

回归分析

相关分析描述分析项之间是否有关系,回归分析(线性回归分析)研究影响关系情况,回归分析实质上就是研究 X(自变量,通常为量数据)对 Y(因变量,定量数据)的影响关系情况,有相关关系不一定有回归影响关系。

分析步骤共为四步,分别是:

- 第一步:首先对模型情况进行分析

包括模型拟合情况(比如 R 平方为 0.3,则说明所有 X 可以解释 Y 30%的变化原因),模型共线性问题(VIF 值小于 5 则说明无多重共线性),是否通过 F 检验(F 检验用于判定是否 X 中至少有一个对 Y 产生影响,如果呈现出显著性,则说明所有 X 中至少一个会对 Y 产生影响关系)。

- 第二步:分析 X 的显著性

如果显著(p 值判断),则说明具有影响关系,反之无影响关系。

- 第三步:判断 X 对 Y 的影响关系方向

回归系数 B 值大于 0 说明正向影响,反之负向影响。

- 第四步:其它

比如对比影响程度大小(回归系数 B 值大小对比 X 对 Y 的影响程度大小)

分析项	回归分析说明
网购满意度, 重复购买意愿	网购满意度是否会影响样本重复购买意愿?网购满意度越高,是否重复购买意愿也会越高?

如果多个量表题表示一个维度,可使用“生成变量”的平均值功能。将多个量表题合并成一个整体维度。

分析结果表格示例如下:

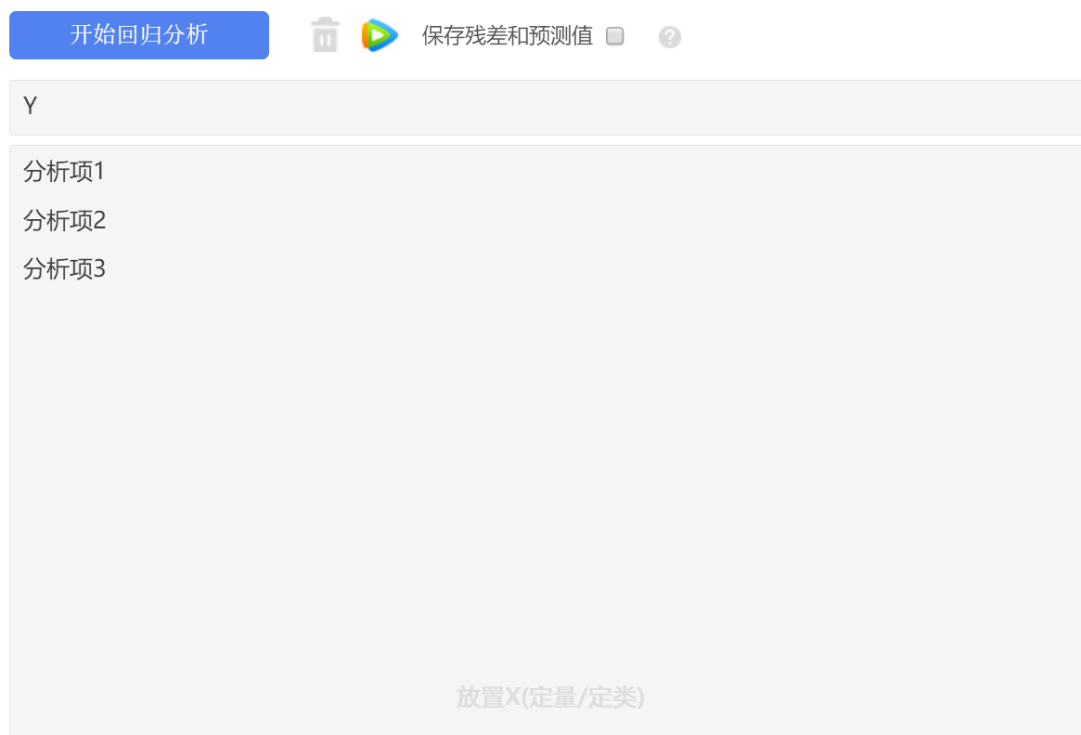
	非标准化系数		标准化系数	t	p	VIF 值	R ²	调整 R ²	F
	B	标准误	Beta						
常数	0.774	0.384	-	2.014	0.047*	-	0.351	0.326	14.188**
分析项 1	0.198	0.099	0.202	1.998	0.048*	1.320			
分析项 2	0.437	0.124	0.374	3.519	0.001**	1.320			

分析项 3	0.004	0.124	0.004	0.034	0.973	1.230			
-------	-------	-------	-------	-------	-------	-------	--	--	--

* $p < 0.05$ ** $p < 0.01$

备注:F 对应的 P 值是由*号表示 (* $p < 0.05$ ** $p < 0.01$, 如果 F 值右上角没有*号, 则说明 F 值对应的 P 值大于 0.05)

SPSSAU 操作截图如下:



回归分析案例

Contents

1 背景.....	3
2 理论.....	3
3 操作.....	4
4 SPSSAU 输出结果.....	5
5 文字分析.....	5
6 剖析.....	6

1 背景

比如想研究“淘宝客服服务态度”，“淘宝商家服务质量”对于“淘宝忠诚度”的影响关系情况，此句话中明显的可以看出“淘宝客服服务态度”，“淘宝商家服务质量”这两项为 X ；而“淘宝忠诚度”为 Y 。

