

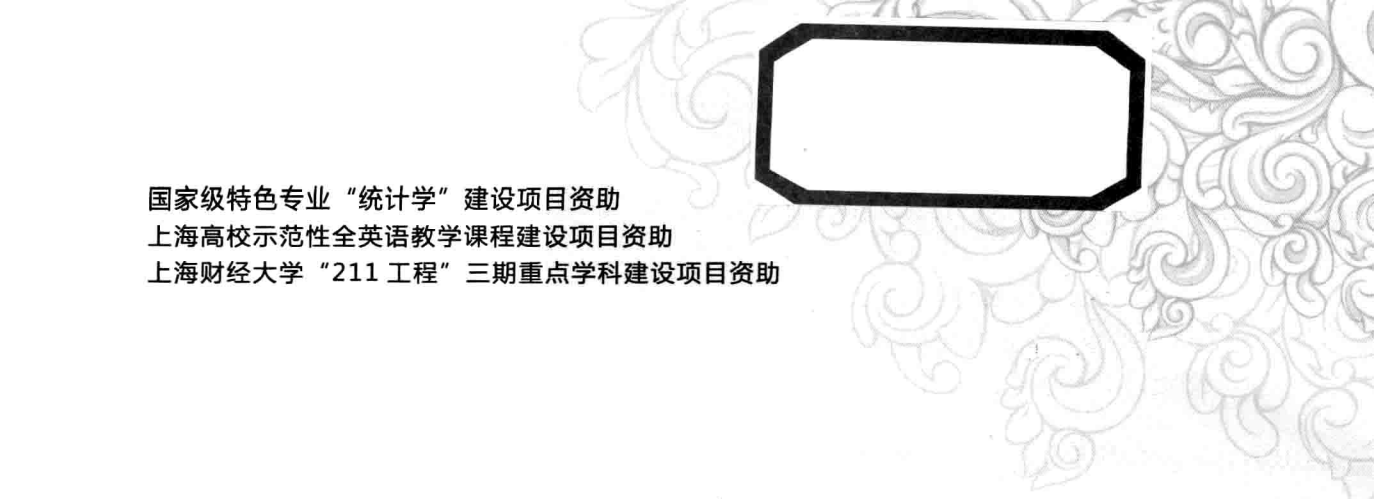
国家级特色专业“统计学”建设项目

Business Statistics *Practical Cases Analysis*

商务统计 双语案例精析

杨楠◎编著

上海财经大学出版社



国家级特色专业“统计学”建设项目资助
上海高校示范性全英语教学课程建设项目资助
上海财经大学“211工程”三期重点学科建设项目资助

Business Statistics *Practical Cases Analysis*

商务统计 双语案例精析

杨楠◎编著

图书在版编目(CIP)数据

商务统计双语案例精析/杨楠编著. —上海:上海财经大学出版社, 2012. 2

ISBN 978-7-5642-1253-7/F · 1253

I. ①商… II. ①杨… III. ①商业统计—案例—分析—
双语教学—教材 IV. ①F712.3

中国版本图书馆 CIP 数据核字(2011)第 273082 号

☐ 责任编辑 李宇彤
☐ 封面设计 张克瑶
☐ 责任校对 王从远 林佳依

SHANGWUTONGJISHUANGYUANLIJINGXI

商务统计双语案例精析

杨楠 编著

上海财经大学出版社出版发行

(上海市武东路 321 号乙 邮编 200434)

网 址: <http://www.sufep.com>

电子邮箱: webmaster@sufep.com

全国新华书店经销

上海华教印务有限公司印刷装订

2012 年 2 月第 1 版 2012 年 2 月第 1 次印刷

787 mm×1092 mm 1/16 12.75 印张(插页:1) 279 千字

印数:0001—3000 定价 32.00 元

发现统计的商务价值

在市场咨询与民意研究行业工作了十多年,这使得我有机会与成百上千的外企、民企、国企、政府机关、民间机构等各种类型的客户进行广泛的接触。在这过程中,我深深感到,统计之数字之美,只有在实践中得到应用,其价值之美才能得到充分的展现,否则就是一堆枯燥的数据。我也曾在一些高校中多次进行过“市场研究在商业中的应用”之类的讲座,此中我也能感受到,在与学生的演讲中,讲概念时同学们的眼神多有游离,而一旦进入案例,尤其说到一些关键数据如何对客户的商业决策产生支撑作用,并获得商业的巨大成功时,同学们的眼神多是热烈而有互动。我也曾应杨楠博士之邀在上海财经大学进行相关讲座,她曾代表学生就讲题提出要求:多讲案例。所有这些让我对统计有了一个基本的认识,那就是,统计的价值之美在于应用,而应用的范例最为直观而有趣。

