

分析步骤

1. 对数据进行Shapiro-Wilk（小数据样本，一般样本数5000以下）或者Kolmogorov–Smirnov（大数据样本，一般样本数5000以上）检验，查看其显著性；
2. 若不呈现出显著性（p值大于0.05或0.01，严格为0.05，不严格为0.01），说明符合正态分布，反之说明不符合正态分布；
- PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布。

正态性检验分析结果

输出结果1：总体描述结果

复制

变量名	样本量	中位数	平均值	标准差	偏度	峰度	S-W检验 ?	K-S检验 ?
话题热点词个数	21	0.667	0.702	0.196	0.098	-0.795	0.931(0.142)	0.126(0.8505604819879713)
讨论量	21	16	139.429	510.621	4.563	20.871	0.266(0.000***)	0.473(0.00007789757125447454)
阅读量(万)	21	51.6	341	1320.688	4.575	20.948	0.254(0.000***)	0.508(0.000015259576791886003)
评论量	21	22	60.952	137.314	4.165	18.173	0.414(0.000***)	0.384(0.002668558797805587)

注：***、**、*分别代表1%、5%、10%的显著性水平

图表说明：

上表展示了话题热点词个数、讨论量、阅读量(万)、评论量描述性统计和正态性检验的结果，包括中位数、平均值等，用于检验数据的正态性。

1. 通常正态分布的检验方法有两种，一种是Shapiro-Wilk检验，适用于小样本资料（样本量≤5000）；另一种是Kolmogorov–Smirnov检验，适用于大样本资料（样本量>5000）；

2. 若呈现显著性（p<0.05或0.01），则说明拒绝原假设（数据符合正态分布），该数据不满足正态分布，反之则说明该数据满足正态分布。

PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布。

智能分析

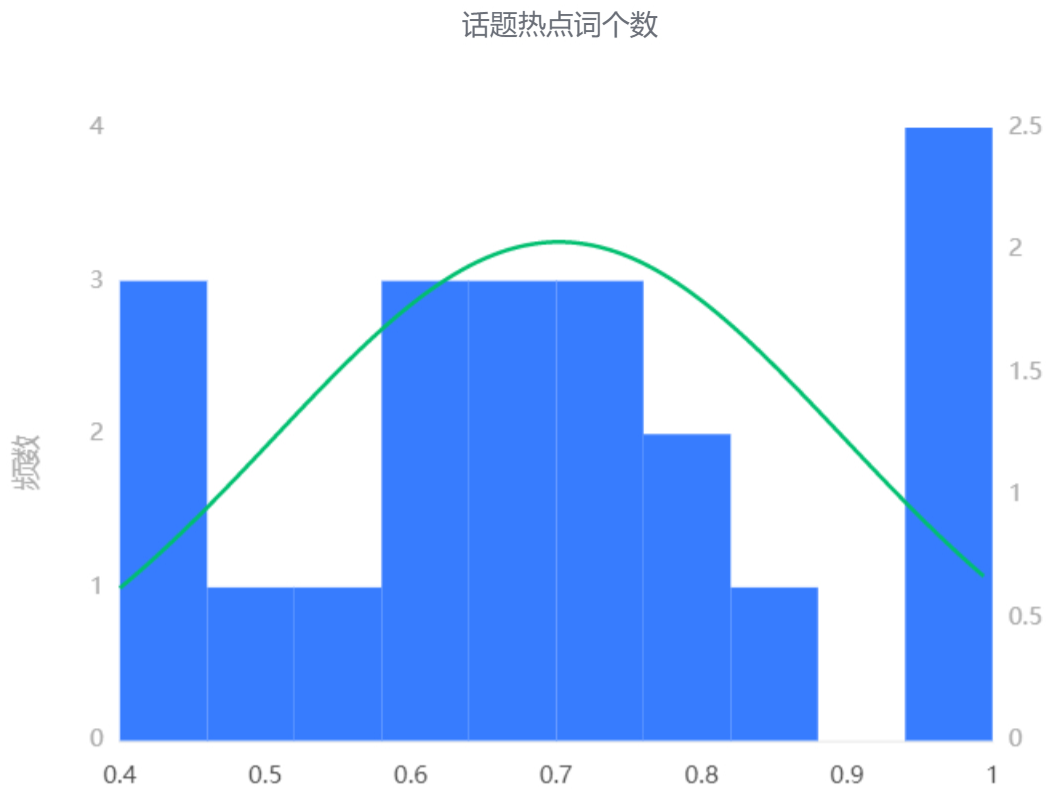
分析项：话题热点词个数样本N < 5000，采用S-W检验，显著性P值为0.142，水平不呈现显著性，不能拒绝原假设，因此数据满足正态分布。

