

应用统计

Applied Statistics

主 编 ◎ 胡卫中

副主编 ◎ 吴步昶 孙 钰 袁 静



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

014035213



高等院校经济管理类规划教材

C8-43
223

应用统计

Applied Statistics

主 编 ◎ 胡卫中

副主编 ◎ 吴步昶 孙 钰 袁 静



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社



北航

C1715425

C8-43
223

图书在版编目 (CIP) 数据

应用统计 / 胡卫中主编. —杭州: 浙江大学出版社, 2014. 1

ISBN 978-7-308-12762-2

I. ①应… II. ①胡… III. ①应用统计学
IV. ①C8

中国版本图书馆 CIP 数据核字 (2014) 第 002408 号

应用统计

主 编 胡卫中

副主编 吴步昶 孙 钰 袁 静

责任编辑 朱 玲

封面设计 续设计

出版发行 浙江大学出版社

(杭州市天目山路 148 号 邮政编码 310007)

(网址: <http://www.zjupress.com>)

排 版 杭州中大图文设计有限公司

印 刷 浙江云广印业有限公司

开 本 787mm×1092mm 1/16

印 张 13

字 数 325 千

版 印 次 2014 年 1 月第 1 版 2014 年 1 月第 1 次印刷

书 号 ISBN 978-7-308-12762-2

定 价 29.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部联系方式: 0571-88925591; <http://zjdxcbbs.tmall.com>

前言

本教材为本科应用型人才培养而编写,可作为高等院校经济管理类各专业学生应用统计课程的教材。我们的目的是引导从未接触过统计工作的经管类学生对统计学的基本理论形成正确的理解,并在此基础上利用所学的统计方法,借助于专业的统计软件,完成经管专业的基本统计分析任务。

统计包括数据搜集、数据处理、数据分析、数据解释并从数据中得出结论等环节,本书内容基本上就是按照这一过程来组织的。全书共9章,包括绪论、数据搜集、数据的图表展示、数据的概况性度量、统计推断基础、列联分析、方差分析、相关与回归分析、时间序列分析和预测等内容。本书较系统地介绍了统计学的基本概念、基本理论和基本方法,并指导学生使用专业统计软件SPSS实现基本的统计分析功能。

本教材具有以下三方面的特点:(1)内容编排模块化,在枯燥的理论阐述过程中穿插了阅读案例、知识链接、经典人物、专栏等内容,提高了教材的可读性;(2)强化案例教学,教材中安排了数量较多的案例,包括导入案例、阅读案例和案例分析等,培养学生运用统计原理和统计方法解决实际问题的能力;(3)知识体系实用有效,简化统计理论讲解,着重阐述应用型人才培养所需的关键内容以及统计软件的操作要点等。

本书的编写思路由胡卫中提出,编写大纲由胡卫中和吴步昶共同商定,孙钰、袁静和董美双负责教材的通稿工作。本书的编写分工如下:第1章、第2章、第4章和第7章由胡卫中完成;第3章、第5章和第6章分别由孙钰、袁静和吴步昶编写;第8章和第9章由董美双负责编写,其中部分案例、例题和习题等由胡卫中提供。

由于编者水平有限,书中不足之处在所难免,敬请同行专家和读者批评指正。

编者

2013年11月

目 录

第 1 章 绪 论	1
1.1 统计及其应用	2
1.2 统计数据类型	3
1.3 统计中的几个基本概念	4
1.4 SPSS 软件简介	5
第 2 章 数据搜集	10
2.1 数据的来源	11
2.2 统计调查方案设计	12
2.3 统计调查方法	15
2.4 统计数据搜集方法	19
2.5 调查问卷设计	23
2.6 建立 SPSS 数据文件	26
第 3 章 数据的图表展示	34
3.1 品质数据的图表展示	35
3.2 数值型数据的图表展示	39
第 4 章 数据的概况性度量	53
4.1 集中趋势的度量	54
4.2 离散程度的度量	59
4.3 偏态与峰态的度量	64
4.4 使用 SPSS 软件对数据进行描述分析	67
第 5 章 统计推断基础	72
5.1 抽样分布	73
5.2 参数估计	77

