**一、数据处理&模型简介&状态的描述**

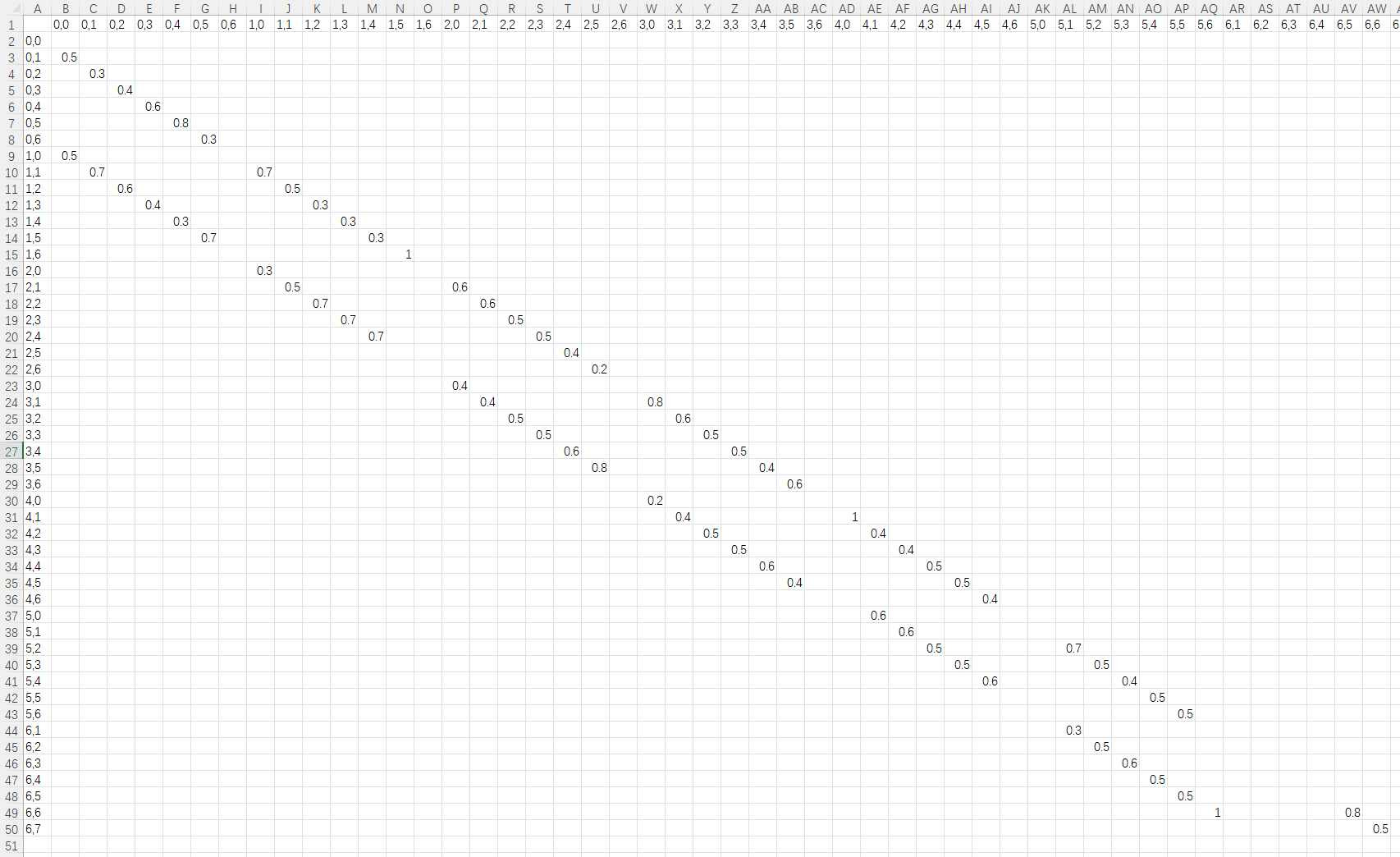
数据分为两类：一类描述具体比赛得分状态，一类描述本球技术类型

* **~~1.1Data Cleaning