分析项：讨论量样本N < 5000，采用S-W检验，显著性P值为0.000***，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。（其峰度（20.871）绝对值大于10并且偏度（4.563）绝对值大于3，可以结合正态分布直方图、PP图或者QQ图进行进一步分析。）

分析项：阅读量(万)样本N < 5000，采用S-W检验，显著性P值为0.000***，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。（其峰度（20.948）绝对值大于10并且偏度（4.575）绝对值大于3，可以结合正态分布直方图、PP图或者QQ图进行进一步分析。）

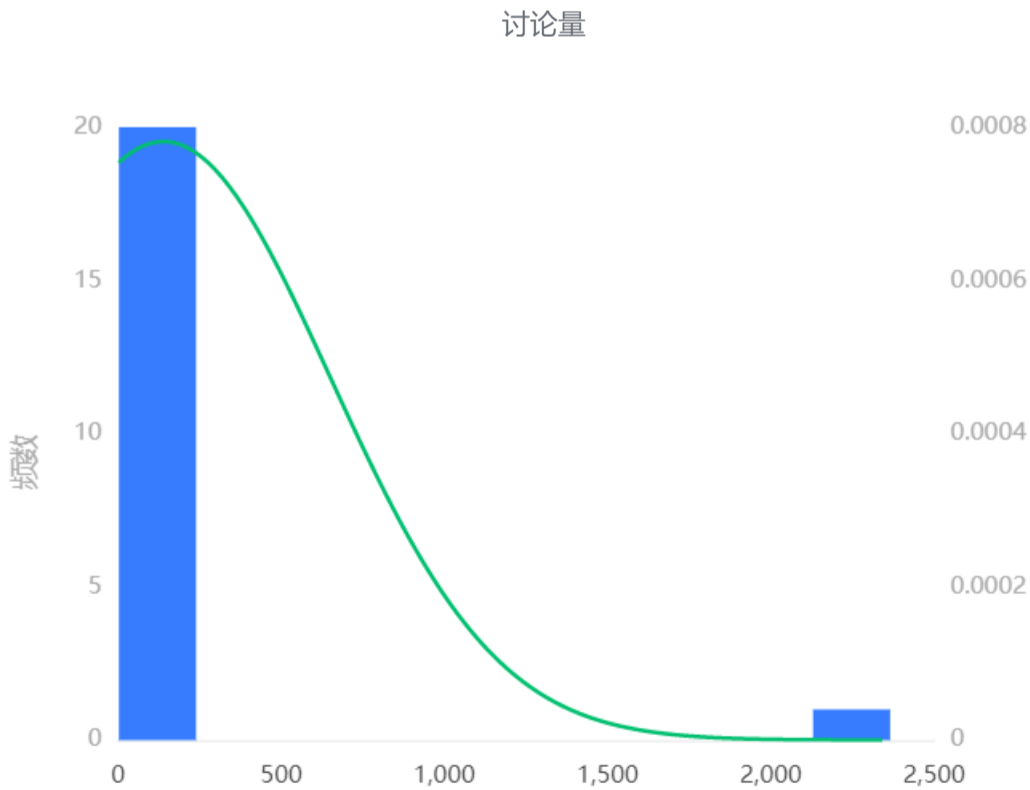
分析项：评论量样本N < 5000，采用S-W检验，显著性P值为0.000***，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。（其峰度（18.173）绝对值大于10并且偏度（4.165）绝对值大于3，可以结合正态分布直方图、PP图或者QQ图进行进一步分析。）

