

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. **正常情况下,效度分析仅仅针对量表数据,非量表题目比如多选,单选性别之类的题目不能进行效度分析。如果一定想分析效度,建议使用‘内容效度’,即用文字详细描述问卷设计的过程,用文字的形式描述清楚问卷是做什么,有什么用处,为什么合理,而且有专家认证,这样就说明问卷设计合理有效。**

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(维度个数),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(维度或变量);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),3 个数字**绝对值**大于 0.4, 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

特别提示:上表格中“分析项 5”分别对应“因子 1”和“因子 2”时,因子载荷系数值分别是 0.501 和 0.457,说明“分析项 5”分别与“因子 1”或“因子 2”的关联性都较高(相对来看与“因子 1”更高一点),但都比较高,所以“分析项 5”属于“因子 1”或“因子 2”均可。此类现象属于‘纠缠不清’。类似上表格中“分析项 9”也属于“纠缠不清”。针对“纠缠不清”现象,SPSSAU 建议暂不处理,一般可接受。

还有一类现象叫做“张冠李戴”,即对应关系出现错误,假设‘分析 1’应该和‘分析 9’,‘分析 10’是同一个维度;明显的上表格中,‘分析项 1’放在‘因子 1’下面,但是‘分析 9’,‘分析 10’都放在‘因子 3’下面。说明‘分析项 1’对应错了位置,本应该在‘因子 3’下面却跑到‘因子 1’下面,因而‘分析项 1’就属于‘张冠李戴’。针对‘张冠李戴’,一定需要将题目删除重新进行分析。

● 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(此现象称作‘张冠李戴’)。

● 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.
--------------	--

SPSSAU 操作截图如下：（如果专业预期量表题目可分为 5 个维度，则首先设置“维度个数”为 5）

开始效度分析



维度个数,建议设置



分析项1

分析项2

分析项3

分析项4

分析项5

分析项6

分析项7

分析项8

分析项9

分析项10

分析项(定量)

WWW.SPSSAU.COM

效度分析案例

Contents

1 背景.....3

2 理论.....3

3 操作.....3

4 SPSSAU 输出结果7

5 文字分析.....7

6 剖析.....7

案例数据下载 <http://study.spssau.com/data-validity.xlsx>

1 背景

当前有一份数据,共涉及 A1~A4,B1~B4,C1~C3,D1~D3 共 14 个量表题,此 14 个题目共分为 4 个维度,分别称作 A, B, C 和 D 维度。现希望对此份数据效度情况进行分析,如果有不合理题项将其进行删除处理。

2 理论

效度是一种概念,其是指研究项(通常比如为问卷量表题目)的设计是否具有科学合理性。通常情况下效度可分为内容效度和结构效度。内容效度是指研究项的设计具有专业性,结合研究目的进行,专家认可等,即使用文字直接描述研究项为什么要这样设计,有什么参考依据,或者有什么专业上的意义,内容效度直接用文字详细描述即可。除开内容效度,还可以使用研究方法进行效度验证,此类效度称做结构效度。

一般情况下,结构效度是使用因子分析方法进行研究。使用研究方法去探索研究项的内在逻辑结构,通常逻辑结构的判断,最终验证效度水平情况。比如本案例中共有 14 个量表题项,4 个维度,而且 14 个题项与 4 个维度之间均有着专业意义上的对应关系情况。那么如果使用研究方法即因子分析进行效度分析时,因子分析结果也显示 4 个维度分别与 14 个量表题项的对应关系,与专业上基本吻合,此时此说明研究方法识别出的逻辑内在结构,与专业意义上的内在逻辑结构有着一致性,此时则说明数据具有有效性。

如果 14 个量表题项与 4 个维度之间的对应关系情况,与专业知识情况不符合,比如 A1 这个题目被划分到了 C 维度下面,此时则说明可能 A1 这个题项应该被删除处理,其出现了‘张冠李戴’现象。因而在进行分析时很可能对部分不合理的题项进行删除处理。

另单独提示:有时候会出现‘纠缠不清’现象,比如 A1 可归属为 A 维度,同时也可归属到 C 维度,这种情况较为正常(称作‘纠缠不清’),需要结合实际情况处理即可,可将 A1 删除,也可不删除,带有一定主观性。效度分析是一个多次重复的过程,比如删除某个或多个题项后,则需要重新再次分析进行对比选择等。最终目的在于:维度与题项对应关系,与专业知识情况基本吻合。

关于维度(因子)和题项对应关系,维度(因子)和题项之间交叉会得到一个‘因子载荷系数’,此系数绝对值大于 0.4,则说明二者有着较强关联性,该题项可以与该维度(因子)有着对应关系。

