

---

# 线性回归中标准化回归系数是做什么用的？

作者：陶立元 赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1675.html>

*本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！*

多因素线性回归是我们在临床研究中经常用到的，在做完回归后大家往往喜欢问这些影响因素中谁的影响效应最大呢？同时，在SPSS线性回归分析的结果中有标准化回归系数这一列，这一列又是用来做什么的呢？本文就跟大家一起来讨论这一问题。

首先我们来看一下数据，在这个例子中有5个变量，其中y是因变量，x1至x4是自变量。想探索x1-x4对y的影响大小，同时比较不同自变量x对y的作用谁大谁小。



5:

	y	x1	x2	x3	x4
1	24.22	10.00	5.75	13.60	29.36
2	24.22	3.00	9.32	6.20	14.31
3	19.03	15.00	2.50	11.10	26.08
4	23.39	3.00	5.66	9.70	19.62
5	19.49	4.00	2.83	7.30	42.82
6	24.38	6.00	6.86	7.30	22.76
7	19.03	2.90	3.22	7.70	31.00
8	21.11	9.00	4.90	6.00	17.28
9	23.32	5.00	3.54	6.70	30.25
10	24.34	2.00	4.51	7.20	24.28
11	23.82	8.00	8.47	9.10	18.94
12	22.86	20.00	9.92	8.10	16.08
13	24.49	12.00	6.01	7.00	29.50
14	23.37	6.00	4.31	6.30	25.64
15	20.81	7.00	3.46	7.10	32.26
16	24.14	5.00	10.21	7.40	16.01
17	26.45	4.00	19.31	5.10	19.03
18	25.22	2.30	8.65	7.60	17.46
19	27.22	3.00	8.54	8.60	20.36
20	25.93	6.00	7.21	8.90	15.92
21	26.99	12.00	8.75	7.00	15.34
22	25.71	7.00	13.07	13.50	8.05
23	28.41	4.00	8.90	13.50	12.31
24	26.39	4.00	23.26	8.20	5.59

这时候我们往往会做多因素线性回归，其结果如下：

---

我们可以看到在上述结果中有未标准化回归系数和标准化回归系数，除此之外还有t值和p值(当然t和p是用来判断该因素对y的影响是否有统计学意义的，不是本文讨论的重点)。上表中的未标准化回归系数是回归方程中不同自变量的系数，通过该系数和常数项咱们可以因变量的预测值。同时还有标准化回归系数，它便是用来比较不同自变量(x)对因变量(y)的作用大小的。标准化回归系数的觉得值越大，说明该自变量x对因变量y的作用愈大。

那么标准化回归系数是怎么算出来的呢?它就是将上述的 $x_1$ - $x_4$ 和y的值都标化到标准正态分布对应的值，即均数为0，标准差为1。对上述数据进行标准化后，它们之间便可以相互比较了。假设因变量是身高和体重，增加1cm的身高和增加1kg的体重，对因变量y的影响谁大谁小无法比较。但

---

是如果我们将身高、体重和因变量都标准化到标准正态分布后，这时候身高增加1个单位和体重增加1个单位对与y的影响就可以比较了。

标化的过程很简单，就是将一个正态分布转化成标准正态分布，即用每个指标的数据减去该指标的均数，再除以该指标的标准差：

然后再用转化后的参数之间去做回归。我们进行了转化并做了计算，变量名称带有n的为标准化后的新变量，其结果如下：

	y	x1	x2	x3	x4	yn	x1n	x2n	x3n	x4n
1	24.22	10.00	5.75	13.60	29.36	-0.22	.70	-.68	2.51	1.15
2	24.22	3.00	9.32	6.20	14.31	-0.22	-.81	-.12	-.85	-.49
3	19.03	15.00	2.50	11.10	26.08	-1.93	1.78	-1.19	1.38	.79
4	23.39	3.00	5.66	9.70	19.62	-.50	-.81	-.69	.74	.09
5	19.49	4.00	2.83	7.30	42.82	-1.78	-.59	-1.13	-.35	2.62
6	24.38	6.00	6.86	7.30	22.76	-.17	-.16	-.50	-.35	.43
7	19.03	2.90	3.22	7.70	31.00	-1.93	-.83	-1.07	-.17	1.33
8	21.11	9.00	4.90	6.00	17.28	-1.24	.48	-.81	-.94	-.17
9	23.32	5.00	3.54	6.70	30.25	-.52	-.38	-1.02	-.62	1.25
10	24.34	2.00	4.51	7.20	24.28	-.18	-1.03	-.87	-.40	.60
11	23.82	8.00	8.47	9.10	18.94	-.35	.27	-.25	.47	.01
12	22.86	20.00	9.92	8.10	16.08	-.67	2.00	-.25	-.25	.00
13	24.49	12.00	6.01	7.00	29.50	-.13	1.13	-.64	-.49	1.17

这时候就会发现，未标准化的回归系数和上次计算结果中的标准化回归系数是一样的。它的绝对值的大小就可以比较x1-x4这四个变量对y的作用大小了。

最后再说一下未标准化回归系数和标准化回归系数的转换关系， $b_j$ 是未标化的系数， $b_j'$ 是标化的系数：

上式中 $S_j$ 是自变量j的标准差， $S_Y$ 是因变量Y的标准差。在本例中x1的标准差为4.64，y的标准差为3.047。x1标准化回归系数=  
 $x_1$ 未标准化回归系数\*(4.64/3.047)=-0.101\*(4.64/3.047)=-0.154，与SPSS的计算结果一致。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发