输出结果2：正态性检验直方图



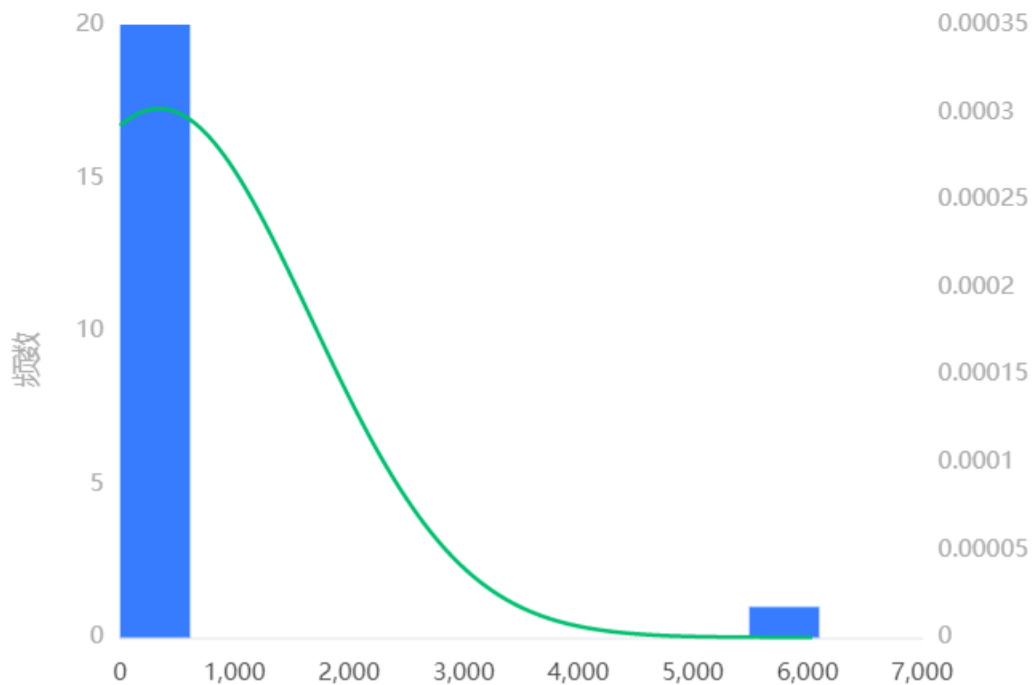
图表说明：

上图展示了话题热点词个数数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。



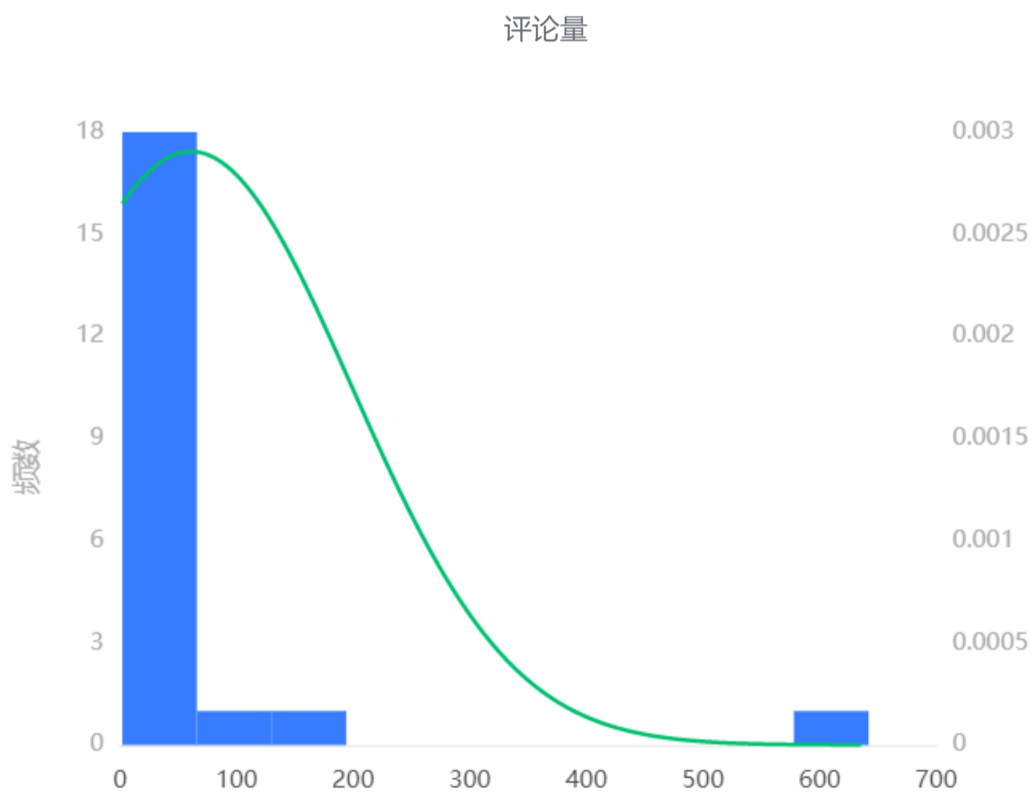
图表说明：