3 操作

Step1: 第一次分析

本例子中 A1~A4,B1~B4,C1~C3,D1~D3 共 14 个量表题,此 14 个题目共分为 4 个维度,因此在分析前可主动告诉 SPSSAU,此 14 项是四个维度,否则 SPSSAU 会自动判断多少个维度(通常软件自动判断与实际情况有很大出入,所以建议主动设置维度个数)。如下图:

开始效度分析

4

A1

A2

A3

A4

B1

B2

B3

B4

C1

C2

C3

D1

D2

D3

完成后结果如下：

	因子载荷系数				共同度
	因子1	因子2	因子3	因子4	
A1	0.568	0.334	0.440	0.182	0.662
A2	0.401	0.337	0.560	0.226	0.639
A3	0.230	0.115	0.891	0.147	0.882
A4	0.224	0.258	0.816	0.167	0.812
B1	0.282	0.620	0.280	0.245	0.602
B2	0.508	0.508	0.288	0.252	0.663
B3	0.701	0.363	0.210	0.220	0.716
B4	0.740	0.335	0.345	-0.001	0.779
C1	0.325	0.703	0.273	-0.023	0.675
C2	0.325	0.801	0.124	0.138	0.781
C3	0.120	0.760	0.118	0.326	0.712
D1	0.205	0.262	0.261	0.842	0.888
D2	0.596	0.243	0.178	0.585	0.788
D3	0.731	0.171	0.179	0.218	0.643

从上图中可以看出：

- ✓ A1~A4 这 4 项，它们全部对应着因子 3 时，因子载荷系数值均高于 0.4，说明此 4 项应该同属于一个维度，即逻辑上 A1~A4 这 4 项，并没有出现‘张冠李戴’现象。但是 A1 和 A2 这两项出现‘纠缠不清’现象，A1 和 A2 除了可以对应因子 3，也可以放在因子 1 下面。一般出现‘纠缠不清’现象时，暂时保留，先处理清楚‘张冠李戴’问题更好。
- ✓ B1~B4 共 4 项，B2,B3,B4 这 3 项对应着因子 1 下面，但是 B1 却对应着因子 2，因此 B1 这项属于‘张冠李戴’，应该将 B1 删除。B2 同时对应因子 1 和因子 2 均可，属于‘纠缠不清’，暂不处理 B2。
- ✓ C1~C3 共 3 项，此 3 项均对应着因子 2，此 3 项并没有出现‘纠缠不清’或者‘张冠李戴’问题。
- ✓ D1~D3 共 3 项，D3 出现了‘张冠李戴’问题，应该进行删除处理。D2 出现了‘纠缠不清问题’，应该给予关注。

总结上述分析可知：B1 和 D3 这两项出现‘张冠李戴’，应该首先将此两项删除；而 A1，A2，B2，D2 共四项有出现‘纠缠不清现象’，暂时不处理（进行关注即可）。将 B1 和 D3 这两项删除后，进行第二次分析。

Step2: 第二次分析

将 B1 和 D3 这两项删除后，进行第二次分析。结果如下：

	因子载荷系数				共同度
	因子1	因子2	因子3	因子4	
A1	0.655	0.398	0.240	0.235	0.700
A2	0.412	0.555	0.313	0.257	0.642
A3	0.240	0.893	0.095	0.157	0.888
A4	0.241	0.823	0.232	0.178	0.821
B2	0.607	0.254	0.388	0.300	0.674
B3	0.758	0.178	0.254	0.291	0.756
B4	0.801	0.311	0.249	0.073	0.806
C1	0.375	0.274	0.704	0.010	0.712
C2	0.377	0.126	0.778	0.172	0.794
C3	0.104	0.149	0.799	0.340	0.788
D1	0.191	0.255	0.220	0.858	0.886
D2	0.525	0.184	0.222	0.640	0.768
特征根值(旋转前)	6.677	1.071	0.769	0.716	-
方差解释率%(旋转前)	55.643%	8.928%	6.408%	5.963%	-
累积方差解释率%(旋转前)	55.643%	64.571%	70.979%	76.942%	-
特征根值(旋转后)	2.906	2.344	2.334	1.648	-
方差解释率%(旋转后)	24.214%	19.537%	19.454%	13.737%	-
累积方差解释率%(旋转后)	24.214%	43.751%	63.204%	76.942%	-

从上图可知：

- ✓ A1 出现‘张冠李戴’现象，应该删除，A2 出现‘纠缠不清’现象，暂不处理，但应该给予关注。
- ✓ B2~B4 共 3 项没有问题
- ✓ C1~C3 共 3 项没有问题
- ✓ D2 出现‘纠缠不清现象’