5.3 假设检验	87
第6章 列联分析	102
6.1 列联表和列联表分布	104
6.2 χ^2 检验	107
6.3 使用 SPSS 进行列联分析	110
第7章 方差分析	124
7.1 单因素方差分析	125
7.2 无重复双因素方差分析	131
7.3 重复双因素方差分析	139
第8章 相关与回归分析	146
8.1 相关分析	147
8.2 一元回归分析	153
8.3 多元线性回归	163
第9章 时间序列分析和预测	174
9.1 时间序列及其类型	175
9.2 时间序列的描述性分析	177
9.3 平稳序列的预测	181
9.4 趋势型序列的预测	186
9.5 复合型序列的分解预测	189
9.6 使用 SPSS 进行时间序列分析	193
参考文献	199

第1章 绪论

【教学目标】

通过本章学习,了解统计的含义,掌握统计学的基本概念及统计数据类型。

【教学要求】

知识要点	能力要求	相关知识
统计学及常用软件	理解统计的含义及应用范围	概念:统计学、描述统计、推断统计
基本概念	理解统计学的常用概念	概念:总体、样本、参数、统计量
数据类型	能对数据类型作出判断	概念:分类数据、顺序数据、数值型数据

【导入案例】

经济低迷使美国人离不起婚

根据婚姻顾问和律师们的经验,在经济困难的时代里,离婚率通常都要上升,比如1997年经济衰退期间,美国离婚率上升了17%。但2008年的趋势则完全相反。人们估计造成这种变化的原因是经济动荡的严重程度。

伊利诺依州的婚姻顾问杰夫·格朗雷说,在最近几个月里,找他进行婚姻咨询的家庭数量增长了25%,因为越来越多的夫妻尽力在挽救自己的婚姻,以及金钱。格朗雷这里的咨询要花费大约1000美元,并不是一笔小数目,但比起离婚的数千甚至上万美元花费来说,还是比较划算的。

离婚其实很昂贵。根据Divorce360网站的调查数据,一起经过法庭辩论的离婚案件中,一对至少拥有一个孩子的夫妇需要花费5.3万~18.8万美元不等,这其中包括律师费、金融顾问费、咨询费,以及离婚后另买或另租房屋的费用。通常情况下,这些费用中的大部分是夫妻出售了原有房屋和分割了财产后再来填补,但如今疲软的房地产市场使许多房屋所有者的资产大幅缩水,房产这个在不少人看来是婚姻财产最大块头的“肥肉”如今却变成了“拖油瓶”。

由于权威的2008年度美国人离婚数据还未出炉,因此目前谈论离婚率下降的问题可能还缺乏有力证据。不过,在美国一些地区已发布的半年期的数据中,我们可以看到,这种趋势已经不可避免。

在芝加哥,库克县巡回法院的报告称,在2008年前9个月里,该地区离婚和分居案件比2007年同期下降5%,减少了600起。加利福尼亚州弗雷斯诺县、俄克拉荷马州科曼奇县等地也有离婚案件减少的数据报告。

在南佛罗里达,迈阿密达德县的报告称2008年1月至5月间,此地的离婚案件数量同比下降了18%。也许并不是巧合,该地区的房地产价格也在同期下降了大约20%。

佛罗里达律师协会家庭法务部主席、婚姻家庭律师斯科特·鲁宾指出：“司法部门官员告诉我们的是，想来打官司的人说，他们负担不起分开组建两个家庭的费用——有些人几乎连原来的家庭都负担不起。”他举例说：“房产市场低迷，房子卖不出去，即便卖出去也是以极低的价格。因此许多人决定……‘现在卖房不划算，咱们还是一起过吧，尽力解决咱们之间的问题，同时希望经济还能反弹。’”

纽约精神治疗医师罗比·路德维格的一名客户就是对离婚有心无力的人。“有一名女客户对我说，‘我的天啊，我不能忍受我丈夫，每天我都想离开他，不过我负担不起离婚的代价。’于是他们夫妻俩又决定继续一起生活。”

以上统计数据能支持美国人离不起婚的结论吗？

资料来源：上海财经大学《社会统计学》省级精品课程网站（有删节）。

在日常生活中我们经常会接触到“统计”这一术语，在媒体上也经常会看到一些报道使用统计数据或图表等。在本章中我们将讨论统计学的一些基本问题，包括统计学的定义、统计数据及其分类、统计学的一些基本概念等。