上图展示了讨论量数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。



图表说明：

上图展示了阅读量(万)数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

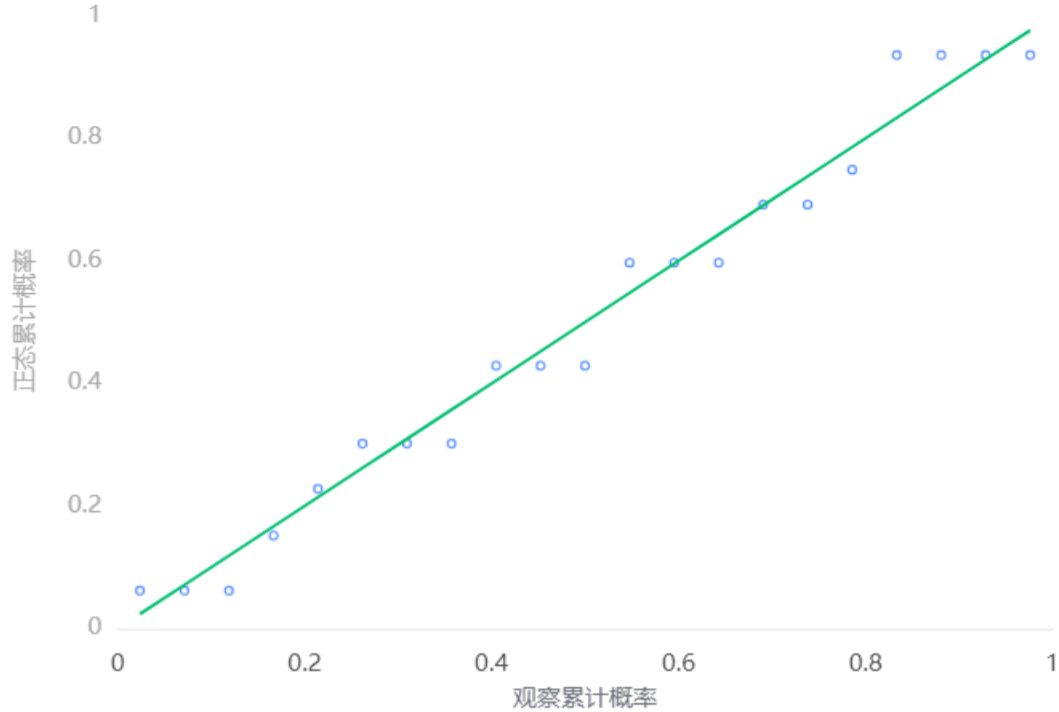


图表说明：

上图展示了评论量数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

