

Group 8

Speed Dating Experiment



陳浩欣 農經四 B03607054

葉明俊 資工三 B04902122

Background

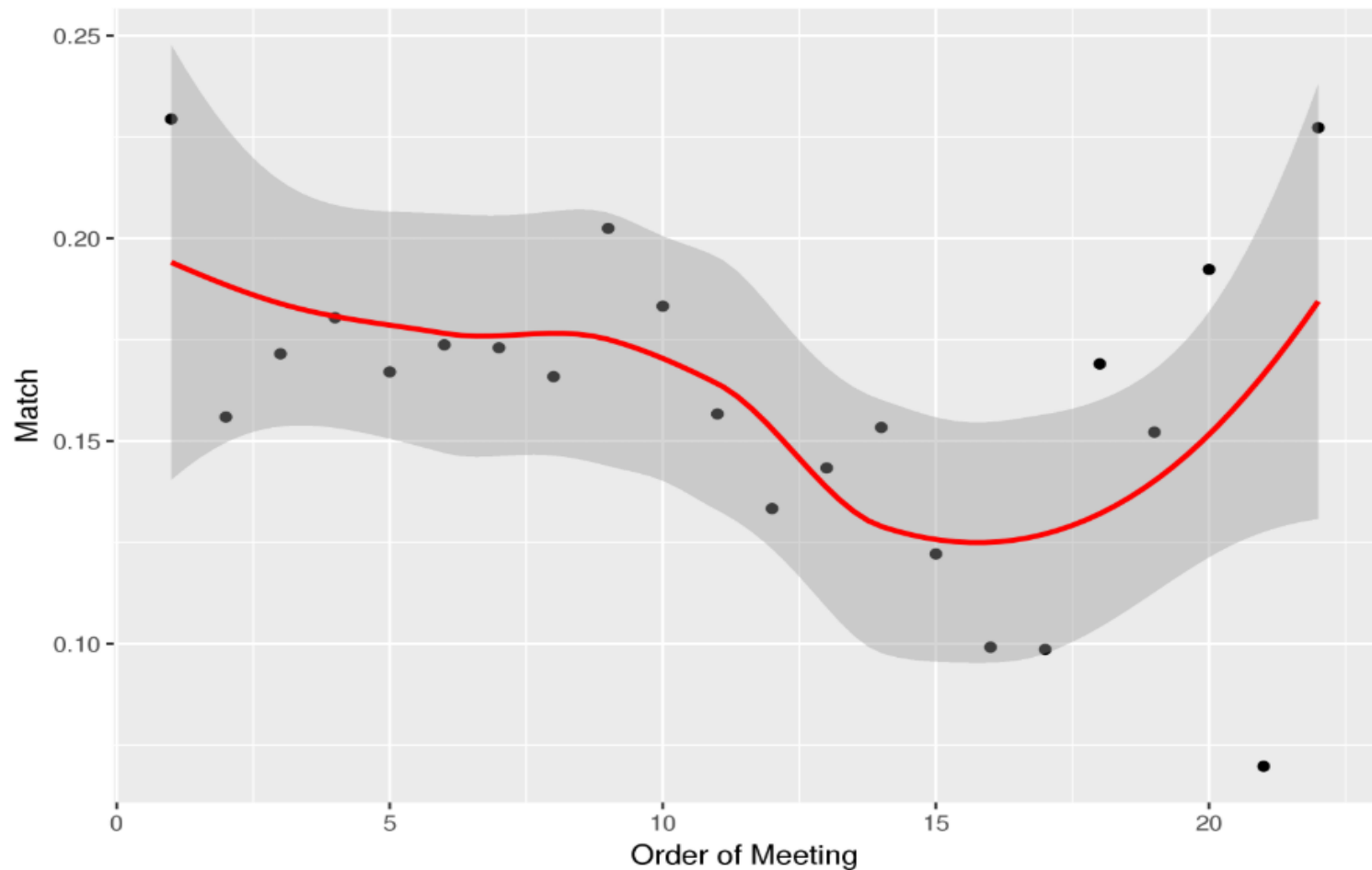
- 人們對於速配行為的基本了解，與成功約會最相關的屬性.
- Dataset 來自2002-2004的速配約會中收集而來.
- Dataset 總共有8378筆紀錄，參與者多來自於Colombia Business School的學生.
- 調查時程:
 - 約會前
 - 追蹤調查在約會後隔天
 - 第二次追蹤調查是在配對後3-4周進行.

Dataset

| 變數 | 變數定義 | 註 |
|----------|-------------------|--------------|
| wave | 活動的波數 | |
| order | 約會次數 | |
| match | 配對成功 | 1=成功，0=失敗 |
| dec_o | Positive response | 1=pos, 2=neg |
| attr_o | 魅力 | |
| sinc_o | 真誠 | |
| intel_o | 智慧 | |
| fun_o | 有趣 | |
| amb_o | 有志向 | |
| shar_o | 分享興趣 | |
| age | 年齡 | |
| field_cd | 職場 | |
| race | 種族 | |