1.1 统计及其应用

无论从事生产、科研或社会活动，人们总会碰到各种各样的数据。你需要分析这些数据，从中得出结论，帮助你作出决策。例如，人们在购买住房、股市交易、假日旅游和求学求职等方面的许多决策都需要依靠有关的统计信息；企业对人力、物力和财力实施有效管理也离不开准确的统计信息。

1.1.1 什么是统计学

统计学是一门关于数据的科学，它提供了一套数据搜集、处理、分析、解释并从数据中得出结论的方法。数据搜集就是取得统计数据。数据处理是将数据用图表等形式展示出来。数据分析则是选择适当的统计方法研究数据，从数据中提取有用信息进而得出结论。

统计分析数据的方法大体上可分为描述统计和推断统计两大类。描述统计是研究数据搜集、处理和描述的统计学方法。包括如何取得研究所需数据、如何用图表形式对数据进行处理和展示、如何通过数据汇总与分析得出数据的特征等内容。推断统计则是研究如何利用样本数据来推断总体特征的统计学方法，又可分为参数估计和假设检验两类。参数估计是利用样本信息推断所关心的总体特征，假设检验则利用样本信息判断检验总体的某个假设是否成立。例如，要了解一个地区的人口特征，不可能对每个人的特征一一进行测量；对成品质量进行检验，往往是破坏性的，也不可能对每个产品进行检验。这就需要抽取样本个体进行测量，根据获得的数据对总体特征进行推断，这就是推断统计要解决的问题。

1.1.2 统计学的应用

统计学的应用领域极其广泛，凡是有数据存在的学科和领域都会用到统计学。统计学

与各领域学科交叉形成了许多交叉性质的学科,如卫生统计学、气象统计学、管理统计学、经济统计学、环境统计学、工程统计学、空间统计学和人口统计学等。

需要指出的是,尽管统计学非常有用,但单纯依靠统计学并不能解决实际工作中的一切问题。统计工具解决实际问题需要具备两方面的知识:一方面需要统计学的基本理论与方法;另一方面需要具有实际问题的知识背景。这样才能把定量分析与定性分析结合起来,在应用统计方法进行定量分析的基础上,再应用各学科的专业知识对统计分析结果做出合理的解释,从而得出满意的结论。

专栏

中国的统计学是在20世纪前期形成和发展起来的。在此之前,统计活动在中国始终处于有“实”无“名”的状态。在考察我国统计史的时候,我们不得不面对这样一种特殊的演变模式:一方面,中国是世界上最早进行统计活动的国家之一,统计实务源远流长,内容丰富,《尚书·禹贡》中的九州表是国际学术界所公认的最早的统计史料,即为古代国势调查的肇端;另一方面,尽管历代学者相继提出了不少朴素的统计观点,应用了多种多样的统计方法对社会经济现象进行调查研究和分析,但是,中国本土的统计思想自先秦以后却发展缓慢,从古代到近代,始终未能够形成一个专门的知识系统。

19世纪末,西学东渐,近代科学被引入中国。1896年,华蘅芳与英国学者丁·弗勒合译的第一部中文版概率论著作《决疑数学》译成刊行,揭开了统计学在中国传播的序幕。随后,越来越多的欧美和日本的统计理论著作相继被译成中文。1930年,中国统计学社成立,标志着我国最早的统计学术族群的形成。

资料来源:马骏.二十世纪中国社会经济统计学的变革与发展.经济问题探索,1997(12).

1.2 统计数据类型

统计数据是对客观现象进行计量的结果,是统计总体单位标志或统计指标的具体数量表现,也是构成统计资料的基本要素。根据采用计量尺度的不同,统计数据可以分为分类数据、顺序数据和数值型数据三类。

分类数据是对事物进行分类的结果,数据表现为类别,是用文字来描述的。例如,反映人口特征的性别、民族等数据,就是通过对人口按不同类别加以分类所获得的数据。为便于统计处理,分类数据可以用数字符号来表示,如用“1”表示男性,“2”表示女性。但这些数字仅是一个符号,不能比较大小或进行任何数学运算。

