

前 言

一、在改革开放中设置和不断发展的《应用统计》课程

80年代初，对外经济贸易大学在创办国际教育时，即将“应用统计”(Applied statistics)定为一门主干课程。十余年来，本着以国际企业需求为导向、中西融通、文理渗透的方针，经老中轻三代统计学教师不断探索，已建立起有经贸特色的统计学教学体系，包括一整套教学大纲、教科书、教学参考资料、计算机应用软件和相应的教学方法。并有幸获得北京市优秀教学成果一等奖(1993年)。本课程所用教材，原为李志伟先生等编著的《统计分析概论》(对外贸易教育出版社，1984年初版，1989年修订再版)。

建设社会主义市场经济必然对经贸高等教育，当然也包括对“应用统计”的教学提出新的要求，特别是要求经贸工作者提高在国内外市场的复杂环境中运用统计信息对经营管理进行分析，并作出科学决策的能力。为此，我们决定编写这部《应用统计》，作为各经贸专业本科“应用统计”课程的教科书，并力求使之更好地适应新形势下培养经贸人才的需要。这部教材对“应用统计”的知识重新作了安排，在广度和深度上都较原用教材有所发展。

本教材由贾怀勤主编。其中，绪论、第一章、第五章、第十三章和第十四章由贾怀勤执笔；第八章、第九章、第十章、第十一章和第十二章由张杰执笔；第二章、第三章和第六章由朱雅华执笔；第四章由杜学孔执笔；第七章由王玉荣执笔。

1997年把“应用统计”列入本校15门重点建设课程。根据我

们制订的课程建设计划，对本教材进行了修订，修订本已于 1998 年 2 月出版。此次修订工作包括：（1）对部分内容进行了更改；（2）对一些例题及相关统计图表进行了更替；（3）订正了原书中的由于排版而造成的错误。

近二十年来，本着以国际商务需求为导向、中西融通、文理渗透的方针，经过三代统计学教师不断探索，我校已建立起具有经贸特色的应用统计教学体系，包括一整套教学大纲、教科书、教学参考资料、计算机应用软件和相应的教学方法，并在原有基础上不断加以充实提高。

二、本课程的性质、研究对象和教学目的。

应用统计是一门专业基础课。它向学生讲授对统计资料进行搜集、整理和分析的科学方法论，使其具有运用统计信息分析经贸问题，作出科学推断的能力。应用统计，既是一门方法论的独立学科，也为学生学习相关专业课程打下了定量分析的基础。

三、本课程的主要内容

1. 有关统计学的一般概念和研究方法。
2. 统计资料的搜集、整理和表述的方法和原则。
3. 横断面数列的基本特征值----
 频数分布数列的集中趋势量数和离中趋势量数；
 相对数。
4. 总体现象发展变化的描述----
 时间数列；

指数。

5. 统计推断的预备知识----

概率基本知识；

典型概率分布；

抽样分布规律简介。

6. 关于单一总体的平均数和比率的推断----

参数估计；

假设检验。

7. 变量间协变关系的描述与推断----

简单回归和相关分析；

多元回归和相关分析。

四. 对学生的要求

要求掌握应用统计的基本理论、基本方法，并且知道如何应用于实践。对于重要概念、重要公式和重要分析程序，要求达到记忆清楚、运用熟练；对于一般概念、次要公式和一般分析程序，要求理解会用。

教授方法，以讲授为主，有的章节可以安排学生讲，每章安排适量课后作业或课堂练习，通过学生动脑动手加深对所学知识的理解、记忆和提高运用能力。

为了增强课堂讲授效果，要求做好预习和复习。

学生应掌握统计软件的应用。

五、参考书

1. 《管理统计学》。徐国强主编，上海财大出版社
2. 《统计学》。(美) David Freedman等人著
魏宗舒等人译。中国统计出版社出版
3. 《社会经济统计学概要》。袁寿庄主编。
中国人民大学出版社
4. 《商务统计》M. C. 弗莱明、J. G. 纳里斯,中信出版社,1999

年

六、本课程教学时数分布方案

课时数	讲授的内容
3	绪论
6	设计、搜集、整理
3	对比分析
6	单变量截面数列的描述性分析
3	时间数列
3	指数
4	概率和概率分布
6	抽样和抽样分布
4	参数估计
6	假设检验
6	简单回归和相关分析
4	多元回归和相关分析
54	

第一章 绪论

教学目的和要求

了解应用统计的学科性质和研究对象，以及它在管理中的广泛应用，从而明确本课程的意义。了解统计研究的基本程序和本课程内容的基本框架。

基本概念

统计学、统计资料和统计活动

大量观察法，总体与样本

统计描述与统计推断。

教学内容

一、应用统计的学科性质和研究对象

1. 统计资料(statistics)，又称统计数据或统计数字(statistical data)。

所谓统计资料指的是，运用科学的调查和整理方法取得的、反映客观总和现象(Aggregate Phenomena)特征的数据型资料(Numerical data)。

具有数据性、综合性、客观性和具体性四个特点。

2. 统计活动

泛指一切搜集、加工和分析关于客观总和现象数据型资料的研究活动。其中，政府统计机构组织的调查，整理社会经济统计资料的系统活动，称为统计工作，国外叫官方统计活动。

3. 统计学已发展成为一个科学体系。

应用统计学是研究搜集、整理和分析客观总和现象数据资料

方法的科学，是一门通用的方法论科学。

数理统计学以概率论为基础，以随机现象为对象，研究有关统计方法的数学理论基础。

前得在适用对象和方法的涵盖面上均宽余后者。前者正是因其以应用导向，才得名。

各专门领域统计学把一般统计方法应用于各自专门实践与各门实质性科学相结合、研究专用指标体系和统计方法。相对于这些专门领域统计学，应用统计又可称为统计学原理(Pinciples of Statistics)。