Match by Number of date

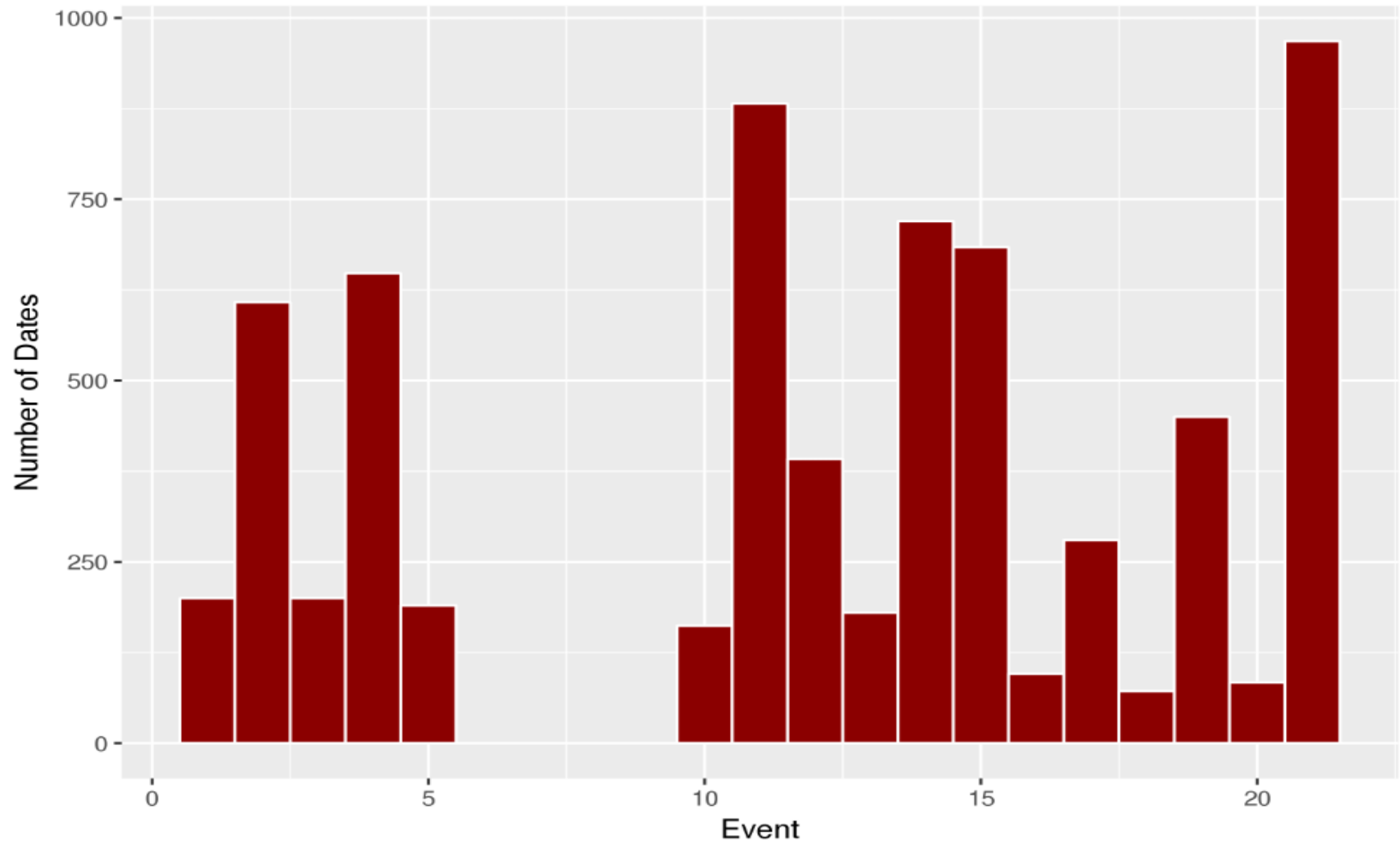
One of the explorations we found to have a relation was the wave. The further in the event the lower the chance of match gets. Until we arrive at the last 4 dates where it seems people are getting nervous to get there dates and the chance of a match increases again.



- (X=約會的次數， Y=成功在一起的幾率 用的變數：
match, order)

Number of dates per event

We looked at the number of dates there were per event. As you can see we took out 6-9 because of the data structure.