顺序数据是只能归于某一有序类别的非数字型数据。顺序数据虽然也是类别,但这些类别是有顺序的。例如,一个人的受教育程度可一次分为小学、初中、高中、大学及以上等;顺序数据也可以用数字代码来表示,如“1”——小学、“2”——中学、“3”——高中、“4”——大学及以上。顺序数据可以排序,但无法知道不同顺序数据的具体差距有多大。

数值型数据是按数字尺度测量的观察值,其结果表现为具体的数值,如物体的重量、人的身高、某学校学生数量等。

不同数据需要采用不同的统计方法来分析处理。例如,分类数据通常统计其频数或频率,计算众数和异众比例,进行列联分析和卡方检验等;对顺序数据,可计算其中位数和四分位差,及等级相关系数等;对数值型数据可计算各种统计量,进行参数估计和检验等。

1.3 统计中的几个基本概念

和其他任何学科一样,统计学也有自己的基本概念。正确理解这些概念,对统计学的学习和应用十分重要。

1.3.1 总体和样本

总体是指所要研究对象的全体。它是由客观存在的具有某种共同性质的许多个别单位组成的集合体。而构成总体的每一个单位就是总体单位,也称个体。例如,在多个消费者构成的总体中,每一个消费者就是一个个体。

总体根据其包含的单位是否可数可以分为无限总体和有限总体。有限总体是指总体的范围能够确定,而且元素的数目是有限可数的。例如,某市的工业企业、某学校的学生等构成的总体。无限总体是指总体所包含的元素是无限或不可数的。例如,流水线上的产品、水库中的鱼等。

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容量。抽样的目的是根据样本提供的信息推断总体的特征。例如,从某社区有投票资格的选民中抽出 100 人的选民样本,然后根据这 100 人的投票意愿去推断整个社区选民的投票意愿。

阅读案例

没有披露的数据

某位统计专家曾经建议,在被告知某个调查的结果时,你需要做的就是反问一句:为了得出这个结论,你调查了多少名被访者?

采用严重有偏的样本几乎能够产生任何人需要的任何结果。只要样本容量足够小,或者你尝试足够多的次数,正确的随机样本也可以达到上述效果。

“用户反映使用多克斯(Doakes)牌牙膏将使蛀牙减少 23%”,大字标题历历在目。你希望减少 23% 的痛苦,于是接着往下读。你发现这些结论出自一家信誉良好的“独立”实验室,并且还经过了注册会计师的证实。有了这些,你还想知道什么呢?

然而,如果你不是特别容易轻信他人,或者不是一个盲目乐观的人,经验将告诉你:一种牙膏很难比其他牙膏好。那么多克斯公司是怎样得出上述结论的?如果是说谎,用大字标题报道这些谎言,他们又如何能够逃避责任呢?事实上,他们根本无需说谎,下面便是简单而有效的方法。

这里的主要把戏是不充分的样本——统计角度的不充分。但对于多克斯公司来说已经足够了。只有当你读小字体的文字时才会发现:被测试的用户仅由 12 人组成。单

凭这点,你便不得不佩服多克斯公司,它留给你一个可能知道全部情况的机会。有的广告商索性将类似的文字都略去,留给读者——即便他是一个老练的统计专家——一个猜想:这里面到底玩了什么把戏?从这个角度来说,多克斯公司由12个人组成的样本还不算太坏。几年前,一种叫做可尼斯博士(Dr. Cornish)的牙粉上市了,并宣传“在治疗龋齿方面获得了极大成功”,因为该牙粉中含有尿素,而经过实验室的证明,尿素对于治疗龋齿是有益的。然而,由于实验室的工作只是刚刚起步,仅仅建立在6个案例之上,毫无疑问这个结论是缺乏意义的。

下面,让我们再回头看看,多克斯公司是怎样轻易地获得一个不存在漏洞并经得起检验的标题。让规模不大的一组人连续记录6个月的蛀牙数,接着使用多克斯牙膏。之后一定会发生以下一种结果:蛀牙明显增多,蛀牙明显减少,或者蛀牙数量无显著变化。如果是第一种或者是第三种结果,多克斯公司编档保存好,当然最好是藏在别人找不到的地方,然后重新实验。由于机遇的作用,迟早有一组被测试者将证明有很好的效果,并且这个结果足以好到作为标题直至引发一场广告战。事实上,不管实验者使用的是多克斯牙膏,还是发酵粉,或者还是继续使用原来的品牌,上述结果都会发生。

