我们使用斯皮尔曼模型来测量评级与评论的相关性，斯皮尔曼模型计算最后的结果是一种秩相关系数，它是跟据原始数据的排序进行求解。

1.我们将每一条数据根据评论时间进行排序

2.将每一条数据的star\_rating和review的得分情况提取出来作为两个列向量R(Xi)和R(Yi)

3.将R(Xi)和R(Yi)输出到下面公式中：

d = \sum_{i=1}^{N} \left | R\left ( X_i\right )-R\left ( Y_i \right ) \right |^{2}

4.最后，根据下面公式计算出两个列向量之间的相关性Rs,得到结果如表1

Rs = 1-\frac{6 \times d}{N\times\left (N^2 -1 \right )}

|  |  |
| --- | --- |
| product | coefficient |
| hair\_dryer | 0.437746 |
| microwave | 0.559344 |
| pacifier | 0.372332 |

表1.系数

从表中可以得出，三种产品的相关系数大于0.05，所以，specific star ratings 会影响用户去编写某种类型的评论。并且我们在数据中发现部分评论确实会受到之前的评论的影响。并且我们在reviews中发现部分评分如表2中的评论。

|  |  |
| --- | --- |
| Review\_id | reviews |
| R3NQ2GXMX8FDFI | Here's a novel idea!! Read the reviews BEFORE ... |
| |  |  | | --- | --- | | |  | | --- | | RK1L6FAK78AN4 | | | - DO NOT BUY - the 5 star reviews are fake |
| |  | | --- | | ROTACUL0CWK71 | | I don't understand the great reviews for this dryer |
| |  | | --- | | R3EPB70VVPJALX | | Believe the Reviews!!! |
| ⋯ | ⋯ |

表2 部分评论

从表中可以看出，确实有写review会受到其他评论的影响。