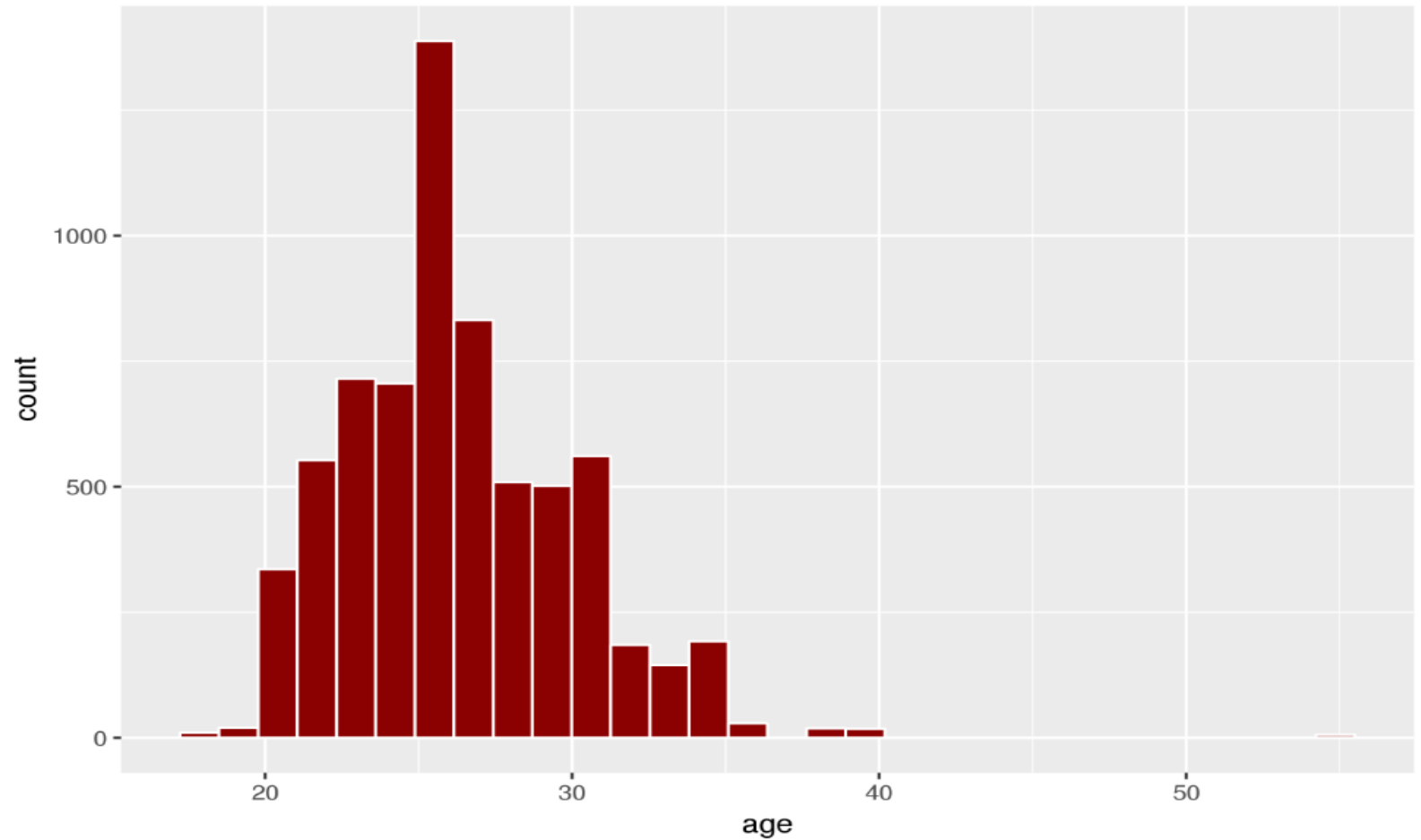


□ (X=event, Y=約會總次數 用的變數：wave)

Age

We took a look at the average age and the distribution of age. We found that the average age is 26.2817425

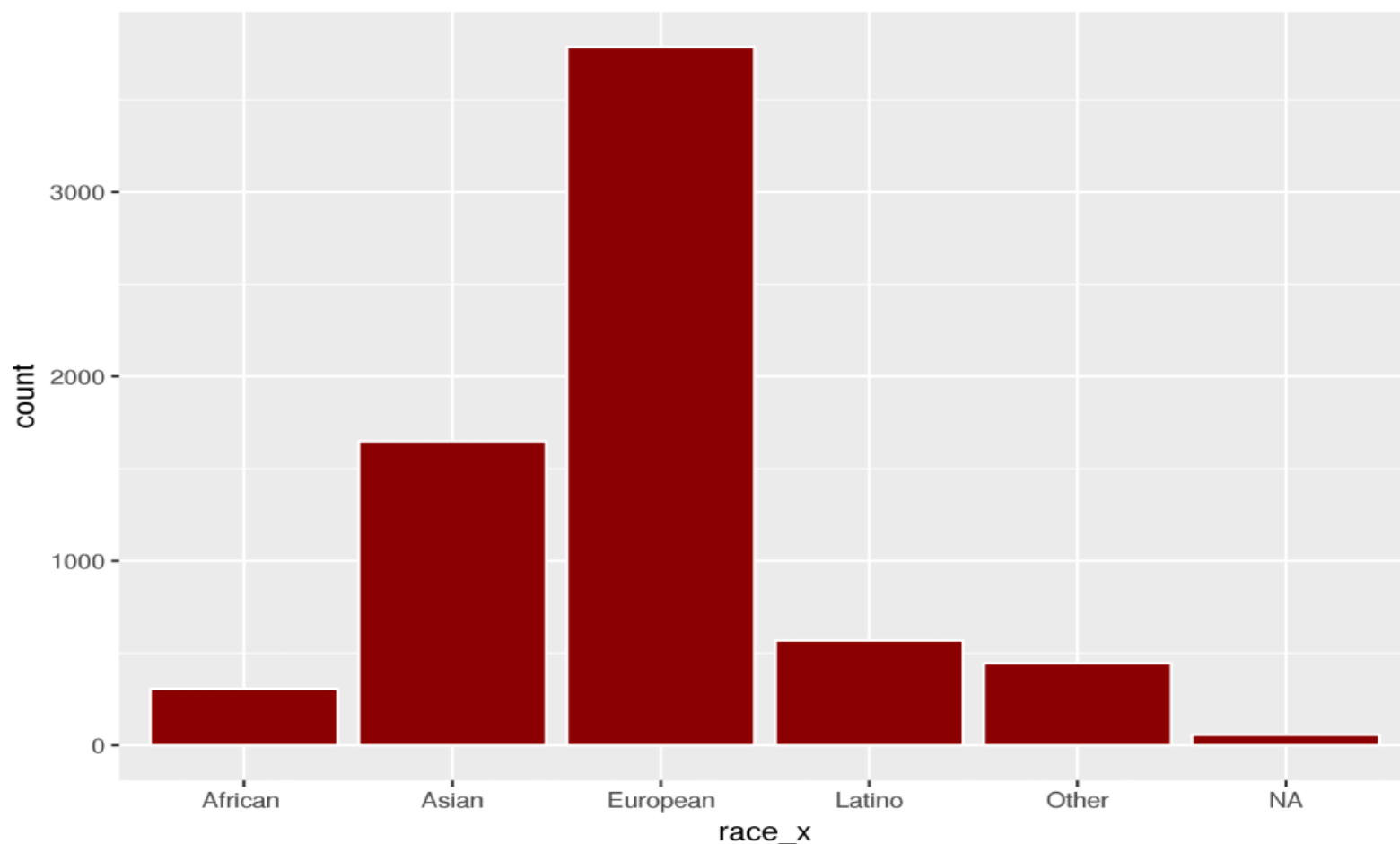
And our age distribution:



□ (X=年齡，Y=人數 用的變數：age)

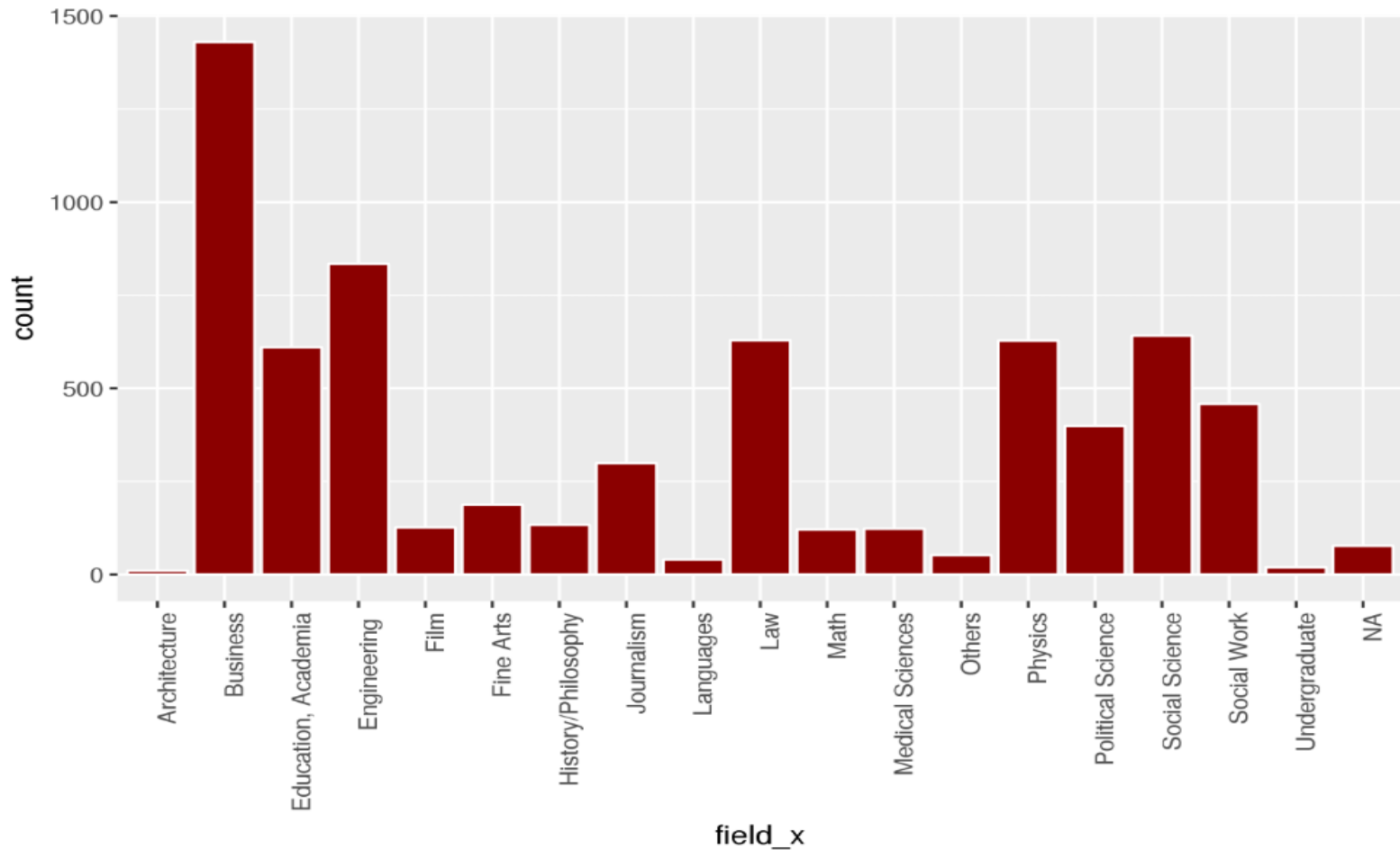
Race

Out of curiosity we took a look at the racial distribution of the data set. Later we found that the variable race had neglect able impact on the outcome of the date.