资料来源:上海财经大学《社会统计学》省级精品课程网站(有删节)。

1.3.2 参数和统计量

参数是用来描述总体特征的概况性数字度量,它是研究者想要了解的总体的某种特征。研究者关心的参数通常有总体平均数、总体标准差和总体比例等。总体参数通常用希腊字母表示。总体平均数用 μ 表示,总体标准差用 σ 表示,总体比例用 π 表示。

由于总体数据通常是不知道的,参数通常是一个未知常数。例如,我们不知道某一地区消费者的收入或支出水平,不知道一批产品的合格率等。正因为如此,所以才需要抽样,通过样本计算估计总体参数。

统计量是用来描述样本特征的概括性数字度量,是估计样本数据计算出来的。由于抽样是随机的,统计量是样本的函数。样本统计量通常用英文字母表示。例如,样本平均数用 $\bar{x}$ 表示,样本标准差用 s 表示,样本比例用 p 表示等。

1.4 SPSS 软件简介

SPSS(Statistical Package for the Social Science)社会科学用软件包是世界上著名的统计分析软件之一。SPSS冠名为社会学统计软件包,是为了强调其社会科学应用的一面(社会科学研究中的许多现象都是随机的,要使用统计学和概率论的定理来进行研究),而实际上它在社会科学、自然科学的各个领域都能发挥巨大作用,并已经应用于经济学、生物学、教育学、心理学、医学以及体育、工业、农业、林业、商业和金融等各个领域。SPSS有如下特点:

(1)操作简单。除了数据录入及部分命令程序等少数输入工作需要键盘键入外,大多数操作可通过“菜单”、“按钮”和“对话框”来完成。

(2)无需编程。具有第四代语言的特点,告诉系统要做什么,无需告诉怎样做。只要了

解统计分析的原理,无需通晓统计方法的各种算法,即可得到需要的统计分析结果。对于常见的统计方法,SPSS 的命令语句、子命令及选择项的选择绝大部分由“对话框”的操作完成。因此,用户无需花大量时间记忆大量的命令、过程、选择项。

(3)功能强大。具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。自带 11 种类型的 136 个函数。SPSS 提供了从简单的统计描述到复杂的多因素统计分析方法,比如数据的探索性分析、统计描述、列联表分析、二维相关、秩相关、偏相关、方差分析、非参数检验、多元回归、生存分析、协方差分析、判别分析、因子分析、聚类分析、非线性回归、Logistic 回归等。

(4)方便的数据接口。能够读取及输出多种格式的文件。比如由 dBASE、FoxBASE、FoxPRO 产生的 *.dbf 文件,文本编辑器软件生成的 ASC II 数据文件,Excel 的 *.xls 文件等均可转换成可供分析的 SPSS 数据文件。能够把 SPSS 的图形转换为 7 种图形文件。结果可保存为 *.txt 及 html 格式的文件。

(5)灵活的功能模块组合。SPSS for Windows 软件分为若干功能模块。用户可以根据自己的分析需要和计算机的实际配置情况灵活选择。

本书的各种分析均通过 SPSS 18.0 中文版进行。SPSS 18.0 的主界面如图 1-1 所示。SPSS 软件主界面包括窗口名显示栏、编辑显示区、编辑区选择栏等,具体说明如下。

