

# 从一处方法错误说起

——小议定序变量的统计分析

游 正 林

《社会学研究》1995年第2期刊登了《外商投资企业职工对劳动条件的评价》一文(以下简称《外》文)。《外》文以职工对劳动时间、劳动强度、劳动场所安全和劳保福利的“满意程度”来评价外商投资企业的劳动条件。在统计分析“满意程度”时,《外》文是这样处理的:“先将满意程度分为5个等级,很满意得2分,满意得1分,一般得0分,不满意得-1分,很不满意得-2分。如果以 $m_i$ ( $i=1,2,3,4,5$ )表示从很不满意顺序到很满意的得分,则 $m_i=i-3$ 。我们再以 $Q_i$ 表示选择某一 $m_i$ 的人数之百分比,则定义满意度 $M = \sum_{i=1}^5 M_i \cdot Q_i$ 。实际上, $M$ 只是以百分比 $Q_i$ 为权重的关于 $m_i$ 的一个加权平均值,我们用它来反映职工对某一劳动条件的满意程度。当 $M=0$ 时,表示处于一般水平, $M>0$ 表示职工对此劳动条件倾向于满意, $M<0$ 则表示职工对此劳动条件倾向于不满意。”之后,进而根据 $M$ 值的大小来比较不同类型企业职工之间的“满意度”的差别。如根据独资、合资、合作企业职工对劳动时间长短的“满意度( $M$ )”分别为-0.111、-0.041和-0.362,得出合作企业职工的满意度比独资企业低,而独资企业又比合资企业低的结论。

这种看上去很“量化”的统计分析,实际上它的统计分析方法是错误的。笔者在此想以《外》文为例,谈谈定序变量的统计分析问题。

在社会研究中,所要测量的变量可分为定类变量、定序变量、定距变量和定比变量四种。定类变量是测量层次中最低的一种,它只能将调查对象分类,如将性别分为男性和女性。定序变量比定类变量高一个层次,它可以按照某种逻辑顺序将调查对象排列出高低、强弱或大小,确定其等级或次序。如《外》文关于劳动条件的评价就是一个定序变量,依满意程度之强弱依次分为“很满意”、“满意”、“一般”、“不满意”和“很不满意”五个等级。定距变量不仅能将变量区分类别和等级,而且可以确定变量之间的数量差别和间隔距离,如我们可以说温度40度比20度高20度,但定距变量没有绝对的零点,因此,其值只能做加减而不能做剩除等数学运算,如我们不能说40度比20度的气温热一倍。只有定比变量所测得的数据既能作加减,又能作剩除运算。如张三的年龄是40岁,李四的年龄是20岁,便既可说张三比李四大20岁,也可以说张三的年龄是李四的两倍,还可以说张三与李四的平均年龄是30岁。

定序变量的数学特性虽然比定类变量高一个层次,即它不仅能区别异同,而且能确定其高低、强弱、大小等,但它并不能确定变量之间的数量差别和间隔距离。如我们虽然知道“很满意”比“满意”的满意度高,但我们并不能确定到底高多少,也不能确定“很满意”与“满意”之间的差距是否与“满意”与“一般”之间的差距(或“一般”与“不满意”、“不满意”与“很不满意”之间

的差距)相等。因此,对定序变量并不能做加、减、剩除等数学运算。打个比方说,我们并不能认为张三的“很满意”的满意度是李四的“一般”的两倍,也不能把张三的“很满意”与李四的“一般”相加,然后得出张三与李四的“平均满意度”是“满意”的结论。

在问卷调查中,经常要统计分析定序变量。为了计算机处理的方便,常给定序变量“打分”,就如《外》文所做的那样,依次给“很满意”、“满意”、“一般”、“不满意”、“很不满意”打“2”、“1”、“0”、“-1”、“-2”分。但这样做,并不意味着就把定序变量“升格”为定距或定比变量了。在这里“2、1、0、-1、-2”仍然只是一个具有顺序意义的识别标志,跟用“5、4、3、2、1”或“A、B、C、D、E”或别的具有顺序意义的符号来代表是一样的,它并不因此而反映事物本身的数量状况,所以,仍不能对2、1、0、-1、-2 这些数字作加减乘除等数学运算。如我们不能因此就认为一个“很满意”(2分)等于两个“满意”(1分)之和,“很满意”与“满意”之间的差距(2分-1分)就与“满意”与“一般”之间的差距(1分-0分)相等,更不能象《外》文那样用  $M = \sum_{i=1}^5 M_i Q_i$  来计算“满意度”,同样,也不能用 M 值的大小来判定“满意度”的强弱及其“倾向”。

对定序变量,到底如何统计分析?

如果是单变量分析,一般只需研究它有哪几种可能变动的情况,每种情况有多少人次(频次)或比例(相对频次)就行了。如《外》文中关于职工对劳动时间的评价,只需统计得出多少人及占总数百分之几的人分别回答“很满意”、“满意”、“一般”、“不满意”和“很不满意”就可以了,没有必要再画蛇添足地给它“打分”,然后再做一番“深入”的统计分析。

如果是双变量分析,则除了单变量分析外,还要研究两个变量之间是否存在关系,再在变量间存在关系的条件下,讨论变量间关系的密切程度问题。

《外》文旨在分析研究不同类型企业(如独资、合资和合资或港资、台资、日资、美资和欧资等)职工对劳动条件的评价是否相同。其实,这是一个定类变量(企业类型)与定序变量(评价即满意程度)的关系分析问题。如果要检验二者之间是否存在关系,即确定不同类型企业职工对劳动条件的评价是否有显著性差异,我们应该用非参数检验方法进行检验,而不是象《外》文那样错误地算出一个 M 值,然后,再根据 M 值的大小轻率地得出不同类型企业职工之间的满意度如何如何不同的结论。至于如何进行非参数检验、如何研究双变量之间的关系的密切程度问题,有关社会统计学已有详细介绍,无须在此赘述。

作者单位:中国工运学院工会学系  
责任编辑:范广伟