输出结果3：正态性检验P-P图

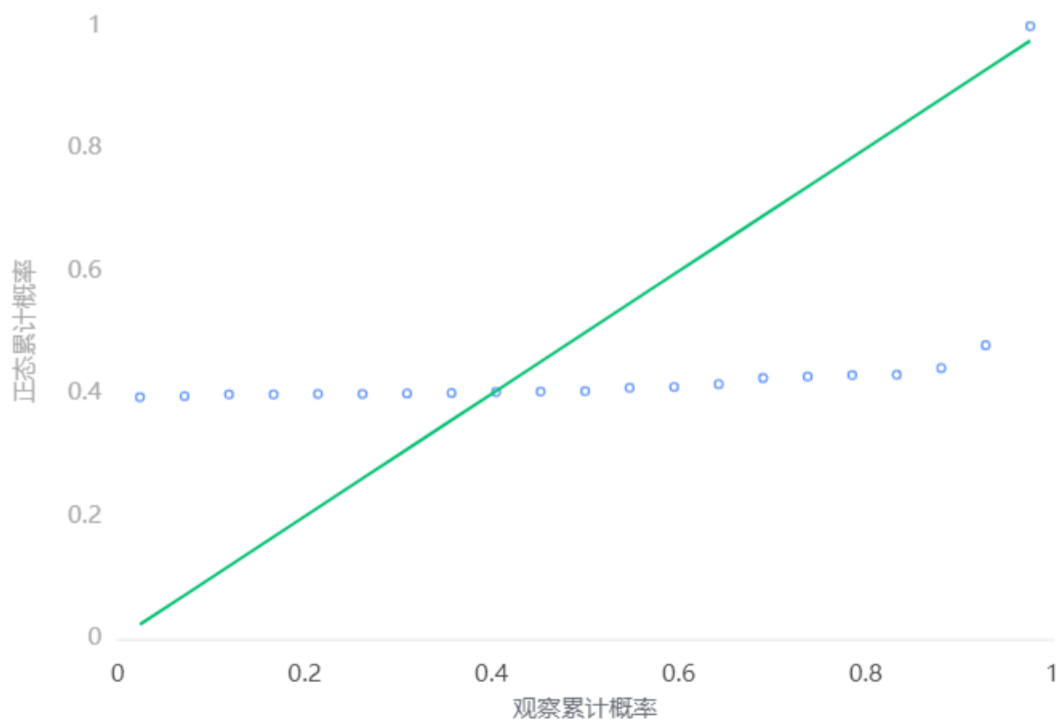
话题热点词个数



图表说明：

上图是话题热点词个数计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

讨论量

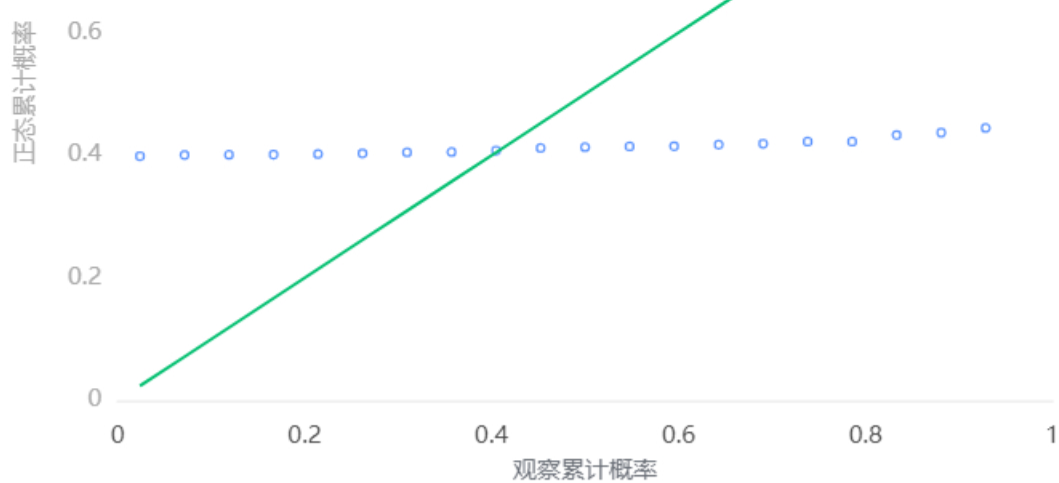


图表说明：

