、数据处理时间等都可以大大缩减，从而提高数据的时效性。

③适应面广。抽样调查可以获得更广泛的信息，它适用于各个领域、各种问题的调查。从适用范围和问题来看，它的适用面要广于全面调查的适用面。抽样调查还适用于一些特殊现象的调查，比如产品质量检验、农产品实验、医药的临床实验等。从调查的项目和指标来看，抽样调查的内容和指标可以更详细、更深入。

④准确性高。由于抽样调查的工作量较全面调查的工作量小，因此，它可以减少由于工作量大、环节多而造成的误差。当然，用样本数据去推断总体时会不可避免地出现推断误差，但这种误差的大小是可以计算并加以控制的，因此推断的结果通常是可靠的。



(2) 误差与样本量

$$\frac{e}{B+A} = \text{误差系数}$$

1-5图



## 第2章 统计数据的描述

### 2.1 复习笔记

#### 一、统计数据的整理

##### 1. 统计数据的分组

统计分组：指按照统计研究的目的，将数据分别列入不同的组内。

按品质标志分组是指按照性别、质量等级等定性指标进行的分组；按数量标志分组是指按照数量或数值等定量指标进行的分组。

##### 2. 次数分配

分组原则：在分组时，要遵循“不重不漏”的原则。

不重是指任一个单位数值只能分在其中某一组中，不能同时分在两组中。当相邻两组的上下限重叠时，应遵循“上组限不在内”的原则；不漏是指任一数值必须分在某一组内，不能遗漏。

##### 3. 次数分配图形

(1) 直方图：在平面直角坐标系上，将分组标志作为横轴，并将各组次数作为纵轴，画出各组的长方形图即直方图。

(2) 折线图：以各组标志值中点位置作为该组标志的代表值，然后用折线将各组次数连接起来，形成折线图。

(3) 折线图与直方图的关系：直方图与折线图的面积是相等的。折线图的折线将直方图的直角切下，正好补在旁边较低的直方图上。这样，直方图与折线图所表示的分布规律是相同的，是两种面积相同但表示形式不同的次数分配图示法。

(4) 次数分布曲线：当所观察的次数越多，组距越小且组数越多时，所给出的折线图就会越光滑，逐渐形成一条光滑的曲线，即次数分布曲线。它反映了数据或统计量的分布规律。

##### 4. 洛伦茨曲线与基尼系数

(1) 洛伦茨曲线(如图2-1所示)是描述收入和财富分配性质的曲线。

图2-1中，横轴是累积的人口百分比，纵轴是累积的收入或财富百分比。如果一个国家或地区的收入分配完全按人口平均分配，则此时同一累积百分比的人口就一定占有相同累积百分比的收入。这时，该国的收入分配程度曲线就与绝对平均的对角线重合。如果某国绝大多数人口占有很少的财富和收入，而一小部分人口占有了绝大部分的收入，则该国的曲线就靠近下横轴和右纵轴。

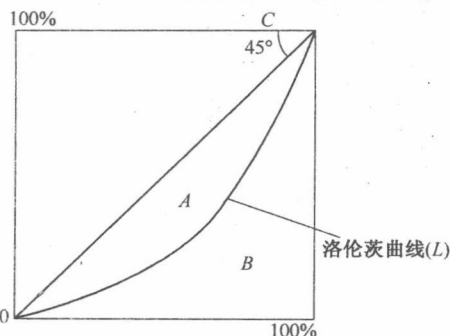


图 2-1

国家或地区的收入分配完全按人口平均分配，则此时同一累积百分比的人口就一定占有相同累积百分比的收入。这时，该国的收入分配程度曲线就与绝对平均的对角线重合。如果某国绝大多数人口占有很少的财富和收入，而一小部分人口占有了绝大部分的收入，则该国的曲线就靠近下横轴和右纵轴。

(2) 基尼系数(衡量收入分配平均程度)

$$\text{基尼系数} = \frac{A}{A+B}$$

式中:  $A$  表示实际收入  $L$  曲线与绝对平均线(对角线)之间的面积;  $B$  表示实际收入  $L$  曲线与绝对不平均线之间的面积。

如果  $A=0$ , 则基尼系数  $=0$ , 表示收入绝对平均; 如果  $B=0$ , 则基尼系数  $=1$ , 表示收入绝对不平均。基尼系数在 0 和 1 之间取值, 一般认为, 基尼系数若小于 0.2, 表明分配平均但缺乏效率; 基尼系数在 0.2~0.4 之间是比较适当的, 即一个社会既有效率又没有造成极大的分配不公; 基尼系数为 0.4 时, 被认为是收入分配不公平的警戒线; 基尼系数超过了 0.6, 则表示可能由于收入分配不公导致社会不稳定。

