

SSPSSAU 分析手册

SPSSAU 分析方法	1
基本统计说明	1
1 频数分析	3
2 交叉(卡方)分析	4
3 描述分析	4
4 分类汇总	5
5 相关分析	5
6 回归分析	6
7 聚类分析	7
8 因子分析	7
9 信度分析	8
10 效度分析	9
11 方差分析	10
12 T 检验	11
13 多重响应	11
14 事后检验	12
15 单样本 T 检验	12
16 配对 T 检验	13
17 逐步回归分析	13
18 分层回归分析	14
19 正态性检验	15
20 非参数检验	16
21 双因素方差分析	16
22 二元 Logit 分析	17

SPSSAU 分析方法

针对 22 大类分析方法进行说明,分别包括每类分析的功能,指标和举例说明等。

基本统计说明

数据类型是每类分析方法的基石,区分好数据类型,便可找到合适的分析方法。基本统计名词概念,可有助于理解分析结果指标意义。

1. 数据分两类,定量和定类,区别为数字大小是否具有比较意义

术语	说明	举例
定量数据	数字大小具有比较意义	您对天猫的满意度情况(非常不满意,比较不满意,中立,比较满意,非常满意)

定类数据	数字大小代表分类	性别(男和女),专业(文科、理科、工科)
------	----------	----------------------

- 定量:数字有比较意义,比如数字越大代表满意度越高,量表为典型定量数据
- 定类:数字无比较意义,比如性别,1代表男,2代表女

用户可结合数据类型及相应需求选择使用对应的分析方法,如下表所示:

分析方法	功能	一句话说明	数据类型
频数	百分比	男女比例分别多少	定类
交叉(卡方)	差异关系	不同性别人群是否抽烟差异	<ul style="list-style-type: none"> ● X(定类) ● Y(定类)
描述	平均值	平均身高	定量
分类汇总	差异关系		<ul style="list-style-type: none"> ● X(定类) 【可选】 ● Y(定量/定类) 【可选】
相关	相关关系	身高和体重有没有关系	<ul style="list-style-type: none"> ● Y(定量) 【可选】 ● X(定量) 【可选】
回归	影响关系	身高影响体重吗?	<ul style="list-style-type: none"> ● Y(定量) ● X(定量/定类)
聚类	人群分类	300个人分成几类?	定量
因子	<ul style="list-style-type: none"> ● 浓缩 ● 权重 	<ul style="list-style-type: none"> ● 30句话概述成5个关键词(因子) ● 5个关键词(因子)分别代表30句话的信息比重? 	定量
信度	可靠性	数据真实吗?	定量
效度	有效性	数据有效吗?	定量
方差	差异关系	不同收入群体的身高是否有差异?	<ul style="list-style-type: none"> ● X(定类) ● Y(定量)
T检验	差异关系	不同性别群体的身高是否有差异?	<ul style="list-style-type: none"> ● X(定类) ● Y(定量)
多重响应	百分比	多选题的选择比例情况如何	<ul style="list-style-type: none"> ● X(定类) 【可选】 ● 多选题选项
事后检验	差异关系	不同收入群体的身高详细差异情况?	<ul style="list-style-type: none"> ● X(定类) ● Y(定量)
单样本T检验	差异关系	身高是否明显等于1.8	定量
配对T检验	差异关系	注射新药和没有注射的两组老鼠,血压一样吗	<ul style="list-style-type: none"> ● 配对1(定量) ● 配对2(定量)
逐步回归	影响关系	帮我自动找出影响身高的因素	<ul style="list-style-type: none"> ● Y(定量) ● X(定量/定类)
分层回归	影响关系	身高对于体重的影响,再加入饮食习惯,看看饮食习惯对体重的影响有多严重	<ul style="list-style-type: none"> ● Y(定量) ● 分层1(定量/定类) ● 分层2(定量/定类) ● 分层3(定量/定类) ● 分层4(定量/定类)
正态性检验	正态检验	数据正态吗	定量
非参数检验	差异关系	身高数据不正态时,我想研究收入与身高的差异关系	<ul style="list-style-type: none"> ● Y(定类) ● X(定量)
双因素方差	差异关系	性别和地区对于身高的差异	<ul style="list-style-type: none"> ● Y(定量) ● X(定类,2个)

二元 Logit	影响关系	哪些因素影响人们是否购买电影票	<ul style="list-style-type: none"> ● Y(定类) ● X(定量/定类)
----------	------	-----------------	---