二、统计学的基本研究方法—大量观察法

1. 统计学的基本研究方法是大量观察法(Method of Mass Observation)。其他重要的方法有综合指标法、分组法等。
2. 大量观察法的含义是，对客观总和现象组成的集团中的大量个体逐一观察。
3. 总体与总体单位的概念
总体的特点：同质性、大量性、变异性。
4. 大量观察的意义
相互抵消个别的、偶然的差异性，显示集体的、必然的规律性。
显示个体间差异程度。

三、统计活动的基本程序

资料搜集—资料整理—综合分析

四、统计描述和统计推断

1. 样本的概念、作用
2. 统计描述、统计推断的含义，两者间关系

五、统计方法的应用

1. 统计方法在管理中的广泛应用
2. 定性分析和定量分析相结合

第二章 统计资料的搜集整理和表述

教学目的和要求

明确各种调查方式的含义和使用场合。掌握统计资料和表述的要求和格式，特别是频数分布表和图的基本概念和绘制反方法。

基本概念

初级资料与次级资料。

普查、抽样调查和重点调查。

离散型数据和连续型数据，横断面数列与时间数列；质别分组数列与量别分组数列；时点数列与时期数列。

频数分布、组距、组限、组中值、相对频数与累积频数。

教学内容

第一节 统计资料的搜集

一、统计资料的搜集渠道

1. 统计资料分为初级资料和次级资料。
2. 自然科学研究中级资料的搜集方式是观察和实验。
3. 某些社会经济资料可就行政机构的强制性登记之便获得。
4. 本章只讲社会经济初级资料的搜集—调查(survey)。

二、统计调查方法和方式

1. 调查方式

- (1) 全面调查：普查(census)；全面统计报表。

(2) 非全面调查

i. 抽样调查

随机抽样(random sampling)

方便抽样(convenient sampling)

立意抽样(judgement sampling)

