

## 利用Logistic回归分析研究影响应届毕业生就业的因素

2008-7-7

### 一、问题的提出

应届毕业生就业问题关系到千家万户,是一个重要的民生问题。应届毕业生的就业率也是衡量整个社会就业形势以及经济状况的一个重要参考指标。但是我国现行的应届毕业生就业情况统计是由各类学校各自完成的,而由于种种原因其统计并不完整、准确。为了较准确地反映浦东新区应届毕业生的就业情况,浦东新区从2004年起开展了每年一次的应届毕业生就业情况抽样调查。对常住在浦东的当年应届毕业生(指全日制中专/职校/技校、大专/高职、大学本科、各类研究生选择就业的应届毕业生)进行抽样调查,根据调查结果推算浦东新区应届毕业生的当年就业率。

应届毕业生就业情况抽样调查的结果,不仅可用来推算应届毕业生就业率,而且对应届毕业生的基本背景情况、择业心态等有较全面的反映。因此调查数据可供进一步挖掘,以对各种因素与毕业生就业的关系进行研究,从而可为如何进一步改善应届毕业生就业情况提供参考。

虽然利用一般的数据分析手段(如数据交叉表等),可以直观地看出各种因素对就业的影响,但不能深入反映各种因素与就业的数学关系,也就是说不能定量地反映各种因素对就业的影响程度,不能比较各因素对就业影响的大小。

这就要求引入一定的统计学方法。如果把应届毕业生的背景情况、择业心态等因素看作自变量,把就业情况看作应变量,那么很显然,可以应用回归的方法,得出各种因素与就业情况的数学关系。

但是这个数学关系不能应用一般的回归方法得出,因为作为应变量的就业情况不是一个定距变量,而是一个定性变量。也就是说就业情况的可能性只有两个:即就业或未就业。如果把就业定义为 $y=1$ ,未就业即为 $y=0$ ,那么, $y$ 只能有1或0这两个取值。面对这样的情况,就必须引入特殊的回归方法,即Logistic回归分析。

### 二、实证分析

#### 1、Logistic回归分析简介

Logistic回归分析是专门针对应变量为定性变量的统计分析方法。而根据应变量取值的多少,又可分为适用于应变量只有两个取值的Binary Logistic回归分析和适用于应变量有多个取值的Multinomial Logistic回归分析。由于本文研究的应届毕业生就业情况应变量只有两个取值,所以适用Binary Logistic方法。下文中的Logistic回归分析都是指Binary Logistic。通过Logistic回归分析可以得出自变量与取值为0或1的应变量之间的回归方程。

#### 2、对应届毕业生就业情况的Logistic回归分析

本文运用统计分析软件SPSS的Logistic回归分析功能对2007年浦东新区应届毕业生抽样调查的数据进行分析,以研究对应届毕业生就业的影响因素。

将2007年浦东应届毕业生就业情况抽样调查的数据输入SPSS软件。设应届毕业生就业情况为应变量y, y=1为就业, y=0为未就业。设自变量x1表示应届毕业生的性别 (x1=1为男性, x1=2为女性); x2表示应届毕业生的学历 (x2=1为中专/技校/职校, x2=2为大专/高职, x2=3为本科, x2=4为研究生); x3表示应届毕业生是否有专业资格证书 (x3=1为有资格证书, x3=2为无资格证书); x4表示应届毕业生期望月收入 (单位: 元); x5表示应届毕业生期望工作地点 (x5=1为期望在浦东工作、x5=2为期望在浦西工作、x5=3为期望在其他地区工作); x6表示应届毕业生最主要找工作方式 (x6=1为到人才市场, x6=2为通过网上的招聘信息, x6=3为通过学校的信息栏, x6=4为单位到学校招聘, x6=5为亲戚朋友介绍, x6=6为学校老师推荐联系, x6=7为考公务员, x6=8为其他)。然后进行logistic回归分析, 得到以下结果:

### (1) 模型检验

表1 Omnibus Tests of Model Coefficients (模型检验)



模型检验结果 $\chi^2=924.654$ ,  $P=0.000<0.001$ , 表明模型有统计意义。

### (2) 拟合效果

表2 Classification Table(a) (拟合效果)



a The cut value is .500

对就业情况进行拟合, 以预测概率0.5为判别是否就业分界点。对于y=0(未就业)有51.3%的准确性, 对于y=1(就业)有79.4%的准确性。对于所有个案有67.7%的准确性。