2. P值:显著性值或 Sig 值,描述某事情发生的概率

如果 P 值小于 0.01 即说明某件事情的发生至少有 99% 的把握,如果 P 值小于 0.05 (并且大于 0.01) 则说明某件事情的发生至少有 95% 的把握。

研究人员想研究不同性别人群的购买意愿是否有明显的差异,如果对应的 P 值小于 0.05,则说明呈现出 0.05 水平的显著性差异,即说明不同性别人群的购买意愿有着明显的差异,而且对此类差异至少有 95% 的把握。绝大多数研究均希望 P 值小于 0.05,即说明有影响,有关系,或者有差异等。

- 常见标准:0.01 和 0.05,分别代表某事情发生至少有 99% 或 95% 的把握
- 语言表述:0.01 或 0.05 水平显著
- 符号标示:0.01 使用 2 个*号表示,0.05 使用 1 个*号表示

3. 量表,通常指李克特量表,测量样本对于某构念(通俗讲即某事情)的态度或看法

量表答项类似于“非常同意”、“同意”、“不一定”、“不同意”、“非常不同意”,也或者“非常满意”、“比较满意”、“中立”、“比较不满意”、“非常不满意”等。大多数统计方法均只能针对量表,比如信度分析,效度分析,探索性因子分析(Exploratory Factor Analysis, EFA)等。

量表的尺度形式有多种,常见是五级量表,即五个答项,另外还会有七级量表,九级量表或者四级量表等。

1 频数分析

频数分析用于计算定类数据的选择频数和比例,频数分析常用于样本基本背景信息统计,以及样本特征和基本态度情况分析。

分析项	频数分析说明
性别	数字 1 代表男,数字 2 代表女,数字分别表示两个类别;男和女分别的选择百分比多少?
年龄	数字 1 代表 20 岁以下,数字 2 代表 20~30 岁,数字 3 代表 30 岁以上;数字代表三个不同类别,三个类别人群的百分比是多少?

分析结果表格示例如下:

问题	选项	频数	百分比 (%)
性别	男	82	41.4
	女	116	58.6
年龄	20 以下	62	31.3
	20 到 30 岁之间	33	16.7
	30 岁以上	103	52.0

2 交叉(卡方)分析

卡方分析(交叉表分析,列联表分析),用于分析定类数据与定类数据之间的关系情况.例如研究人员想知道两组学生对于手机品牌的偏好差异情况.

首先判断 p 值是否呈现出显著性,如果呈现出显著性,则说明两组数据具有显著性差异,具体差异可通过选择百分比进行对比判断.

分析项	卡方检验说明
学历,网购平台偏好	不同学历样本人群,他们网购平台偏好是否有差异?

分析结果表格示例如下:

	选项	学历		总计	X^2	p
		本科以下	本科			
分析项 1	选项 1	1(1.8)	0(0.0)	1(0.9)	13.251	0.017*
	选项 2	6(10.5)	6(10.3)	12(10.4)		
	选项 3	10(17.5)	11(19.0)	21(18.3)		
	选项 4	30(52.6)	36(62.1)	66(57.4)		
总计		57(100.0)	58(100.0)	115(100.0)		

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

3 描述分析

对于定量数据,比如量表评分(非常不满意,不满意,非常满意等),或者身高体重的值,可以通过描述性分析计算数据的集中性特征(平均值)和波动性特征(标准差值),同时 SPSSAU 系统还提供最大最小值,以及中位数值.描述性分析通常用于研究量表数据的基本认知情况分析,使用平均值去表述样本对于量表数据的整体态度情况.

分析项	描述分析说明
网购满意度	数字 1 代表非常不满意,2 代表比较不满意,3 代表一般,4 代表比较满意,5 代表非常满意;则可通过描述分析计算平均得分,描述整体满意情况如何.

