

职业紧张研究中的数据处理与统计方法

余善法

摘要: 职业紧张已成为职业卫生学科研究的热点问题,但目前国内研究者在研究数据处理和统计分析方法应用方面仍存在较多的错误和不足,影响正确、科学地呈现研究结果,难以如实反映研究水平。针对此现状,本文着重介绍问卷条目应答不全资料的处理及依据评分确定组别的界值等常用的数据处理方法;常模应用应注意的问题;研究结果数据的分布特征;以及相关分析、配对资料分析、协方差分析、分层多元回归分析、logistic 回归分析和结构方程模型等常用统计分析方法使用过程中的关键环节。

关键词: 职业紧张; 数据处理; 统计方法

Data Processing and Statistical Methods in Occupational Stress Research YU Shan-fa (Henan Provincial Institute for Occupational Health, Henan 450052, China) · The author declares he has no actual or potential competing financial interests.

Abstract: Occupational stress has been a hot spot in occupational health research. However, the mistakes and defects in research data processing and statistical analysis by domestic researchers have adversely affected the results presented in a correct and scientific manner, and derogated study performance. With regard to the above mentioned aspects, this article introduced common data processing methods including: handling missing variables in questionnaire data and marginal values of category determined by scoring; considerations in norm application; data distribution characteristics; and the key points of statistical methods including correlation analysis, matched data analysis, analysis of covariance, hierarchical multiple linear regression analysis, logistic regression analysis, and structural equation modeling.

Key Words: occupational stress; data processing; statistical analysis methods

近年来,国内的研究者越来越关注职业紧张领域的研究,在国内外期刊上发表的论文逐年增多,但笔者发现,在职业紧张相关的调查研究中,对于数据处理和统计方法应用方面常存在较多错误和不足,致使研究结果难以正确、科学、充分地呈现。笔者根据多年来的研究和阅读文献的体会,对有关职业紧张研究的数据处理和统计方法在应用中的常见问题加以梳理,本文予以报道,供同行参考。

1 问卷条目应答不全资料的处理

为了调查到真实的情况,职业紧张问卷调查一般采用匿名调查方式,存在调查对象的依从性和配合问题,被调查对象对条目的含义不理解或理解片面等都

会影响对调查问卷条目的应答,致使条目的应答不全。如果将部分应答不全的资料全部删去,则会丢失大量的样本,产生选择偏倚。目前国际上对应答不全资料主要采取下列方式处理,当量表或子量表由4个条目组成时,一个条目的无应答是可以接受的,缺失的数据用该条目有应答数据的均值替代^[1];对于工作紧张模式问卷,多数作者采取当子量表有1~2个条目无应答时,缺失的数据用该条目有应答数据的均值替代^[2-3];对于付出-回报失衡模式问卷,建议子量表无应答条目在50%及以下,缺失的数据用该条目有应答数据的均值替代,超过50%无应答条目,则这个个体的数据不纳入统计处理。余善法等^[4-5]在研究中采取的处理方法是如果一个多条目(10个条目以上)量表或子量表的无应答条目在3个及以下,则缺失的数据用该条目有应答数据的均值替代,超过4个无应答条目,则这个个体的数据不纳入统计处理。对于条目数少于10的量表或子量表有1~2个条目无应答时,缺失的数据用该条目有应答数据的均值替代。

[基金项目] 河南省医学科技攻关计划重大项目(编号: 201001009);
河南省卫生科技领军人才基金(编号: 3022)

[作者简介] 余善法(1964—),男,博士,主任医师;研究方向:职业
心理与工效学;E-mail: yu-shanfa@163.com

[作者单位] 河南省职业病防治研究院,河南 450052

2 依据评分确定组别的界值

在职业紧张调查数据的分析时,经常要依据评分的高低对调查对象进行分组,目前有3种界值:(1)评分的均值;(2)评分的中位数;(3)三分位数或四分位数。这是导致研究结果无法比较的一个主要问题,因此在比较不同研究结果的异同时,应注意不同研究间评分界值是否相同。这3种界值无优劣之分,研究者可根据自己的研究目的、样本量等进行选择,如样本量大可考虑三分位数或四分位数,如样本量较小则考虑均值或中位数。近年来,对不同的变量进行组合分析逐渐增多,如依据工作紧张模式和付出-回报失衡模式主要因子评分分别将研究对象分为4个不同的组别,即:(1)低需求(付出)高控制(回报),(2)高需求(付出)高控制(回报),(3)低需求(付出)低控制(回报),(4)高需求(付出)低控制(回报)^[4-8]。此外,根据工作紧张模式和付出-回报失衡模式的理论,分别计算心理需求与控制水平的比值、躯体需求/(控制水平 $\times 0.5556$)值和付出/(回报 $\times 0.5454$)值,以衡量工作紧张和付出-回报失衡的程度(紧张水平),两者均以1为界值对研究对象进行分组^[4-5,8]。

