鲁东大学学报(自然科学版)
Journal of Ludong University(Natural Science Edition) 2022,38(4):337—341

基于 Logistic 回归的短缺药品预警模型构建

曲 帅,魏新江

(鲁东大学 数学与统计科学学院,山东 烟台 264039)

摘要: 为了建立合理有效的短缺药品预警模型,本文使用 K-S 检验和 Mann-Whitney U 检验对指标进行筛选,对经过筛选处理后的预警指标进行因子分析,确保在消除指标间多重共线性的同时,保留尽可能多的初始指标信息。对提取的公因子使用 Logistic 回归分析,得到短缺药品预警模型,检验结果显示,训练样本和测试样本的预测正确率分别为 97.06%和 91.18%,表明该模型可以有效预测药品短缺。

关键词: 因子分析; Logistic 回归; 预警模型; 药品短缺

中图分类号: O213 **文献标志码:** A **文章编号:** 1673-8020(2022)04-0337-05

药品短缺问题是世界难题。美国食品药品监督管理局(Food and Drug Administration,FDA)将药品短缺定义为受到 FDA 监管的药品,在临床上可互换种类且总体供应不能满足当前或预计的用户水平的需要[1]。FDA 制定了处理药品短缺问题的相关程序,能够根据短缺药品的属性采取应对措施[2]。加拿大医院的药品供应链中也存在药品短缺问题,其中企业范畴有致使药品短缺的不确定因素[3—4]。当前我国药品短缺的品种繁多,覆盖的疾病种类广泛[5]。鉴于药品短缺问题的重要性及复杂性,完全由市场调节难以解决,需要了解药品短缺的状况,建立健全国家短缺药品管理制度,为解决我国药品短缺问题提供切合实际的方法和途径[5—7]。

在处理药品短缺问题上,监测预警已成为主要的应对方法,被列入我国短缺药品供应保障的政策中。2018 年 9 月,国务院办公厅发布的《关于完善国家基本药物制度的意见》明确提出,要建立健全全国短缺药品监测预警系统,加强短缺药品预警应对措施[8]。目前,关于短缺药品监测预警的研究主要分为定性分析和定量分析。在定性分析方面,文献[9]从供给和需求两方面分析我国药品短缺的警源,为我国短缺药品监测预警系统的建立提出了解决方案;文献[10]对比多个国家监测预警系统建立流程的异同,结合我国药品短缺相关政策,为改进我国短缺药品监测预警系统提出建议。在定量分析方面,文献[11]通过展示复方丹参片的实例,以原料成本和最高零售限价的比例作为预警指标,对基本药物范畴的中药供应进行预警;文献[12]对辽宁省短缺药品进行多方面的描述性分析和评分,根据评分结果构建辽宁省短缺药品预警模型;文献[13]将层次分析法、因子分析和多重线性回归分析等方法相结合,开展短缺药品监测预警平台的研究。

事实上,使用多重线性回归分析仅能得到因子影响程度的排序。相较于多重线性回归模型,Logistic 回归模型的预测效果和适用性更好,不仅能够判别药品是否短缺,而且可以衡量发生药品短缺的概率[14]。因子分析可以将多个原始变量浓缩成少数的因子变量,在达到降维的同时保留原始变量的信息,消除原始变量之间的多重共线性[15]。本文将因子分析和 Logistic 回归分析法相结合,构建短缺药品预警模型,探索适用于短缺药品预警的指标体系。

1 样本与指标选取

1.1 样本选取

本文从山东省公共资源交易中心 2020 年的

收稿日期: 2021-08-13; **修回日期:** 2022-06-05

基金项目: 国家自然科学基金(61973149);山东省自然科学基金重点项目(ZR2020KF029);2021 年度省社科规划研究项目(21CSDJ20)

第一作者简介: 曲帅(1997—),男,山东德州人,硕士研究生,研究方向为经济与社会统计。E-mail: ld_qushuai@163.com

通信作者简介: 魏新江(1977—),男,山东东营人,教授,硕士研究生导师,博士,研究方向为非线性系统控制、鲁棒控制等。E-mail: weixinjiang@163.com