ii. 重点调查

iii. 典型调查

2、调查方法：直接调查法、采访法、通讯法、报告法。

三、统计调查设计应注意的问题

1. 调查对象

调查对象的规定：定义法与列举法。

2. 调查表、调查项目和填表说明

调查表按格式分为统计调查表与问卷。

调查表设计原则：简明性、必要性、可能性、规范性、循环性、非诱导性。

3. 调查内容的时间规定

存量：标准调查时间

流量：调查起讫时间。

4. 调查目的决定调查内容。方法和方式。

第二节 统计资料的整理

一、统计资料的审查与汇总

1. 审查：逻辑审查与计算审查
2. 汇总：逐级汇总与集中汇总（超级汇总）；手工汇总与计算机汇总。

二、统计数据与数据的类型

1. 数据的类型与计值方法

离散型数据：点数（count）。

连续型数据：量度（measure）。

2. 数列

统计数据的特点。

横断面数列（cross-section data series）：

质别分组；量别分组数列（频数分布数列）；

时间数列（time series）：

时点数列；时期数列。

三、质别数列的分组

1. 分组原则：穷尽与互斥（同样使用于量别分组）。
2. 分类的概念

第三节 统计资料的表述

一、统计表与统计图

1. 对于统计表、图一般要求：序号、标题、计量单位、资料来源和注释。
2. 统计表体的格式要求：

(1) 行标题（行目）与列标题（栏目）；

(2) 单位标注方法；3种；

(3) 表心数字

“不足以列出”和“不详”的表示方法。

3. 统计图

(1) 种类：条形图、扇形图、线形图和频数分布图（频数多边形与直方图）；

(2) 要求：横纵轴标明单位、变量名和尺度；图形与变量值成比例；正确使用图例。

二、频数分布表的编制方法

1. 基本概念：组距、组限、（上限与下限）、组中值、频数。

2. 组距表写形式

离散型变量

连续型变量

10-19

10-（不足）20

10-20（以下）

20-29

20-（不足）30

或 20-30（以下）

30-39

30-（不足）40

30-40（以下）

不等距和开口组的使用。

3. 分组方法：技术家经验。

4. 相对频数的概念和作用。

5. 累积频数的概念和作用。

较大制累积频数（累积低于上限）

较小制累积频数（累积等于和高于下）

累积相对频数。

三、频数分布图及其作用

直方图与频数多边形；

累积频数曲线与累积相对频数曲线。

第三章 数据的对比分析

教学目的和要求

掌握各种相对数的定义、适用情况及应用；学会用相对数来研究事物及现象的发展变化规律。

基本概念

相对数，结构相对数，比例、类比相对数，强度相对数、计划完成相对数，动态相对数

教学内容

一、结构相对数

结构相对数是将总体按某一标志分组，以各组的总量指标数值与总体总量指标数值对比求得比重或比率，来反映总体内部组成状况的相对指标。

二、比例相对数

比例相对数是将总体按某一标志分组，用各组的总量指标数值相对比而求得比值或比例。

三、类比相对数

类比相对数是把同一指标在相同时间、不同空间条件下的数值相对比的比值或比率。

四、强度相对数

由同一总体的两个性质不同但有一定联系的总量指标相对比而得到的相对数，称为强度相对数。

五、计划完成相对数

计划完成相对数就是把某项工作在一段时间内的实际完成数与计划任务数对比，以表明计划的完成程度。

六、动态相对数

动态相对数就是将不同时期或不同时点的同一经济现象的数量表现进行对比，以观察现象的动态变化情况。

第四章 单变量截面数据的描述性分析

教学目的和要求

变量数量集中趋势指标的算术平均数、中位数、众数，以作为变量数列离散趋势的标准概念、具体计算方法和作用，学会运用平均数来研究事物或现象的总体特征和变化规律。

教学内容

简单算术平均数、加权算术平均数、中位数、众数、几何平均数、标准差及变异系数。

第一节 频数分布的描述性量数

对于按数量分组得到的频数分布，需要一些可以用数字集中地、明确地描述频数分布的量数（本大纲所说的分组数列量数，均指等距数列），主要有两种：

一、趋中量数

用于描述频数分布集中趋势。

常用的趋中量数有：算术平均数、中位数、众数等。

二、离中量数

频数分布中的变量不仅有集中趋势，同时也有偏离趋中量数的离中趋势，态度这种离散趋势的量数叫做离中量数，其中最重

要的有全距、标准差等。

第二节 集中趋势量数

一、算术平均数

算术平均数是一种常用的趋中量数，一般称为平均数。

(一) 简单算术平均数

观察值的总和除以观察值个数的商

简单算术平均数的特点：

1. 它是根据总体或样本的全部数值计算得来。它的数值是唯一的。
2. 各观察值与平均数之间差距的平均和小于任何其他数值之间差距的平方和，即为最小值。
3. 在观察值数目较小的情况下，它易受极端值的影响。

(二) 加权算术平均数：

根据各变量值的重要程度不同，将各个变量值乘以代表个变量值重要程度的权数，最后以乘积总和被权数总和去除，所得之商就是加权平均数。公式为：

$$\mu = \frac{\sum Wx}{\sum W}$$

分组资料算术平均数按下面的公式计算：

$$\mu = \frac{\sum fx}{\sum f} (N = \sum f)$$

二、中位数

(一) 原始数据中位数

将数值按递升或递降的次序排列，“居中”的数即为中位数。

(二) 分组资料的中位数按下面的公式计算：

$$Md = l + \frac{(N/2 - F)}{f} i$$

(三) 中位数的特点

1. 中位数有特殊用途；
2. 各观察值与中位数之间绝对差距之和小于其他任何数值之间的绝对差距之和，既是最小值；
3. 不受极端值影响。

三、众数

(一) 原始数据的众数

是一系列数值中出现次数最多的数值。

(二) 分组资料的众数按下面的公式计算：

$$Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} i$$

(三) 众数的特点

1. 众数有特殊用途
2. 不受极端值的影响

四、平均数、中位数、众数之间的关系

1. 再偏态分布的偏斜程度不太显著时，三量数大致有一个固定的关系，即中位数与算术平均数的差距约等于众数与平均数距全长的 1/3，用公式表示如下：

$$\mu - M_0 = 3(\mu - md) \text{ 或写成}$$

$$M_0 - \mu = 3(Md - \mu)$$

2. 对对称分布： $\mu = Md = M_0$
3. 对于含有开口组的频数分布，可以根据三者关系由 M_0 和 Md 推算 μ 。

五、几何平均数

- (一) 按下面公式计算所得结果成为几何平均数。

$$Mg = \sqrt[n]{x_1 x_2 \cdots x_n}$$

- (二) 再经济中，主要用于求时间序列的平均发展率和平均增长率，

公式为：

$$\text{平均发展率} = \sqrt[n]{\frac{\text{期末水平}}{\text{期初水平}}} * 100\%$$

平均增长率=平均发展率-100%

(三) 进行简单推算

第三节 离中量数

离中量数，用来量度全部数据偏离趋中量数的程度。主要有下面两种类型：

(一) 全距：全距=最大值-最小值

(二) 标准差

标准差是度量全部数据偏离算术平均数的程度。

1. 原始数据的标准差按下面公式计算：

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

2. 分组资料的标准按下面公式计算：

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}}$$

3. 标准差的意义及特点

(1) 对正态分布标准差的数据概念：如果一个总体具有正态分布，则平均数与标准差之间有如下关系：

从 $\mu - \sigma$ 到 $\mu + \sigma$ 之间，约有 68.3% 的观察值在次范围之内

从 $\mu - 2\sigma$ 到 $\mu + 2\sigma$ 之间，约有 95.5% 的观察值在次范围之内

从 $\mu - 3\sigma$ 到 $\mu + 3\sigma$ 之间，约有 99.7% 的观察值在次范围之内

(2) 标准差受极端值的影响

(三) 样本平均数与标准差

以上所说的是总体的方差和标准差，但如果统计资料来源于样本，就要计算样本平均数和标准差。

1. 样本平均数

(1) 原始资料:

$$\bar{x} = \frac{\sum x}{n}$$

(2) 分组资料

$$\bar{x} = \frac{\sum fx}{\sum f}$$

2. 样本标准差

(1) 原始资料

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

(2) 分组资料

$$S = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f - 1}}$$

(四) 变异系数

有时需要对较大差别的样本的分散程度进行比较, 此时可用变异系数这种相对数进行比较, 其公式为:

$$V = \frac{S}{\bar{x}} * 100\%$$

第六章 时间序列分析

教学目的和要求

了解构成时间数列和四个因素的概念，掌握最小二乘法配合线性长期趋势的要领。会运用趋势方程式对时间数列的趋势变化进行分析。了解用移动平均比率法求季节指数的一般步骤，会运用季节指数对时间数列的季节性进行分析。

基本概念

时间数列因素分解，长期趋势分析，季节性变动的测量。

教学内容

第一节 时间数列的变动因素

一、四因素分解

1.长期趋势(T)