总结可知：应该将 A1 先删除后再次进行第 3 次分析，另对 A2 和 D2 这两项给予关注。

Step3: 第三次分析

将 A2 先删除后再次分析结果如下：

	因子载荷系数				共同度
	因子1	因子2	因子3	因子4	
A2	0.377	0.273	0.576	0.298	0.637
A3	0.100	0.213	0.900	0.162	0.891
A4	0.217	0.255	0.826	0.169	0.823
B2	0.395	0.605	0.278	0.301	0.690
B3	0.254	0.785	0.206	0.281	0.801
B4	0.266	0.791	0.347	0.075	0.822
C1	0.721	0.335	0.290	0.023	0.716
C2	0.789	0.348	0.133	0.178	0.794
C3	0.787	0.107	0.135	0.333	0.760
D1	0.227	0.164	0.253	0.864	0.889
D2	0.226	0.526	0.203	0.641	0.780
特征根值(旋转前)	6.075	1.064	0.769	0.696	-
方差解释率%(旋转前)	55.226%	9.670%	6.987%	6.330%	-
累积方差解释率%(旋转前)	55.226%	64.897%	71.883%	78.213%	-
特征根值(旋转后)	2.355	2.340	2.289	1.620	-
方差解释率%(旋转后)	21.407%	21.277%	20.807%	14.723%	-
累积方差解释率%(旋转后)	21.407%	42.684%	63.491%	78.213%	-
KMO值	0.876				-
巴特球形值	1320.467				-
df	55				-
p值	0.000				-

从上表可知：D2 可同时出现在因子 2 和因子 4 下面，但考虑到 D 维度当前仅余下 2 项，因而表示可以接受，A，B，C 和 D 共 4 个维度，它们与题项对应关系良好，与专业情况完全相符。因而最终效度分析结束，总共删除掉 B1，D3 和 A1 共 3 项。

4 SPSSAU 输出结果

	因子载荷系数				共同度
	因子1	因子2	因子3	因子4	
A2	0.377	0.273	0.576	0.298	0.637
A3	0.100	0.213	0.900	0.162	0.891
A4	0.217	0.255	0.826	0.169	0.823
B2	0.395	0.605	0.278	0.301	0.690
B3	0.254	0.785	0.206	0.281	0.801
B4	0.266	0.791	0.347	0.075	0.822
C1	0.721	0.335	0.290	0.023	0.716
C2	0.789	0.348	0.133	0.178	0.794
C3	0.787	0.107	0.135	0.333	0.760
D1	0.227	0.164	0.253	0.864	0.889
D2	0.226	0.526	0.203	0.641	0.780
特征根值(旋转前)	6.075	1.064	0.769	0.696	-
方差解释率%(旋转前)	55.226%	9.670%	6.987%	6.330%	-
累积方差解释率%(旋转前)	55.226%	64.897%	71.883%	78.213%	-
特征根值(旋转后)	2.355	2.340	2.289	1.620	-
方差解释率%(旋转后)	21.407%	21.277%	20.807%	14.723%	-
累积方差解释率%(旋转后)	21.407%	42.684%	63.491%	78.213%	-
KMO值	0.876				-
巴特球形值	1320.467				-
df	55				-
p值	0.000				-

将最终处理好的结果作为最终结果，并且进行描述即可，由于中间过程可能会非常多，因而不需要具体报告中间过程表格。

5 文字分析

使用探索性因子分析进行效度分析，14 个量表题目共分为 4 个维度；在进行第一次分析时发现 B1 和 D3 这两项，它们与维度对应关系情况，和专业情况不符合（另外一种写法是因子载荷系数值小于 0.4，本应该对应该维度时因子载荷系数大于 0.4，但是却低于 0.4 了），因而删除此两项后进行第二次分析，第二次分析时删除掉 A1 这个题项。最终余下 11 个题项，此 11 项与维度对应关系情况良好，与专业预期相符。并且从上表可知：KMO 值为 0.876>0.6，通过巴特球形检验，累积方差解释率值为 78.213%，说明 4 个维度可以提取出大部分题项信息。因而综合说明研究数据具有良好的结构效度水平。

6 剖析

效度分析涉及以下几个关键点，分别如下：

- ✓ 区分‘张冠李戴’和‘纠缠不清’这两种现象；
- ✓ 通常先处理掉‘张冠李戴’现象，然后自然的到后面‘纠缠不清’问题会减少，后续即使有个别‘纠缠不清’现象时，也基本可以接受。
- ✓ 通常需要首先设置好维度个数，而不是让软件自动识别。
- ✓ 需要来回多次进行对比，找出最佳结果。
- ✓ 同样的样本数据，不同的分析思路，有可能出现结果不相同，但只要符合效度的思维概念即可。
- ✓ 如果无论如何效度分析均不达标，可考虑以单个维度为准，比如本例中有 4 个维度，则进行 4 次效度分析，然后将 4 次分析结果合并规范。（PS:此类做法的思路是，题目同属于 1 个维度，仅需要删除掉因子载荷系数值低于 0.4 的题项即可，不用考虑多个维度间的逻辑对应关系。）