and Visualization~~**

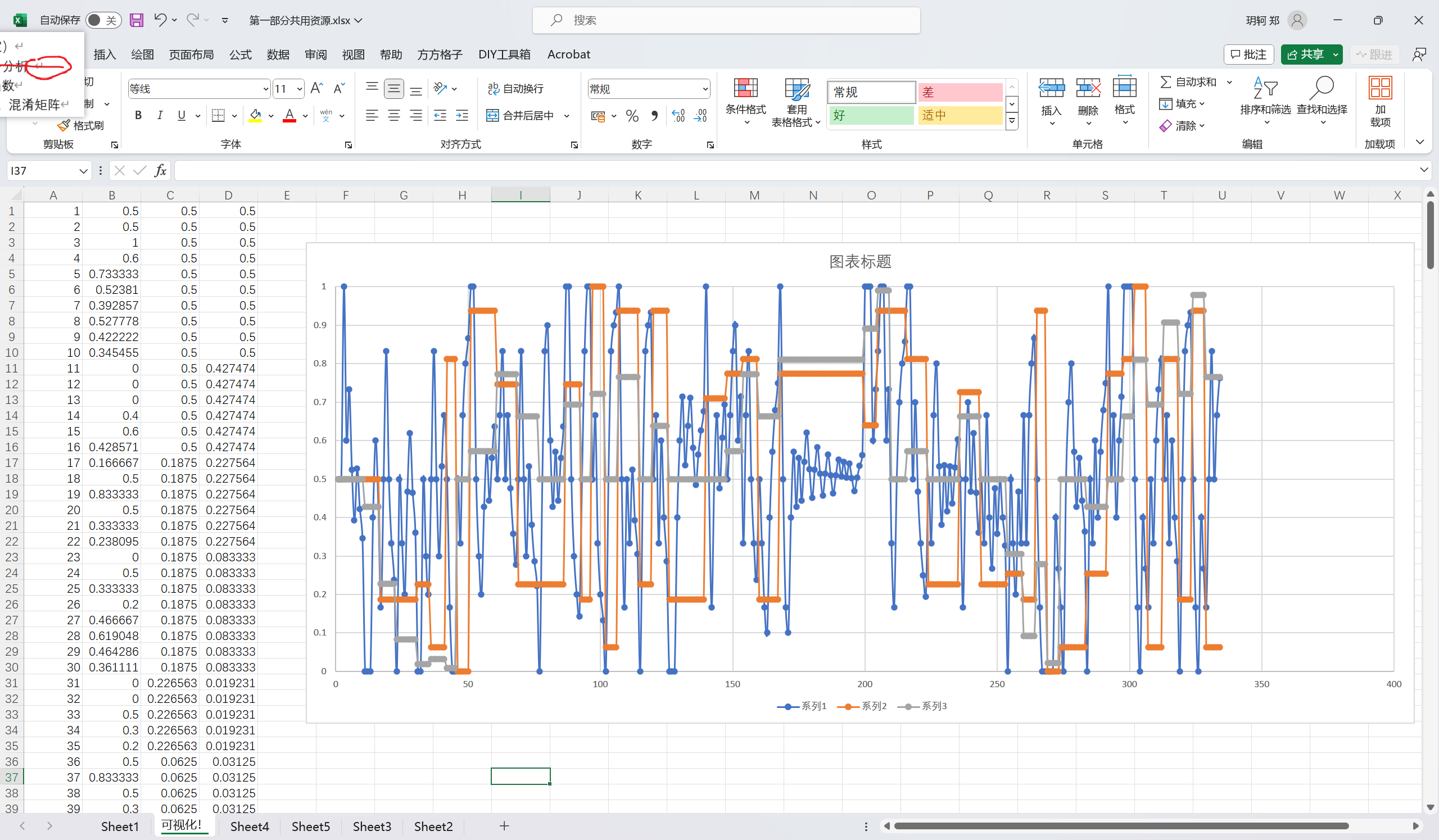
**1.2 V数据挖掘&特征工程：**

马尔可夫链模型（Markov Chain Model）