2.季节性变动(S)

3.周期性变动(C)

4.不规则变动(I)

二、时间序列分析模式

1.加法模式 $Y=T+S+C+I$ 2.乘法模式 $Y=T.S.C.I$

本大纲使用乘法模式

第二节 长期趋势分析

一、线性图

用于观察变量随时间变化的趋势。

二、趋势分析方法

用最小二乘法求拟合方程式，称为趋势方程式。

三、趋势方程式

1. 线性趋势方程式 ($Y_c = a + bx$)

$$b = \frac{\sum XY}{\sum X^2}$$

$$a = \frac{\sum Y}{n}$$

2. 非线性趋势方程式

(1) 指数趋势 ($Y_c = ab^x$)

$$\log a = \frac{\sum \log Y}{N}$$

$$\log b = \frac{\sum (X \log Y)}{\sum X^2}$$

(2) 抛物线趋势 ($Y_c = a + bx + cx^2$)

$$\sum Y = Na + c \sum X^2$$

$$\sum XY = b \sum X^2$$

$$\sum X^2Y = a\sum X^2 + c\sum X^4$$

3. 趋势方程变量注释
4. 趋势方程系数的经济意义

四、趋势值的求法

第三节 季节性变动的测

一、季节指数的概念

用于表示具有典型季节变动的现象在每个季节（或每个月）的变动方向和幅度的百分数。

二、季节指数的求法：移动平均比率法。

三、季节指数的应用

1. 消除季节因素的影响（称为季节性调整）
经季节性调整值 = 实际观察值 / 季节指数

2. 用于预测

预测值 = 趋势值 * 季节指数

第七章 指数

教学目的和要求

明确指数的定义和编写原理，比率加权平均指数与综合加权指数间的关系和前者独立存在的意义。掌握拉氏公式、派氏公式编制综合加权指数的方法，学会运用指数解释物价和物量的变化方向和幅度，极其对物值变动的影晌。

基本概念

简单指数、加权综合指数、比率加权平均指数。

教学内容

第一节 指数基本概念

一、指数

是用来反映不能直接相加的、涉及到多种因素组成的社会经济现象综合变动方向和幅度的相对数。

二、指数的种类

1. 按所描述的经济现象分为物价指数、物量指数和物值指数。
2. 按权数所属的时间分为基期加权指数和报告期加权的指数等。
3. 按公式形式分为综合加权指数和比率加权综合指数。

第二节 简单指数

一、简单价格指数

$$P_n = \frac{\sum p_n}{\sum p_o} * 100\%$$

二、比率简单平均数

物价指数：

$$P = \frac{1}{N} \sum \frac{p_n}{p_o} * 100\%$$

物数指数：

$$Q = \frac{1}{N} \sum \frac{q_n}{q_o} * 100\%$$

三、简单指数的问题

简单（不加权）指数的都不顾及商品在指数中的比重，难以反映对消费者的实际影响，且易受极端值的影响。由此引出加权指数。

第三节 加权综合指数

一、以基期加权的指数，称为拉斯拜尔(Laspeyres)指数, 简称拉氏指数。公式为：

拉氏指数：

$$P_{on} = \frac{\sum p_n q_o}{\sum p_o q_o} * 100\%$$

拉氏物量指数:

$$Q_{on} = \frac{\sum p_o q_n}{\sum p_o q_o} * 100\%$$

二、以本期加权指数的指数, 称为派氏(paasche)指数, 公式为:

派氏物价指数:

$$P_{nn} = \frac{\sum p_n q_n}{\sum p_o q_n} * 100\%$$

派氏物量指数:

$$Q_{nn} = \frac{\sum p_n q_n}{\sum p_n q_o} * 100\%$$

三、拉、派氏指数的优缺点比较

四、物价指数

$$V = \frac{\sum p_n q_n}{\sum p_o q_o} * 100\%$$

五、指数体系

价格指数、物量指数与价格指数三者之间关系为:

1. 物值指数=拉氏物价指数*派氏物量指数
2. 物值指数=派氏物价指数*拉氏物量指数

3. 因素分析

第四节 比率加权平均数

一、拉氏比率加权平均数

物价指数：

$$P_{on} = \frac{\sum (p_o q_o) \frac{p_n}{p_o}}{\sum p_o q_o} * 100\%$$

物量指数：

$$Q_{on} = \frac{\sum (p_o q_o) \frac{q_n}{q_o}}{\sum p_o q_o} * 100\%$$

二、派氏加权比率平均指数

物价指数：

$$P_{nn} = \left[\frac{\sum (p_n q_n) \left(\frac{p_n}{p_o}\right)^{-1}}{\sum p_n q_n} \right]^{-1} * 100\%$$

物量指数：

$$Q_{nn} = \left[\frac{\sum (p_n q_n) \left(\frac{q_n}{q_o}\right)^{-1}}{\sum p_n q_n} \right]^{-1} * 100\%$$

