

统计软件包在经济、社会统计中的应用

郭 华

(一)

经济、社会统计主要是通过一些指标来说明经济、社会现象的数量特征,即对经济、社会现象进行定量分析。列宁曾明确指出:社会统计是“社会认识的最有力的武器之一。”^①通过这个武器我们可以对经济和社会的发展现状与趋势有个明晰的认识,并为科学决策提供依据。

统计作为一种从数量方面认识事物的科学方法,在经济、社会分析中最大的困难有3个:(1)要处理的数据多,不论是经济统计还是社会统计,要说明一个现状并不是几个数据就能描述清楚,必须要有大量的数据。且结构比较复杂,既有能进行数据处理的变量,如产值、产量、收入等,还有一些非数值型的变量,如性别、种族等。特别是对于有一定规模的经济或社会调查,如果通过手工运算统计的话,将耗费极大的人力、物力和时间。(2)当对一些数据进行探索性分析时,要经过不同方法的反复试验比较,才能找到有效的、综合性的处理手段,手工统计分析就很难做到这一点。(3)对一组数据要进行综合统计分析,且分析的问题角度比较多,既要进行一些简单的统计分析,又要进行一些较为复杂的多变量分析及各种非参数检验,同时一些数学模型也比较复杂,根本难以用手工来计算。上述问题给统计分析带来了一定难度。现代科学技术的发展,特别是电子计算机的广泛运用,为解决上述难题,创造了条件。

(二)

为了让社会科学工作者和没有经过专门计算机训练的人也能方便地使用计算机来进行统计分析,计算机上配备的各种统计软件包应运而生。统计软件包不同于我们对某一特殊问题或某一特殊统计方法而编制的专项程序,它包含了由覆盖面大、使用频率高的各种统计方法组成的统计命令。统计软件包也不同于由若干统计程序组成的统计库,统计库必须依附于编写的语言系统才能工作,而统计软件包则有自己的数据与文件管理系统和操作,这就使得它能更方便地处理统计数据。目前国际上比较通用的统计软件包有3种:(1)统计系统SYSTAT(The

system for statistics), (2)统计分析系统SAS(Statistical Analysis System)、(3)社会科学统计软件包SPSS(Statistical Package for the Social Sciences)。在这些统计软件包里,除了可以方便地建立数据集文件外,还有一些统计命令,其中包括:(1)数据排序分类;(2)打印图表和统计量;(3)进行简单的统计分析(如单变量的均值、标准差、方差、极大值、极小值、全距等描述统计);(4)双变量的相关与回归(简单回归、多元回归和时间序列)等;(5)方差分析;(6)非参数检验(比如频率分布、卡方分布等);(7)更复杂的多变量分析(比如:因素分析、判别分析、聚类分析等)。

经济、社会统计属于社会科学领域,作为社会科学研究最常用的统计软件是社会科学统计软件包SPSS,因为这个统计软件包十分注意反映社会科学统计问题的特殊性。对于社会科学所需处理的各种数据均能解决。它包括:(1)数值型变量(亦称高测度水平的变量),如产量、产值、收入等;(2)非数值型变量(亦称一般测度水平的变量),如学历、级别等;(3)只能分成类目的变量称为低测度水平变量,如性别、民族等。SPSS统计软件包针对不同测度水平的变量提供不同的统计方法。特别是处理非数值型变量其统计功能很强,一方面它强调直接可用于处理这类变量的非参数统计分析方法的作用;另一方面,它又非常注意采用非数值型变量数量化的办法,所谓带虚设变量回归就是如此。由于社会科学的数据包括经济调查和社会调查的数据容易造成数据的缺损,SPSS规定一整套处理缺损数据的办法,几乎在它的每一个统计命令之后都可用MISSING或OPTIONS子命令来说明对缺损数据的处理方式,使之对于缺损的数据,也往往能挖掘出有用的信息。这样使得社会科学数据的复杂性或多值性,以及抽样数据的不确定性都能方便地利用社会科学统计软件包SPSS加以解决,从而使社会科学工作者能方便地利用现代化的计算手段——计算机进行各种统计分析,诸如分组分析法、对比分析法、动态分析法、指数分析法、平衡分析法和图示分析法等。