□ (X=種族，Y=人數 用的變數：race_o)

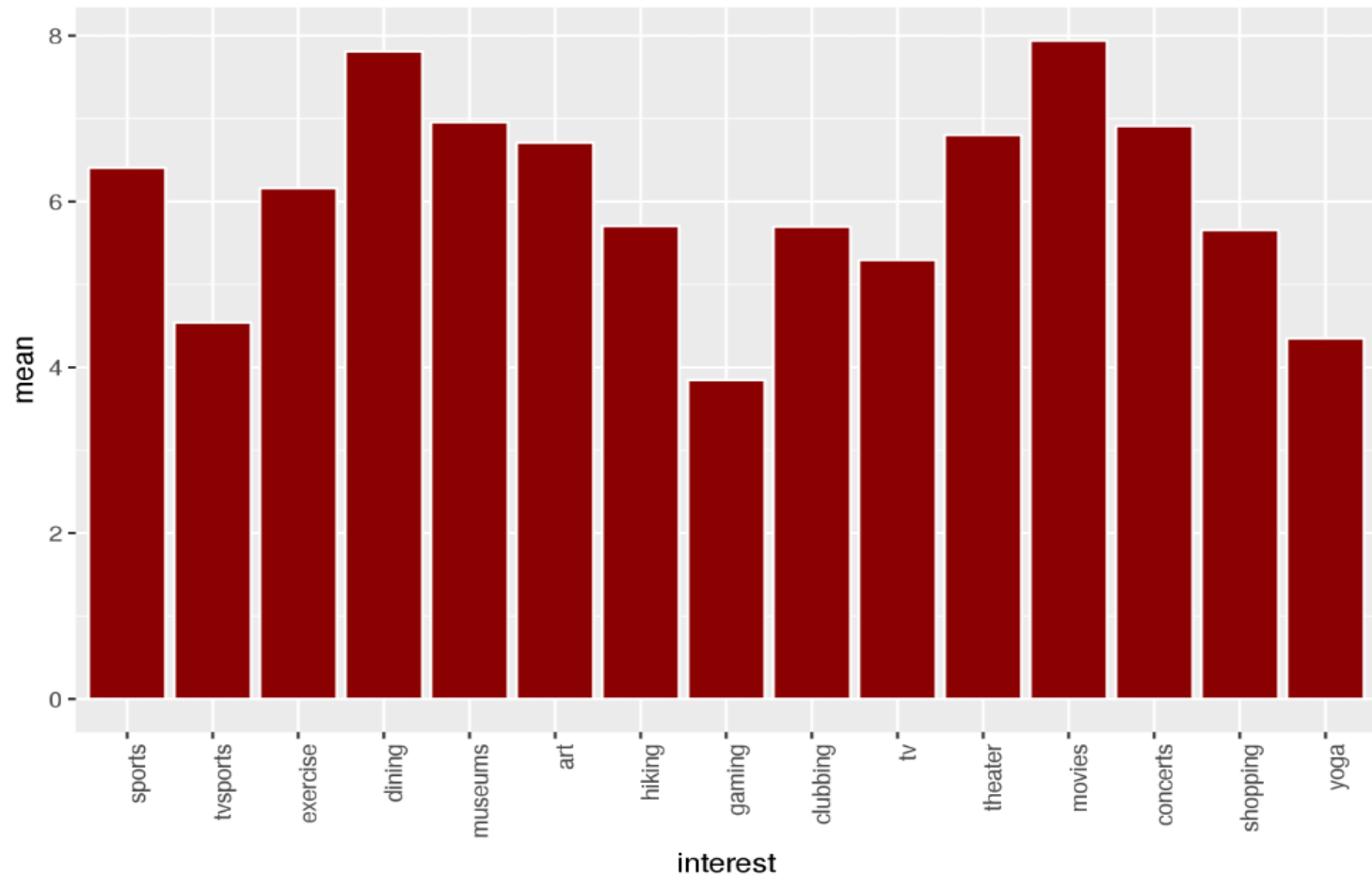
Field of study



□ (X=專業領域，Y=人數 用的變數field_cd)