三、比率加权平均指数四个公式，与综合加权指数的四个公式从形

式上是相同的，但有独立存在的意义。

第五节 环比指数

一、环比指数的基本概念

采用固定加权，每期指数都以上一期为基期，连环对比，以便在临近各期之间比较。

二、环比指数与固定基期指数的关系

三、连锁

第六节 指数的应用

一、随价调整

二、货币购买力指数=消费价格指数的倒数

第七节 指数编制的一些问题

一、抽选样本问题

二、选择基期问题

三、加权问题

四、指数公式的选择问题

第八章 概率

教学目的

使学生了解随机事件的含义，明确概率的基本概念，掌握概率的基本运算规则。

基本概念

随机现象、事件、概率模式、概率的加法规则、条件概率，事件的独立性。

教学内容

第一节 随机现象和概率的基本概念

一、随机现象与事件

1. 随机现象

经济活动中存在着大量结果无法预先确定的随机现象。概率论是从数量角度来研究随机现象规律性的，是数理统计的基础。

2. 随机事件

随机现象所能出现的不同结果称为随机事件。

二、概率的基本概念

概率是用来描述随机事件发生的可能性的。对于一个事件发生的概率的理解和计算，可以从三个不同的角度来解释：
(1) 古典概率论以这样的解释为基础，即随机现象所能发生的事件是有限的，互不相容的，而且每个基本事件发生的可能性相

等。

(2) 相对频率论认为一个事件出现的概率能通过相同条件下反复试验确定。

(3) 凭主观的概率称为主观概率。

第二节 概率的计算规则

一、加法规则是推算合并事件概率的基本规律。其计算公式为：

$$P(A+B) = P(A) + P(B) - P(AB)$$

当事件互不相容时，计算公式为：

$$P(A+B) = P(A) + P(B)$$

二、条件概率表明在某一事件发生的条件下，另一事件发生的概率。

公式表示为：

$$P(A/B) = \frac{P(AB)}{P(B)}$$

三、事件的独立性。如果一个事件的出现不影响另一事件出现的概率，则称这两个事件为相互独立事件，否则为相依事件。

事件相互独立时：

$$P(A) = P(A/B)$$

$$P(B) = P(B/A)$$

四、概率的乘法规则是计算联合事件出现概率的规则，其公式为：

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2 / A_1) P(A_3 / A_1 A_2) \cdots P(A_n / A_1 A_2 \cdots A_{n-1})$$

如果事件之间相互独立，则公式为：

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n)$$

第九章 概率分布

教学目的和要求

介绍随机变量的基本概念及其数值特征，使学生能够了解不同类型的随机变量的主要分布规律并能够掌握其计算方法及应用。

基本概念

随机变量、概率分布、期望值、方差、二项分布、正态分布。

第一节 随机变量和概率分布

一、 随机变量

表示随机现象各种可能结果的变量。随机变量分为离散型随机变量和连续型随机变量。

二、 概率分布

是与随机变量数值相联系的一系列概率称为概率分布。概率分布可以用表格、图形和公式来表示。

三、 随机变量的期望值

随机变量的期望值是概率最重要的集中趋势量数，在统计分析中有广泛的应用。

离散型随机变量期望值的计算公式为：

$$E(X) = \sum_{i=1}^n x_i p(x_i)$$

随机变量的方差是反映随机变量概率分布的离散趋势，计算公式为：

$$\sigma^2(X) = \sum_{i=1}^n (x_i - \mu)^2 p(x_i)$$

第二节 典型的离散型随机变量的概率分布----二项分布

一、二项试验的条件

- 1.每次试验只能有正反两种互相排斥的结果
- 2.每次试验的成功的概率固定不变
- 3.每次试验相互独立

二、二项变量

是二项试验中成功的次数。

三、二项分布的概率值，计算公式为：

四、二项分布的期望值和标准差

二项分布的期望值 $E(X) = n\pi$

二项分布的方差

$$\sigma^2(X) = n\pi(1-\pi)$$

二项分布成功比率的期望值

$$E\left(\frac{X}{n}\right) = E(p) = \pi$$

比率的方差为

$$\sigma^2(p) = \frac{\pi(1-\pi)}{n}$$

第三节 正态分布

一、正态分布在统计学中的地位----是统计和抽样的理论基础

二、正态分布的特点

1. 正态随机变量X 的取值域为整个X轴,X轴

为正态曲线的渐近线.

2. 正态曲线与X轴围成的面积为1.

3. 正态曲线的分布中心为

4. 正态曲线的形状由随机变量的标准差决

定. 标准差越大, 正态曲线矮胖; 标准差

越小, 正态曲线高瘦

二、正态分布的概率密度函数

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

三、标准正态分布及应用

$$z = \frac{X - \mu}{\sigma}$$

$$P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

四、一般正态分布和标准正态 分布的关系

五、利用正态分布推算二项分布概率

利用正态分布推算二项分布概率的条件是

1、

$$\begin{aligned} n\pi > 5 \\ n(1-\pi) > 5 \end{aligned}$$

2、利用正态分布推算二项分布概率时一般要进行连续性校正。

第十章 抽样和抽样分布

教学目的与要求

介绍有关抽样的一些基本概念，使学生对常用的抽样方法有初步的了解，说明主要样本统计量概率分布的基本规律。

基本概念

随机抽样、抽样误差、基本抽样设计、样本统计量、抽样分布、中心极限定理。

教学内容

第一节 有关抽样的一些基本概念

一、有关总体的概念

抽样调查的目的在于广泛用样本的特征值去推断总体特征值。

1. 作业总体：直接从中抽取样本的总体。
2. 目标总体：抽样调查目标所在的整体。
3. 抽样框架：由调查对象的单位名单或图册组成，以供从中抽取样本的基础资料。

二、有关抽样的概念

1. 随机抽样：根据随机原则从总体中抽取一定数量的调查单位的抽样方法。
2. 参数与统计量

总体参数：反映总体数量特征的指标。如总体平均数，总体比率，

总体标准差等.

