



人大经济论坛

www.pinggu.org

公司财务预警数据分析报告

分析员：某某某

TEL:010-68454276

邮箱：resrarch@pinggu.org

*基于人大经济论坛的强大统计计量和数据处理方面的实力，我们基本可以承担所有形式的数据处理和分析，如果您有需要解决的数据处理和分析的问题，请发邮件到：research@pinggu.org 或电话：010-68454276 说明您的问题，我们会在 1-2 个工作日内给予您回复。

更多精彩 就在 人大经济论坛 www.pinggu.org



目录

第一：软件介绍.....	2
第二：客户需求与数据情况.....	3
第三：数据处理过程.....	4
第四：工作进度安排.....	4
第五：数据分析及结果.....	5
1. 相关分析：计算其中连续变量之间的相关系数.....	5
2. 列联分析.....	6
3. 是否困境与连续型变量的两独立样本 t 检验.....	7
4. logistic 回归.....	8
5. 结论.....	11
第六：所用方法及原理小结.....	12

第一：软件介绍



SPSS 是世界上最早的统计分析软件，由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制。

SPSS 大部分功能可以菜单操作实现，由于其操作简单，深受非统计专业人士以及部分统计专业人士的欢迎。SPSS 已经在我国的社会科学、自然科学的各个领域发挥了巨大作用。该软件还可以应用于经济学、生物学、心理学、地理学、医疗卫生、体育、农业、林业、商业、金融等各个领域。



SPSS 软件的主要特点有：

1、操作简便

界面非常友好，除了数据录入及部分命令程序等少数输入工作需要键盘键入外，大多数操作可通过鼠标拖曳、点击“菜单”、“按钮”和“对话框”来完成。

2、功能强大

具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。SPSS 提供了从简单的统计描述到复杂的多因素统计分析方法，比如数据的探索性分析、统计描述、列联表分析、二维相关、秩相关、偏相关、方差分析、非参数检验、多元回归、生存分析、协方差分析、判别分析、因子分析、聚类分析、非线性回归、Logistic 回归等。

3、全面的数据接口

能够读取及输出多种格式的文件。比如由文本文件，Excel 的*.xls 文件等均可转换成可供分析的 SPSS 数据文件。结果可保存为*.txt, word, PPT 及 html 格式的文件。

4、针对性强

SPSS 针对初学者、熟练者及精通者都比较适用。并且现在很多群体只需要掌握简单的操作分析，大多青睐于 SPSS，而那些熟练或精通者也较喜欢 SPSS，因为他们可以通过编程来实现更强大的功能。

第二：客户需求与数据情况

在上市公司中，经常会有公司发生财务困境，而发生财务困境不是偶然的和突发的，这跟这些公司的所有权结构、管理成本、审计等有很密切的关系。本文的目的是通过选取一些与公司财务困境有关联的变量，选取 50 个已经发生财务困境和 50 个从未发生过财务困境的公司，建立模型。通过这个模型可以得出，究竟哪一些变量会对公司是否发生财务困境有显著的影响，并且通过模型可以对公司的财务状况进行预测，也就是建立公司财务预警机制。本文选取的因变量是本公司是否发生财务困境，这里共选取了 100 个公司。自



变量的选取情况如表 2.1：

表 2.1 自变量列表

最大股东持股比率	第一大股东持股比例
Herfindahl10 指数，股权集中度	Herfindahl 10 指数
国有持股比率	国有持股比率
公司管理层持股比率	1. 董事会持股比例 2. 监事会持股比例 3. 高管人员持股比例 4. 董事长持股比例 5. 监事会主席的持股比例 6. 总经理持股比例
独立董事的数量，和比率	1. 独立董事总人数 2. 董事会的规模（通过和 1 计算得到比率）
公司管理成本比率	1. 管理费用 2. 管理费用/销售收入（即营业收入）
审计会执行效率	是否设置审计委员会

第三：数据处理过程

1. 分析客户需求，整理数据成为 spss 可供分析的形式。
2. 对数据异常值的处理（将异常值用缺失值代替）：最大股东持股比率、Herfindahl10、国有持股比率、公司管理层持股比率取值应该在 0 与 1 之间；审计会执行效率，是否困境这两个变量只能取值为 0 或 1；对公司管理成本比率、独立董事比率异常大的值的剔除
3. 分析方法选择。本文选取的主要分析方法是 logistic 模型。
4. 数据分析过程及结果展现。