上图是讨论量计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

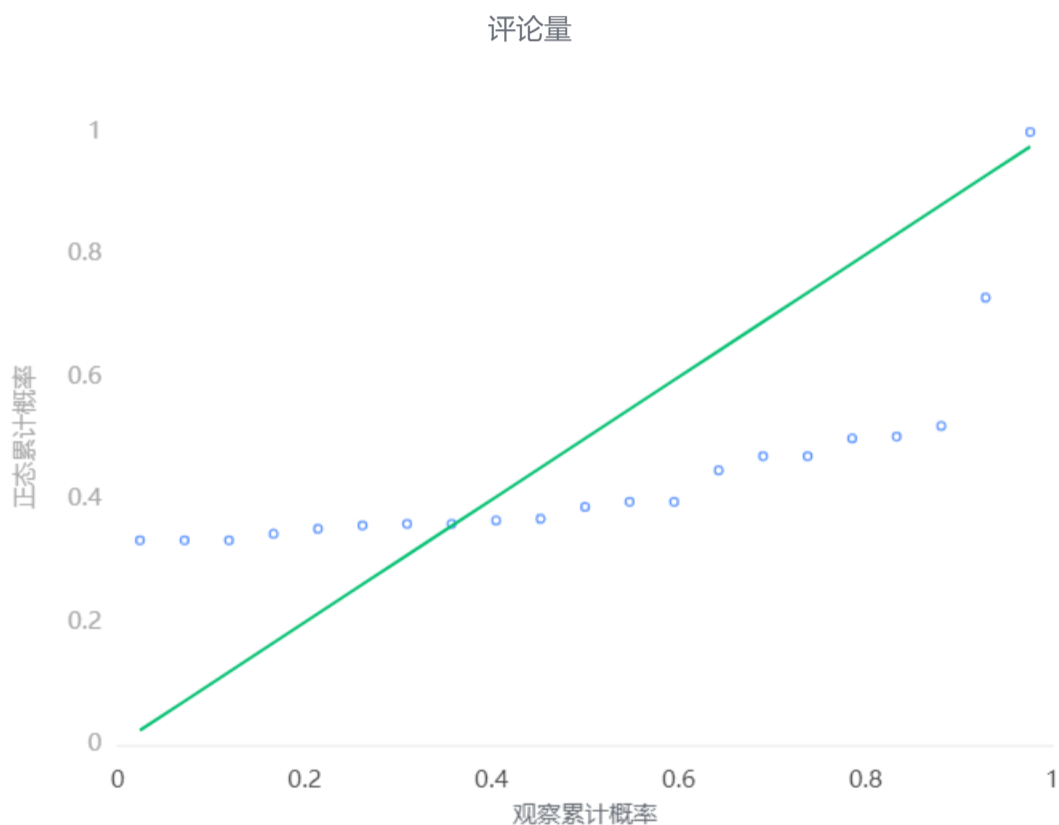
阅读量(万)





图表说明：

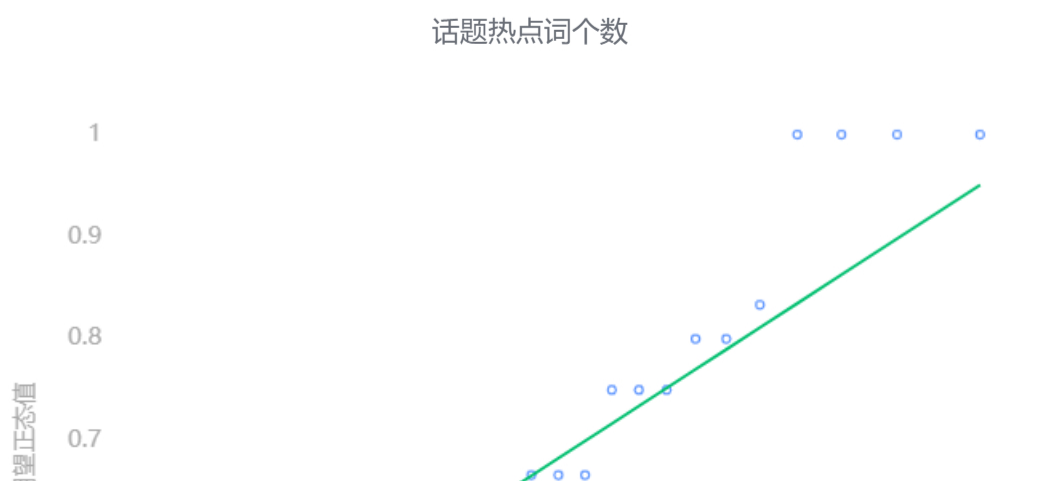
上图是阅读量(万)计算观测的累计概率 (P) 与正态累计概率 (P) 的拟合情况。拟合程度越高越服从正态分布。