样本统计量:反映样本数量特征的指标,如样本平均数,样本比率
样本标准差等。

3. 抽样误差: 随机抽取只限于总体的一部分样本而引起的误差。
4. 非抽样误差: 人为因素造成的样本统计量与总参数之间的差距。
包括偏差与差错。

抽样误差可以通过调节样本量的大小加以控制。

第二节 基本抽样设计

一、简单随机抽样: 从总体中不加任何分组、划类、排队等, 完全随机的抽样调查单位, 每一个总体单位被抽中的概率为 , 具体方法有随机号码法抽签法等。

二、系统抽样: 总体各单位按某一标志顺序排列, 按一定间隔抽取调查单位的抽样方法。如果总体单位不能整除时, 一般用环行排列。

三、分层抽样: 将总体单位按其属性特征分为若干类型或层, 然后在类型或层中抽取样本单位。

四、聚点抽样: 从总体中成群成组地抽取调查单位。

分层抽样要求各层内部相对差距不大, 各层之间差别较大。聚点抽样则要求各组之间差别不大, 各组内部差别较大。

第三节 抽样分布

一、抽样分布

是指一个样本统计量的全面概率分布。

二、样本平均数的抽样分布

1. 样本平均数的期望值等于总体平均数，即

$$E(\bar{x}) = \mu$$

2. 样本平均数的标准误差等于总体标准差除以样本量的算术平方根。即：

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

3. 正态总体的样本平均数的分布也是正态的。
4. 按照中心极限定理，即使是从非正态总体中抽样，只有样本量足够大，样本平均数的分布也接近正态分布。
5. 对 u 差距的标准正态化。求标准正态差用公式：

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

三、比率的抽样分布

1. 样本比率：样本中成功次数与样本量的比率。公式为：
2. 样本比率的期望值等于总体比率。即

$$E(p) = \pi$$

3. 样本比率的标准差公式为：

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

3. 在重复抽样的条件下：样本比率的分布服从二项分布。当 $n\pi$ 和 $n(1-\pi)$ 均大于等于 5 时，逼近正态分布。求标准正态差用公式：

$$Z = \frac{p - \pi}{\sigma_p}$$

四、在有限总体中进行无退还抽样时，如果样本量与总体量相比所占比重相对较大（超过 1/10 时），为了消除先后抽取的基本元素间的影响，要用有限总体校正系数进行校正。样本平均数的

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

标准差为：

样本比率标准差为：

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \sqrt{\frac{N-n}{N-1}}$$

第十一章 参数估计

教学目的和要求

解释点估计的概念及其评选标准,说明区间估计的基本原理,介绍区间估计在实际工作中的应用,帮助学生熟练掌握对总体平均数和比率的区间估计的计算方法以及样本量的确定方法。

基本概念

点估计,点估计的评选标准,对总体平均数的区间估计,t分布,对总体比率的区间估计,样本量的确定。

教学内容

第一节 点估计

一. 点估计

用一个单一估计值对总体参数进行评估。

二. 点估计的评选标准

1. 无偏性: 如果一个估计量的期望值等于该统计量所要估计的总体参数,则称其为无偏估计量。
2. 一致性: 如果一个估计量的取值随着样本的增加趋近于所要估计的总体参数的真值,则称其具有一致性。
3. 有效性: 离散程度最小的估计量称为有效估计量。
4. 充分性: 如果一个估计量已包含了样本提供的全部的关于所估计的总体参数的信息,则称其为充分估计量。

三. 按照上述标准,我们分别用样本平均数,样本比率,样本标

准差作为总体平均数，总体比率和总体标准差的估计量。

其中，作为总体标准的无偏估计量的样本标准差的计算公式为：

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

计算样本比率的标准误差时，由于总体比率不知，用样本比率标准差作为估计量，公式为：

$$S_p = \sqrt{\frac{p(1-p)}{n}}$$

第二节 区间估计

一. 区间估计

用一个在一定概率保证下的置信区间对总体参数进行估计。

二. 对总体平均数和总体比率的区间估计

1. 关于区间估计的推理：总体水平数或总体比率的真值在一定概率的保证下包含在以样本平均数或比率为中心的某一区间内，即用一定可靠程度保证样本平均数或比率与总体平均数或比率的离差不超过某一限定范围，这个范围即置信区间，

2. 总体标准差明确时的区间估计

对总体平均数估计的置信区间的公式为：

$$\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}}$$

3. 总体标准差不明确时的区间估计

- (1) 大样本：当总体标准差不明确时，用样本标准差来取代。如果样本量足够大，区间仍可用正态分布进行推算。
- (2) 小样本：如果样本量较小，则应用 t 分布进行推断，公式为：

$$\mu = \bar{x} \pm t \frac{\sigma}{\sqrt{n}}$$

t 分布与正态有共同点，它也是一种对称的分布，且 t 的取值范围也是；但分布比正态分布更分散，其陡缓程度取决于自由度 ($v=n-1$)