第四：工作进度安排



1. 前期数据处理：

时间	工作内容
当天	①了解客户需求，判断能否处理并报价
一个工作小时内	②搜集确定客户数据，与客户达成一致

2. 数据处理阶段：

时间	工作内容
当天 三个工作小时	①按照客户需求，制定工作计划，选择合适的软件 ②用 spss 对数据进行异常值处理和格式整理
第二天 三个工作小时	①用 spss 进行相关分析、列联分析 ②用 spss 对数据两样本 t 检验和 logistic 回归分析，初步撰写分析报告

3. 数据分析后的处理

时间	工作内容
第二天 一个半工作小时	①由数据测试员对所得数据结果进行检验 ②分析结果的描述、解释，辑写分析报告
第二天 一个半工作小时	①将数据分析结果提交客户，向其描述，解释结果 ②客户满意度评价

第五：数据分析及结果

1. 相关分析：计算其中连续变量之间的相关系数

表格 1 中红色的字体标出来的是两个变量的相关系数大于 0.6，并且显著相关的相关系数。

从中我们需要特别注意的是最大股东持股比率与 Herfindahl 10 指数之间的相关系数达到了 0.969，这样我们在做回归分析的时候讲它们一起加入模型中可能会出现多重共线性问题。



表 1 连续变量之间的相关系数

		最大股东 持股比率	Herfindahl 10 指数	国有持股 比率	管理费用/ 销售收入 (即营业 收入)	独立董事 比例	管理层持 股
最大股东 持股比率	Pearson Correlatio n	1	.969**	.646**	0.114	.164*	-0.066
	Sig. (2- tailed)		0	0	0.12	0.024	0.368
Herfinda hl 10 指 数	Pearson Correlatio n	.969**	1	.671**	0.116	.180*	-0.08
	Sig. (2- tailed)	0		0	0.115	0.013	0.275
国有持股 比率	Pearson Correlatio n	.646**	.671**	1	0.101	0.045	-0.108
	Sig. (2- tailed)	0	0		0.168	0.537	0.14
管理费用 /销售收 入 (即营 业收入)	Pearson Correlatio n	0.114	0.116	0.101	1	.626**	-0.015
	Sig. (2- tailed)	0.12	0.115	0.168		0	0.844
独立董事 比例	Pearson Correlatio n	.164*	.180*	0.045	.626**	1	-0.043
	Sig. (2- tailed)	0.024	0.013	0.537	0		0.556
管理层持 股	Pearson Correlatio n	-0.066	-0.08	-0.108	-0.015	-0.043	1
	Sig. (2- tailed)	0.368	0.275	0.14	0.844	0.556	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

2. 列联分析

表格 2 是一个基本的统计情况。表格 3 中我们可以看出，在 0.05 的显著性水平下，是否困境与是否设置审计委员会之间没有显著的相关关系。



表 2 是否困境与是否设置审计委员会的列联分析

		是否设置审计委员会		Total
		无	有	
是否困境	非困境	4	86	90
	困境	10	89	99
Total		14	175	189

表 3 卡方检验

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.199 ^a	1	0.138		
Continuity Correction ^b	1.452	1	0.228		
Likelihood Ratio	2.279	1	0.131		
Fisher's Exact Test				0.17	0.113
Linear-by-Linear Association	2.188	1	0.139		
N of Valid Cases	189				
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.67.					
b. Computed only for a 2x2 table					

3. 是否困境与连续型变量的两独立样本 t 检验

表 4 是连续型变量按照是否困境进行分类的统计信息，从中可以看出，在两个水平上，各个变量的值是有差异的，但是该差异是否是由随机因素引起的，还需要进行显著性检。

表 4 是否困境与连续型变量的描述性统计

	是否困境	N	Mean	Std. D	Std. Error M
最大股东持股比率	非困境	90	0.4448776	0.20510536	0.02162
	困境	99	0.2895973	0.13598675	0.0136672
Herfindahl 10 指数	非困境	90	0.2772186	0.1791512	0.0188842
	困境	99	0.1205657	0.10843315	0.0108979
国有持股比率	非困境	90	0.38	0.286	0.03
	困境	99	0.12	0.199	0.02
管理费用/销售收入	非困境	89	0.082713	0.1321125	0.0140039
	困境	97	1.543189	10.5558374	1.0717829
独立董事比例	非困境	90	0.5892	0.25715	0.02711
	困境	99	0.6463	0.3058	0.03073
管理层持股	非困境	88	0.0153879	0.0905469	0.0096523
	困境	99	0.0047955	0.04676309	0.0046999