药品集中采购数据中,分别选取51种短缺药品和51种非短缺药品作为配对样本进行研究。样本数量大于30,根据中心极限定理和统计知识,该样本可以代表总体数据。为检验模型预测的准确性,本文将51对药品分为两组,其中34对药品是训练样本,17对药品是测试样本。

1.2 指标选取

根据普遍性、全面性、相关性和可比性的指标选取原则,参照文献[13]提出的短缺药品监测指标体系,并结合山东省公共资源交易中心样本数据的实际情况,本文选取了16个建模指标,分别为规格、中标价、转换比、最小包装单位中标价格、需求量、采购量、采购金额、撤单量、撤单金额、拒单量、拒单金额、到货量、到货金额、退货量、退货金额、到货率。

2 实证分析

2.1 指标筛选

由于本文选取的初始指标数较多,为降低计算复杂程度并确保模型的预测效果,对上述16个指标进行筛选。筛选流程如下:首先对各指标进行K-S检验,检验其正态性,若指标服从正态分布,则进行T检验,若指标不服从正态分布,则进行Mann-Whitney U检验;最后,根据显著性水平来确定模型所用指标。

2.1.1 正态性检验

采用K-S检验,对选取的16个指标作正态性检验,根据检验结果选择显著性检验的方法。K-S检验结果如表1所示,可以看到,16个指标的P值都远小于显著性水平0.05,表明这些指标均不符合正态分布。

2.1.2 Mann-Whitney U检验

为得到短缺药品与非短缺药品之间有显著性差异的指标,需对指标进行显著性检验。若其中某个指标不具备显著性差异,应将其剔除。下面使用Mann-Whitney U检验对16个指标进行显著性检验,检验结果见表2。

从表2可以得出:在0.05的显著性水平下,到货率没有通过检验,说明到货率在两组药品样本之间不存在显著差异;其他指标均通过检验,说明这些指标在短缺药品样本和非短缺药品样本之间存在显著差异。因此,在后续分析中将不考虑到货率对短缺药品预警的影响,仅考虑其他15个指标的影响。

表1 K-S检验结果

Tab.1 K-S test results

指标名称	表示变量	检验值	P值
规格/mg	X_1	0.283	0.000
中标价/元	X_2	0.247	0.000
转换比	X_3	0.259	0.000
最小包装单位中标价格/元	X_4	0.430	0.000
需求量	X_5	0.297	0.000
采购量	X_6	0.247	0.000
采购金额/元	X_7	0.263	0.000
撤单量	X_8	0.361	0.000
撤单金额/元	X_9	0.286	0.000
拒单量	X_{10}	0.375	0.000
拒单金额/元	X_{11}	0.453	0.000
到货量	X_{12}	0.250	0.000
到货金额/元	X_{13}	0.278	0.000
退货量	X_{14}	0.342	0.000
退货金额/元	X_{15}	0.325	0.000
到货率/%	X_{16}	0.167	0.007

注:需求量、采购量、撤单量、拒单量、到货量和退货量均采用最小包装单位。

表2 独立样本的Mann-Whitney U检验结果

Tab.2 Mann-Whitney U test results for independent samples

指标名称	表示变量	W统计量	P值
规格/mg	X_1	840.5	0.002
中标价/元	X_2	1 704.5	0.007
转换比	X_3	415.0	0.000
最小包装单位中标价格/元	X_4	2 480.0	0.000
需求量	X_5	0.0	0.000
采购量	X_6	219.0	0.000
采购金额/元	X_7	365.0	0.000
撤单量	X_8	403.0	0.000
撤单金额/元	X_9	454.0	0.000
拒单量	X_{10}	279.0	0.000
拒单金额/元	X_{11}	411.0	0.000
到货量	X_{12}	276.0	0.000
到货金额/元	X_{13}	478.0	0.000
退货量	X_{14}	355.5	0.000
退货金额/元	X_{15}	430.5	0.000
到货率/%	X_{16}	1 086.0	0.152

注:需求量、采购量、撤单量、拒单量、到货量和退货量均采用最小包装单位。

2.2 因子分析

Logistic回归模型假设自变量之间不存在多重共线性。本文用于药品短缺预警的15个指标之间可能存在多重共线性,这会降低模型的稳健性,易使模型产生错判。因此,为了消除多重共线