4. 对总体比率估计的置信区间公式为：

$$\pi = P \pm Z S_p = p \pm Z \sqrt{\frac{p(1-p)}{n}}$$

第三节 样本量的确定

一. 求估计平均数所需的样本量

公式为：

$$n = \frac{Z^2 \sigma^2}{e^2}$$

其中总体标准差的大小可利用过去的资料确定，也可用抽样所得到的样本标准差代替，还可以用全距的 $1/6$ 作为总体标准差的估计量。

二. 求估计比率所需的样本量

公式为:

$$n = \frac{Z^2 p(1-p)}{e^2}$$

其中 p 可根据过去的资料来定, 如无资料一般可用保守的假定 0.5 来作为估计值。

第十二章 假设检验

教学目的和要求

解释假设检验的基本概念及原理，说明假设检验的程序及应用，使学生掌握对总体平均数及比率进行假设检验的方法及含义，了解 II 型误差的计算方法及假设检验的效率。

基本概念

拟定假设，对立假设，显著系数，单尾检验，双尾检验，临界值，检验统计量，I, II 型误差，功效曲线。

教学内容

第一节 假设检验的基本概念和程序

- 一. 有关假设检验的概念
 1. 假设检验：在两种互相对立的行动之间，通过抽样试验来进行选择的统计分析方法。
 2. 拟订假设：作为被检验主体的假设，无充分根据不能轻易推翻。
 3. 对立假设：与拟订假设对立的假设。
 4. α 型假设：拒绝一个正确的拟订假设，其概率用来表示。此概率称为显著性水平。
 5. β 型误差：接受一个错误的拟订假设，其概率用表示。
 6. 临界值：确定抽样结果是否显示与拟订假设有明显差异的界

限。

二. 假设检验的程序

1. 作出拟订假设和对立假设；
2. 规定显著性水平；
3. 确定检验用的统计量；
4. 根据显著性水平计算临界值；
5. 建立决策规则；
6. 用抽样结果与决策规则比较；
7. 作出决策。

三. 假设检验的类型

1. 单尾检验：检验是单方向的，即只注意假设值是否偏高（或偏低）。
2. 双尾检验：检验是双方向的，即检验的目的在于观察样本统计量与假设参数有无差距（无论正负）。

第二节 对总体平均数和比率的假设检验

一、对总体平均数的假设检验

利用样本平均数对假设总体平均数进行检验。

1. 若总体是正态或接近正态的，总体标准差已知，或者总体标准差未知而用样本标准差替代，但样本量足够大，使用正态分布进行推断。
2. 若总体是正态的，但总体标准差未知，而且样本量较小，使用

t分布进行推断。

二、对总体比率的假设检验

利用样本比率对假设总体比率进行检验，如果样本量较大，使用正态分布进行推断。

第三节 II型误差概率的计算

一、II型误差的大小是通过计算得到的，它取决于参数真值与假设值的差距。

二、功效曲线

1. 检验的功效：拒绝一个错误的拟定假设的概率。此概率等于 $1-\beta$ 。
2. 功效曲线：将实际参数取不同值时，检验的功效用图形显示出来，所得曲线即功效曲线。

三、样本量与两种误差的关系

1. 当样本量固定时，减少一种误差发生的概率会导致另一种误差发生的概率增大。
2. 当样本量增加时，可同时减少两种误差发生的概率。

第十三章 简单回归分析和相关分析

教学目的与要求

明确回归与相关的涵义。掌握根据观察值拟合样本简单回归方程，并由样本回归方程对总体进行推断的原理和方法，能够解析其中的量数的意义。掌握可决系数和相关系数和计算方法，并会对其作出解析。

基本概念

样本回归方程；最小二乘法；正规方程；回归系数；可决系数；相关系数。

教学内容

第一节 回归分析

一、回归与相关分析的基本概念

1. 函数型关系和统计型关系；
2. 回归分析和相关分析的不同作用；
3. 回归模型分类：

按自变量个数分为简单回归和多元（复）回归；

按因变量对自变量的协变方式分为线性和非线性，线性中又分为正回归与负回归。

二、简单线性回归方程的拟合

1. 总体回归模式 $Y=A+Bx+e$

假设条件：

- (1) Y值随机出现，

(2) x 给定 Y 可能取值服从正态分布,

(3) 对于不同的 X 给定值, Y 可能取值的分布方差恒定, 有唯一的方差

总体回归方程:

2. 线形回归方程的拟合

样本回归方程: $Y_c = a + bx$

最小二乘法确保 $\Sigma(Y - Y_c)^2$ 为最小, 逐有正规方程:

$$\Sigma Y = na + b \Sigma x$$

回归系数:

$$\Sigma XY = a \Sigma X + b \Sigma X^2$$

斜率:

$$b = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{x}^2}$$

截距:

$$a = \bar{Y} - b\bar{X}$$

3. 在样本回归范围内对 Y 值的估计: 由 X 求 Y .

估计的标准误差:

$$S_{Y.X} = \sqrt{\frac{\sum (Y - Y_c)^2}{n-2}} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

三、有关回归分析的统计推断

1、对因变量的估计和推算

对总体条件平均数的区间估计

$$\mu_{Y.X} = Y_c \pm t_{n-2} S_{Y.X} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x - \bar{x})^2}}$$

对总体单个值的推算

$$Y_i = Y_c \pm t_{n-2} S_{Y.X} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x - \bar{x})^2}}$$

2、对总体斜率的假设检验

H₀: β = 0 (总体回归不显著)

H₁: β ≠ 0 (总体回归显著)

检验统计量 $t_{n-2} = (b - 0) / s_b$ 其中:

$$s_b = \frac{S_{Y.X}}{\sqrt{\sum X^2 - nx^2}}$$

3、对总体斜率的区间估计

$$B = b \pm t_{n-2} S_b$$

第二节 相关分析

一、可决系数的含义和计算方法

1、总变差、明确变差和不明确变差的概念及相互关系。

2、可决系数和定义公式和简算公式

$$r^2 = \frac{\sum(Y_c - \bar{Y})^2}{\sum(Y - \bar{Y})^2} = \frac{a\sum Y - b\sum XY - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2}$$

二、相关分析

1.相关系数和计算

$$r = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{(\sum X^2 - n\bar{X}^2)(\sum Y^2 - n\bar{Y}^2)}}$$