* **。。。（结果**状态转移矩阵**）**



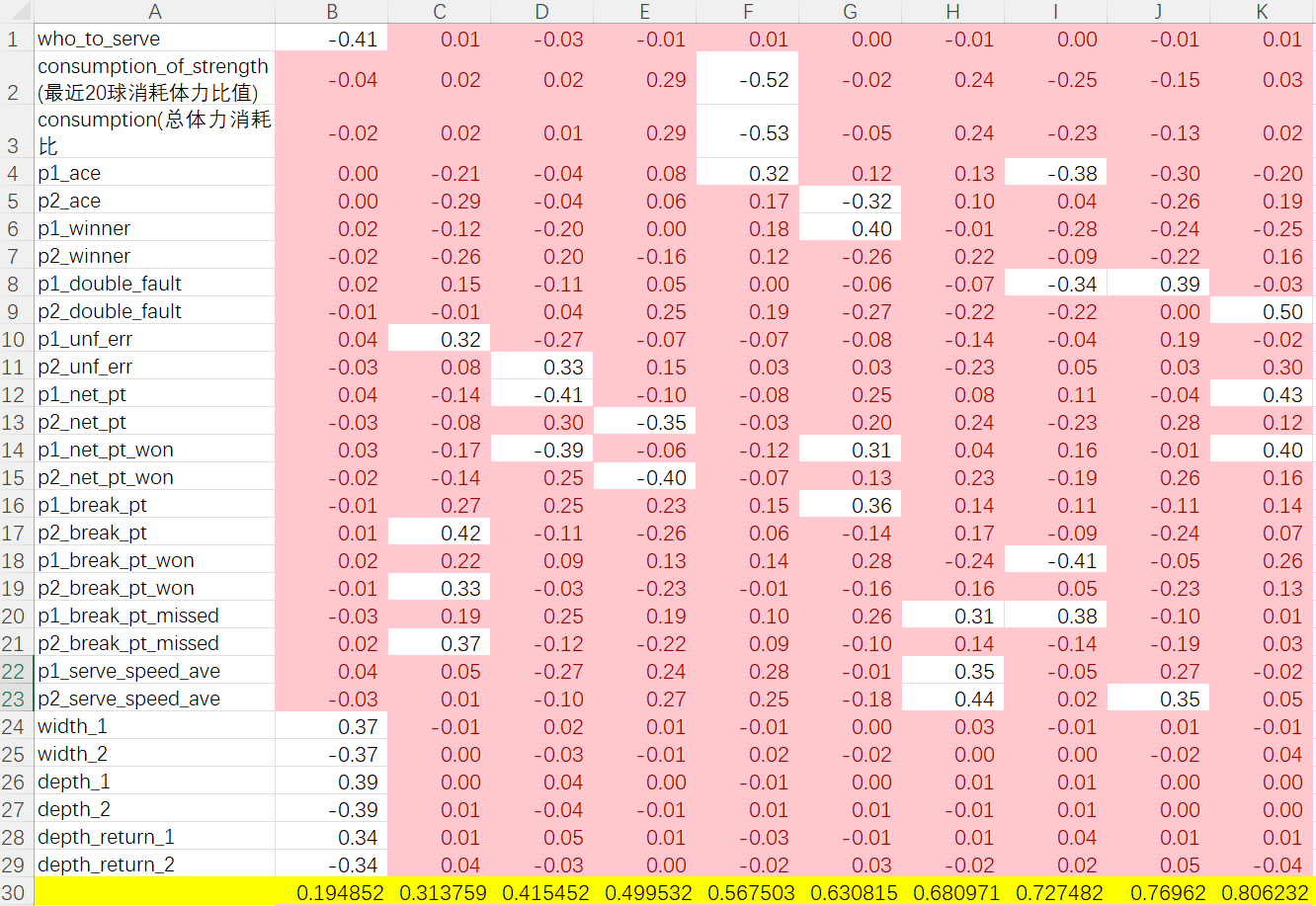
* **~~。。。（结果~~**~~每种状态的胜率~~**~~）~~**
* **。。。（V的折线可视化结果）**