性的干扰,且尽可能减少信息损失,本文使用因子分析提取公因子,将 15 个指标综合成数量较少且信息不重合的公共指标,用以建立短缺药预警模型。

2.2.1　KMO 和 Bartlett 球形检验

KMO 和 Bartlett 球形检验可以判断因子分析是否适用于样本。对标准化后的数据进行 KMO 和 Bartlett 球形检验,结果如表 3 所示。由表 3 可以得到,KMO 统计量为 0.566,Bartlett 球形检验的 P 值为 0.000,远小于显著性水平 0.05,说明因子分析适用于该样本。

表 3　KMO 和 Bartlett 球形检验结果

Tab.3 KMO and Bartlett spherical test results

KMO 检验	KMO 统计量	0.566
	近似卡方值	1 878.592
Bartlett 球形检验	自由度(df)	105.000
	显著性(Sig.)	0.000

2.2.2　因子旋转

采用因子旋转的目的是使因子载荷矩阵中的因子载荷绝对值向 0 和 1 分化,进而更容易解释因子。因子旋转有多种方法,本文采用最大方差正交旋转法,旋转前后公因子的方差贡献率如表 4 所示。

表 4　旋转前后公因子的方差贡献

Tab.4 Variance contribution of common factors before and after rotation

公因子	旋转前			旋转后		
	总方差	方差贡献率/%	累计方差贡献率/%	总方差	方差贡献率/%	累计方差贡献率/%
F_1	5.800	38.669	38.669	4.189	27.930	27.930
F_2	2.407	16.044	54.713	2.614	17.429	45.358
F_3	1.534	10.229	64.942	2.252	15.013	60.371
F_4	1.118	7.955	72.897	1.437	10.082	70.453
F_5	1.024	7.326	80.223	1.390	9.769	80.223
F_6	0.942	6.283	86.506			
F_7	0.691	4.106	90.112			
F_8	0.538	3.089	93.700			
F_9	0.398	2.656	96.356			
F_{10}	0.267	1.783	98.139			
F_{11}	0.160	1.066	99.205			
F_{12}	0.065	0.431	99.636			
F_{13}	0.047	0.313	99.949			
F_{14}	0.007	0.045	99.994			
F_{15}	0.001	0.006	100.000			

注: 因子旋转后仅保留旋转前总方差大于 1 的 5 个公因子。

公因子的选取是从累计方差贡献率和特征值两个方面考虑。由表 4 可知,经过因子旋转后,仅前 5 个公因子的特征值大于 1,且这 5 个公因子的累计贡献率为 80.223%,大于 80%,因此本文用前 5 个公因子代替原 15 个指标。

2.2.3　公因子表达式

因子旋转后保留了综合众多指标且指标信息不重合的 5 个公因子,每个公因子可以用原 15 个指标线性表示,指标系数为各公因子得分系数。各公因子得分系数见表 5。

由表 5,得到关于标准化的 15 个指标的公因子线性表达式:

$$F_1 = 0.054X_1 + 0.046X_2 + 0.003X_3 + 0.005X_4 - 0.027X_5 + 0.057X_6 + 0.227X_7 - 0.176X_8 + 0.041X_9 - 0.158X_{10} - 0.123X_{11} + 0.161X_{12} + 0.252X_{13} + 0.227X_{14} + 0.351X_{15},$$

$$F_2 = 0.193X_1 - 0.018X_2 - 0.105X_3 + 0.048X_4 + 0.118X_5 + 0.019X_6 + 0.069X_7 + 0.034X_8 - 0.011X_9 + 0.428X_{10} + 0.433X_{11} - 0.015X_{12} + 0.050X_{13} - 0.175X_{14} - 0.165\ X_{15},$$

$$F_3 = -0.274X_1 + 0.069X_2 - 0.023X_3 + 0.098X_4 + 0.197X_5 + 0.218X_6 - 0.112X_7 + 0.521X_8 + 0.305X_9 + 0.129X_{10} + 0.047X_{11} + 0.035X_{12} - 0.162X_{13} + 0.000X_{14} - 0.219\ X_{15},$$

$$F_4 = -0.111X_1 + 0.170X_2 + 0.695X_3 - 0.358X_4 + 0.288X_5 - 0.035X_6 + 0.001X_7 - 0.045X_8 + 0.006X_9 - 0.055X_{10} - 0.086X_{11} - 0.026X_{12} + 0.005X_{13} - 0.034X_{14} + 0.011X_{15},$$

$$F_5 = -0.211X_1 + 0.682X_2 + 0.124X_3 + 0.469X_4 + 0.090X_5 - 0.076X_6 + 0.099X_7 + 0.014X_8 + 0.148X_9 - 0.004X_{10} + 0.055X_{11} - 0.100X_{12} + 0.078X_{13} - 0.051X_{14} + 0.020\ X_{15}。$$

