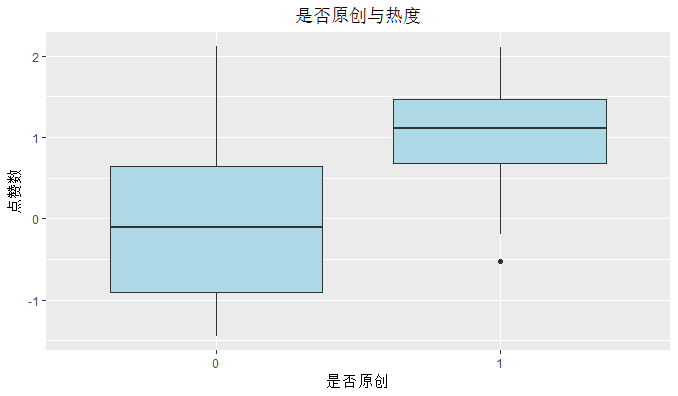
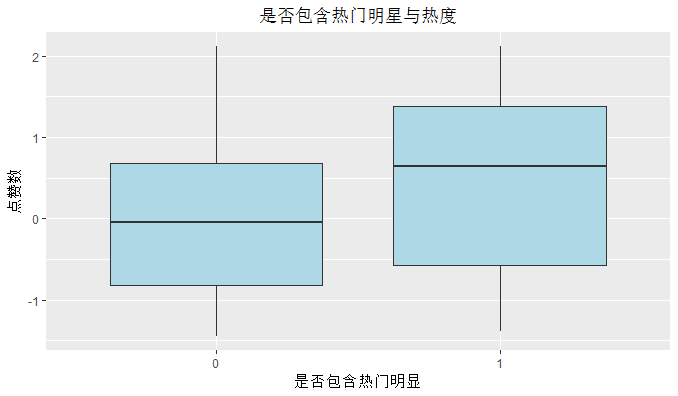
方法介绍：

我们对于不同类别的数据进行多因素方差分析，希望观察公众推送的热度与不同的主题以及公众推送是否是原创是否存在明显差异，我们采用每篇公众文章的点赞数量作为对于推送热度的度量。自变量我们采用了三种不同的因子型数据，分别为