分析结果表格示例如下:

	样本量	最小值	最大值	平均值	标准差	中位数
分析项 1	198	1.57	5.00	3.43	0.76	0.76
分析项 2	198	2.00	5.00	3.93	0.86	0.86
分析项 3	198	2.00	5.00	3.84	0.90	0.90
分析项 4	198	1.00	5.00	3.32	1.01	1.01

4 分类汇总

分类汇总用于交叉性研究,结合 X 的放置情况,以及汇总类型的选择情况,最终涉及四种情况,如下表。

X 的放置情况	汇总类型	对应方法
X (不放置)	平均值(默认)	描述分析
X (不放置)	百分比	频数分析
X (放置)	平均值(默认)	方差分析
X (放置)	百分比	交叉(卡方)分析

备注:如果用户,建议用户分别参考对应四种方法即可。

5 相关分析

相关分析用于研究定量数据之间的关系情况,包括是否有关系,以及关系紧密程度等.此分析方法通常用于回归分析之前;相关分析与回归分析的逻辑关系为:先有相关关系,才有可能有回归关系。

用户可自由拖拽分析项进入分析列表框,区别仅在于输出格式不同。

相关分析使用相关系数表示分析项之间的关系;首先判断是否有关系(有*号则表示有关系,否则表示无关系);接着判断关系为正相关或者负相关(相关系数大于 0 为正相关,反之为负相关);最后判断关系紧密程度(通常相关系数大于 0.4 则表示关系紧密);相关系数常见有两类,分别是 Pearson 和 Spearman, SPSSAU 系统默认使用 Pearson 相关系数。

分析项	相关分析说明
网购满意度,重复购买意愿	网购满意度和重复购买意愿之间是否有关系,关系紧密程度如何?

格式 1 (当仅放入一个框中时):

	平均值	标准差	分析项 1	分析项 2	分析项 3	分析项 4	分析项 5
分析项 1	3.43	0.76	1				
分析项 2	3.93	0.86	.673**	1			
分析项 3	3.84	0.90	.740**	.712**	1		
分析项 4	3.32	1.01	.681**	.705**	.642**	1	
分析项 5	3.03	1.09	.520**	.666**	.489**	.604**	1

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

格式 2 (两个框均放置项时):

	分析项 1	分析项 2	分析项 3
分析项 4	.641**	.705**	.242**
分析项 5	.520**	.656**	.589**

分析项 6	.321**	.875**	.242**
分析项 7	.611**	.705**	.342**
分析项 8	.530**	.826**	.589**

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

备注：通常情况下会使用格式 1，如果希望格式 2，则右侧两个框中均需要放置分析项。单从相关分析方法角度看，其并不区分 X 和 Y，但从实际意义上看，通常是研究 X 和 Y 的相关关系。

6 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到样本重复购买意愿?网购满意度越高,是否重复购买意愿也会越高?

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			

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

7 聚类分析

聚类分析用于将样本进行分类处理,通常是以定量数据作为分类标准;用户可自行设置聚类数量,如果不进行设置,系统会提供默认建议;通常情况下,建议用户设置聚类数量介于 3~6 个之间.

- 第一步:进行聚类分析设置
- 第二步:结合不同聚类类别人群特征进行类别命名

分析项	聚类分析说明
网购满意度 20 个题项	根据网购满意度情况判定,当前市场上共有几类人群?比如满意度差,一般,满意度高三类人群

分析结果表格示例如下:

聚类类别	频数	百分比 (%)
类别 1	82	41.4
类别 2	61	30.8
类别 3	55	27.8
合计	198	100.0

	聚类类别(平均值±标准差)			F	p
	类别 1(N=82)	类别 2(N=61)	类别 3(N=55)		
分析项 1	3.23±1.33	2.88±0.73	2.63±0.81	3.73	0.03*
分析项 2	2.62±1.48	2.57±1.21	2.32±0.76	0.56	0.58
分析项 3	2.14±1.10	2.16±0.76	2.25±0.95	0.13	0.88
分析项 4	3.31±1.12	3.32±1.02	3.82±0.85	2.67	0.07
分析项 5	3.75±1.06	3.56±0.80	3.82±0.76	0.97	0.38
分析项 6	4.56±0.72	4.42±0.61	4.57±0.68	0.72	0.49
分析项 7	4.45±0.84	4.46±0.66	4.55±0.83	0.19	0.83
分析项 8	4.18±0.96	4.24±0.67	4.36±0.74	0.46	0.63

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

8 因子分析

因子分析(探索性因子分析)用于探索分析项(定量数据)应该分成几个因子(变量),比如 20 个量表题项应该分成几个方面较为合适;用户可自行设置因子个数,如果不设置,系统会以特征根值大于 1 作为判定标准设定因子个数.

因子分析通常有三个步骤;第一步是判断是否适合进行因子分析;第二步是因子与题项对应关系判断;第三步是因子命名.

- 第一步:判断是否进行因子分析,判断标准为 KMO 值大于 0.6.
- 第二步:因子与题项对应关系判断.