利用线性表达式计算公因子 $F_i(i=1,2,\cdots,5)$，并用于构建 Logistic 预警模型。

2.3 短缺药品预警模型构建

2.3.1 模型建立

下面基于训练样本数据，使用 Logistic 回归分析5个公因子。由于 F_2 和 F_3 的 P 值大于显著性水平0.05，说明 F_2 和 F_3 作用效果不显著，故将其去除，最终得到关于 F_1，F_4 和 F_5 的 Logistic 预警模型。模型参数估计如表6所示。

根据表6中参数估计值，建立短缺药品预警模型如下:

$$P^*=\frac{1}{1+\exp\left[-(0.337-10.638F_1-5.569F_4+8.390F_5)\right]}, \tag{1}$$

其中，P^* 表示药品发生短缺风险的概率。

表5 公因子得分系数

Tab.5 Common factor score coefficient

指标	公因子					指标	公因子				
	F_1	F_2	F_3	F_4	F_5		F_1	F_2	F_3	F_4	F_5
X_1	0.054	0.193	−0.274	−0.111	−0.211	X_9	0.041	−0.011	0.305	0.006	0.148
X_2	0.046	−0.018	0.069	0.170	0.682	X_{10}	−0.158	0.428	0.129	−0.055	−0.004
X_3	0.003	−0.105	−0.023	0.695	0.124	X_{11}	−0.123	0.433	0.047	−0.086	0.055
X_4	0.005	0.048	0.098	−0.358	0.469	X_{12}	0.161	−0.015	0.035	−0.026	−0.100
X_5	−0.027	0.118	0.197	0.288	0.090	X_{13}	0.252	0.050	−0.162	0.005	0.078
X_6	0.057	0.019	0.218	−0.035	−0.076	X_{14}	0.227	−0.175	0.000	−0.034	−0.051
X_7	0.227	0.069	−0.112	0.001	0.099	X_{15}	0.351	−0.165	−0.219	0.011	0.020
X_8	−0.176	0.034	0.521	−0.045	0.014						

表6 模型参数估计

Tab.6 Model parameter estimation

变量	参数估计值	标准差	Wald 统计量	自由度(df)	显著性水平(Sig.)	优势比
F_1	−10.638	3.956	7.230	1	0.007	0.000
F_4	−5.569	2.104	7.004	1	0.008	0.004
F_5	8.390	3.164	7.032	1	0.008	4 403.217
常数	0.337	0.724	0.217	1	0.641	1.401

2.3.2 模型检验

将训练样本和测试样本数据分别代入模型(1)进行验证，其中分割点设置为0.5。若检验药品对应的概率 P^* 大于0.5，该药品判定为短缺药品，否则该药品为非短缺药品。对比训练样本和测试样本实际值和预测值，结果见表7。

表7 模型预测结果

Tab.7 Model prediction results

组别	训练样本预测值		测试样本预测值	
	短缺药品	非短缺药品	短缺药品	非短缺药品
短缺药品	33	1	17	0
非短缺药品	1	33	3	14
准确率/%	97.06	97.06	85.00	100.00

基于表7，得到训练样本和测试样本数据的预测正确率分别为97.06%、91.18%。从模型在训练样本和测试样本中的预测精度来看，模型(1)对历史数据有很好的拟合效果，且对未来数据的预测也同样准确。文献[15—17]所用预警模型和本文研究模型相同，其中，文献[16]在企业财务风险领域的总预测正确率为82.50%，文献[15,17]在企业信用风险领域的总预测正确率分别为85.70%和81.42%，而本文短缺药品预警的总预测正确率为95.10%，说明本文建立的短缺药品预警模型具有良好的预测性能。