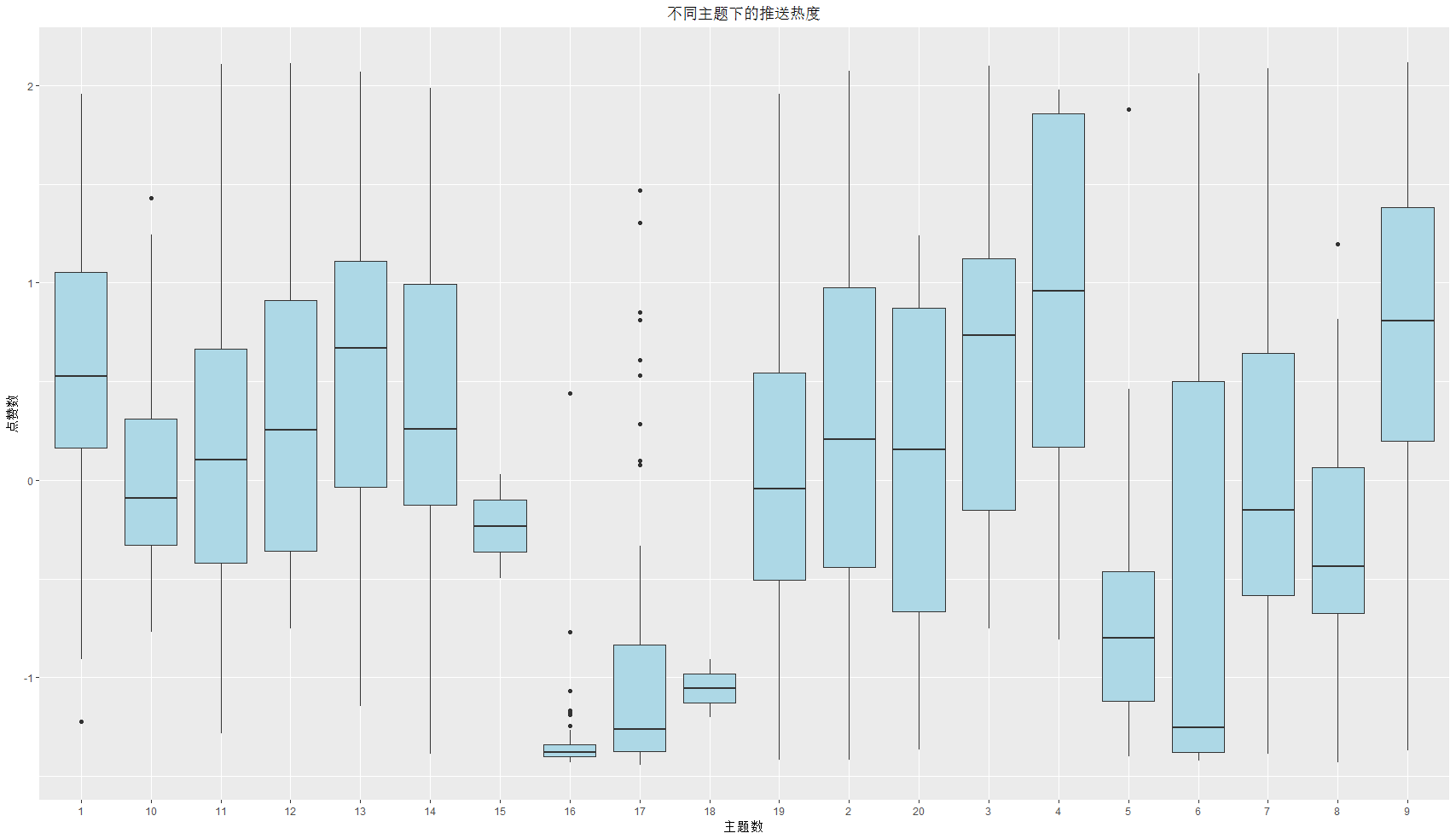
1.数据预处理

我们首先观察对于是否为原创以及是否包含排名前10位的明星做分组的箱型线，通过下图我们发现原创以及不同的主题是具有明显区别的。





然后，我们观察不同主题箱型线，我们发现16、17两个类别中的异常值现象比较严重，观察这两个主题，我们发现这两类主题对应的主题词类别并不明显，因此我们选择删掉这两个主题。同时在异常值处理中我们删掉样本主题的样本数量小于10的样本。完成异常值处理。