因子与题项对应关系判断:假设预期为 3 个因子(变量),分析题项为 10 个;因子与题项交叉共得到 30 个数字,此数字称作“因子载荷系数”(因子载荷系数值表示分析项与因子之间的相关程度);针对每个因子(变量),对应 10 个“因子载荷系数”,针对每个分析项,则有 3 个“因子载荷系数值”(比如 0.765,-0.066,0.093),选出 3 个数字绝对值最大的那个值(0.765),如果其对应因子 1,则说明此题项应该划分在因子 1 下面。

对不合理题项进行删除,共有三种情况;第一类:如果分析项的共同度值小于 0.4,则对应分析项应该作删除处理;第二类:某分析项对应的“因子载荷系数”的绝对值,全部均小于 0.4,也需要删除此分析项;第三类:如果某分析项与因子对应关系出现严重偏差,也需要对该分析项进行删除处理。

● 第三步:因子命名

在第二步删除掉不合理题项后,并且确认因子与题项对应关系良好后,则可结合因子与题项对应关系,对因子进行命名。

分析项	因子分析说明
网购满意度 20 个题项	网购满意度由 20 个题表示,此 20 个题项可浓缩成几个大方面?

分析结果表格示例如下:

	因子载荷系数			共同度
	因子 1	因子 2	因子 3	
分析项 1	0.765	-0.066	0.093	0.598
分析项 2	0.676	0.081	-0.017	0.464
分析项 3	0.657	0.207	-0.205	0.517
分析项 4	0.645	0.271	0.089	0.497
分析项 5	0.501	0.457	0.085	0.467
分析项 6	0.311	0.697	-0.005	0.583
分析项 7	0.226	-0.669	0.130	0.516
分析项 8	0.191	0.644	0.046	0.453
分析项 9	0.476	-0.187	0.542	0.555
分析项 10	0.001	-0.048	0.968	0.939

9 信度分析

信度分析用于测量样本回答结果是否可靠,即样本有没有真实作答量表类题项;信度分析仅针对定量数据.信度系数(Alpha 值,下同)如果在 0.8 以上,则该测验或量表的信度非常好;信度系数在 0.7 以上都是可以接受;如果在 0.6 以上,则该量表应进行修订,但仍不失其价值;如果低于 0.6,量表就需要重新设计题项。

校正的项总计相关性:此指标用于判断题项是否应该作删除处理,如果值小于 0.3,通常应该考虑将对应项进行删除处理;**项已删除的 Alpha 值:**此指标用于判断题项是否应该作删除处理,如果该值明显高于“Alpha”值,此时应该考虑将对应项进行删除处理。

分析项	信度分析说明
-----	--------

网购满意度四个题项	网购满意度由四个题表示,分别是产品满意度,快递满意度,售后满意度,价格满意度(并且均为量表数据);样本对此变量(网购满意度)的回答是否真实可靠?
-----------	--

分析结果表格示例如下:

	校正项总计相关性(CITC)	项已删除的 α 系数	α 系数
分析项 1	0.619	0.872	0.883
分析项 2	0.621	0.872	
分析项 3	0.674	0.865	
分析项 4	0.696	0.863	
分析项 5	0.665	0.866	
分析项 6	0.659	0.867	

10 效度分析

效度用于测量题项(定量数据)设计是否合理,通过因子分析(探索性因子分析)方法进行验证;研究人员心中预期着变量与题项对应关系;进行因子分析后,因子(即变量,使用因子分析时称因子)与题项对应关系;二者预期基本一致时,则说明具有良好效度水平。

如果用户预期分析项可分为几个方面(变量),则用户可自行设置因子个数,如果不设置,系统会以特征根值大于 1 作为判定标准设定因子个数;

● 因子与题项对应关系判断

假设分析题项为 10 个,预期分为 3 个因子(变量);因子与题项交叉共得到 30 个数字,此 30 个数字称作“因子载荷系数”(因子载荷系数值表示分析项与因子之间的相关程度);针对每个分析项对应行,则有 3 个“因子载荷系数值”(比如 0.765,-0.066,0.093),选出 3 个数字绝对值最大的那个值(0.765),如果其对应因子 1,则说明此题项应该划分在因子 1 下面,如下表所示:

	因子载荷系数			共同度
	因子 1	因子 2	因子 3	
分析项 1	0.765	-0.066	0.093	0.598
分析项 2	0.676	0.081	-0.017	0.464
分析项 3	0.657	0.207	-0.205	0.517
分析项 4	0.645	0.271	0.089	0.497
分析项 5	0.501	0.457	0.085	0.467
分析项 6	0.311	0.697	-0.005	0.583
分析项 7	0.226	-0.669	0.130	0.516
分析项 8	0.191	0.644	0.046	0.453
分析项 9	0.476	-0.187	0.542	0.555
分析项 10	0.001	-0.048	0.968	0.939

● 效度分析对不合理题项进行删除

共有三种情况; 第一类:如果分析项的共同度值小于 0.4,则对应分析项应该作删除处理;第二类:某分析项对应的“因子载荷系数”的绝对值,全部均小于 0.4,也需要删除此分析项;第三类:如果某分析项与因子对应关系出现严重偏差,也需要对该分析项进行删除处理.

● 效度分析的其余判断指标

特征根值(通常使用旋转后,以大于 1 作为标准),方差解释率(意义较小),累积方差解释率(通常使用旋转后,以大于 50%作为标准),KMO 值(大于 0.6 作为标准),巴特球形值对应的 sig 值(小于 0.01 作为标准).

分析项	效度分析说明
网购满意度 四个题项	网购满意度由四个题表示,分别是产品满意度,快递满意度,售后满意度,价格满意度(并且四个题项均为量表数据);此四个题表示网购满意度是否合理科学?
网购满意度 20 个题项	网购满意度由 20 个题表示,并且可分为四个大方面,每方面与题项均有着对应关系预期;此种预期与软件出来的结果是否基本一致,如果一致则说明有效,否则需要对题项进行删减处理,删除不合理题项,保留有效题项.

11 方差分析

方差分析(单因素方差分析),用于分析定类数据与定量数据之间的关系情况.例如研究人员想知道三组学生的智商平均值是否有显著差异.方差分析可用于多组数据,比如本科以下,本科,本科以上共三组的差异;而下述 T 检验仅可对比两组数据的差异.

首先判断 p 值是否呈现出显著性,如果呈现出显著性,则说明不同组别数据具有显著性差异,具体差异可通过平均值进行对比判断.

分析项	方差分析说明
学历,网购满意度	不同学历的人群,他们网购满意度是否有差异?

分析结果表格示例如下:

	学历(平均值±标准差)			F	p
	本科以下(N=67)	本科(N=53)	硕士及以上(N=28)		
分析项 1	3.23±1.33	2.88±0.73	2.63±0.81	3.73	0.03*
分析项 2	2.62±1.48	2.57±1.21	2.32±0.76	0.56	0.58
分析项 3	2.14±1.10	2.16±0.76	2.25±0.95	0.13	0.88
分析项 4	3.31±1.12	3.32±1.02	3.82±0.85	2.67	0.07
分析项 5	3.75±1.06	3.56±0.80	3.82±0.76	0.97	0.38

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

12 T 检验

T 检验(独立样本 T 检验),用于分析定类数据与定量数据之间的关系情况.例如研究人员想知道两组学生的智商平均值是否有显著差异.T 检验仅可对比两组数据的差异,如果为三组或更多,则使用方差分析.如果刚好仅两组,建议样本较少(低于 100 时)使用 T 检验,反之使用方差分析.

首先判断 p 值是否呈现出显著性,如果呈现出显著性,则说明两组数据具有显著性差异,具体差异可通过平均值进行对比判断.

分析项	T 检验说明
性别,网购满意度	不同性别的两类人群,他们网购满意度是否有差异?

分析结果表格示例如下:

	性别(平均值±标准差)		t	p
	男(N=67)	女(N=53)		
分析项 1	3.23±1.33	2.88±0.73	3.73	0.03*
分析项 2	2.62±1.48	2.57±1.21	0.56	0.58

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

13 多重响应(单-多)

SPSSAU 系统中多重响应可以研究 单选题(X)和多选题(Y)之间有关系情况.如果不放置单选题(X),仅放置多选题(Y),则时直接对多选题进行统计.

分析项	多重响应(单-多)分析说明
性别,网购在乎因素(多选)	不同性别人群网购在乎因素的差异情况如何?