(三)

在计算机上用社会科学统计软件包SPSS 进行统计分析,大致可分为以下二步:

第一,建立数据文件(或子集)。一个完整的数据文件应包括在统计分析时,所要用的所有变量,即基本状况变量和测度变量。由这些变量所组成每个CASE(或叫观察值或个案)的数据组成一个数据集,这个数据集被称为数据文件。如有一个样本容量为200人的抽样调查,每个个案由个人的年龄(AGE)、性别(SEX)、受教育年限(EDUCATION)、职业(PROFESSION)、收入(INCOME)等5个变量组成。整个数据文件可以通过SPSS统计软件包所提供固定格式输入方式加以建立,即:

```
DATA LIST /AGE 1—2
SEX 4 EDUCATION 6—7
PROFESSION 9—10 INCOME
12—15
数据(略)
END DATA.
```

通过以上命令的操作,然后存入一个“EXAMP.INC”的文件名中,便建立了这个调查的数据文件。

第二,根据研究需要进行各种统计分析,即通过SPSS统计软件包所提供的一系列统计命令和各种不同参数的选择项来进行各种统计分析。SPSS统计软件包中共包括了20多条统计命令,可以完成对单变量、双变量和多变量的统计分析,并作出各种统计图表。

1.在统计分析中,最简单的统计是单变量统计,即作出各变量的频率分布和一些描述性统计。它包括数据的均值、中位值和众值等统计量。如我们需分析以上数据中收入这个变量的均值、方差和分布情况,其运算程序如下:

```
INCLUD "EXAMP.INC"
```

(调用数据文件的数据)

```
DESC INCOME /STAT=1,6
```

```
/OPT=5
```

(计算变量INCOME的均值、方差)

```
RECO INCOME (...)
```

(对变量INCOME进行分组)

```
FREQ VAR=INCOME/FORM=NEWP
```

```
HIST,NORM
```

(对变量INCOME作频率表、画直方图,以了解其分布)

倘若有若干个变量都需进行这些分析时,只需

在变量INCOME的地方,再增加需进行分析的变量名即可。

2.在统计分析中,除了对单变量进行分析外,还需要进行双变量分析时,SPSS统计软件包可以提供列联、回归、相关及独立性 X^2 检验等统计分析。例如,若对以上这组数据,列出收入与性别的列联表,并检验收入与性别是否独立。只需在以上的数据和分组基础上,再加上一条命令,即:

```
CROS SEX BY INCOME/STAT=1,2,
3,4 /OPT=1,5.
```

(列出性别与收入的列联表,并进行独立性 X^2 检验)

若需要对收入(INCOME)与若干个变量交叉分析时,只需在性别(SEX)后面再加上其他需要与变量INCOME交叉的变量名,便可列出与收入(INCOME)交叉的若干个列联表和 X^2 检验。

3.经济、社会是一个复杂系统,其变迁常常是多种因素共同作用的结果。因此,在经济、社会统计中,我们常常做的是多变量统计分析,而在这方面,SPSS统计软件包便体现出独到之处。在多变量分析时,常用的有因素分析、判别分析及多元回归等。例如我们需要了解收入(INCOME)与年龄(AGE)、受教育年限(EDUCATION)和职业(PROFESSION)的统计关系时,可用:

```
REG VAR=AGE INCOME EDVATION
PROFESSION
```

```
/STA=ALL /DEP=INCOME
```

```
/MET=STEP /RES=HIST.
```

这条回归命令是以收入作为因变量,年龄、受教育年限、职业为自变量,采取逐步回归进入方程的方式,并对回归剩余作直方图。由此便可以得出一个以收入为因变量、年龄、受教育年限、职业为自变量的回归方程。

即: $Y = ax_1 + bx_2 + cx_3$

同时还可进行一些可行性检验,如F检验、T检验等等。

由于篇幅有限,在这里我们不可能一一列举所有的统计命令。但通过以上事例,我们可以看出社会科学统计软件包在经济、社会统计中的作用,它为社会科学工作者进行量化分析提供了有力武器。

注释:

①《列宁全集》第16卷,人民出版社,1959年版,第431页。

(责任编辑 徐云鹏)