Interest

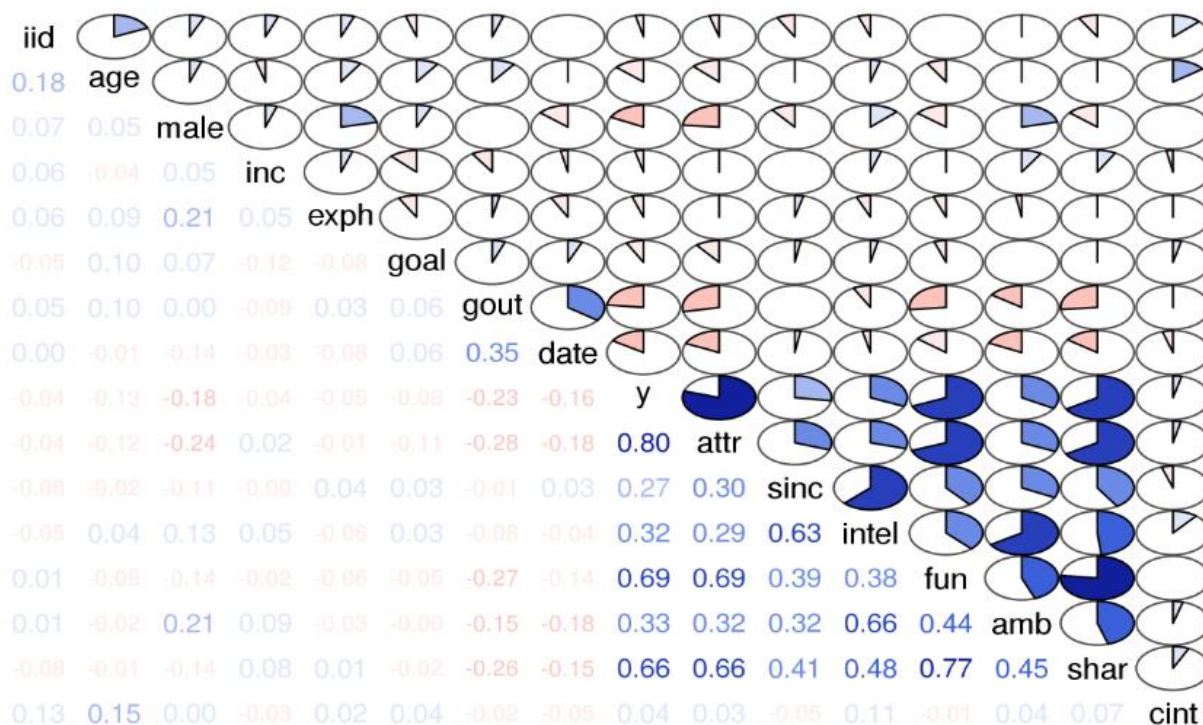
And we wanted to know where the people of our data set are interested in. We did this because we had a suspicion that shared interest could be an important variable.



- (X=興趣，Y=平均數 (滿分=10分) 用的變數：
x軸上的全部)

