



统计数据分析软件教程

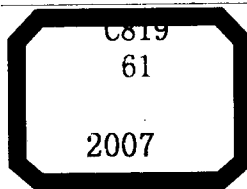
——SPSS13.0（试用版）及EXCEL的应用

王玉荣 编著



对外经济贸易大学出版社

University of International Business and Economics Press



统计数据分析软件教程

——SPSS13.0（试用版）
及 EXCEL 的应用

王玉荣 编著

对外经济贸易大学出版社

图书在版编目 (CIP) 数据

统计数据分析软件教程——SPSS13.0 (试用版) 及 EXCEL 的应用/王玉荣编著. —北京: 对外经济贸易大学出版社, 2007

ISBN 978-7-81078-835-9

I. 统… II. 王… III. ①统计分析-软件包, SPSS13.0-教材 ②电子表格系统, Excel-教材 IV. C819 TP391.13

中国版本图书馆 CIP 数据核字 (2007) 第 046111 号

© 2007 年 对外经济贸易大学出版社出版发行

版权所有 翻印必究

统计数据分析软件教程

——SPSS13.0 (试用版) 及 EXCEL 的应用

王玉荣 编著

责任编辑: 宋志红

对外经济贸易大学出版社

北京市朝阳区惠新东街 10 号 邮政编码: 100029

邮购电话: 010-64492338 发行部电话: 010-64492342

网址: <http://www.uibep.com> E-mail: uibep@126.com

唐山市润丰印务有限公司印装 新华书店北京发行所发行

成品尺寸: 185mm × 230mm 15.25 印张 305 千字

2007 年 5 月北京第 1 版 2007 年 5 月第 1 次印刷

ISBN 978-7-81078-835-9

印数: 0 001 - 5 000 册 定价: 24.00 元

前言

Foreword



在经济全球化进程不断加快,世界经济联系越趋紧密,市场竞争越来越激烈的今天,一个企业要想赢得市场,求得生存和发展,必须最大限度地减少决策失误的概率,为此,决策者仅凭个人的经验、知识和感觉是很难做到这一点的。在决策过程中,必须充分利用集体的经验、知识、智慧和以科学的分析方法,对收集到的数据做出准确、及时的分析并制定正确的决策。

本书坚持以案例为依托,应用“统计学”课程中讲授的基本统计方法,利用当前世界上最流行的统计分析软件 SPSS 以及最普遍的 Excel 软件来解决案例中的问题,使使用者能更好地利用统计分析方法解决实际问题,使统计方法在决策中能发挥重要的作用,也使学生更加对统计学感兴趣。

本书分为两部分;第一部分介绍如何应用 SPSS 进行数据分析;第二部分介绍如何应用 Excel 进行数据分析。两个软件各有特点,但是对于数据分析功能,Excel 是无法与 SPSS 相比的,所以本书的重点在第一部分。当然由于 Excel 软件的普及性,并能完成最基本的统计分析,所以对它的介绍也很有必要。本书对于应用 Excel 进行分析时很繁琐的方法的介绍尽量略去。本书包括的内容主要有:数据文件的建立、数据的编辑、数据的整理、数据的描述性分析、参数估计与假设检验、概率和概率分布、方差分析、非参数检验中的卡方检验、二项分布检验及正态性检验、相关与回归分析、时间数列分析。

本书重点突出以下几个特点:

1. 把 SPSS 和 Excel 放在一起,SPSS 的版本较新;
2. 对统计学分析方法的介绍,力求通俗易懂、简明扼要;
3. 应用实际案例,从研究问题出发,重点介绍软件是如何帮助解决现实中的问题的,并不强调对软件中每个细节的介绍。
4. 针对每一个研究问题,给出了解决问题的统计方法。
5. 在介绍应用 SPSS 软件进行统计分析时,给出了详细的步骤,并展示了每一步的对话框,对对话框中的选项作了简明扼要的介绍,对 SPSS 的输出结果作了详尽的解释。
6. 每章提供了实际案例及相关数据,供读者练习;