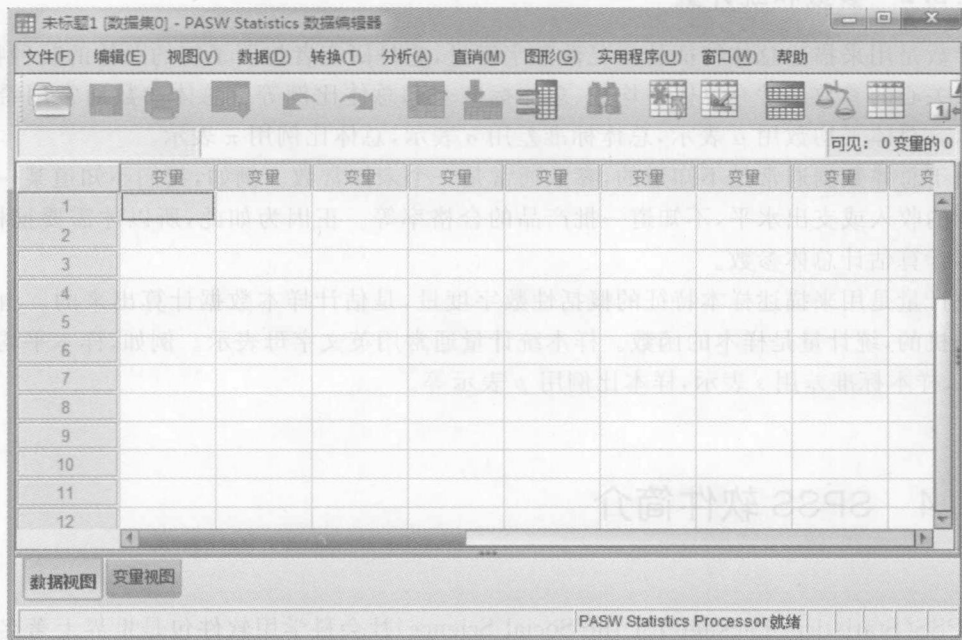


图 1-1 SPSS 软件主界面

窗口名显示栏:在窗口的顶部,显示窗口名称和编辑的数据文件名,没有文件名时显示为“未标题 1 [数据集 0]”。

窗口控制按钮:在窗口顶部的右上角,第一个按钮是窗口最小化,第二个按钮是窗口最大化,第三个按钮是关闭窗口。

SPSS 菜单:在窗口显示的第二行上,有文件、编辑、视图、数据、转换、分析、直销、图形、

实用程序、窗口、帮助等 11 个子菜单。

常用工具按钮:SPSS 有 19 个常用工具按钮,在窗口显示的第三行上,包括:打开数据文档、保存该文档、打印、检索最近使用的对话框、取消用户操作、重新执行用户操作、转向个案、转向变量、变量、查找、插入个案、插入变量、分割文件、加权个案、选择个案、值标签、使用变量集、显示所有变量、拼写检查。

数据单元格信息显示栏:在编辑显示区的上方,左边显示单元格和变量名(单元格:变量名),右边显示单元里的内容。

编辑显示区:在窗口的中部,最左边列显示单元序列号,最上边一行显示变量名称,默认为“变量”。

编辑区选择栏:在编辑显示区下方,“数据视图”在编辑显示区中显示编辑数据,“变量视图”在编辑显示区中显示编辑数据变量信息。

状态显示栏:在窗口的底部,左边显示执行的系统命令,右边显示窗口状态。

【本章小结】

本章介绍了统计学的概念。统计学是搜集、处理、分析、解释并从数据中得出结论的科学。统计学的应用范围非常广泛,目前已经被应用到自然科学和社会科学的众多领域。统计学与各领域实质性学科交叉发展成为若干领域的应用统计学。

统计数据是对客观现象进行计量的结果,是统计总体单位标志或统计指标的具体数量表现。根据采用计量尺度的不同,统计数据可以分为分类数据、顺序数据和数值型数据三类。

统计学的常用概念包括总体与样本、参数和统计量等。总体是指所要研究对象的全体,而构成总体的每一个单位就是个体。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容量。参数是用来描述总体特征的概况性数字度量,它是研究者想要了解的总体的某种特征。统计量是用来描述样本特征的概括性数字度量,是估计样本数据计算出来的。

复习思考题

1. 怎样理解统计学的含义?
2. 总体与样本有何差别? 又有何联系?
3. 举例说明参数和统计量的关系。
4. 统计数据可分为哪三类?
5. SPSS 软件的主要特点是什么? 其主界面由哪些部分组成?