作为一个研究咨询行业中天天与统计数字打交道的从业者,即使在我自己眼中,统计都曾经是一门枯燥艰涩的学科。看见统计数字头晕的情形想来是很多人都有过的体验。因此,如何让统计更加有趣而易懂,易懂而有用,是很多与统计分析有关的从业人士一直想解决的问题。

有趣而易懂,易懂而有用,当看到这本《商务统计双语案例精析》时,我意识到杨楠博士在多年的教学与研究工作中,已然成为这一逻辑的践行者。我仔细拜读了这本书,全书以商务统计的概念、案例、统计操作为主线,围绕统计中常用的概念,从描述性分析到整合性分析,从参数统计到非参数统计,从数值统计到假设检验,进行系统而图文并茂的阐述。尤其是其中的案例,涉及金融、制造、医疗、大宗商品等众多行业,从商务的角度展现了统计在各行各业商业价值。并且,其双语编排,使得读者在阅读的时候能够英汉对照,更有助

于读者有效地汲取西方商务统计方面的知识体系。

这本书对于读者而言,个人觉得至少有三方面的作用:第一,读完此书,读者应该能够快速掌握商务统计中的基本概念,因为书中即使是枯燥的概念,也做得图文并茂;第二,读完此书,读者甚至能够依葫芦画瓢地将统计概念应用到实践中去,因为书中的案例,是商业世界中活生生的案例;第三,读者还可以把它作为工具书,因为在每一章,都有一个如何应用 Excel 和 SAS 等软件进行统计实现的操作小技巧,这一点在以往的统计类书籍中不多见。

实现了这三方面的作用,其实就是发现了统计的价值。通常统计类的书很难体现商务价值,商务类的书着眼于统计的不多。这本《商务统计双语案例精析》,则将两者进行了有机的结合,让我们一起开启商务统计价值的发现之旅!

零点研究咨询集团副总裁

周林古

2012年1月

P 前言

REFACE

全英语及双语教学是高等教育适应经济全球化挑战、培养高级经济管理类人才的重要途径。近年来,师资的国际化、生源的国际化 and 教材的国际化已经使统计学这门经济管理专业人才培养中的核心数量方法课程面临前所未有的挑战与发展机遇。如何更为有效地汲取西方经济管理统计的积淀成果,服务于国内经济建设与对外经济交往,已经成为我们迫切需要解决的问题。

培养学生准确观察数据、进行科学分析和正确决策的能力往往很难完全通过统计学课程教学来实现。理解统计思想、掌握统计方法的关键在于结合现实问题进行思考和分析。对此,案例教学不失为事半功倍的学习途径。具有生动性、启发性和可操作性的案例能够作为统计理论课程教学内容的有益补充,而小组合作讨论、运用软件实施、撰写结果分析报告的过程可以使学生留下深刻的印象,促进其快速掌握通过统计分析进行判断与决策的思路与方法,以更好地解决实际问题。

本案例集撰写的初衷来自于作者在过去多年中为经济、金融、会计、工商管理等本科中外合作专业学生,以及本科统计学双学位学生、MBA 学员讲授双语统计学课程的案例教学环节。作为国外原版教材的有益补充及案例教学环节的重要参考,作者深刻感受到了编撰一本视野开阔、资料丰富、通俗易懂的双语统计学案例集的必要性。对于采用中英文对照的双语形式编著,作者经过了较长时间的考虑,这种方式既可以准确呈现英文原有含义和表达句式、国际标准统计软件输出结果,也可以方便各个层面的读者同时通过对应中英文来精确理解各部分内容。

本案例集所呈现的案例,有的是作者从国外畅销的经济管理统计学、商务统计学经典教材中精选出来的典型案例,也有的是作者基于美国劳工部、世界黄金协会、《福布斯》杂志等网站上的热点问题与数据自行编撰的与现实生活密切相连的案例,基本涵盖了基础统计学的各部分内容,并具有实用性、权威性和前沿性的显著特点。这些案例多为开放式,即使在原版教材中也未给出