7. 内容安排与大部分“统计学”课程中讲授的内容一致；

本书可作为高等院校本科生统计分析软件的教材，也可以作为 MBA 学生、研究生以及从事数据分析人士的参考书。

由于编写时间仓促，书中难免有疏漏，敬请国内同行及广大读者批评指正。

王玉荣

2007 年 2 月

目 录

Contents



第一部分 统计数据分析与 SPSS	(1)
第一章 概述	(3)
第二章 数据录入及编辑	(6)
第三章 数据的整理和表述	(32)
第四章 数据的描述性分析	(77)
第五章 参数估计与假设检验	(92)
第六章 方差分析	(107)
第七章 非参数检验	(121)
第八章 相关与回归分析	(134)
第九章 时间数列分析	(161)
第二部分 统计数据分析与 Excel	(183)
第一章 数据整理	(185)
第二章 数据的描述性分析	(195)
第三章 概率及概率分布	(197)
第四章 参数估计与假设检验	(205)
第五章 方差分析	(214)
第六章 相关与回归分析	(220)
第七章 时间数列分析	(227)
附录 “我看中日经贸关系” 数据库	(235)

第一部分

统计数据分析与 SPSS



第 1 章

概 述

1.1 SPSS 简介及特点

SPSS 是英文 Statistical Package for the Social Science 的缩写，即：社会科学统计分析软件包。20 世纪 60 年代末由美国斯坦福大学的三位研究生研制，1975 年在芝加哥组建 SPSS 总部。它是当今世界上公认和流行的综合统计分析软件包。在我国，SPSS for Windows 以其强大的统计分析功能、方便的用户操作界面、灵活的表格式分析报告及其精美的图形展现，受到了社会各界统计分析人员的喜爱。目前，SPSS 软件使用已经成为许多高等院校统计学、经济学、管理学等专业本科学生的必修课程。

SPSS for Windows 系统特点是操作简单，统计方法齐全，图形、表格绘制方便，输出结果直观。最新版为 15.0 版，网址为 <http://www.SPSS.com/>。本书的各种分析结果均是通过 SPSS for Windows 13.0 全模块英文试用版的运行得到的。

1.2 SPSS 主界面

单击 Windows 的“开始”菜单，在“程序”菜单里点击“SPSS 13.0 for Windows”按钮，即可启动 SPSS。SPSS 主界面有两个，一个是 SPSS 数据编辑窗口（见图 1-1-1），另一个是 SPSS 结果输出窗口（见图 1-1-2）。

1.2.1 SPSS 数据编辑窗口

SPSS 数据编辑窗口主要由以下几个部分组成：标题栏、菜单栏、工具栏、编辑栏、

变量名栏、内容栏、窗口切换标签、状态栏 (见图 1-1-1)。

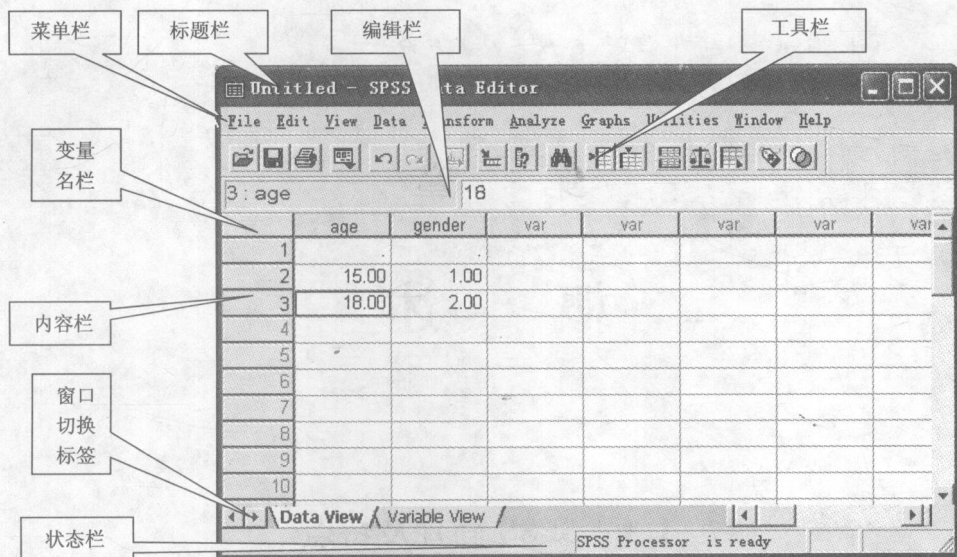


图 1-1-1 SPSS 数据编辑窗口

(1) 标题栏显示数据编辑的数据文件名。

(2) 菜单栏中包括 SPSS 的 10 个命令菜单，每个菜单对应一组相应的功能。File 是文件的操作菜单；Edit 是文件的编辑菜单；Data 是数据的建立与编辑菜单；View 是用户界面设置菜单；Transform 是数据基本处理菜单；Analyze 是统计分析菜单；Graphs 是统计图形菜单，输出各种分析图形；Utilities 是相关应用和设置菜单；Windows 是 SPSS 显示窗口的切换菜单；Help 是 SPSS 帮助菜单。