### (3) 回归方程

表3 Variables in the Equation (方程中的变量)



其中B为参数估计, S.E. 为其标准误, Wald为Wald $\chi^2$ 检验统计量, Sig. 为其相伴概率值, Exp(B) 为其条件不变时对应变量每增加一个单位后就业概率与原就业概率相比的倍数近似值。

由此得到相应的logistic回归方程:

$$\ln(P/(1-P)) = -2.99599 + 0.35664x_1 + 0.10063x_2 - 0.25496x_3 + 0.00036x_4 + 0.39861x_5 + 0.37971x_6$$

其中P为y=1的概率

## 三、若干结论

### 1、找工作方式在考察诸因素中对应届毕业生就业影响最大

在logistic回归分析中, 要比较自变量在回归方程中的重要性, 一般直接比较Wald统计量的大小, 自变量的Wald统计量越大, 显著性就高, 也就更重要。

从表3可知, 在应届毕业生logistic回归方程中各自变量Wald统计量大小比较情况为:  $x_6 > x_4 > x_1 > x_5 > x_3 > x_2$ 。也就是说在考察诸因素中, 对应届毕业生就业影响程度从大到小分别为: 找工作方式、期望月收入、性别、希望单位地点、是否有专业资格证书、学历。找工作方式对应届毕业生就业影响

最大，而学历影响最小。

## 2、应届毕业生就业中的主观选择比自身客观条件更重要

在考察诸因素中，可以分为两类。一类是应届毕业生就业中的主观选择，包括找工作方式、期望月收入、希望单位地点；一类是应届毕业生自身的客观条件，包括性别、是否有专业资格证书、学历。

由表3可知，主观选择中的“找工作方式”、“期望月收入”的Wald统计量远大于其他自变量，也就是说找工作方式、期望月收入对应届毕业生就业影响程度最大。而“希望单位地点”的Wald统计量也大于“是否有专业资格证书”和“学历”，也就是说比后两者对应届毕业生就业影响程度更大。综合起来看，应届毕业生的主观选择比其自身客观条件对其就业影响更大。

这表明应届毕业生要能实现就业，最重要的是要选择合适的找工作方式、要有符合实际的心理期望，而自身的客观条件不是影响就业的最重要因素。自身条件较差的应届毕业生可以通过正确的主观选择来弥补自身的劣势。因此要提高应届毕业生的就业率，政府和社会需更加关注应届毕业生的择业心态和技巧。

## 3、各因素对应届毕业生就业的具体影响

### (1) 找工作方式

由表3可知，“找工作方式”的Wald统计量最大，因此对应届毕业生就业的影响程度最大。

从2007年浦东应届毕业生就业情况抽样调查数据中可知，应届毕业生中未找到工作者的求职方式与已找到工作者有明显差异。未找到工作者求职方式比较单一，超过半数（52.8%）选择上网浏览招聘信息为最主要的求职方式，选择到人才市场求职的占21.1%；而已找到工作者中各种求职方式的分布则比较均匀，最主要有求职方式选择上网浏览招聘信息的比前者低18.4个百分点、选择到人才市场求职的低13.1个百分点，其余大多数求职方式的比例则都比前者明显要高。这表明“找工作方式”单一对应届毕业生就业有不利影响。

### (2) 期望月收入

由表3可知，“期望月收入”Wald统计量大小居第二位，因此对应届毕业生就业的影响程度仅次于找工作方式。

从2007年浦东应届毕业生就业情况抽样调查数据中可知，新区应届毕业生期望收入普遍较高，不符合实际情况。已确定工作单位的应届毕业生平均期望月收入比首份工作平均月收入要高66.9%。未找到工作的应届毕业生期望收入也偏高，平均期望月收入为2538元，是已确定工作单位应届毕业生首份工作平均月收入的1.3倍。这表明应届毕业生对收入期望过高，与实际薪酬情况不符，会对就业产生不利影响。