多重响应时涉及两个术语名词,分别是响应率和普及率。

- 响应率分析多选题(Y)各选项的相对选择比例情况;比如共 100 个样本,平均每个样本选择 3 项,则总共 100 个样本共选择了 300 个选项.如果某某个选项有 60 个人选择,则时响应率为 $60/300=20\%$
- 普及率分析多选题(Y)各选项的选择普及情况;比如共 100 个样本,某个选项有 60 个人选择,则时普及率为 $60/100=60\%$
- 响应率和普及率的区别在于被除数不一样.响应率加和一定为 100%,普及率加和通常会高于 100%
- 绝大多数情况下是对**普及率**进行分析,如果想深入分析,则可以对响应率也进行分析;

分析结果表格示例如下(SPSSAU 同时会生成饼图/圆环图/柱形图/条形图/折线图等):

交叉汇总表

项	性别 (%)		汇总 (N=177)
	男 (N=149)	女 (N=28)	
多选题选项 1	92 (61.7)	20 (71.4)	112 (63.3)
多选题选项 2	107 (71.8)	17 (60.7)	124 (70.1)
多选题选项 3	85 (57.0)	18 (64.3)	103 (58.2)
多选题选项 4	68 (45.6)	11 (39.3)	79 (44.6)

14 事后检验

事后检验基于方差分析基础上进行;用于分析定类数据与定量数据之间的关系情况.例如研究人员想知道三组学生(本科以下,本科,本科以上)的智商平均值是否有显著差异.比如分析显示三组学生智商有着明显的差异,那具体是本科以下与本科这两组之间,还是本科以下与本科以上两组之间的差异;即具体两两组别之间的差异对比,则称为事后检验;事后检验的方法有多种,系统默认使用常见的 LSD 事后检验法.

首先判断 p 值是否呈现出显著性,如果呈现出显著性,则说明不同组别数据具有显著性差异,具体差异可通过平均值进行对比判断;以及可具体通过事后检验判断具体两两组别之间的差异情况.

分析项	事后检验说明
学历,网购满意度	不同学历人群(本科以下,本科,本科以上)网购满意度不同前提下,具体那两组之间呈现出差异性?

分析结果表格示例如下:

	分析项 X(平均值±标准差)			F	p	事后检验
	选项 1(N=67)	选项 2(N=53)	选项 3(N=28)			
分析项 1	3.23±1.33	2.88±0.73	2.63±0.81	3.73	0.03*	选项 1>选项 2>选项 3
分析项 2	2.62±1.48	2.57±1.21	2.32±0.76	0.56	0.58	
分析项 3	2.14±1.10	2.16±0.76	2.25±0.95	0.13	0.88	
分析项 4	3.31±1.12	3.32±1.02	3.82±0.85	2.67	0.07	

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

15 单样本 T 检验

单样本 T 检验用于分析定量数据是否与某个数字有着显著的差异性,比如五级量表,3分代表中立态度,可以使用单样本 T 检验分析样本的态度是否明显不为中立状态;系统默认以 0 分进行对比.

首先判断 p 值是否呈现出显著性,如果呈现出显著性,则分析项明显不等于设定数字,具体差异可通过平均值进行对比判断.

分析项	单样本 T 检验说明
网购满意度	样本的网购满意度是否明显不为中立状态,以及是否有明显的满意?

分析结果表格示例如下:

	样本量	最小值	最大值	平均值	标准差	<i>t</i>	<i>p</i>
分析项 1	198	1.57	5.00	3.43	0.76	3.73	0.03*
分析项 2	198	2.00	5.00	3.93	0.86	0.56	0.58
分析项 3	198	2.00	5.00	3.84	0.90	0.13	0.88

16 配对 T 检验

配对 T 检验,用于配对定量数据之间的差异对比关系.例如在两种背景情况下(有广告和无广告);样本的购买意愿是否有着明显的差异性;配对 T 检验通常用于实验研究中.

首先判断 *p* 值是否呈现出显著性,如果呈现出显著性,则说明配对数据具有显著性差异,具体差异可通过平均值进行对比判断.

分析项	配对 T 检验说明
有广告时的购买意愿; 无广告时的购买意愿	有广告和无广告两种背景情况下时,购买意愿是否有着显著性差异?

分析结果表格示例如下:

	配对(平均值±标准差)		差值	<i>t</i>	<i>p</i>
	配对 1	配对 2			
分析项 1 配对 分析项 2	3.23±1.33	2.88±0.73	0.35	3.73	0.03*
分析项 3 配对 分析项 4	2.62±1.48	2.57±1.21	0.05	0.56	0.58
分析项 5 配对 分析项 6	2.14±1.10	2.16±0.76	-0.02	0.13	0.88

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

17 逐步回归分析

逐步回归分析研究 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,X 可以为多个,但并非所有 X 均会对 Y 产生影响;当 X 个数很多时,可以让系统自动识别出有影响的 X;这一自动识别分析方法则称为逐步回归分析;如果全部 X 均没有显著性,此时系统默认返回 回归分析结果.