表 5 是否困境与连续型变量的两独立样本 t 检验

Independent Samples Test						
		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
最大股东持股比率	Equal variances assumed	32.871	0	6.185	187	0
	Equal variances not assumed			6.071	152.263	0
Herfindahl 10 指数	Equal variances assumed	32.957	0	7.346	187	0
	Equal variances not assumed			7.185	143.679	0
国有持股比率	Equal variances assumed	28.283	0	7.232	187	0
	Equal variances not assumed			7.112	156.819	0
管理费用/销售收入（即营业收入）	Equal variances assumed	4.409	0.037	-1.305	184	0.194
	Equal variances not assumed			-1.363	96.033	0.176
独立董事比例	Equal variances assumed	0.243	0.623	-1.381	187	0.169
	Equal variances not assumed			-1.393	185.9	0.165
管理层持股	Equal variances assumed	4.005	0.047	1.021	185	0.309
	Equal variances not assumed			0.987	126.815	0.326

从 t 检验的结果上来看，在 0.05 的显著性水平下，管理费用/销售收入、独立董事比例、管理层持股在是否出现困境两个水平上的均值没有显著差异。所以这三个变量对因变量的影响可能不是显著的。

4. logistic 回归 我们把所有自变量加入模型中，看一下回归模型的情况

首先是模型的拟合情况，Cox & Snell 和 Nagelkerke R 方分别为 .422、.563，不过这两个统计量主要用于不同模型之间的比较。

表 6 模型总体情况

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	153.787 ^a	.422	.563
a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.			



模型内部的拟合情况上看，有 82.1% 的样本进行了正确的样本内预测。

表 7 拟合情况

Classification Table ^a					
	Observed		Predicted		
			kun		Percentage Correct
			非困境	困境	
Step 1	kun	非困境	66	21	75.9
		困境	12	85	87.6
	Overall Percentage				82.1

a. The cut value is .500

模型参数估计情况，告诉我们自变量对因变量的影响情况，sig 大于 0.05 的参数估计值，我们认为这些自变量对因变量的影响不是显著的。而 sig 小于 0.05 的变量，我们认为这些自变量对因变量的影响是显著的。

其中，最大股东持股比例、管理费用/销售收入的参数估计值为正，也就是这两个变量的值越大，则企业发生困境的概率越大。Herfindahl10 指数、国有持股比例的参数估计值为负，也就是这两个变量的值越大，则企业发生困境的概率越小。

表 8 参数估计

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	最大股东持股比例	17.828	4.907	13.198	1	.000	5.530E7
	Herfindahl10 指数	-28.434	6.549	18.849	1	.000	.000
	国有持股比例	-2.375	1.014	5.491	1	.019	.093
	是否设置审计委员会	-.042	.786	.003	1	.958	.959
	独立董事比例	1.534	1.126	1.858	1	.173	4.638
	管理层持股	-3.953	2.561	2.383	1	.123	.019
	管理费用销售收入（即营业收入）	3.917	1.591	6.064	1	.014	50.244
	Constant	-2.015	1.327	2.305	1	.129	.133

a. Variable(s) entered on step 1: 最大股东持股比例, Herfindahl10 指数, 国有持股比例, 是否设置审计委员会, 独立董事比例, 管理层持股, 管理费用销售收入（即营业收入）.



模型 1：

$$\ln\left(\frac{p}{1-p}\right) = -2.015 + 17.828 * x_1 - 28.434 * x_2 - 2.375 * x_3 - 0.042 * x_4 + 1.534 * x_5 - 3.953 * x_6 + 3.917 * x_7$$

其中： $p = P(y=1|x)$

表 9 变量对应

最大股东持股比率	X1
Herfindahl10 指数	X2
国有持股比率	X3
是否设置审计委员会	X4
独立董事比例	X5
管理层持股	X6
管理费用销售收入（即营业收入）	X7

由于最大股东持股比率、Herfindahl10 指数测量的是同一个指标，二者的参数估计值正负号相反，这是不合理的，可能是二者之间的很强的相关性引起了多重共线性问题，现在我们分别把这两个变量纳入到模型中。

$$\ln\left(\frac{p}{1-p}\right) = 0.814 - 6.558 * x_2 - 2.03 * x_3 - 0.337 * x_4 + 1.136 * x_5 - 3.129 * x_6 + 3.904 * x_7$$