3 常模应用需注意的问题

职业紧张研究涉及两个方面的常模,一个是职业紧张问卷常模,一个是心理健康问卷常模。关于职业紧张问卷常模,尽管国际上使用的职业紧张问卷有几十种,但目前的文献仅见少数问卷有科学合理的常模数据,如英国曼彻斯特大学 COOPER 等编著的职业紧张指标(OSI)的英国常模^[9],大多数问卷无常模数据,可能与研究常模需要有充分代表性的不同人口统计学特征分布的大样本而这一点往往很难做到有关。关于心理健康问卷的常模,存在的主要问题是一些常模数据调查的年代已距今二三十年,加之心理健康受社会经济等因素影响较大,已丧失可比性,故比较的结果也无科学性和实际意义,应慎用。即便是近年的常模,也应考虑常模制定的背景、有无可比性、比较结果的正确解释等问题。

4 研究结果数据的分布特征

职业紧张研究中有心理变量和生理变量。目前心理变量多视为正态分布,但 NAKATA 等^[10]在最近的一项研究中发现,抑郁症状评分不呈正态分布。生理

变量如体液指标,尤其是唾液、尿液指标多呈非正态分布。因此,在数据分析时,应首先对心理变量和生理变量进行正态性检验,将非正态数据转换成正态分布数据,以适应统计方法对数据分布的要求或使用非参数统计方法。NAKATA 等^[10]在分析工作满意感、协变量和免疫标志的关系时,对所有免疫标志、年龄、体质指数(BMI)及抑郁症状评分进行以10为底的对数转换,使数据更符合正态分布。余善法等^[11-12]在装配线工人职业紧张与唾液考的松、sIgA 和溶菌酶水平关系的研究中发现,唾液考的松、sIgA 和溶菌酶浓度呈正偏态分布,于是在数据分析前,即对原始数据进行常用对数转换。

5 相关分析方法的应用

在研究职业紧张的文献中,多使用相关分析来考察职业紧张因素、个体特征变量与紧张结局变量之间的关系。但在具体的使用过程中,应根据数据的分布情况选择正确的相关分析方法,如果数据呈正态分布,则使用 Pearson 相关分析;若数据呈非正态分布,则使用 Spearman 相关分析,或将原始数据转换成正态分布数据后,使用 Pearson 相关分析;若是等级资料,则使用等级相关分析。国外的文献多使用 Spearman 相关分析。对结果的解释应考虑相关系数的大小,若相关系数较小,即使有统计学意义,其实际的意义也是有限的,这时结论的推论应留有余地。

6 配对资料分析

为了消除或控制混杂因素的影响,有的研究需要对数据进行配对分析,其结果更有科学价值和说服力。但配对因素的选择要考虑样本量,样本量较大时可考虑多个配对因素,样本量较小时则配对的因素不宜过多,总的原则是保证配对后各组有足够的例数。余善法等^[13]在分析性别对职业紧张测试结果的影响时,考虑到年龄、文化程度和所从事的职业对研究结果可能存在影响,不同性别组间进行上述三因素配比,使得各组间构成差异无统计学意义,则结果提示,性别对某些职业紧张测试结果有影响。在一项火车司机职业紧张因素分析的研究中,作者以某铁路局1116名男性火车司机为研究对象,严格按照年龄(95%相等,其余相差 <0.5 岁)、文化程度和婚姻状态等同配对的原则选择其中客车和货车司机各241名进行分析,得出客车司机的紧张度大于货车司机的结论^[14]。