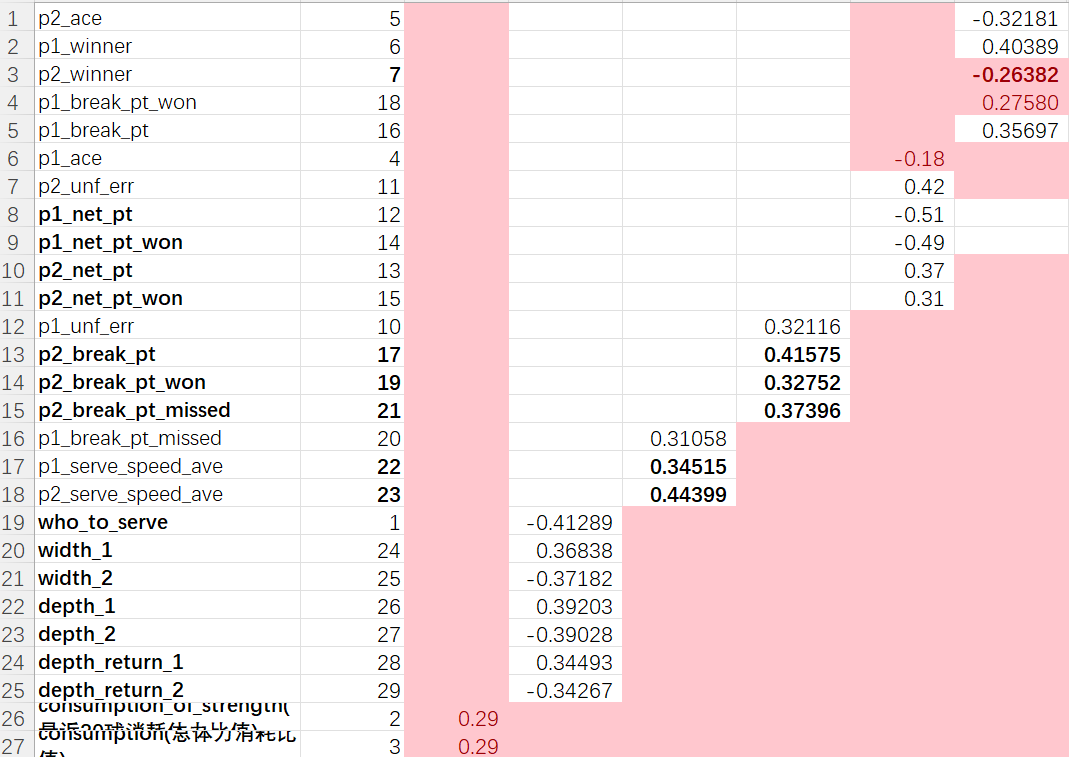
1.3 Q：PCA

为了保证准确性，拟打算取10项主成分，得到80%的贡献

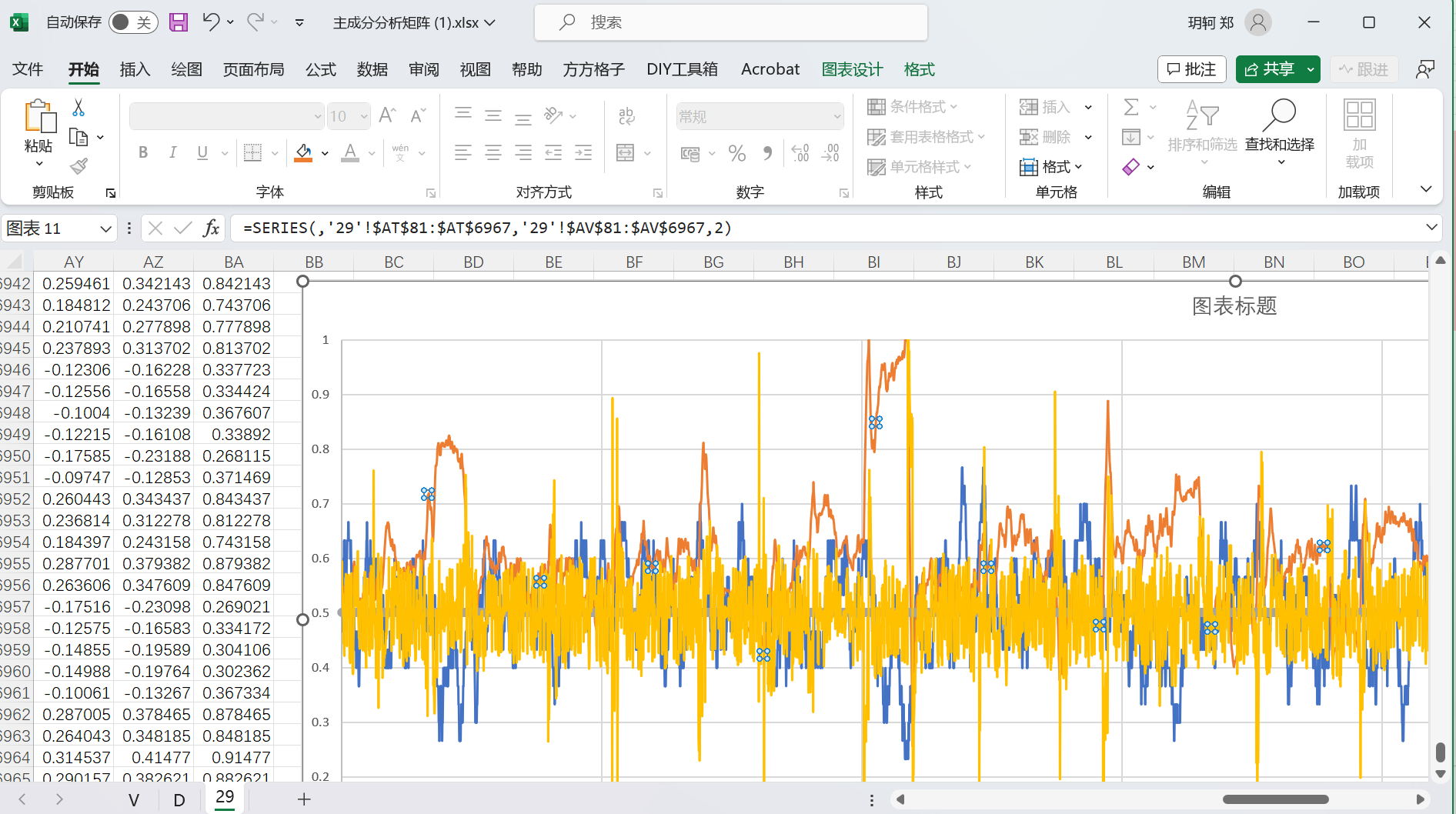
* **（**特征值的贡献值+主成分向量内容**可视化结果）——热力+柱状**



* **（**修剪后**可视化结果）——雷达**

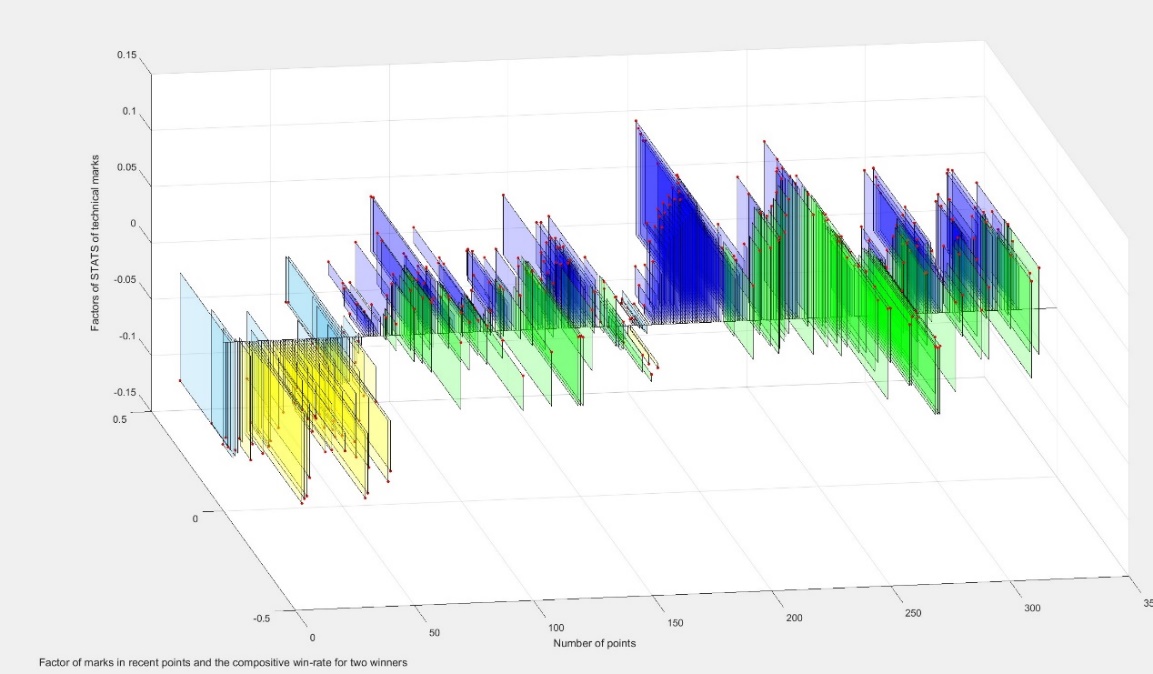


* **Q折线图**



* **V、Q可视**

最后可视化：



**二、数据处理的评估&教练的反驳**

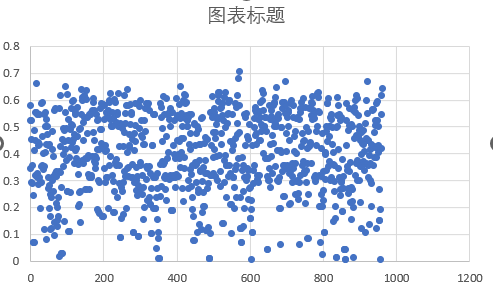