### (3) 性别

由表3可知，“性别”Wald统计量大小居第三位，因此对应届毕业生就业的影响程度在考察诸因素中居第三位。

由表3可知，“性别”的Exp(B)值为1.429，也就是说其他条件不变，当“性别”取值由1变为2时，应届毕业生的就业概率是原就业概率的1.429倍，即其他条件不变，应届毕业生中女性的就业概率是男性的1.429倍，女性比男性更易实现就业。

从2007年浦东应届毕业生就业情况抽样调查数据中可知，女性就业率为76.8%，比男性高6.8个百分点，也验证了这个结果。

### (4) 希望单位地点

由表3可知,“希望单位地点”Wald统计量大小居第四位,因此对应届毕业生就业的影响程度在考察诸因素中居第四位。

由表3可知,“希望单位地点”的Exp(B)值为1.49,也就是说其他条件不变,当“希望单位地点”取值增加1时,应届毕业生的就业概率是原就业概率的1.49倍。即其他条件不变,应届毕业生中“希望单位地点”在浦西的就业概率是浦东的1.49倍,“希望单位地点”在浦西更易实现就业。

从2007年浦东应届毕业生就业情况抽样调查数据中可知,应届毕业生希望在浦东就业的占85.2%,希望在浦西就业的占14.7%,希望在其它地区就业的占0.1%。在已确定工作单位的应届毕业生中,有61.2%在浦东就业,在浦西就业的占38.0%,另有0.8%在其他地区就业。这表明“希望单位地点”在浦西和其他地区的更易就业,应届毕业生应打破狭隘的地域观念,到更广阔的就业市场上去闯荡。

#### (5) 有无专业资格证书

由表3可知,“有无专业资格证书”Wald统计量大小居第五位,因此对应届毕业生就业的影响程度在考察诸因素中居第五位。

由表3可知,“有无专业资格证书”的Exp(B)值为0.775,也就是说其他条件不变,当“有无专业资格证书”取值由1变为2时,应届毕业生的就业概率是原就业概率的0.775倍,即其他条件不变,应届毕业生中无专业资格证书的就业概率是有专业资格证书的0.775倍,有专业资格证书比无专业资格证书更易实现就业。

从2007年浦东应届毕业生就业情况抽样调查数据中可知,毕业生中拥有专业资格证书的就业率为75.7%,比没有专业资格证书者的就业率高18个百分点,也验证了这个结果。

#### (6) 学历

由表3可知,“学历”Wald统计量最小,因此对应届毕业生就业的影响程度在考察诸因素中最小。

由表3可知,“学历”的Exp(B)值为1.106,也就是说其他条件不变,当“学历”取值增加1时,应届毕业生的就业概率是原就业概率的1.106倍。因此学历更高的应届毕业生更易实现就业。

从2007年浦东应届毕业生就业情况抽样调查数据中可知,三校生、大专生、本科生和研究生就业率分别为63.5%、68.1%、83.6%和87.1%,学历越高就业率就越高,也验证了这个结论。

### 四、推广应用的建议

上述分析表明引入logistic回归方法对比较各种因素对应届毕业生就业影响的大小、考察各种因素对应届毕业生就业的具体影响都有着较好的效果。

因此,为考察应届毕业生就业的影响因素及其变化,今后每年的应届毕业生抽样调查都应把logistic回归纳入数据分析计划,并对历年的分析结果进行对比。这将有可能为有关部门改善应届毕业生就业竞争力、提高就业水平提供决策参考。

此次应用2007年浦东应届毕业生就业情况抽样调查数据求出的logistic回归方程,其拟合效果为67.7%的个案预测准确性,应该说还不是很高。这表明,还有重要的影响因素没有包括在回归方程中。今后的应届毕业生就业情况抽样调查应进一步改进问卷设计,以包括更多对就业有影响的因素,从而使logistic回归方程的拟合效果能够进一步提高。这样就能够更加准确和完整地考察应届毕业生就业的影响因素。

在社会统计中,像应届毕业生就业情况调查这样,其关键统计指标是定性变量的情况有很多。对这些统计数据,都可以应用logistic回归分析,来考察各种因素对统计指标的影响。而为使logistic回归方程的拟合效果更好,在此类统计调查中,都应注意尽可能多地包括对各种相关情况的统计调查,这样才能

准确、完整地考察统计指标的各种影响因素。

作者：周示莹 单位：浦东新区社会经济调查中心