完整的解答分析。作者在教学过程中,经过案例教学的探索实践,总结出科学合理的解答过程,经过精心编排,呈现在这本案例集中。

本案例集中每一章内容的展现形式如下:首先是该章所对应统计方法的概括性描述,接着是一个具体案例问题的背景描述和数据呈现,之后是对该案例的分析框架、计算过程与结果以及对结果的延伸讨论和软件应用实施过程解释,最后是关键公式与注解。通过这样一种结构设计,希望读者能从具体的案例问题和现实数据出发来考虑如何分析,而不是从理论推导出发,因而更加生动、具体、真实,也更加符合从事商务工作的人员在现实工作中使用统计学方法解决现实问题的逻辑。

本案例集可以作为经济管理类非统计学专业的本科生学习统计学课程的辅助教材,也可以作为 MBA、MPA 学生学习统计学课程的辅助教材,还可以作为商务领域的广大实际工作者学习统计方法应用的参考书。

本案例集的编著是多位成员协同努力的结果,其中张希参与了最初的策划与设计,以及探索性数据分析之数值法、区间估计、方差分析、多元回归部分的工作,胡瑶参与了假设检验、一元线性回归部分的工作,龚逸君参与了非参数统计、时间序列分析部分的工作,蔡志文参与了探索性数据分析之图示法部分的软件输出及图表设计的工作,在此,对团体成员的辛勤工作表示深深的谢意!由于作者的水平有限,本书难免会存在不妥之处,恳请同行专家和广大读者给予我们宝贵的批评和建议。

杨楠

2012年1月

于上海财经大学

目 录



第一章	探索性数据分析之图示法	1
1.1	定量数据的图示分析	2
1.2	定性数据的图示分析	4
案例 1	年轻人暑期工作情况分析	5
案例 2	福布斯全美 400 富豪榜	14
第二章	探索性数据分析之数值法	27
2.1	集中趋势的测定	28
2.2	离散趋势的测定	28
案例	全美护士工作满意度调查研究	29
第三章	区间估计	39
案例	关于改进伯克投资服务公司简讯内容的调查研究	40
第四章	假设检验:单个总体情形	49
案例	生产规格检验分析	50
第五章	假设检验:两总体情形	59
案例	佛罗里达的小时工资差异	60
第六章	方差分析	68
案例	医疗中心患者流量调查分析	69
第七章	一元线性回归	77
案例	校友捐赠额影响因素分析	78

第八章	多元回归	94
	案例 1 棉织物销量影响因素分析	95
	案例 2 黄金价格主要影响因素分析	102
第九章	非参数方法	116
	案例 购买产品时的选择偏好	117
第十章	时间序列分析	137
	案例 预测食品和饮料的销售额	138
	参考文献	168
附录		169
	A. 1 福布斯全美 400 富豪榜数据集	169
	A. 2 金价数据	180
	A. 3 标准正态分布表	187
	A. 4 t 分布临界值表	188
	A. 5 F 分布临界值表	191
	A. 6 Wilcoxon 临界值下界表	195

Contents



Chapter 1	Displaying and Exploring Data: Graphical Presentations	1
1.1	Graphs to describe quantitative data	2
1.2	Graphs to describe qualitative data	4
Case 1	School's Out	5
Case 2	Forbes 400 Richest Americans	14
Chapter 2	Displaying and Exploring Data: Numerical Measures	27
2.1	Numerical Measures of Location	28
2.2	Numerical Measures of Dispersion	28
Case	National Health Care Association	29
Chapter 3	Interval Estimation	39
Case	Bock Investment Services	40
Chapter 4	Hypothesis Test: One Population	49
Case	Manufacturing Processes Control	50
Chapter 5	Hypothesis Test: Two Populations	59
Case	Hourly Wage Rate in Florida	60
Chapter 6	Analysis of Variance	68
Case	Bell Grove Medical Center	69
Chapter 7	Simple Linear Regression	77
Case	Alumni Donation	78

Chapter 8	Multiple Regression	94
	Case 1 Sales of Cotton Fabric	95
	Case 2 Main Influential Factors of Gold Price	102
Chapter 9	Nonparametric Methods	116
	Case Product Preference	117
Chapter 10	Time Series and Forecasting	137
	Case Forecasting Food and Beverage Sales	138
References		168
Appendix		169
	A. 1 Data Set-Forbes 400 Richest Americans	169
	A. 2 Data Set-Gold Price	180
	A. 3 Standard Normal Distribution	187
	A. 4 t Distribution Critical Value	188
	A. 5 F Distribution Critical Value	191
	A. 6 Wilcoxon Critical Value	195