案例分析

2008 年上海毕业生月工资中位数 2783 元

当前即将进入毕业生求职高峰,为给用人双方合理确定工资水平提供参考,上海市人力资源和社会保障局对 2008 年毕业并从事工作半年以内的毕业生工资水平进行了调查,在此基础上,发布了毕业生工资指导价位。本次调查范围涵盖研究生、本科、大专、中职等各学历层次毕业生。

一、毕业生工资增幅减缓

2008 年,毕业生月工资中位数为 2783 元,比上年增长 11.7%,但增幅较上年(18.3%)回落了 6.6 个百分点;高位数为 6328 元,增长 4.4%,增幅回落 1.8 个百分点;低位数为 1426 元,增长 5.7%,增幅回落 1.6 个百分点。可见,毕业生工资水平虽然继续保持增长,但增幅出现了回落。这主要是受 2007 年下半年以来的金融危机影响,部分企业生产经营压力增大,在招聘新进员工时加强了人工成本的控制和管理。

二、多数月薪 1501~2500 元

2008 年,月工资在 1500 元及以下的毕业生占 7.1%,1501~2500 元的占 38.9%,2501~3500 元的占 26.5%,3501~4500 元的占 14.7%,4501 元及以上的占 12.8%。可见,对于刚工作半年以内的毕业生来说,其工资主要集中在 1501~2500 元,其次是 2501~3500 元,两者合计占到 65.4%。

三、各学历层次工资差距缩小

从各学历层次看,2008 年,月工资中位数最高的是研究生,为 4634 元;其次是大学本科,为 3000 元;第三是大专(含高职),为 2237 元;第四是中职(含中专、技校、职校),为 2013 元;各学历层次毕业生的月工资中位数差距比例为 2.3 : 1.5 : 1.1 : 1,而上年为 2.7 : 1.5 : 1.2 : 1。

可见,在影响毕业生工资水平上,虽然学历仍是重要因素之一,但各学历层次毕业生的工资差距有所缩小。这反映出,企业在看重学历的同时,更看重实际工作能力。

四、薪酬行情

1. 建筑行业毕业生工资增长

受金融危机、消费需求下降、企业生产经营压力大等影响,汽车制造、对外贸易、住宿餐饮等行业的毕业生工资水平有所下降。如汽车制造业中,车辆零部件设计人员和汽车电机电器设计人员的月工资中位数为 3464 元和 3778 元,分别比上年下降了 4% 和 19.7%;对外贸易行业中,进出口业务人员的月工资中位数为 2414 元,下降了 13.9%;住宿餐饮行业中,总台管理员和餐饮服务员的月工资中位数为 1600 元和 1750 元,分别下降了 2.7% 和 7.8%。

但与此同时,为应对金融危机影响,上海市出台了各项促进经济发展的措施,特别是加强了基建投资,以加快城市各类基础设施的建设。

因此,从工资水平看,建筑行业的毕业生工资水平有所增长。如建筑设计工程技术人员月工资中位数为 2538 元,比上年增长 8.1%;土木建筑工程技术人员为 2220 元,增长 12.9%;给排水工程技术人员为 2100 元,增长 15.3%。

在就业形势严峻的情况下,由于一般管理和销售类岗位的就业“门槛”较低,毕业生往往争相应聘,因此企业选择余地较大。对于一些供大于求的岗位,提供的工资水平不高,且出现了一定下降。如前台服务接待的月工资中位数为 1793 元,比上年下降了 5.4%;销售助理为 2000 元,下降了 17.4%;采购助理为 2546 元,下降了 3.1%。

资料来源:上海财经大学《社会统计学》省级精品课程网站。

根据以上案例回答下列问题:统计研究的对象是什么?本例提到了什么统计总体?本例采用了什么研究方法?

第2章 数据搜集

【教学目标】

通过本章学习,了解数据搜集方法和调查组织形式;根据调查目的和任务设计调查方案和调查问卷;学会建立 SPSS 数据文件。

【教学要求】

知识要点	能力要求	相关知识
数据来源	了解数据直接来源和间接来源的特点	数据的直接来源、数据的间接来源
调查方案设计	(1)了解调查方案设计的基本内容 (2)具备调查方案撰写能力	调查方案的 10 项基本内容
调查组织形式	(1)了解基本的统计调查方式 (2)了解抽样调查的组织形式	(1)统计报表制度、普查、抽样调查、重点调查、典型调查 (2)概率抽样、非概率抽样
统计调查方法	了解询问调查与观察实验的特点	询问调查、观察与实验
问卷设计	(1)能根据调查目的设计问卷 (2)了解调查问卷设计的注意事项	问卷结构、调查项目
SPSS 数据文件	能建立 SPSS 数据文件	定义变量、数据输入