## 二、分布集中趋势的测度

### 1. 众数

众数是总体单位中标志值出现次数最多的那个数值。它是一种位置代表值。

对于分组数据, 计算众数的下限公式为:  $M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$ ; 上限公式为  $M_o = U - \frac{\Delta_2}{\Delta_1 + \Delta_2} \times i$ 。

### 2. 中位数

指数数据排序后, 位置在最中间的数值, 可表示为  $M_o$ 。中位数将数据分成两半, 一半数据比中位数大, 一半数据比中位数小。

要求得中位数, 首先要确定中位数的位次。未分组资料时, 中位数位次  $= \frac{N+1}{2}$ ;

当总体位数  $N$  为奇数时, 中位数就是中位数位次上的那个数据; 当  $N$  为偶数时, 中位数是中位数位次上两项数据的算术平均数。

分组资料时, 中位数位次  $= \frac{\sum f}{2}$ ; 计算中位数的下限公式为:

$$M_e = L + \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \times i$$

上限公式为:

$$M_e = U - \frac{\frac{\sum f}{2} - S_{m+1}}{f_m} \times i$$

中位数具有的性质: 数据值与中位数之差的绝对值之和最小, 即  $\sum_{i=1}^n |x_i - M_e| = \min$  (最小)。

### 3. 四分位数

四分位数就是将数据分布 4 等分的三个数值, 其中中间的四分位数就是中位数。

(1) 由未分组资料计算四分位数, 首先是确定四分位数的位次, 再找出对应位次的标志值即为四分位数。设样本容量为  $n$ , 则:

$$M_1 \text{ 的位次} = \frac{(n+1)}{4}$$

$$M_3 \text{ 的位次} = \frac{3(n+1)}{4}$$

如果  $\frac{(n+1)}{4}$  计算出来的位次恰好是整数, 这时各位位次上的标志值即为相应的四分位

数。如果 $\frac{(n+1)}{4}$ 计算出来的位次不是整数，这时可用插值法计算四分位数，即与该位次相邻的两个整数位次上的标志值的加权算术平均数，权数的大小取决于两个整数位次与四分位次距离的远近，距离越近，权数越大。

(2)由分组资料计算四分位数。第 $i$ 四分位数的计算公式为：

$$M_i = L_i + \frac{\frac{i \sum f}{4} - S_{m_i-1}}{f_{m_i}} \times d_i \quad (i=1, 2, 3)$$

#### 4. 均值

即算术平均数，是数据集中趋势的最主要测度值。

(1)对于未分组数据采用算术平均数，其计算公式为：

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

(2)对于分组数据采用加权算术平均数，其计算公式为：

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$$

#### 5. 几何平均数

指 $n$ 个比率连乘积的 $n$ 次方根，即

$$G = \sqrt[n]{a_1 \times a_2 \times \cdots \times a_n} = \sqrt[n]{\prod_{i=1}^n a_i}$$

#### 6. 切尾均值

指去掉大小两端的若干数值后计算中间数据的均值。其计算公式为：

$$\bar{x}_\alpha = \frac{x_{(n\alpha+1)} + x_{(n\alpha+2)} + \cdots + x_{(n-n\alpha)}}{n - 2 \times n\alpha}$$

式中： $n$ 表示观察值的个数； $\alpha$ 表示切尾系数， $0 \leq \alpha < \frac{1}{2}$ ； $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 是数据 $x_1, x_2, \cdots, x_n$ 经过排队后由小到大形成的顺序统计量值。

改变切尾系数 $\alpha$ 的值可以选择集中趋势的测度值。当 $\alpha$ 取0或 $\alpha$ 接近1/2时，切尾均值公式变成算术平均数和中位数的公式，这是切尾均值的两种特例。

#### 7. 众数、中位数和均值的关系

在对称的次数分配和统计分布中，众数、中位数和均值都是同一数值。在尾巴拖在右边的右偏(正偏)分布中，众数 $<$ 中位数 $<$ 均值；而在尾巴拖在左边的左偏(负偏)分布中，均值 $<$ 中位数 $<$ 众数。

### 三、分布离散程度的测度

#### 1. 极差(全距)

即数据最大值与最小值之差，它是数据离散或差异程度的最简单测度值，即

$$R = \max(x_i) - \min(x_i) \text{ 或 } R = x_{(n)} - x_{(1)}$$

#### 2. 内距

即两个四分位数之差，即

$$\text{内距} = \text{上四分位数} - \text{下四分位数} = Q_3 - Q_1$$