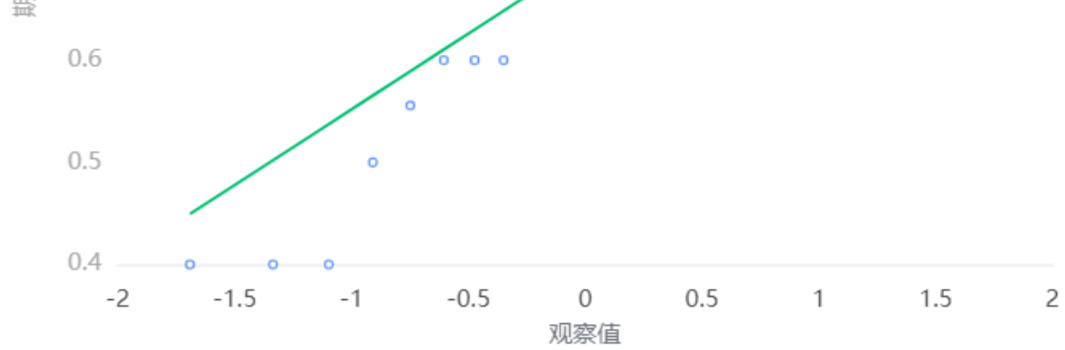


图表说明：

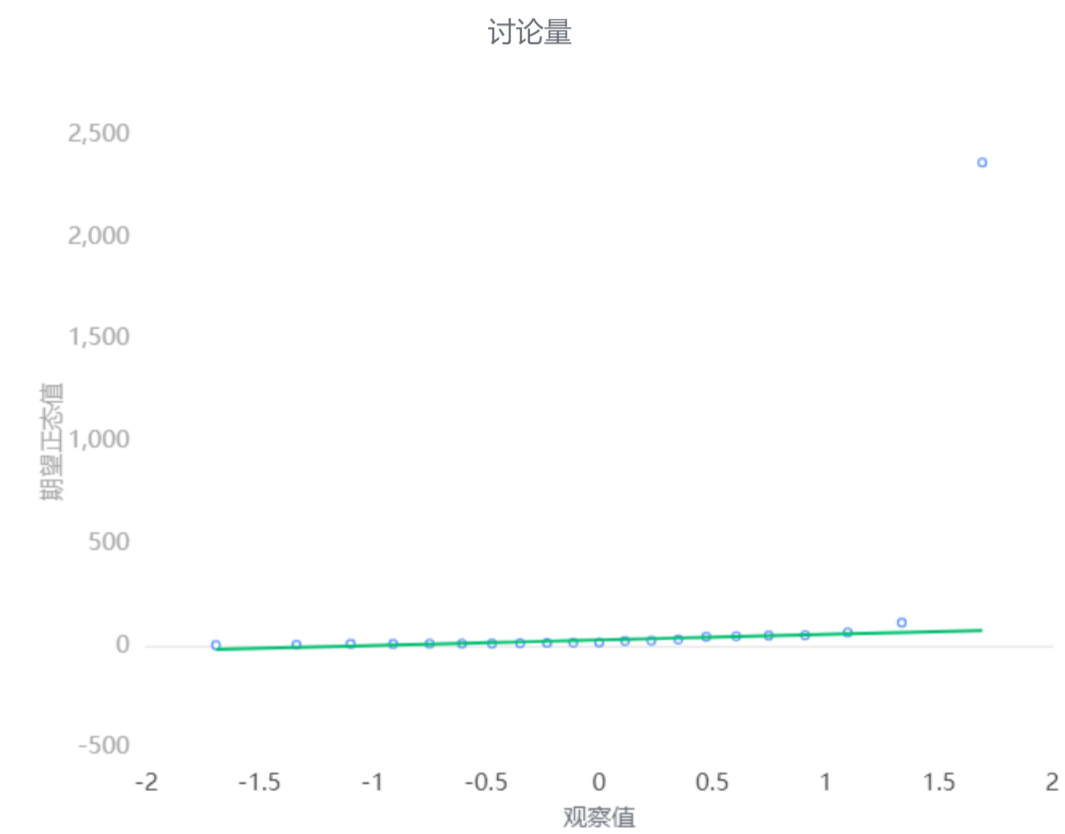
上图是评论量计算观测的累计概率 (P) 与正态累计概率 (P) 的拟合情况。拟合程度越高越服从正态分布。

