

固定效应模型与混合效应模型在测量误差估计中的应用^{*}

姚树祥¹ 巫秀美² 倪宗瓚² 乔友林³ 姜 勇¹

提 要 探讨疾病危险因素暴露剂量的测量误差评价方式,以云南锡业公司所属锡矿井下作业环境有害物质监测资料为例,分别进行固定效应和混合效应的方差成分模型拟合,估计该资料的测量误差大小。固定效应的方差分析结果显示氡、氡子体和粉尘的测量信度均接近1,混合效应模型也显示资料的测量质量较佳(信度系数估计值均>0.8)。提示固定效应模型和混合效应模型均可应用于测量质量的评价,应根据资料的特性选择正确的统计模型,以得到较为准确的测量误差估计值。
关键词 测量误差 固定效应模型 混合效应模型 方差成分模型

The Application of Fix-effect and Mixed-effect Models to the Measurement Error Estimation Yao Shuxiang, et al. Labor Protection Institute of Yunnan Tin Cooperation(Gejiu 661000)

To investigate on statistical methods for the assessment of exposure measurement error. Concentrations of radon, radon daughters and dust in underground mines measured by the Labor Protection Institute of Yunnan Tin Cooperation as an example, a fixed-effect model and a mixed-effect model were fitted to calculate the reliability coefficients. All of the estimated reliability coefficients were close to 1, which showed that the data were under well controlled. It indicates that both fixed-effect and mixed-effect models could work as useful methods for the assessment of measurement error if applying with caution.
Key words Measurement errors Fixed-effect model Mixed-effect model Variance component model

职业性暴露与疾病间关系的正确评价,是基于观察对象接受暴露剂量的准确估计。工人接触职业性暴露的剂量大小,往往经由环境中有害物质浓度的测量值换算得到。任何测量皆不可避免地存在着误差,因此正确估计测量误差的大小是保证数据质量的重要手段之一。测量值、真值与测量误差间的关系可表达为测量误差模型,即 $X_s = X_t + \epsilon_{s,t}$,其中, X_s 为实测值,有 $X_s \sim (0, \sigma_s^2)$; X_t 为真值,有 $X_t \sim (0, \sigma_t^2)$; $\epsilon_{s,t}$ 为测量误差,具有独立于 X_t 及 $\epsilon_{s,t} \sim (0, \sigma_{\epsilon}^2)$ 的性质。测量误差的大小通常用信度系数($0 < R < 1$)表示^[1]。在职业暴露剂量的测量中,通常以方差分析方法将重复测量值的总变异分解为测量点间的变异,以及同一测量点的重复测量值间的变异(即误差项),然后按公式 $R = 1 - 1/F$ (其中 F 即为方差分析的 F 统计量)计算测量信度系数的估计值。

本文拟对云南锡业公司劳动防护研究所(云锡劳研所)在1989年到1993年的生产环境有害物质监测资料,进行固定效应与混合效应的方差分析以评价数据的质量,并探讨此二统计

模型在矿工职业性暴露测量误差估计中的应用。现报道如下。

1 资料来源与分析方法

1.1 资料来源 云锡劳研所每年在该公司所属各矿坑井下的作业环境中随机选取测定点,对职业性暴露浓度进行数次重复测量,本研究采用1989~1993年于23个单位中的随机测量点上进行的氡、氡子体与粉尘浓度的测量资料。原始测量值显示,大多数矿坑井下的暴露剂量测量值服从对数正态分布,因此采用了对数变换对资料进行预处理。

此外,该公司自1975年起先后在各所属矿坑建立了17条机械式通风管道,并持续研究改善井下通风设施,有效地逐年降低了作业环境氡气、氡子体→潜能值和粉尘的浓度^[2]。为避免不同测量年份间作业环境中有害物质浓度的变动影响测量误差的准确估计,本研究是在控制了测量年份的影响后再探讨其测量信度。

1.2 测量误差的估计

1.2.1 固定效应模型 按资料收集的方式,本文所使用的数据属于嵌套设计资料,即各单位的测量点嵌套在年份之中。此时,测量值的总变异可分解为不同年份测量值间的变异、单位间的变异与单位的变异。按模型拟合结果,可以得到信度系数的无偏估计值: $R = (MS_{\text{单位间}} - MS_{\text{单位内}}) / AMS_{\text{单位内}} = 1 - 1/F^{[3]}$ 。

1.2.2 混合效应模型 上述模型假定测量点间的变异是一固定效应,但在实际测量时测量点是随机选择的。若在方差成分模型中拟合随机效应项,则可以提供更接近真实信度系数的估计值,为此将测量误差模型进一步扩展为:

$$X_{(ij)k} = \mu_x + \alpha_i + \gamma_{(ij)} + \epsilon_{(ij)k} \quad (i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n_{(ij)}) \quad (1)$$