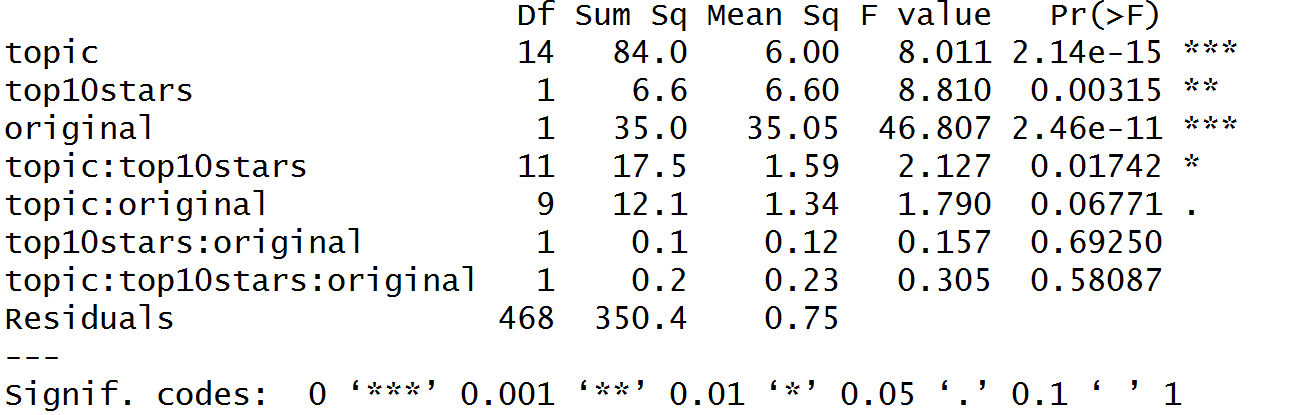
2、假设检验

首先，我们需要观察数据是否符合方差分析的假设：

方差齐性检验是指需要检验不同水平下的数据方差是否相等。我们采用Bartlett方法进行检验，得出这三种不同的分组下P值均大于0.05，得出方差具有齐次性的结论。

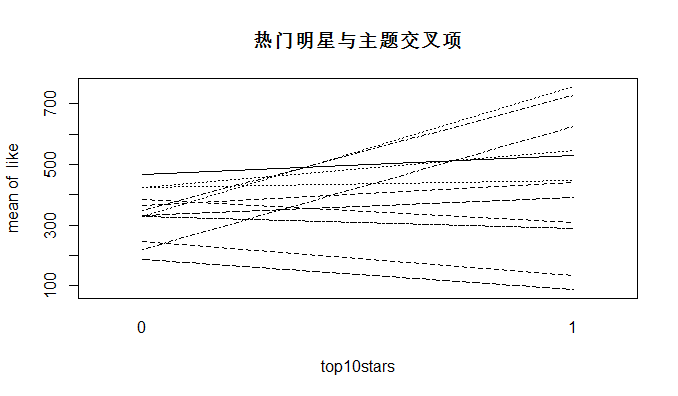
3、三因素方差分析

我们对于不同的主题、是否为原创、以及是否包含最热门的10位明星这三个变量进行三因素方差分析，同时引入交叉项：

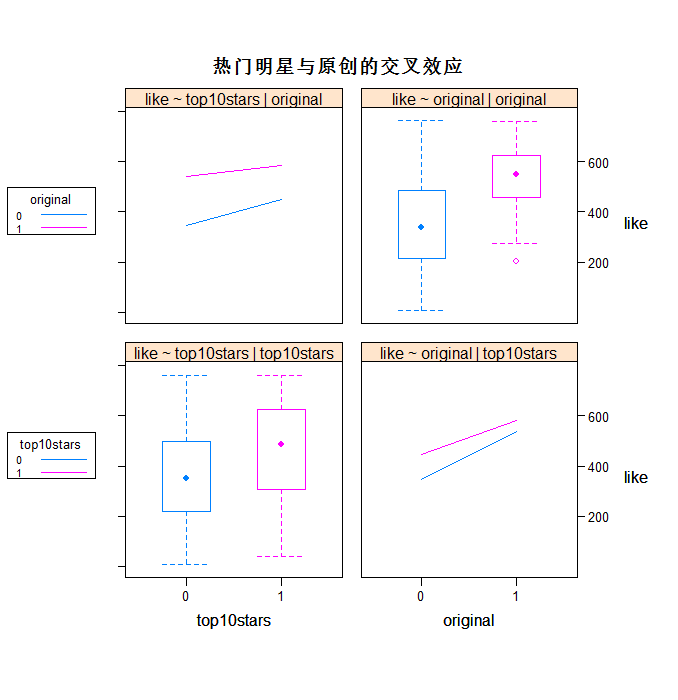


从结果中我们可以发现点赞数量与主题、原创性以及是否包含前10位明星这三个因子均有显著的相关关系，并且是否包含热门明星与主题，以及主题与原创性对结果也具有交叉效应。

因此，我们做出交叉效应图，可以观察到，是否包含热门明星与不同的主题确实具有明显的交叉效应。



而对于是否包含热门明星与是否为原创这两个因子交叉效应不显著，如下图示：



从而我们得出结论：明星公众号的点赞数量与原创性及明星热度均有明显关系，原创的推送以及涉及热门明星，以及部门热门主题均能提高微信推文的热度。