**2.1 runs of success分析**

对于player1：

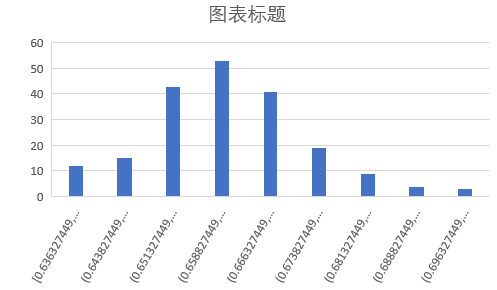
M预测筛选后的均值0.624，方差0.02

R预测筛选后的均值**0.612**，方差0.13

* **（可视化图）散点图：**

**两者对比**

* **将每个RANDOM的均值做分布图（正太）**



方差更小更稳定的预测，趋向也更符合（趋近1）

按照逻辑斯蒂得分预估：

M的得分为5.6，R的得分为3.1

对于player2：

M预测筛选后的均值0.387，方差0.02

R预测筛选后的均值0.468，方差0.08

* **（可视化图同上）**

方差更小更稳定的预测，趋向也更符合（趋近0）

按照逻辑斯蒂得分预估：

M的得分为5.7，R的得分为1.9

**swings in play分析**

对于player1：

M预测筛选后的均值0.654，方差0.02

R预测筛选后的均值0.568，方差0.13

* **（可视化图同上）**

方差更小更稳定的预测，趋向也更符合（趋近1）

按照逻辑斯蒂得分预估：

M的得分为7，R的得分为1.4

对于player2：

M预测筛选后的均值0.317，方差0.012

R预测筛选后的均值0.508，方差0.100

* **（可视化图同上）**