3 结语

山东省公共资源交易中心的药品集中采购数据中包含大量可以预测药品短缺的信息，使用该数据建立预警模型，可以很好地预测药品是否短缺。本文对筛选处理后的指标采用因子分析，进而对提取的因子使用 Logistic 回归分析，使构建的短缺药品预警模型不仅消除指标之间的多重共线性，而且预测准确率较高。由于采购数据中的定

性数据在各药品间区分度不够大,使得本文的指标类型不够全面,在进一步研究中将会丰富指标类型,使用包含定量和定性数据的综合预警指标构建短缺药品预警模型。

参考文献:

[1] 姚立新,BOEHM G,郑强.美国药品短缺及 FDA 采取的应对策略 [J].中国新药杂志,2012,21(20):2359-2367.

[2] Food and Drug Administration.Frequently asked questions about drug shortages [EB/OL].(2021-11-13) [2021-08-10].https://www.fda.gov/Drugs/DrugSafety/DrugShortages/ucm050796.htm.

[3] ZWAIDA T A,BEAUREGARD Y,ELARROUDI K.Comprehensive literature review about drug shortages in the Canadian hospital's pharmacy supply chain [C]//International Conference on Engineering,Science,and Industrial Applications,2019.

[4] BEDARD M.Drug shortages: can we resolve that problem? [J].Canadian Journal of Anesthesia-Journal Canadien D Anesthésie,2013,60(6):523-527.

[5] 武丽娜,方宇,杨才君,等.我国药品短缺问题研究进展评述 [J].中国药事,2016,30(5):458-465.

[6] 乌日图.建立国家短缺药品管理制度 [J].瞭望,2007(16):40-42.

[7] 杨悦,黄果,初智铭,等.美国处理药品短缺问题的经验及其对我国的启示 [J].中国药房,2008,19(28):2173-2176.

[8] 国务院办公厅.关于完善国家基本药物制度的意见(国办发〔2018〕88 号) [EB/OL].(2018-09-19)[2021-08-10].http://www.gov.cn/zhengce/content/2018-09/19/content_5323459.htm.

[9] 郭冬梅.关于构建我国药品短缺风险预警管制体系的思考 [J].广东药学院学报,2015,31(5):642-645.

[10] 刘青泽,韩月,朱虹,等.国内外短缺药品监测预警体系对比分析 [J].中国药业,2019,28(18):1-4.

[11] 杨光,王永炎,陆建伟,等.基于全国中药资源普查的中药基本药物供应预警方法探讨 [J].中草药,2015,46(1):7-10.

[12] 周鹤,赵春阳,陈维媛,等.辽宁省短缺药品预警分析 [J].中国药物警戒,2020,17(1):44-50.

[13] 黄润青.云南省短缺药品供应保障现状调查及监测预警平台的构建研究 [D].昆明:昆明医科大学,2020.

[14] 解秀玉,管西三.企业财务风险预警模型研究:基于制造业数据 [J].南京审计学院学报,2013,10(4):58-68.

[15] 梁琪.企业经营管理预警:主成分分析在 Logistic 回归方法中的应用 [J].管理工程学报,2005,19(1):100-103.

[16] 陈芳,吴杰.中小企业财务危机预警模型比较研究:基于因子分析与 Logistic 回归模型的对比 [J].财会通讯,2017(5):106-108.

[17] 邓晶,秦涛,黄珊.基于 Logistic 模型的我国上市公司信用风险预警研究 [J].金融理论与实践,2013(2):22-26.

Construction of Drug Shortage Warning Model Based on Logistic Regression

QU Shuai, WEI Xinjiang

(School of Mathematics and Statistics Science,Ludong University,Yantai 264039,China)

Abstract: In order to establish a reasonable and effective drug shortage warning model,K-S test and Mann-Whitney U test were used to filtrate the indicators,and factor analysis was performed on the early-warning indicators after filtrating,so as to ensure that the multicollinearity among indicators is eliminated and the maximum initial indicator information is retained as far as possible.Logistic regression was used for the extracted principal factors to obtain the warning model of drug shortage,and the test results show that the prediction accuracy of training samples and test samples are 97.06% and 91.18%,respectively,which indicates that the model can predict drug shortage effectively.

Keywords: factor analysis; Logistic regression; warning model; drug shortage

(责任编辑　顾建忠)