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

首先分析最终余下的 X 情况;以及被模型自动排除在外的 X;接着对模型拟合情况(比如 R 平方为 0.3,则说明所有余下 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).

- 第二步:分析 X 的显著性

模型余下的 X 一定具有显著性;具体分析 X 的影响关系情况即可.

- 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	逐步回归分析说明
网购满意度,重复购买意愿	网购满意度 20 项;其中具体那几项会影响到样本重复购买意愿?20 项过多,让软件自动删除掉没有影响的项,余下有影响的项

分析结果表格示例如下:

	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	0.774	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			

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

18 分层回归分析

分层回归的核心即为回归分析;区别在于分层回归可分为多层;比如第一次放入 4 个 X;第二层放入 3 个 X;第 3 层放入 2 个 X;每一层均在上一层基础上放入更多项;那放入的更多项是否对模型有解释力度,此则为分层回归关心的问题;分层回归通常用于中介作用或者调节作用研究中.

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

分析描述各个模型的拟合情况,以及 R 平方值的变化情况.

- 第二步:分析 X 的显著性

结合自身需要;分析具体 X 的显著性情况.

- 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

结合不同模型的对比,得出相关结论(比如中介作用或者调节作用研究的相关结论).

分析项	分层回归分析说明
网购满意度,重复购买意愿	第一层放入性别,学历,年龄,收入等基本个人信息;第二层放入核心研究项;深入说明核心研究项对于重复购买意愿的影响情况?(核心研究项加入后,R 方有明显变化)

分析结果表格示例如下:

	分层 1		分层 2		分层 3	
	B	标准误	B	标准误	B	标准误
常数	0.212	0.036	0.312	0.041	0.215	0.04
分析项 1	0.308**	0.047	0.316**	0.054	0.223**	0.055
分析项 2	0.177**	0.046	0.219**	0.053	0.165**	0.052
分析项 3	-	-	-0.075	0.067	-0.137*	0.066
分析项 4	-	-	-	-	0.191**	0.058
分析项 5	-	-	-	-	0.192**	1.058
R^2	0.506		0.341		0.386	
调整 R^2	0.501		0.334		0.378	
F	98.287**		49.572**		48.186**	
ΔR^2	-		0.341		0.046	
ΔF	-		49.572**		28.463**	

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

19 正态性检验

正态性检验用于分析数据是否呈现出正态性特质。

分析项	正态性检验说明
购买意愿	样本的购买意愿情况是否符合正态性特质呢？

正态性特质是很多分析方法的基础前提，如果不满足正态性特质，则应该选择其它的分析方法，SPSSAU 将常见的分析方法正态性特质要求归纳如下表（包括分析方法，以及需要满足正态性的分析项，如果不满足时应该使用的分析方法）：

分析方法	说明	正态性条件	如果不满足，分析方法使用	备注
方差分析	X 对于 Y 的差异	Y 需要满足正态性	非参数检验	可考虑对 Y 进行生成变量转换，比如开根号，自然对数等；希望数据满足正态性
相关分析	分析项相关关系情况	分析项均需满足正态性	如果满足，使用 Pearson 相关系数，如果不满足使用 Spearman 相关系数	如果不满足正态性，则使用 Spearman 相关系数

分析结果表格示例如下：

名称	样本量	Kolmogorov-Smirnov 检验		Shapiro-Wilk 检验	
		统计量	p	统计量	p
购买意愿	17402	0.268	0.000**	0.868	0.000**

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

特别提示：

1：如果样本量大于 50，则应该使用 Kolmogorov-Smirnov 检验结果，反之则使用 Shapiro-Wilk 检验的结果。

2: 如果 P 值大于 0.05, 则说明具有正态性特质, 反之则说明数据没有正态性特质。

3: 如果是问卷研究, 数据很难满足正态性特质, 而实际研究中却也很少使用 不满足正态性分析时的分析方法, SPSSAU 认为有以下三点原因,

- 参数检验的检验效能高于非参数检验, 比如方差分析为参数检验, 所以很多时候即使数据不满足正态性要求也使用方差分析
- 如果使用非参数检验, 呈现出差异性, 则需要对比具体对比差异性 (但是非参数检验的差异性不能直接用平均值描述, 这与实际分析需求相悖, 因此有时即使数据不正态, 也不使用非参数检验, 或者 Spearman 相关系数等)
- 理想状态下数据会呈现出正态性特质, 但这仅会出现在理想状态, 现实中的数据很难出现正态性特质 (尤其是比如问卷数据)