(3) 工具栏中列示了一些常用的操作工具的快捷图标。可以根据操作者的需要增减操作工具的快捷图标，以使操作更为方便。

(4) 编辑栏中可以输入数据，以使它显示在内容区指定的方格里。

(5) 变量名栏列出了该编辑文件所包含的变量名。SPSS 自动的变量命名是：var0001、var002…。在本图中有两个变量：age 和 gender。

(6) 内容栏列出了各样本在各变量上的取值。SPSS 内容栏的每一行代表一个样本。内容栏左边是样本的标号。

(7) 窗口切换标签处有两个标签，即“Data View”和“Variable View”。“Data View”对应的表格用于样本数据的查看、录入与修改，而“Variable View”则用于变量定义和属性的输入与修改。这种内容栏的变换与 Microsoft Excel 的输入较为相似。

(8) 状态栏用于说明显示 SPSS 当前的运行状态。当 SPSS 被打开后, 将会显示“SPSS Processor is ready”的提示信息。

1.2.2 SPSS 结果输出窗口

SPSS 结果输出窗口主要由以下四个部分组成: 菜单栏、工具栏、输出部分和状态栏。

菜单栏、工具栏和状态栏与 SPSS 数据编辑窗口类似, 这里就不再赘述。下面主要介绍一下输出部分, 输出部分主要分为: 索引输出区和详细输出区 (见图 1-1-2)。

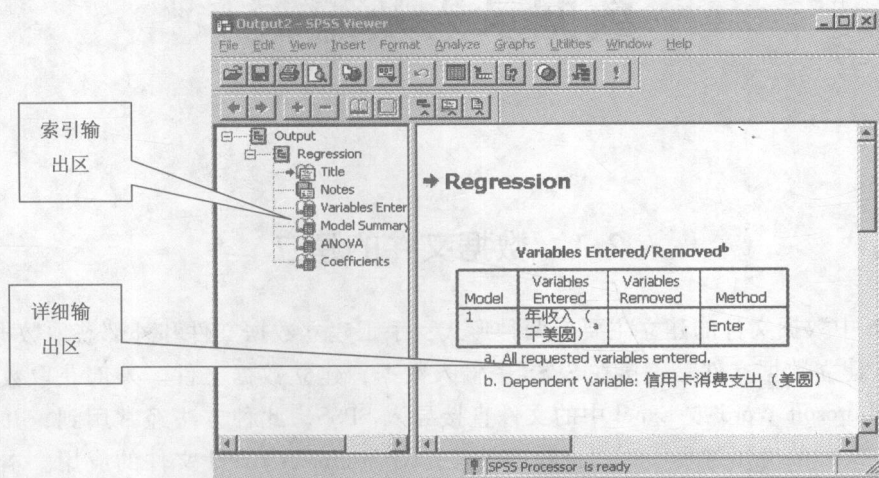


图 1-1-2 SPSS 结果输出窗口

(1) 索引输出区用于显示已有分析结果标题和内容索引。以简洁的方式反映出详细输出区的各个内容项, 便于用户查找和操作。索引输出以一个索引树根结构显示, 当你想要查找需要的输出结果时, 只要单击索引树上相应的图表名称, 就可显示在窗口上, 非常方便。

(2) 详细输出区输出的是研究者所要得到的具体图表, 与索引输出区的结果是一一对应的。详细输出区的表格可以进行编辑等操作。如果要对选中表格进行编辑, 可双击该表格, 当表格四周出现黑色斜线边框时, 即可对表格内数据进行修改。

第2章

数据录入及编辑

2.1 数据文件的建立

SPSS 中数据文件的建立主要有两种形式：手工建立数据文件和外部获取数据文件。

手工建立数据文件是直接在 SPSS 中输入数据，建立数据文件。外部获取数据文件是指将 Microsoft Word 或 Excel 中的文件直接导入 SPSS，此种方法经常用到，并且比较方便。在实际的应用过程中，用得最多的还是对 Microsoft Excel 文件的应用。本章中主要介绍 Excel 数据文件如何转变为 SPSS 数据文件。