表 10 参数估计

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1a	国有持股比率	-2.030	.915	4.917	1	.027	.131
	是否设置审计委员会	-.337	.781	.187	1	.666	.714
	独立董事比例	1.136	.970	1.373	1	.241	3.116
	管理层持股	-3.249	2.571	1.597	1	.206	.039
	管理费用销售收入（即营业收入）	3.904	1.627	5.757	1	.016	49.578
	Herfindahl10 指数	-6.558	1.820	12.987	1	.000	.001
	Constant	.814	1.020	.636	1	.425	2.257
Step 1a	最大股东持股比率	-3.263	1.344	5.896	1	.015	.038
	国有持股比率	-2.680	.866	9.571	1	.002	.069
	是否设置审计委员会	-.388	.774	.251	1	.616	.678
	独立董事比例 管理层	.713	.910	.613	1	.434	2.040
	持股	-2.971	2.535	1.374	1	.241	.051
	管理费用销售收入（即营业收入）	4.221	1.699	6.175	1	.013	68.106
	Constant	1.211	1.032	1.378	1	.241	3.358



模型 2-1

$$\ln\left(\frac{p}{1-p}\right) = 0.814 - 6.558 * x_2 - 2.03 * x_3 - 0.337 * x_4 + 1.136 * x_5 - 3.129 * x_6 + 3.904 * x_7$$

其中： $p = P(y=1|x)$

模型 2-2

$$\ln\left(\frac{p}{1-p}\right) = 1.211 - 3.263 * x_1 - 2.68 * x_3 - 0.388 * x_4 + 0.713 * x_5 - 2.971 * x_6 + 4.221 * x_7$$

其中： $p = P(y=1|x)$

从这两个模型上可以看出，最大股东持股比率、Herfindahl10 指数的参数估计值都成为负值，并且都是显著的。这样我们可以得出结论，最大股东持股比率、Herfindahl10 指数，这两个变量的值越大，则企业发生困境的概率越小。

最大股东持股比率增加一个单位，会使得企业发生困境的发生比变为原来的.038 倍。

Herfindahl10 指数比率增加一个单位，会使得企业发生困境的发生比变为原来的.001 倍。

国有持股比率的参数估计值是负值，其值越大，则越会降低发生困境的概率。管理费用/销售收入的参数估计值是正值，其值越大，则越会增加发生困境的概率。

5. 结论

<i>Hypothesis1</i>	接受原假设
<i>Hypothesis2</i>	接受原假设
<i>Hypothesis3</i>	不能判断，该变量对因变量的影响不是显著的，所以不能判定是正的影响还是负的影响。
<i>Hypothesis4</i>	不能判断，该变量对因变量的影响不是显著的，所以不能判定是正的影响还是负的影响。
<i>Hypothesis5</i>	接受原假设
<i>Hypothesis6</i>	不能判断，该变量对因变量的影响不是显著的，所以不能判定是正的影响还是负的影响。



第六：所用方法及原理小结

以上分析采用了数据分析中基本的相关分析、列联分析、两样本 t 检验和 logistic 回归分析。

通过本例的处理，我们可以更好地理解进行数据处理的一般思路 and 过程。相关分析、列联分析、两样本 t 检验是在前期对数据情况和变量之间的关系进行一个初期的探索。Logistic 模型是最后进行财务预警的模型。

logistic 回归分析是对一般的线性回归分析的扩展，它应用与因变量是分类变量的情形。在本例子中，因变量只会取两个值，0 或者 1，其中 1 代表出现了财务困境。

本例选取的统计软件是 SPSS，实际上大部分的同及软件都可以实现。不过与其他软件相比，SPSS 使用菜单操作，方便简单。并且输出的结果美观、直白。

在上面的分析中，一个很重要的方面是要看显著性。也就是在一定的显著性水平下，该关系是否是显著的。我们只要看伴随概率 sig 就可以了。Sig 小于显著性水平，则该关系是显著的，否则是不显著的。

*作者小注：建立学习统计与学习统计软件结合起来，统计软件的学习能够帮助我们更快的建立统计的学习框架，以及从实用的角度体会统计在实践中的应用。人大经济论坛的计量经济学与统计版面有丰富的学习资源可以免费下载，并且国内外的统计高手在这里云集交流。网址：<http://www.pinggu.org/bbs/forum-5-1.html>