20 非参数检验

非参数检验用于研究定类数据与定量数据之间的关系情况。例如研究人员想知道不同性别学生的购买意愿是否有显著差异。如果购买意愿呈现出正态性, 则建议使用方差分析, 如果购买意愿没有呈现出正态性特质, 此时建议可使用非参数检验。

分析项	非参数检验分析说明
性别, 购买意愿	不同性别人群购买意愿差异情况如何? (如果购买意愿正态则使用方差分析, 如果不正态则使用非参数检验)

分析结果表格示例如下:

	性别(平均值)		MannWhitney 检验统计量	p
	男(N=10607)	女(N=6795)		
购买意愿	3.44	3.21	-16.44	0.00**

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

说明:

如果 X 组别为两组, 比如上表中男和女共两组, 则应该使用 MannWhitney 统计量, 如果组别超过两组, 则应该使用 Kruskal-Wallis 统计量结果。SPSSAU 自动为你选择 MannWhitney 或者 Kruskal-Wallis 统计量。

21 双因素方差分析

双因素方差分析, 用于分析定类数据(2 个)与定量数据之间的关系情况。例如研究人员性别, 学历对于网购满意度的差异性; 以及男性或者女性时, 不同学历是否有着网购满意度差异性; 或者同一学历时, 不同性别是否有着网购满意度差异性。

双因素方差分析是相对于单因素方差分析而言; 区别在于 X(定类数据)的个数; 如果仅为一个称为单因素方差; 两个为双因素方差; 单因素方差分析(即方差分析)的使用非常普遍; 但双因素方差更多用于实验研

究.

首先判断 p 值是否呈现出显著性,如果呈现出显著性,则说明 X 或者交互项对于 Y 有着差异(影响)关系.

分析项	双因素方差分析说明
性别,学历,网购满意度	不同性别或者不同学历的人群,他们网购满意度是否有差异?同一性别时,不同学历人群网购满意度是否有差异?同一学历时,不同性别人群网购满意度是否有差异?

分析结果表格示例如下:

	数字	标签	样本量	平均值	标准差
分析项 1	1	男	50	3.45	0.98
	2	女	50	3.56	0.89
分析项 2	1	本科以下	30	3.12	0.97
	2	本科	30	3.21	0.65
	3	硕士	40	3.78	0.76

差异源	平方和	df	均方	F	p
截距	2.373	1	2.373	2.009	0.103
分析项 1	3.375	1	3.375	4.009	0.003*
分析项 2	68.724	1	68.724	81.646	0.003*
分析项 1*分析项 2	4.479	1	4.479	5.322	0.003*
误差	340.062	404	0.842		

22 二元 Logit 分析

二元 Logit 回归分析用于研究 X (定量/定类数据)对于 Y (定类数据,二类)的影响关系;此处 Y 仅能两个数字(0 和 1, 比如 0 代表不购买, 1 代表购买);(回归分析与二元 Logit 回归分析均研究影响关系,区别在于 Y 的数据类型).

分析步骤共为三步,分别是:

- 第一步:首先对模型情况进行分析

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).

- 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.

分析项	回归分析说明
网购满意度 20 个题项, 是否愿意推荐	网购满意度 20 个题项是否会影响到样本推荐意愿;满意度高者,其推荐意愿更高.

分析结果表格示例如下:

	回归系数	标准误	<i>z</i>	<i>p</i>	<i>OR</i> 值	OR 95% CI (下限)	OR 95% CI (上限)
x1	0.683	0.454	1.506	0.132	0.683	-0.206	1.573
x2	0.005	0.295	0.017	0.986	0.005	-0.573	0.583
x3	0.351	0.301	1.165	0.244	0.350	-0.239	0.940
x4	0.534	0.355	1.501	0.133	0.534	-0.163	1.230
x5	-0.235	0.370	-0.634	0.526	-0.235	-0.960	0.491
x6	0.099	0.357	0.278	0.781	0.099	-0.600	0.798
x7	-0.420	0.423	-0.992	0.321	-0.420	-1.249	0.410
截距	0.945	1.648	0.573	0.566	0.945	-2.286	4.175

因变量: Y

Pseudo *R* 平方: 0.045