方差更小更稳定的预测，趋向也更符合（趋近0）

按照逻辑斯蒂得分预估：

M的得分为8.9，R的得分为0.6

**2.2 momentum也不是随机的，而是与时间/赛程相关（ARIMA）**

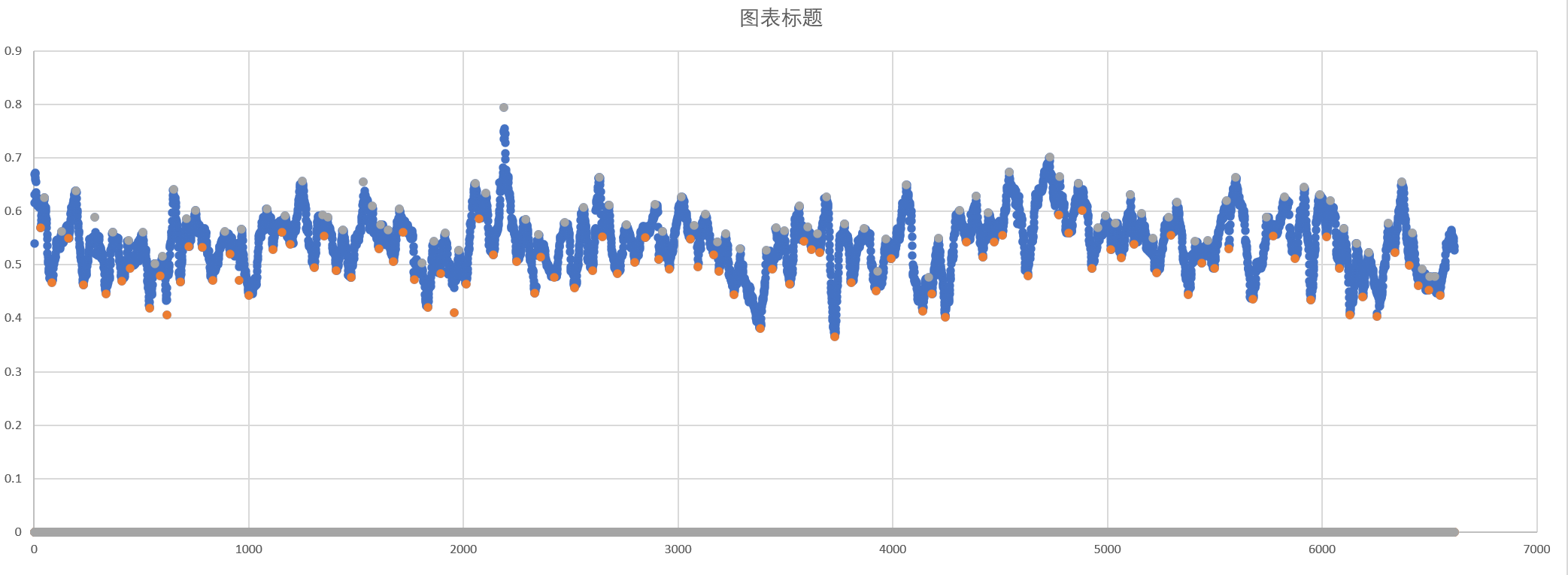
为证明我们momentum与白噪音显著不同，可以通过白噪音和时间序列分解得到说明。

**白噪音证明（找cy）**

* **。。。解释+程序结果**

**三、模型结果&预测**

* ~~3.1 贝叶斯变点~~**~~结果：~~**
* **3.2 卡尔曼滤波分析——可视化**



U\X\Y 列

+

最后决赛的聚焦