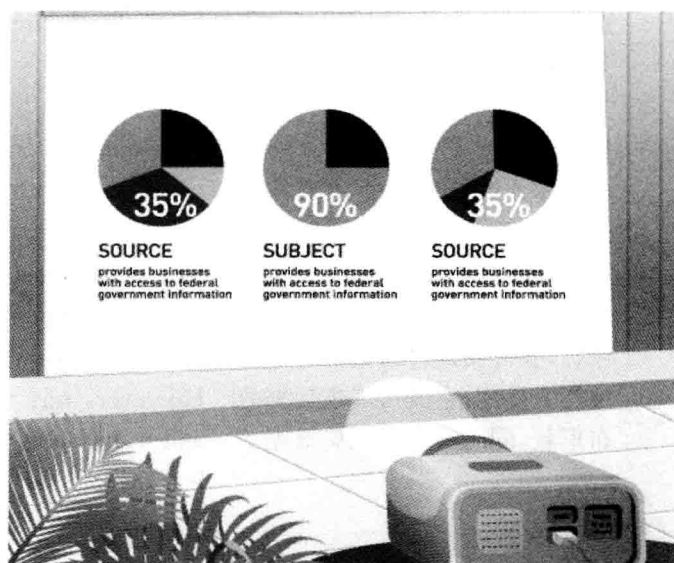
Chapter 1

Displaying and Exploring Data: Graphical Presentations

第一章 探索性数据分析之图示法

There are many ways to present data in pictures. Sometimes a unique picture is used to fit a particular situation. The purpose of a plot, graph, or picture of data is to give you a visual summary that is more informative than simply looking at a collection of numbers. Done well, a picture can convey a message quickly that would take you longer to find if you had to study the data on your own. Done poorly, a picture can mislead all but the most observant of readers. In this chapter, graphical methods of descriptive statistics are introduced. In applied cases, you will learn how to use them to summarize data.

用图形来展现数据的方法有多种,特定情况下要使用特定的方法。借助图形,可以了解数据集的大致特征,形成清晰的视觉化的总结。如果使用得当,图形可以快速准确地传递信息;相反,若使用不当,图形也可以具有相当的欺骗性,误导除行家以外的所有读者。本章将介绍常用的图示法,并通过案例让读者体会如何用图示来展现概括数据集的基本特征。



1.1 Graphs to describe quantitative data

1.1 定量数据的图示分析

直方图

直方图是在每个分组区间上绘制一个长条形而产生的图形,它可以用来描述已表示成频数或频率的数据。从直方图可以看出数据分布的疏密。

Histogram

A **histogram** is a graph that consists of vertical bars constructed on a horizontal line that is marked off with intervals for the variable being displayed. The intervals correspond to those in a frequency distribution table. The height of each bar is proportional to the number of observations in that interval. The number of observations can be displayed above the bars.

绘制直方图

在绘制直方图时,一般用横轴(X轴)表示所研究的变量,用纵轴(Y轴)表示每组观测数据的频数或频率。纵坐标也可用百分比,即把频数除以样本量,这样与用频数得到的形状一样,只是量纲不同。

Creating a Histogram

To create a histogram, simply divide the range of the data into intervals, simply count how many fall into each part of the range. Draw a bar whose height is equal to the count for each part of the range. Or, equivalently, make the height equal to the proportion of the total count that falls in that interval.

茎叶图

探索性数据分析主要使用简单的算术或易画的图形来快速总结数据集特征,很多方法最初使用笔和纸就能完成。时至今日,茎叶图这种古老的方法仍被广为使用,它既可展现数据分布形状,同时又显示原始数据。

Stem-and-leaf Display

The techniques of **exploratory data analysis** consist of simple arithmetic and easy-to-draw graphs that can be used to summarize data quickly. Many of these techniques were initially developed to provide a simple overview of a data set using pencil and paper. However, one technique — referred to as the **stem-and-leaf display** — is still widely used. A stem-and-leaf display can be used to show both the rank order and shape of a data set simultaneously.