【导入案例】

老黄金香烟公司的广告

老黄金(Old Gold)香烟公司曾利用一个毫无价值的抽样调查结论制造了大量喧闹并大赚了一笔。

故事起源于《读者文摘》某编辑的一个偶然想法。该编辑自己抽烟,而且他并不认为各种品牌的香烟完全相同。他的杂志开始行动起来,聘请了一些实验室人员,对不同品牌香烟的烟雾展开了分析。杂志刊登了最终结果,列出每种品牌香烟的烟雾中尼古丁以及其他有害物质的含量。在详尽的数据支持下,该杂志声明:所有品牌的香烟是一样的,无论你吸的是什么的牌子的香烟,不会有任何差异。

也许你会认为,对于香烟厂商和那些为香烟杜撰崭新广告词的人而言,这是一个有力的打击,它还将引爆关于平缓喉痛和对胸部有益的广告索赔案。但某些人却有了其他发现,在一长串具有相同有害物质的品牌名单上,总有一个排在最后,这就是“老黄金”牌香烟。于是,广告漫天飞舞,大幅广告以最大的字体刊登在报纸上。广告的标题和副本仅仅提到,由一家国家级杂志主持的实验证明,“老黄金”牌香烟在不良物质以及尼古丁含量方面“排名最后”,任何关于各个品牌的差异并不显著的文字和暗示都被省略了。

最后,“老黄金”公司被通知“中止并停止”使用这个具有误导性的广告。但这并不要紧,他们早已获得了足够的好处。

提示:差别有意义时才能称为差别。

资料来源:上海财经大学《社会统计学》省级精品课程网站(有删节)。

统计数据搜集是根据统计研究的目的和任务,运用科学的调查方法向调查单位搜集统计资料的过程。数据搜集是数据整理、数据分析和统计推断的基础和前提,在很大程度上决定了整个统计工作的质量水平。

2.1 数据的来源

所有统计数据追踪其初始来源,都来自调查或实验。但从来源渠道看,数据可分为两类:一种是通过直接的调查或实验获得的第一手原始数据,一般称之为数据的直接来源;另一种是别人通过调查或实验获得,并将这些数据进行加工和汇总后公布的数据,通常称之为数据的间接来源。

2.1.1 数据的间接来源

第二手或间接的统计数据主要来自于国内外公开出版或公开报道的各种出版物。公开出版或公开报道的社会经济统计数据主要来自国家和地方的统计部门以及各种报刊媒介。例如,公开的出版物有《中国统计年鉴》、《中国统计摘要》和各种专业统计年鉴,以及各省、市、地区的统计年鉴等。提供世界各国社会和经济数据的出版物也有很多,如《世界经济年鉴》、《国外经济统计资料》、世界银行各年度的《世界发展报告》等。联合国有关部门及世界各国也定期出版各种统计数据。

二手资料比较容易得到,相对来说比较便宜,并能很快地获取。有些二手数据,例如由国家统计局普查结果所提供的数据,是不可能由任何一个调查公司按原始数据去搜集的。尽管二手数据不可能提供特定调研问题所需的全部答案,但二手数据在许多方面都是很有用的。例如,二手数据可以帮助我们寻找处理问题的途径,构造适当的设计方案,回答一些调查问答题,检验某些假设,更深刻地解释原始数据等。因此,考查研究可能得到的二手数据是搜集原始数据的先决条件。一般应从二手数据开始分析。只有当二手数据的来源已经全用完了或者有了一定的盈余以后,才考虑进行调研搜集原始数据。

尽管二手数据对调研很有帮助,但调研者在使用二手数据时应当谨慎,因为二手数据有一定的局限性和缺陷。由于二手数据是为其他目的而不是为手边的问题搜集的,因此,二手数据对当前问题的帮助在一些重要方面是有缺陷的。资料的相关性和准确性都不够。搜集二手数据的目的、性质和方法不一定适合当前的情况。而且,二手数据也可能缺乏准确性,或者有些过时了。在使用二手数据之前,有必要先对二手数据进行评估。