输出结果4：正态性检验Q-Q图



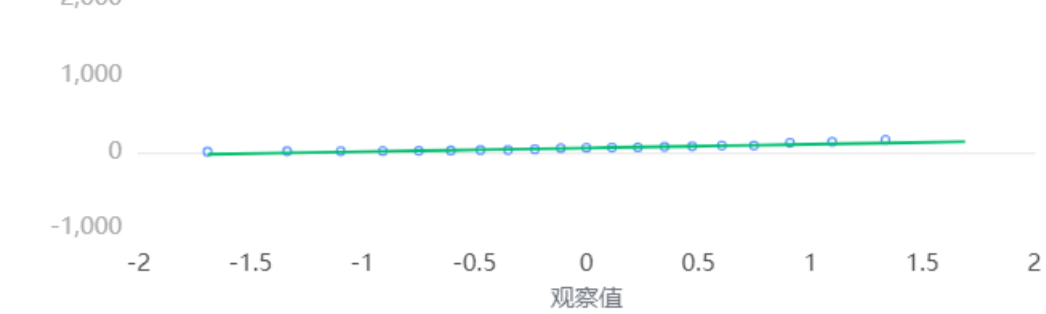


图表说明：
Q-Q图，全称“Quantile Quantile Plot” 用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

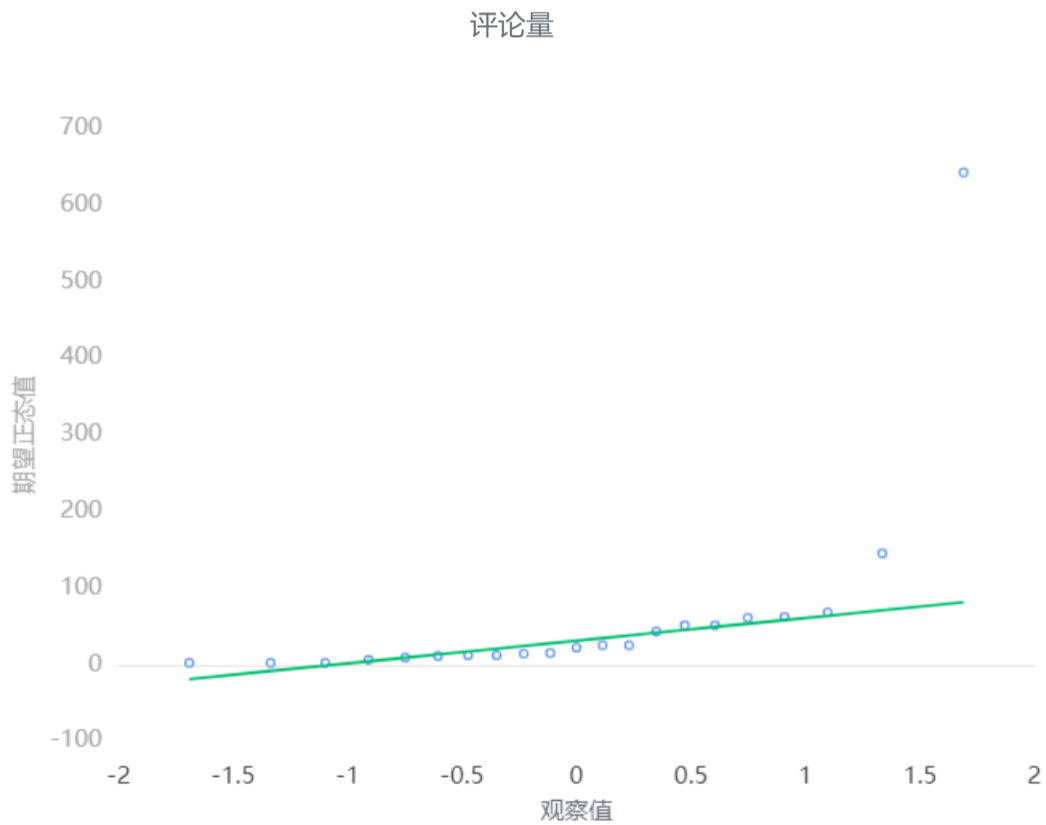


图表说明：
Q-Q图，全称“Quantile Quantile Plot” 用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。





图表说明：
Q-Q图，全称“Quantile Quantile Plot” 用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。



图表说明：
Q-Q图，全称“Quantile Quantile Plot” 用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。