绘制茎叶图

茎叶图像一片带有茎的叶

Creating a Stem-and-leaf Display

Only two steps are needed to create a stem-and-

子。茎为较大位数的数字,叶为较小位数的数字。

步骤 1:把每个数据分解成茎、叶两部分,在一个垂直竖线的左侧将观测值的茎依次列出,在该竖线右侧将对应的叶依次列出。

步骤 2:把所有数据由小到大批序,每个都按步骤 1 分成茎和叶,然后按茎和叶的大小排列成图。

leaf display.

Step 1: First arrange the leading digits of each data value to the left of a vertical line. To the right of the vertical line, record the last digit for each data value as we pass through the observations in the order they were recorded.

Step 2: Sorting the digits on each line into rank order. The second step is to attach a leaf to represent each data point. The next digit in the number is used as the leaf, and the remaining digits are simply dropped.

盒形图

盒形图又称箱形图、箱线图、盒子图。它使用一个矩形盒子形状来展示数据分布中五个最关键的点的位置。使用盒形图便于比较多个数据集的分布情况。

Box-Plot

A visually appealing and useful way to present a **five-number summary**^① is through a **box-plot**, sometimes called a box and whisker plot. This simple picture also allows easy comparison of the center and spread of data collected for two or more groups.

绘制盒形图

步骤 1:画一个封闭矩形,左右两边分别对应于数据集的上、下四分位数(点)。

步骤 2:在盒子的中间、数据集的中位数处,画一条竖线。

步骤 3:盒子的宽度即是数据集的四分位距。

步骤 4:在盒子左右两边,分别向外画出延伸线,长度为 1.5 倍的四分位距。如果最小或最大值在此范围内,则线就画到最小、最大值为止。

Creating a Box-Plot^②

Step 1: Draw a box with ends at the lower and upper quartiles.

Step 2: Draw a line in the box at the median.

Step 3: Compute the width of the box and this is defined to be the interquartile range.

Step 4: Draw whiskers at each end with length equal to 1.5 times the interquartile range; if the minimum or maximum occurs before the full length is used, stop there.

① 五点法(Five-number Summary)中的统计量指的是一组数据的最小值、第一四分位数、中位数、第三四分位数和最大值,它们可以提供该组数据的重要位置及分散程度的信息。

② 不同的软件及不同选项所生成的盒形图两头线长的定义不尽相同,但封闭盒子长度的定义基本一样。

步骤 5: 用星号标出在延伸线范围外的数值。

Step 5: Use an asterisk to indicate any additional data points beyond the range covered by the box and whiskers.

盒形图的释义

盒形图将一个数据集分为四部分,最小的 $1/4$ 位于盒子左边以左,接下来的 $1/4$ 位于盒子内部左半边,第三大的 $1/4$ 位于盒子内部右半边,最大的 $1/4$ 位于盒子右边以右。异常值也很容易识别。

Interpreting Box-Plot

Notice that box-plot essentially divides the data into fourths. The lowest fourth of the data values is contained in the range of values below the start of the box, the next fourth is contained in the first part of the box (between the lower quartile and the median), the next fourth is in the upper part of the box, and the final fourth is between the box and the upper end of the picture. Outliers are also easily identified.

散点图

散点图是展示两个数值型变量间关系的直观图形,其中一个变量在横轴方向,另一个变量在纵轴方向。散点图比单纯的线性图复杂,它提供了更多关于两变量之间关系的信息。

Scatter Diagram

Scatter diagram is a graphical means of showing the relationship between two quantitative variables. One variable is shown on the horizontal axis and the other variable is shown on the vertical axis. Although a scatter diagram can be more difficult to read than a line graph, it displays more information.

1.2 Graphs to describe qualitative data

1.2 定性数据的图示分析

饼图

饼图是较常用的一种展示定性数据频率分布情况的图示。它以整个圆的 360° 代表全部数据的总和,按照各类型组所占的百分比(频率),把一个“饼”切割为各个扇形。

Pie Chart

The **pie chart** is a commonly used graphical device for presenting relative frequency and percent frequency distributions for qualitative data. To draw a pie chart, draw a circle and use the relative frequencies to subdivide the circle into sectors, or parts, which correspond to the relative frequency for each class.