Explanatory Analysis

Correlation of interesting variables



- 使用變數dec_o(positive responses)來找出到底哪些人有興趣再和之前一起參加這個約會event出來見面。

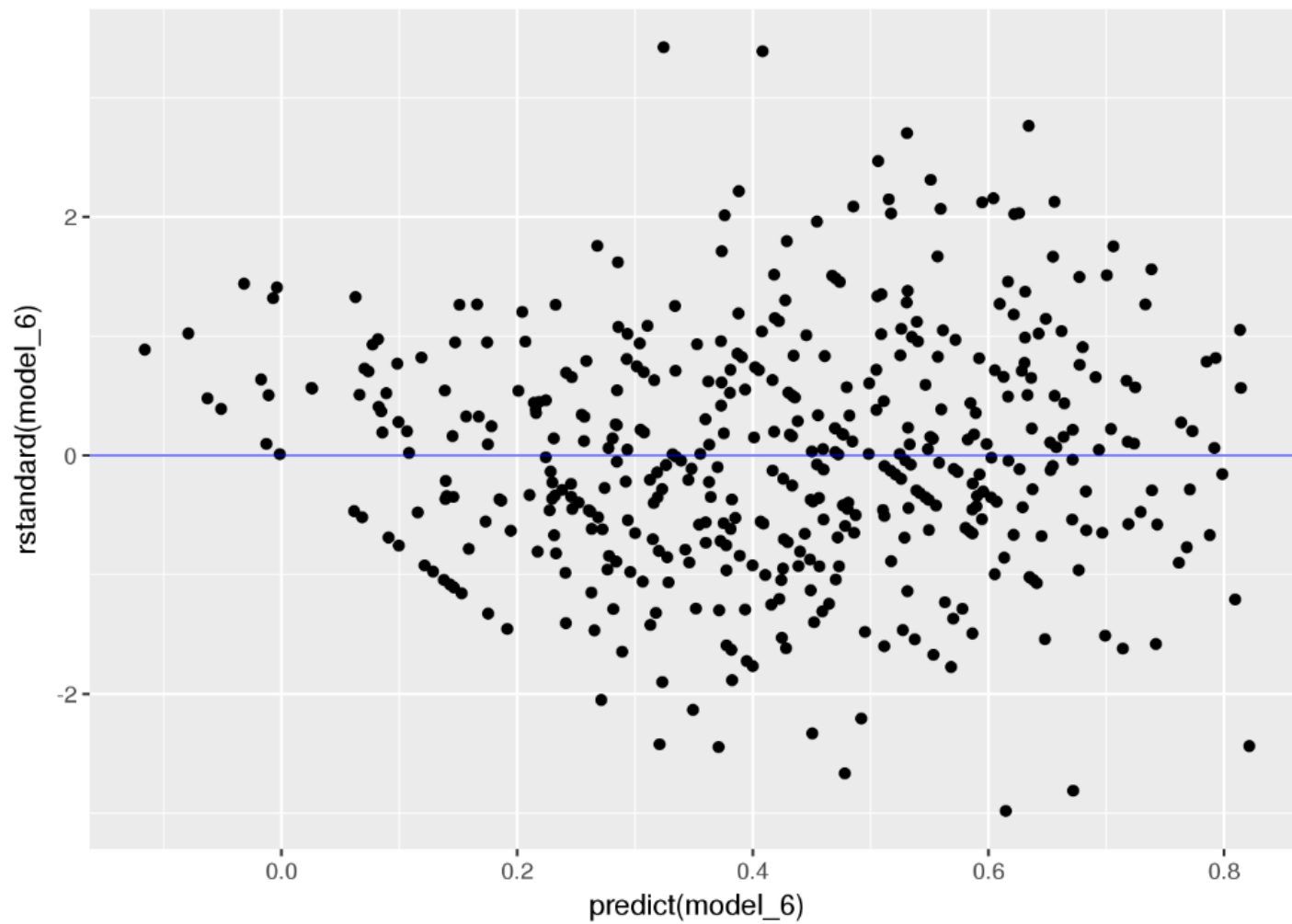
建構模型

- 接下來就是建構模型。我們要從這六個變數組合出有最高的R平方值。
- 一共有六種模型：
 - 魅力
 - 魅力+分享興趣
 - 魅力+有趣
 - 魅力+分享興趣+有趣
 - 魅力*分享興趣*有趣
 - 魅力+分享興趣：有趣

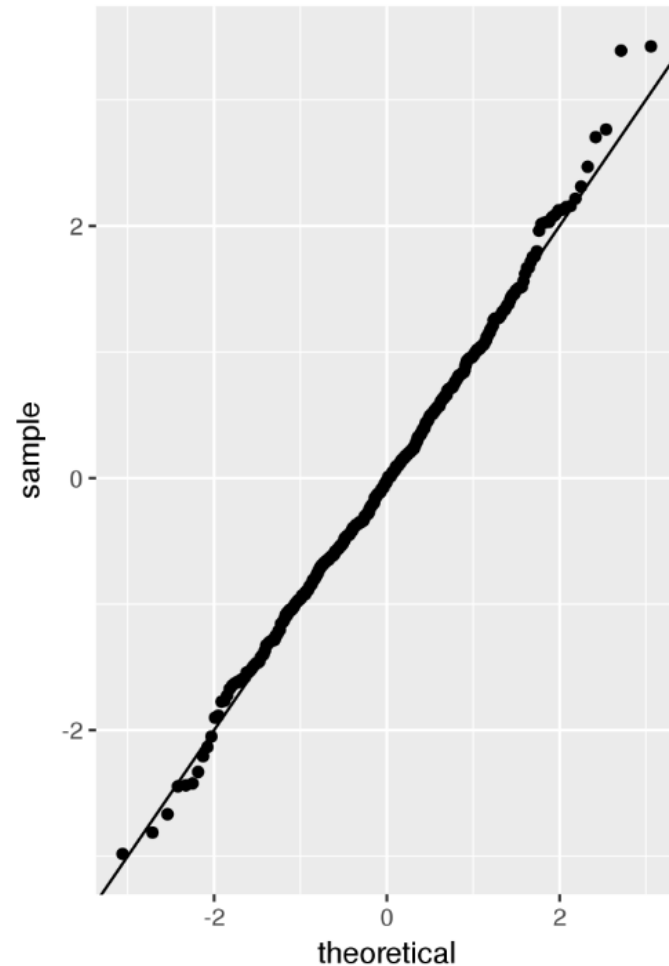
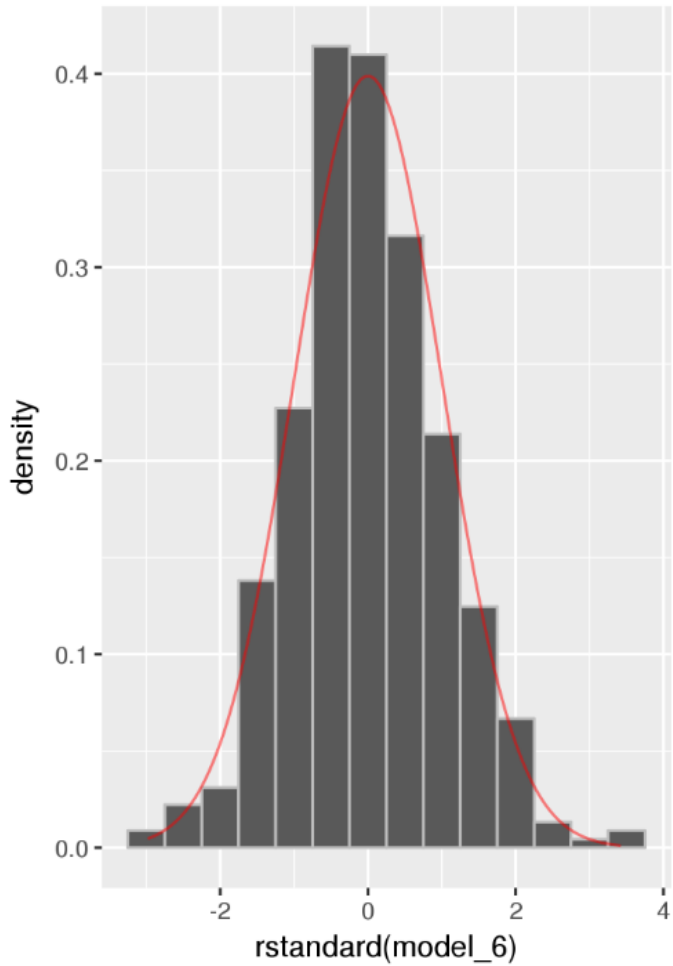
迴歸分析的假設