(另提示：如果“淘宝客服服务态度”由几个题项表示，此时可使用“生成变量”->“平均值”功能，将几项概括成一个整体)

2 理论

回归分析是研究定量数据之间的影响关系情况，通常回归分析是放在相关关系分析之后。

针对回归分析：首先要 F 检验，如果 F 值右上角有*号，说明回归分析通过 F 检验，即说明这个回归分析有意义可以做。然后通常要看以下几个指标： R^2 这个值在 0~1 之间，越大越好，它代表回归方程模型拟合的好坏，如果为 0.5，说明有 50% 的点的都掉在回归方程上面，这个值没有好坏之分，一般只是希望越大越好。同时 VIF 值代表多重共线性，所有的 VIF 值均需要小于 10，相对严格的标准是小于 5。

接着分析具体 X 对 Y 的影响关系，首先判断有没有呈现出显著性，即 P 值是否小于 0.05，如果 $P < 0.05$ 则说明有影响关系，反之则说明没有影响关系。在说明已经有影响关系的前提下，具体是正向或是负向影响关系，则是通过“非标准化系数”或者“标准化系数”进行判断，大于 0 则说明是正向影响，反之则说

明为负向影响。

特别提示：回归研究 X 对 Y 的影响；这里的 X 通常是定量数据，而非定类数据。如果 X 为定类数据，此时一般是作为控制变量(可能对模型有干扰因而放入的项)，即放入模型，但并不对其进行分析。如果 X 为定类数据，而且还一定要分析它对 Y 的影响，此时应该进行虚拟变量（也称哑变量）【SPSSAU 中的生成变量功能可进行设置】设置后再放入。

试想 X 如果为定类数据比如性别； Y 为满意度；如果 X 呈现出显著性，即“越性别，越满意”这样的结论完全让人匪夷所思。所以以下建议：

- ✓ 如果 X 为定类数据，直接放入模型时，一般是不会对其进行分析，而仅仅是作为控制变量(可能对模型有干扰因而放入的项)放入就好。
- ✓ 如果 X 为定类数据，并且想分析 X 对 Y 的影响；比如相对于男性，女性有没有满意度更高。此时则需要进行虚拟变量（也称哑变量）【SPSSAU 中的生成变量功能可进行设置】设置。

关于虚拟变量的更多理论知识，建议参考：

<http://study.spssau.com/dummy.pdf>

<http://wiki.mbalib.com/wiki/%E8%99%9A%E6%8B%9F%E5%8F%98%E9%87>

[%8F](#)

3 操作

“淘宝客服服务态度”，“淘宝商家服务质量”这两项为 X ；而“淘宝忠诚度”为 Y 。操作如下图：



备注：专业用户如果有特别要求，比如希望保存残差值和预测值，则可将“保存

残差和预测值”这项打勾即可，通常情况并不需要。

4 SPSSAU 输出结果

回归4分析结果 ● 导出EXCEL表格 打印PDF结果

	非标准化系数		标准化系数	t ●	p ●	VIF ●	R ² ●	调整R ² ●	F ●
	B ●	标准误 ●	Beta ●						
常数	0.723	0.236	-	3.069	0.002**	-	0.509	0.504	102.055**
淘宝客服服务态度	0.505	0.067	0.482	7.538	0.000**	1.643			
淘宝商家服务质量	0.312	0.066	0.304	4.754	0.000**	1.643			
因变量：淘宝忠诚度									
* p<0.05 ** p<0.01									

5 文字分析

上表使用回归分析去研究“淘宝客服服务态度”，“淘宝商家服务质量”对于“淘宝忠诚度”的影响关系情况。从上表可以看到：

模型通过 F 检验 ($P < 0.01$)，意味着研究模型具有意义，“淘宝客服服务态度”，“淘宝商家服务质量”这两项中至少一项，会对“淘宝忠诚度”产生影响关系。模型 R 平方值为 0.509，意味着“淘宝客服服务态度”，“淘宝商家服务质量”这两项可以解释“淘宝忠诚度”的 50.9% 变化原因。以及模型公式为：淘宝忠诚度 = $0.723 + 0.505 * \text{淘宝客服服务态度} + 0.312 * \text{淘宝商家服务质量}$ 。另外，针对模型的多重共线性进行检验发现，模型中 VIF 值全部均小于 5，意味着不存在着共线性问题，模型较好。具体分析可知：

“淘宝客服服务态度”的回归系数值为 0.505，并且呈现出 0.01 水平的显著性 ($P < 0.01$)，意味着“淘宝客服服务态度”会对“淘宝忠诚度”产生显著的正向影响关系。以及“淘宝商家服务质量”的回归系数值为 0.312，并且呈现出 0.01 水平的显著性 ($P < 0.01$)，意味着“淘宝商家服务质量”会对“淘宝忠诚度”产生显著的正向影响关系。

总结分析可知：“淘宝客服服务态度”，“淘宝商家服务质量”这两项全部均

会对淘宝忠诚度产生显著的正向影响关系。

6 剖析

通常来说,回归分析之前还需要使用相关分析研究相关关系,首先要保证有相关关系之后,再进行回归分析较为科学。 R^2 值越大越好,但此指标相对来讲重要性不是特别高,更多关注于 X 的显著性,便于讨论 X 对于 Y 的影响关系情况。

WWW.SPSSAU.COM