条形图

条形图也是一种展示定性

Bar Graph

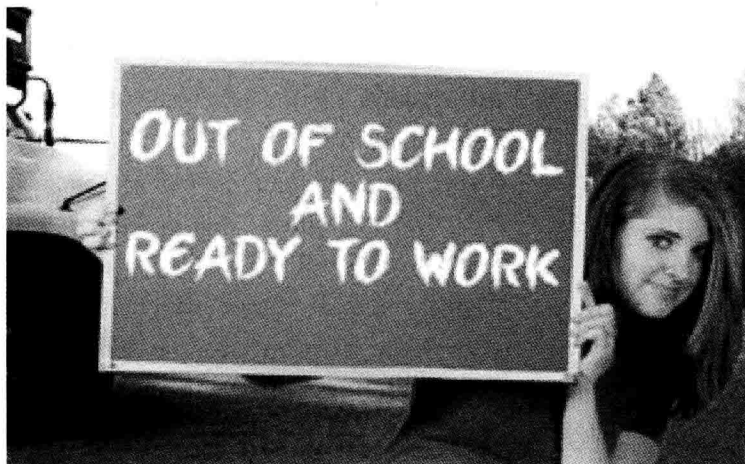
Bar graph also shows percentages or frequencies in

数据频率分布的图形。在条形图中,每一分类组被表示成一个条形,条的长度代表了这个组中所含数据的频数或频率。

various categories, and it can represent two or three categorical variables simultaneously. One categorical variable is used to label the horizontal axis. Within each of the categories along that axis, a bar is drawn to represent each category of the second variable. Frequencies or percentages are shown on the vertical axis.

Case 1 School's Out

案例 1 年轻人暑期工作情况分析



案例介绍

“暑假里,离开学校,
直到秋天,
我们也许就不再回来了。”

当“离开学校”成为爱丽斯·库珀的第一要务时,他说:“一生中最棒的三分钟是何时?也许在学校的最后一天的最后三分钟就是这样的时间,你坐在那里就像保险丝快要慢慢熔断。”



Introduction

Source: www.bls.gov/spotlight

*“Out for summer
Out till fall
We might not be back at all”*

After “School’s Out” became Alice Cooper’s first hit, he said “What’s the greatest three minutes of your life? One is the last three minutes of the last day of school, when you’re sitting there and it’s like a slow fuse burning.”

对上百万的 16~24 岁之间的年轻人来说,很多都像图片中拿着标语牌的女孩一样,已经准备好去工作了,也有很多还在寻找工作。还有一些,也许认为很难找到合适的工作,于是回到学校或是休息一段。

在这个案例中,美国劳工局希望分析近年来年轻人暑期工作的相关情况,如:他们参加暑期学校的比例,他们工作的行业,失业情况,发生人身伤害的情况,等等。相关数据被搜集整理在表 1.1, 1.2, 1.3 及 1.4 中。其中包括从 1986~2010 年的每年暑期学校参加率、2007~2010 年每年 7 月的失业率、2010 年 7 月调查得到的 16~24 岁青年人的就业数据,以及 2003~2009 年暑期工作中发生致命伤害事件的行业分布数据。美国劳工局希望对已有数据进行一些初步分析。

So what do millions of youth age 16 - 24 do once that "slow fuse" is no longer burning and school really is out for the summer? Some, like the young lady holding the placard in the photo, are "ready to work" and even have jobs waiting for them. Others may be ready to work, but will have to find a job. Still others, perhaps believing there is no job out there for them, return to school or just relax.

In this case U. S. Bureau of Labor Statistics takes a look at the number of young people over the past years, the proportion of youth who are working, the industries in which they work, the unemployment rate among youth, and the proportion who are in school in the summer. Four data sets obtained are shown in Table 1.1, 1.2, 1.3 and 1.4. They report yearly percentage of summer school attendance from 1986 to 2010, unemployment rate in July from 2007 to 2010, number of employed persons aged 16 - 24 in July 2010, and distribution of fatal occupational injuries during the summer months by selected occupation groups and age groups from 2003 to 2009. U. S. Bureau of Labor Statistics decided to do some preliminary analysis of the data already collected.

Table 1.1 Percentage of Summer School Attendance (暑期学校的参加率)

July, of year	16 to 24 years (%)	20 to 24 years (%)	July, of year	16 to 24 years (%)	20 to 24 years (%)
1986	9.8	8.4	1993	12.6	11.6
1987	10.6	8.6	1994	16.3	13.8
1988	11.7	10.7	1995	18.0	14.8
1989	12.3	10.5	1996	18.5	15.5
1990	13.3	11.4	1997	19.3	15.3
1991	13.3	12.0	1998	20.4	16.4
1992	12.6	11.9	1999	21.5	16.8