7 协方差分析

心理健康研究结果的变量受人口统计学特征影响较大,单因素分析时由于组间人口统计学特征变量构成不均衡,结果会受混杂因素的影响,可能会出现错误的结果,这时应考虑使用协方差分析。在一项职业紧张与抑郁症状关系的性别差异分析中,作者在进行资料的预分析时,发现年龄、工龄、文化程度和工种对职业紧张因素和抑郁症状评分存在影响,年龄和工龄效应一致,因此人口统计学特征组间差异分析采用协方差分析,以控制比较组间年龄、工龄、文化程度和工种构成的不均衡对结果的影响^[5]。

8 分层多元回归分析

分层多元回归分析(hierarchical multiple linear regression, HMLR)是职业紧张研究常用的统计方法,用于比较或确定不同因素的效应大小,分析交互作用^[8]。NAKATA等^[10]使用分层多元回归分析探讨免疫标志物与工作满意感和协变量的关系,第一步引入工作满意感、年龄、文化程度、吸烟、饮酒量、体育锻炼、睡眠、体质指数、药物使用等变量,第二步引入抑郁症状和工作中的人际冲突不良,第三步引入职业种类、公司类型和工作时间等职业因素。在这个分析中,药物使用、职业种类和公司类型为分类变量,其余因素为连续变量,应变量免疫标志物水平为连续变量。余善法等^[14]使用分层多元逐步回归分析探讨了职业紧张因素、工作心理控制源以及两者的乘积对紧张结局的效应,分析因素的独立和交互作用。第一步引入职业紧张因素,第二步引入工作心理控制源,第三步引入职业紧张因素与工作心理控制源的乘积。

9 logistic 回归分析

logistic 回归分析是职业紧张研究数据分析的另一常用多因素统计方法。国际上早年的文献和国内的作者通常先进行单因素 logistic 回归分析,筛选出有统计学意义的自变量,然后再将这些自变量引入方程进行多因素 logistic 回归分析。但近期国外的文献多直接进行多因素 logistic 回归分析。在 logistic 回归分析时,应注意对分类变量进行哑变量的转换,一般年龄、工龄等应作为连续变量处理,应变量为分类变量,如疾病的有无,评分的高低等。对结果进行解释时,应注意自变量作用的方向、OR 值的大小及可信区间和变量的统计学意义等,变量 OR 值 95% 可信区间畸宽提

示样本量不足、数据质量不高和变量数据分组不合理。余善法等^[4-5, 15]使用 logistic 回归分析对肌肉骨骼系统症状及其相关危险因素、职业紧张与下肢肌肉骨骼系统疾患的关系、职业紧张与抑郁症关系的性别差异进行探讨,得到了较好的结果。LEE 等^[16]使用该方法对上肢肌肉骨骼系统疾患的社会心理危险因素进行分析,也得到较好的结果。

10 结构方程模型的应用

目前在职业紧张研究中,多使用结构方程模型(structural equation modeling, SEM)来分析问卷的结构效度和职业紧张因素与紧张结局之间的关系^[17-18]。在对问卷结构效度进行分析时,SEM 具有如下优点:将测量误差从变量的变异量中分离出来,使得因子负荷具有较高的精确度;可根据理论假设,一个条目可以同时分属于不同因子;自由设定某些因子之间具有或不具有相关;可同时估计因子结构和因子关系。因此,采用 SEM 对问卷结构效度进行分析,具有更高的准确性和自由度,这种方法更能满足现实的应用条件^[18]。在实际的分析过程中,应使用不同的指数检验模型的拟合度,最常见的模型拟合指数有适合度指数,包括拟合指数(GFI)、调整拟合指数(AGFI)、正规拟合指数(NFI)、非正规拟合指数(NNFI)、增量拟合指数(IFI)和替代性指数,包括平均概似平方误根系数(RMSEA)、比较适合度指数(CFI)等几种^[19]。GFI 和 AGFI 一般需要大于 0.90 才可以视为理想的拟合度,NFI、NNFI 和 IFI 数值越大表示拟合度越佳,同时系数值需大于 0.90 也才可以视为理想的拟合度。RMSEA 数值越大代表模型越不理想,数值越小代表模型拟合度越理想,一般以 0.05 为良好拟合的门槛,以 0.08 为可接受的拟合门槛。CFI 一般以 0.95 为通用的门槛,同时在小样本的 SEM 分析中,用来评估模型拟合度十分稳定。在结果的解释时,应注意由于所得的结果受到研究设计的限制,一般只能进行变量的预测或解释,而不能推论因果关系;SEM 一个最主要的目的是检验研究者所提出的理论或概念架构是否具有实证的意义,本身无法创造知识或理论,需要研究者在整理前人的理论或知识的基础上,建构一套适当的概念模型,再使用 SEM 技术协助研究者完成模型的分析与讨论,不能过度依赖技术指标的数据和进行过度推论;SEM 分析的结果应仅就该研究所使用的样本与所检