2.1.1 手工建立数据文件

研究问题 1：下面是笔者为了研究居民收入与生活状况而设计的调查问卷的节选片段，（见表 1-2-1）。我们考虑如何把问卷中的信息转化为 SPSS 数据文件中的内容。

表 1-2-1

问卷节选

一、您的年龄是_____岁

二、您的性别是

① 男 ② 女

三、您的月均收入是

① 300 元以下 ② 300 ~ 500 元 ③ 500 ~ 1 000 元

④ 1 000 ~ 1 500 元 ⑤ 1 500 ~ 3 000 元 ⑥ 3 000 元以上

四、您是否感到生活压力大

- ① 是 ② 否

五、如果您家近年来生活水平提高了，那么主要原因是（任选两项）

- ① 家庭就业人数增加 ② 工资增加
- ③ 奖金和津贴增加 ④ 其他收入来源增加

六、如有存款，请将存款的主要目的按重要性编号填写在下面的表格中，最重要的填在第一格中，最不重要的填在最后一格中。

--	--	--	--	--	--	--	--	--	--

- ① 办婚事 ② 防老 ③ 以备急需
- ④ 添置高档商品 ⑤ 购房 ⑥ 旅游
- ⑦ 为子女上学 ⑧ 不知买什么好，先存起来
- ⑨ 保值生息 ⑩ 其他，请注明_____。

七、你家拥有的投资品种有

- ① 银行理财 ② 股票 ③ 债券
- ④ 股票型基金 ⑤ 货币基金 ⑥ 外汇
- ⑦ 其他，请注明_____。

八、您对实施储蓄实名制的态度是：

反对 1 2 3 4 5 6 7 赞成

九、请用 1-7 的数字表示你对下列每句话的评价

非常不同意	不同意	有点不同意	无所谓	有点同意	同意	非常同意
1	2	3	4	5	6	7

序号	调查项目、	1	2	3	4	5	6	7
1	您家的生活条件非常优越							
2	您对现在的生活状态感到满意							
3	您对现在的工作感到满意							
4	您的家庭非常和睦							
5	您与同事的关系很和谐							

2.1.1.1 确定变量的个数

本问卷节选中共有 9 个问题，但并不一定只设 9 个变量，下面结合以上问卷介绍如何设置变量。在数据编辑窗口中，一列为一个变量。变量个数的确定依赖于问卷中问题

答案的方式。问卷中绝大多数问题采用封闭式答案。作答的基本方式有：

- (1) 是否式 (表 1-2-1 问题四)。
- (2) 单项选择 (表 1-2-1 问题二)
- (3) 多项选择 (表 1-2-1 问题五、七)
- (4) 排名式：要求将所列各个答案排序 (表 1-2-1 问题六)
- (5) 量表：用于测度被访者的态度和认识，将一个测量连续统划分成若干个位点 (通常是 1~7 个点)，由被访者确定一点 (表 1-2-1 问题八、九)
- (6) 数值归档：对于收入、年龄、企业销售额这样的问题，不要求回答具体数值，而是事先划分出数值区间，由被访者确认 (表 1-2-1 问题三)
- (7) 填空 (表 1-2-1 问题一)