其中, $\gamma_{(ij)}$ 与 $\epsilon_{(ij)k}$ 相互独立,且有 $\gamma_{(ij)} \sim N(0, \sigma_b^2)$, $\epsilon_{(ij)k} \sim N(0, \sigma_w^2)$, $E(X_{ijk}) = \mu_x + \alpha_i$,对于 i, j 及 $k \neq k'$,有 $\text{cov}(X_{(ij)k}, X_{(ij)k'}) = \sigma_b^2$, μ_x 是暴露剂量的跨组固定均值(即 $X_{(ij)k}$ 的均值), α_i 是不同年份间的固定效应, $\gamma_{(ij)}$ 是随机效应项,它反映了每个单位的暴露真值减去总体均值($\mu_x + \alpha_i$)的大小, $\epsilon_{(ij)k}$ 则是测量误差。

一般利用极大似然法迭代估计出混合效应模型中的参数,再按式(2)计算信度系数的估计值^[4]。

$$R = \sigma_b^2 / (\sigma_b^2 + \sigma_w^2 / n_{(ij)}) \quad (2)$$

2 结果

在SAS软件中对本资料拟合嵌套设计的固定效应(PROC NESTED 程序)与混合效应(PROC MIXED 程序)的方差分析,得到测量信度估计值如表1与表2所示。

表1 固定效应模型所得氡、氡子体、粉尘的测量信度

	氡	氡子体	粉尘
MS年度间	96.3371	83.0434	174.4717
MS单位间	52.4227	45.2695	1351.7365
MS单位内	3.0101	2.6962	16.4868
F(单位间)	17.42	16.79	81.99
R	0.9426	0.9404	0.9878

^{*} 本课题为国家自然科学基金资助项目
1. 云南锡业公司劳动防护研究所(个旧 661000)
2. 华西医科大学卫生统计学教研室
3. 中国预防科学院肿瘤研究所

表 2 混合效应模型分析结果

测定项目	σ^2_b	σ^2_w	$F_{(年份间)}(P)$	R
氡	1. 44595	3. 13168	4. 69(0. 0009)	0. 8218~0. 9921
氡子体	1. 27071	2. 75529	8. 34(0. 0001)	0. 8218~0. 9921
粉尘	25. 30654	17. 45811	6. 45(0. 0001)	0. 9355~0. 9975

由于云锡劳研所采用非均衡设计对作业环境中的有害物质进行浓度测量,重复测量数 $n_{(ij)}$ 在各年份及各组间不等,因此各测量点的测量信度估计值不同。从表 1 与表 2 中信度系数估计值均大于 0. 8 可知,云锡环境监测数据的整体测量质量较好。

3 讨论

环境中有害物质浓度的测量质量,直接影响工人接受职业性暴露剂量的估计,并可能进而影响了对资料所作统计分析及其所得结果的可靠性。故在进行数据的统计分析以及对研究结果进行讨论前,应先对资料的质量控制进行评价。

方差成分模型是得到职业性暴露测量信度的常用方法,该法的基本原理为对单次调查研究中的数次重复测量值进行相应的方差分析,得到模型参数的估计值和信度系数的估计值^[5]。出于不需另外进行确认研究(validation study)或信度研究,且资料分析方法较为简便,故较为实用。

本研究资料中井下作业环境中粉尘、氡及其子体的含量确实因通风设施的加强而逐年下降,在估计测量误差的大小时不应忽视年份不同所产生的影响,故本文要用嵌套设计的方差分析进行固定效应与混合效应的模型拟合。

固定效应的方差分析是方差成分模型中形式最为简单而又常用的一种,其优点是模型容易理解且大部分统计软件有成熟的分析模块支持。然而,实际的测量数据往往是对随机选择的测量点进行测量,此时不仅误差项为随机变量,分组变量也应视为随机变量处理。

信度系数估计值与样本含量的大小有关,例如,云锡劳研所在 1989~1993 年期间对放射性氡的测量中,重测次数最多的是“老厂锡矿”的某测量点进行了 273 次的重复测量,最少的

是“黄茅山采选厂”的某测量点,仅测量了 10 次,利用式(3)算得其相应的信度系数分别为 0. 9921 与 0. 8218。此结果证实了心理测量学研究中问卷的长度与信度的大小有密切关系的理论,即具有同样稳定度(或内部一致性)的两份问卷,问题多者表现为测量的信度系数估计值较高。在职业性暴露浓度的数据则表现为:对于同样稳定的测量值,重复测量次数越多其信度越高。