验的变量关系来讨论, 分析的证据可以说明某一个因果概念是可能存在的, 但是不能据此排除其他模型的存在。

·作者声明本文无实际或潜在的利益冲突。

参考文献:

- [1] BERMIN P, THEORELL T. Demand-control-support among female and male managers in eight Swedish companies[J]. Stress Health, 2001 17(4): 231-243.
- [2] AHOLA K, HAKANEN J. Job strain, burnout, and depressive symptoms: A prospective study among dentists[J]. J Affect Disord, 2007, 104(1-3): 103-110.
- [3] CLAYS E, De BACQUER D, DELANGHE J. Associations Between Dimensions of Job Stress and Biomarkers of Inflammation and Infection[J]. J Occup Environ Med, 2005, 47(9): 878-883.
- [4] 余善法, 谷桂珍, 周文慧, 等. 职业应激与下肢肌肉骨骼系统疾患的关系[J]. 中华劳动卫生职业病杂志, 2011, 29(12): 882-886.
- [5] 余善法, 谷桂珍, 周文慧, 等. 职业应激与抑郁症关系的性别差异[J]. 中华劳动卫生职业病杂志, 2011, 29(12): 887-892.
- [6] MUHONEN T, TORKELESIN E. The Demand-Control-Support Model and health among women and men in similar occupations[J]. J Behav Med, 2003, 26(6): 603-613.
- [7] STORR CL, TRINKOFF AM, ANTHONY JC. Job strain and non-medical drug use[J]. Drug Alcohol Depend, 1999, 55(1-2): 45-51.
- [8] De JONGE J, BOSMA H, PETER R, et al. Job strain, effort-reward imbalance and employee well-being: a large-scale cross-sectional study[J]. Soc Sci Med, 2000, 50(9): 1317-1327.
- [9] COOPER CL, SLOAN S, WILLIAMS S. Darlington community study, the indicator[M]. Berkshire: NFER-NELSON publishing company limited, 1988.
- [10] NAKATA A, TAKAHASHI M, IRIE M, et al. Job satisfaction is associated with elevated natural killer cell immunity among healthy white-collar employees[J]. Brain Behav Immun, 2010, 24(8): 1268-1275.
- [11] YU SF, JIANG KY, ZHOU WH, et al. Relationship between occupational stress and salivary cortisol, sIgA, and lysozyme in assembly line workers[J]. Chin Med J, 2008, 121(17): 1741-1743.
- [12] 余善法, 姜开友, 周文慧, 等. 装配线工人职业紧张与唾液考的松、sIgA 和溶菌酶水平关系研究[J]. 工业卫生与职业病, 2010, 36(2): 99-104.
- [13] 余善法, 谷桂珍, 闫艳. 性别对职业紧张测试结果的影响[J]. 工业卫生与职业病, 2001, 27(6): 336-338.
- [14] 余善法, 马良庆, 崔明煊, 等. 火车司机职业紧张因素分析[J]. 工业卫生与职业病, 1998, 24(3): 129-133.
- [15] YU S, LU ML, GU G, et al. Musculoskeletal symptoms and associated risk factors in a large sample of Chinese workers in Henan province of China[J]. Am J Ind Med, 2012, 55(3), 281-293.
- [16] LEE HY, YEH WY, CHEN CW, et al. Prevalence and psychosocial risk factors of upper extremity musculoskeletal pain in the industries of Taiwan: a nationwide study[J]. J Occup Health, 2005, 47(4): 311-318.
- [17] 许碧云, 杨惠芳, 陈炳为, 等. 职业紧张与工作能力关系的结构方程模型分析[J]. 卫生研究, 2008, 37(1): 59-61.
- [18] 黄铭仁. 结构方程模式-理论及应用[M]. 北京: 中国税务出版社, 2005.
- [19] 邱皓政, 林碧芳. 结构方程模型的原理与应用[M]. 北京: 中国轻工业出版社, 2009.

(收稿日期: 2013-04-08)

(英文编审: 金克峙; 编辑: 洪琪; 校对: 徐新春)