2.Y的取值范围 r 和 r²的关系

3.关于总体相关系数和假设检验

H₀: P = 0 (总体相关关系不显著)

H₀: P ≠ 0 (总体相关关系显著)

检验统计量

$$t_{n-2} = \frac{r - 0}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

4.P的检验与B检验的关系

三、几个应注意的问题

1、样本回归和相关与总体回归和相关两者间的联系。

2、注意区别因果关系和相伴关系。

3、定量分析与定性分析相结合。

4、回归方程应用的局限性。

5、关于 $r=0$ 的正确解释。

6、b 与 r 的内在联系。

第十四章 多元回归分析和相关分析

教学目的和要求

明确多元回归分析的原理、方法，能够解释复回归系数的含义，明确多元相关分析的原理和方法，能够解释复合可决系数、复相关系数和偏相关系数的含义。重点是掌握含有二个自变量的复回归分析。了解统计软件包在多元回归和相关分析中的作用。

基本概念

偏回归系数，偏相关系数，复可决系数。

教学内容

第一节 含有两个自变量的回归分析

一、含有两个自变量的回归的基础知识

因变量散布的立体空间和总体回归模式

$$Y = A + B_1X_1 + B_2X_2 + e$$

总体回归方程

$$\mu_{Y.12} = A + B_1X_1 + B_2X_2$$

样本回归方程

$$Y_c = a + b_1X_1 + b_2X_2$$

二、偏回归系数和估计标准差的计算与解释

正规方程

$$\sum Y = na + b_1 \sum X_1 + b_2 \sum X_2$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

偏回归系数的含义。

估计标准误差

$$S_{r.12} = \sqrt{\frac{\sum (Y - Y_c)^2}{n-3}} = \sqrt{\frac{\sum Y^2 - a \sum Y - b_1 \sum X_1 Y - b_2 \sum X_2 Y}{n-3}}$$

三、有关含有两自变量回归的统计推断

1、对因变量的估计和推算

$$\mu_{r.12} = Y_c \pm t_{n-3} S_{r.12} \sqrt{\frac{1}{n} + \frac{C}{D}}$$

$$Y_i = Y_c \pm t_{n-3} S_{r.12} \sqrt{1 + \frac{1}{n} + \frac{C}{D}}$$

式中

$$C = (X_1 - \bar{X}_1)^2 (\sum X_2^2 - n \bar{X}_2^2) + (X_2 - \bar{X}_2)^2 (\sum X_1^2 - n \bar{X}_1^2) - 2(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)(\sum X_1 X_2 - n \bar{X}_1 \bar{X}_2)$$

$$D = (\sum X_1^2 - n \bar{X}_1^2)(\sum X_2^2 - n \bar{X}_2^2) - (\sum X_1 X_2 - n \bar{X}_1 \bar{X}_2)^2$$

2、对总体偏回归系数的估计和检验，其中：

$$S_{b_2} = S_{r.12} \sqrt{\frac{\sum X_2^2 - n \bar{X}_2^2}{D}}$$

$$S_{b_1} = S_{r.12} \sqrt{\frac{\sum X_1^2 - n \bar{X}_1^2}{D}}$$

第二节 多元回归和相关分析

一、多元回归总体模式和样本回归方程

$$Y = A + B_1X_1 + B_2X_2 + \cdots + B_mX_m + e$$

$$Y_c = a + b_1X_1 + b_2X_2 + \cdots + b_mX_m$$

(m+1) 个正规方程

$$\sum Y = na + b_1 \sum X_1 + b_2 \sum X_2 + \cdots + b_m \sum X_m$$

$$\sum X_1Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1X_2 + \cdots + b_m \sum X_1X_m$$

$$\sum X_2Y = a \sum X_2 + b_1 \sum X_1X_2 + b_2 \sum X_2^2 + \cdots + b_m \sum X_mX_2$$

... ..

$$\sum X_mY = a \sum X_m + b_1 \sum X_1X_m + b_2 \sum X_2X_m + \cdots + b_m \sum X_m^2$$

二、多元可决系数和相关系数

$$r_{Y.12 \cdots m}^2 = \frac{\sum (Y_c - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$$

$$r_{Y.12 \cdots m} = \sqrt{\frac{\sum (Y_c - \bar{Y})^2}{\sum (Y - \bar{Y})^2}}$$

三、偏可决系数和偏相关系数

引入一个新的自变量，使得原有不明确变异有所减少。这一减少占原有不明确变异的比率，即为新引入自变量的偏可决系数。

$$r_{Y1}^2 = \frac{r_{r,1}^2 - O}{100\%} = r^2$$

第一轮：由零个自变量到一个自变量 X_1

第二轮：继 X_1 之后又引入 X_2

$$r_{Y2,1}^2 = \frac{r_{Y,12}^2 - r_{Y1}^2}{1 - r_{Y1}^2}$$

第 K 轮：继 X_{K-1} 之后又引入 X_K

$$r_{YK,12,\dots,(K-1)}^2 = \frac{r_{Y,12,\dots,K}^2 - r_{Y,12,\dots,(K-1)}^2}{1 - r_{Y,12,\dots,(K-1)}^2}$$

三、逐步回归法

1. 使用统计软件包进行逐步回归

本阶段教学可视条件向学生介绍一些统计软件包的用法，或者向学生展示分析结果的打印输出并作解释。

可能被引入教学的软件包有 Minitab, Spss 等。

2. 方差分析：对总体回归模式的 F 检验方法与意义。
3. 明确计算机只是多元回归和相关分析的工具，而变量的选择，模式的建立在定性分析基础上，经验也是重要的。