测量误差模型不仅可评价某些特殊数据收集方法的整体质量,如空气中粉尘浓度的测量,还可用于评价调查的整体质量。对于测量误差较大的资料,一般应进一步进行离群值诊断及处理,或对原始数据拟合统计模型所得的参数进行校正。国外对于测量误差的估计与相应的校正方法已有广泛的探讨与文献报道^[6,7],但此问题在国内尚未引起足够的认识与重视。其中,职业性暴露浓度的测量值通常服从对数正态分布,故测量误差的相乘模型中的信度系数估计与回归参数的校正方法,应引起更大的重视与更深入的研究。

此外,在测定职业性暴露的浓度时通常并未采用均衡设计,亦即各组的重复测量数往往不等,此时采用不同的回归系数校正法得到的结果会有所差异。因此,如何评价及选择恰当的校正法,还需进一步的研究与尝试。

参 考 文 献

1. Fuller W A. Measurement Error Models. New York: John Wiley & Sons, Inc., 1987

2. 云南锡业公司. 云南锡业公司矿工肺癌防治成果报告(1975~1989 年资料选编) 1990; 12

3. Acebo C, et al. Sleep and apnea in the elderly: Reliability and validity of 24-hour recordings in the home. Sleep 1991; 14: 56

4. Lyles R H, et al. A detailed evaluation of adjustment methods for multiplicative measurement error in linear regression with applications in occupational epidemiology. Biometrics 1997; 53: 1008

5. Kim M Y, Zeleniuch-Jacquotte A. Correcting for measurement error in the analysis of case-control data with repeated measurements of exposure. American Journal of Epidemiology 1997; 145: 1003

6. Lessler J T, Kalsbeek W D. Non-Sampling Error in Surveys. New York: John Wiley & Sons, Inc 1992

7. Longford N T. Random Coefficient Models. Oxford UK: Oxford University Press. 1993

(1999-07-12 收稿 董建群编辑 李溪莹校对)

90 年代初期广西人口自杀死亡
流行病学调查

广西壮族自治区肿瘤防治研究所(南宁 530021)
张振权 余家华 黄天壬 徐庆芬 黄昭东 何振芳

根据卫生部的统一部署,广西开展了 90 年代初期(1990 年 1 月~1992 年 12 月)人口死亡原因的抽样调查研究。调查样本包括 10 个县、市,人口数共计 446 万余,略多于同期全广西人口数的 1/10。调查的死者数合计 80415 人,其中有 1991 例明确归因为自杀致死。根据调查资料计算,自杀所致的粗死亡率为 14. 87/10 万,以 1964 年中国人口标化,则为 12. 37/10 万,世界人口标化,则为 15. 57/10 万。自杀死亡占同期人口死亡总数的 2. 47%,占意外死亡数(9682 例)的 20. 56%,是意外死亡原因中之首位。自杀死亡者的中位年龄为 38. 68 岁;自杀所致标化减寿率为 3. 35‰;平均每一名自杀死亡者减寿 27. 92 岁。广西居民一生(按 0~74 岁计算)死于自杀的危险为

1. 47%。男女两性自杀死亡的危险相当接近,总而言之,女性危险略高。自杀死亡的危险随年龄的增长而不断升高,上升速度以 20 岁~年龄组最令人注意。低年龄组也有自杀死亡现象,自杀死亡者最低年龄为 6 岁。不同婚姻状况人群有不同的自杀死亡率。有配偶者的自杀死亡率低于其他情况,差异有非常显著性($\chi^2=75. 98$ $df=3$, $P<0. 001$)。不同受教育程度对于自杀死亡率也有不同影响,受教育程度越高,自杀死亡率越低,差异也有非常显著性($\chi^2=556. 23$ $df=3$, $P<0. 001$)。自杀的主要方式有:服毒 1353 例(67. 96%),其中服农药 1 134 例;自缢 418 例(20. 99%)。以上两类方式已占全部自杀死亡数的 88. 95%。其他的自杀方式有:投水 91 例(4. 57%),跳楼跳崖 23 例(1. 16%),触电 21 例(1. 05%),利用火器或爆炸物 16 例(0. 80%),利器自裁 11 例(0. 52%),自焚 6 例(0. 30%)。此外,有 18 例自杀者采用的方式不属于上述各类自杀手段,有 27 例自杀者的原始资料未注明自杀方式。另有 7 例死者,原始资料所写死因为病伤情况,但同时已注明其病伤情况实由自杀行为导致。据此,资料分析核定其根本死因为自杀晚期效应。

(1999-04-05 收稿 2000-01-30 修回 董建群编辑 李溪莹校对)