变量个数的确定应根据作答的方式的不同，恰当地选取。此问卷节选中变量的设置个数见 (表 1-2-2)，笔者将在以下的部分对于各问题变量的定义作具体的介绍。

表 1-2-2 问卷变量设置个数

问题编号	一	二	三	四	五	六	七	八	九
变量数量 (个)	1	1	1	1	2 或 4	10	7	1	5

2.1.1.2 定义变量名称和数据的类型属性

在数据编辑窗口中左下角处，点击 Variable View 标签，进入变量浏览界面，见图 1-2-1。

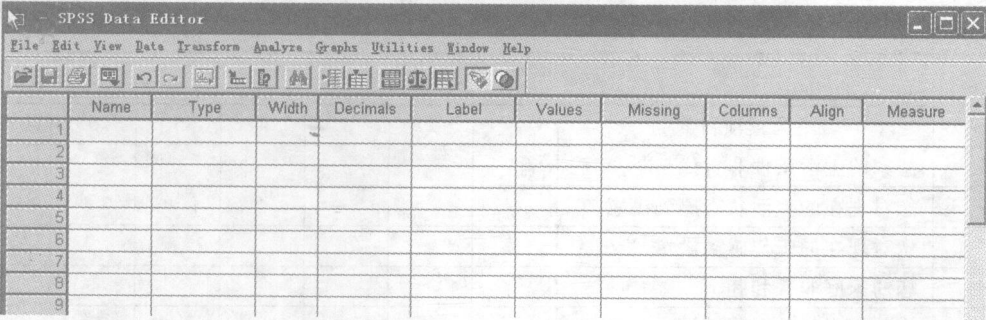


图 1-2-1 变量浏览界面

在打开变量浏览界面后即可对各变量进行定义，定义变量时有 10 种功能选项，分别是 Name、Type、Width、Decimals、Label、Values、Missing、Columns、Align、Measure。下面分别作简要介绍：

(1) Name: 变量名。在此栏目中输入变量名, 如 age、gender 等 (见图 1-2-2)。

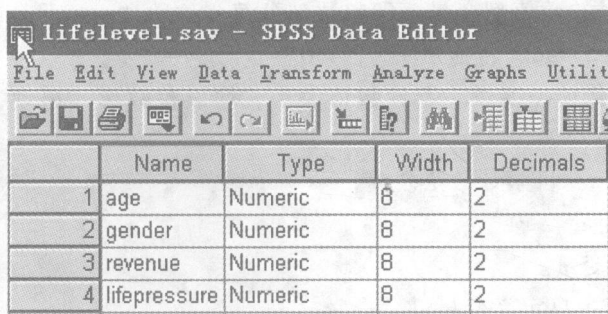


图 1-2-2 变量名的输入展示

定义变量名时应注意以下几点:

第一, 变量名必须以字母为首, 后面跟 A~Z, 0~9 字符, 对于字符数量在 SPSS13.0 版本以后的版本中都没有限制。“?”、“!”、“/”等不能作为变量名, 变量名也不能带拓展名, 如 A1.1。

第二, 有些关键词不能作为变量的名字, 如 AND、OR、NOT、EQ、LT、LE、GT、GE、NE、TO、BY、CROSSTABS、WITH、ALL、THRU、PERCENTAGE。SPSS 中虽不区别大小写字符, 但程序中的命令和关键词要用大写字母、表示系统内定; 变量名等宜用小写字母, 表示人为指定。

第三, 可以用中文作变量名, 但最好不用, 因为涉及到兼容性的问题。由于兼容性的问题很多情况下的输出会产生乱码, 这将给分析造成很大的不便。

(2) Type: 变量类型。点击 Type 按钮, 将会出现  标志, 点击此标志将会出现以下变量类型对话框 (见图 1-2-3), 在此对话框中有 8 种变量类型可供选择。

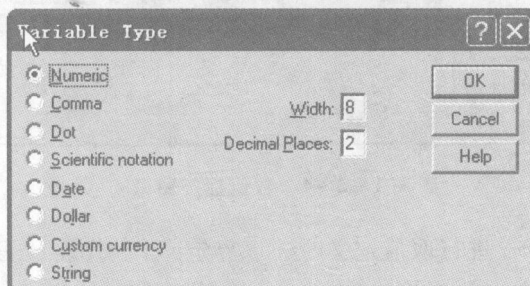


图 1-2-3 变量类型对话框

Numeric: 数值型变量, 系统默认变量为数值型, 系统默认变量长度为 8, 小数位

为 2。这种模式最常用。

Comma: 带逗号的数值型变量, 即整数部分每 3 位数加一逗号, 其余定义方式同数值型。例如: 输入 123456, 将显示 123, 456。

Dot: 带圆点的数值型变量, 不论数值大小, 均以整数形式出现, 每 3 位加一逗号 (但不是小数点)。

Scientific notation: 科学计数法。

Date: 日期型变量。

Dollar: 货币型变量。

Custom currency: 自定义变量。

String: 字符型变量。选中该项后, 可在数据输入时输入中文或者英文字符。一般用的是数值型变量, 字符型变量少用为宜。

(3) Width: 宽度。运算宽度, 默认值是 8, 运算宽度实际上只会改变输出结果的显示宽度, 数据的存储结果与运算的精度不受宽度的影响。

(4) Decimals: 小数位数。默认为 2 位小数。

(5) Label: 变量标签。用来表示变量的含义。例如在此调查问卷中, 我们可以将第一个变量这样表示: Name: age, Label: 年龄。

(6) Values: 数值标签。单击 Value 按钮, 再单击弹出的...按钮, 进入数值标签窗口 (见图 1-2-4)。

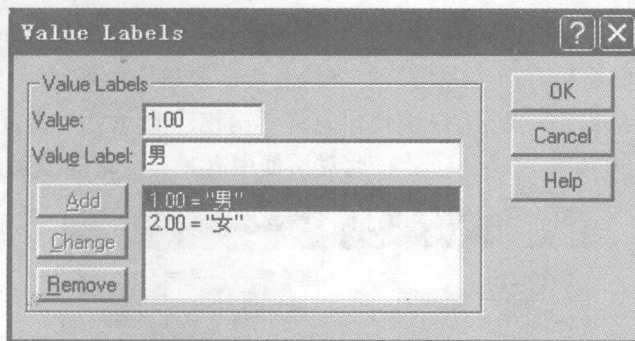


图 1-2-4 数值标签窗口

在对性别变量 gender 进行取值定义时, 可界定“1”代表“男”, “2”代表“女”, 则在第一个“Value”文本框中输入“1”, 再在第二个“Value”文本框中输入“男”, 点 Add 按钮确认, 即可定义“1 = ‘男’”。依次再定义“女”即可, 最后点击 OK 即可。

(7) Missing Value: 缺失值。SPSS 有两类缺失值: 系统缺失值和用户缺失值。在 Data View 窗口中任何空的数字单元都被认为系统缺失值, 用点号 “.” 表示。由于特殊原因造成的信息缺失值, 称为用户缺失值。例如在统计过程中, 可能需要区别一些被调查者不愿意回答的问题, 然后将他们标为用户缺失值, 统计过程可识别这些标识, 带有缺失值的观测被特别处理。

单击 Missing 按钮, 再单击弹出的  按钮, 进入 Missing Values 窗口 (见图 1-2-5)。

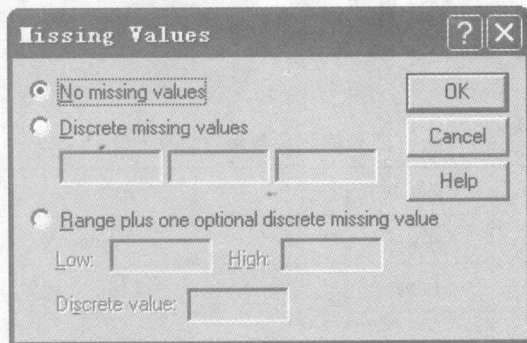


图 1-2-5 Missing Values 窗口

界面上有 3 个单选钮, 默认值为最上方的 No missing Values 即 “无自定义缺失值” 的方式。第二项 Discrete missing values 指定离散的缺失值, 最多可以定义 3 个值。最后一项 Range plus one optional discrete missing value, 指定缺失值所在的区间范围, 并可同时指定一个离散值。

(8) Columns: 输入变量的显示宽度, 默认为 8。

(9) Align: 字符排列方向。有 3 种选择, left 表示左对齐; center 表示居中; right 表示右对齐。

(10) Measure: 数据量度选择。有以下 3 种选择方式:

Scale: 连续型变量。在 SPSS 中连续型变量是定距变量和定比变量的总称, 这类变量的特点是, 取值之间可以比较大小, 可以用加减法计算出差异的大小。例如问题一中的年龄变量就是连续型变量, 确切地说是定距变量。定距变量与定比变量的惟一区别是定距变量中的 “0” 只表示某一取值, 不表示没有。而定比变量则表示没有, 不表示 “0”。通常对定比与定量变量不予区别, 我们在 SPSS 里都用 Scale 来定义。

Ordinal: 定序型变量。其取值的大小能够表示观测对象的某种顺序关系 (等级、方位或大小等)。例如在我们的例子中的问题三月均收入的取值就是① 300 元以下、② 300~500 元、③ 500~1 000 元、④ 1 000~1 500 元、⑤ 1 500~3 000 元、⑥ 3 000 元