- 滿足三個條件：
- 線性
- 誤差值的正常分配
- 誤差值的同質性

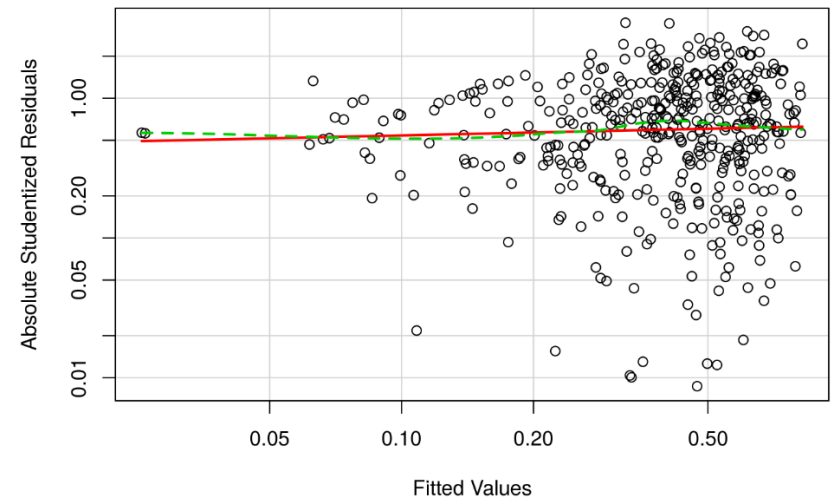
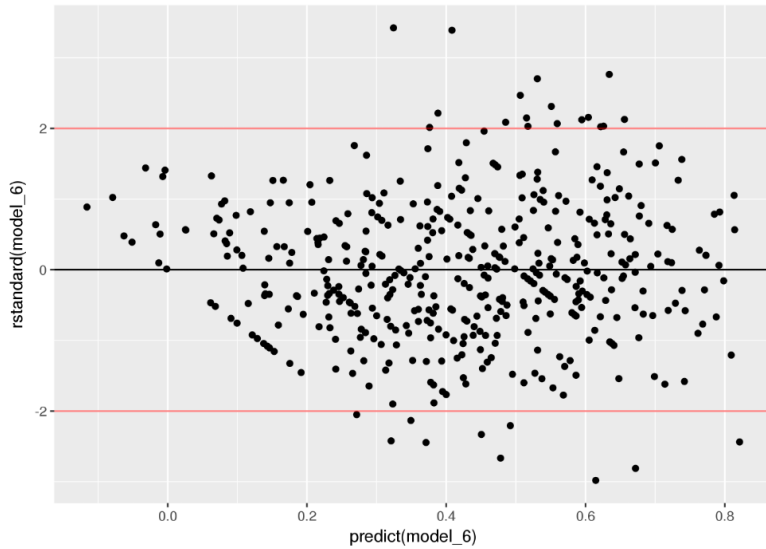
線性



誤差值的正常分配



誤差值的同質性

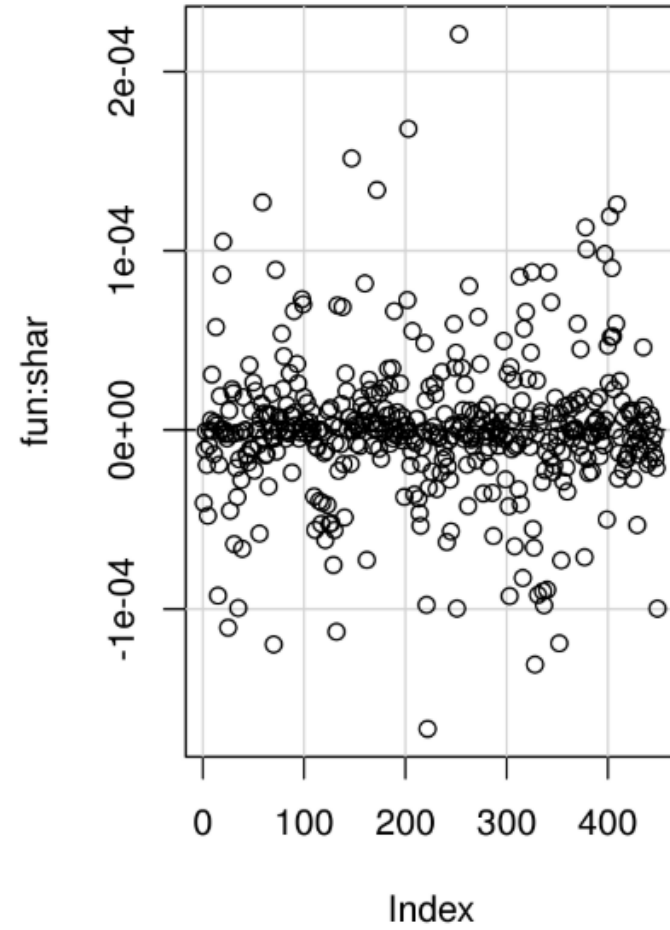
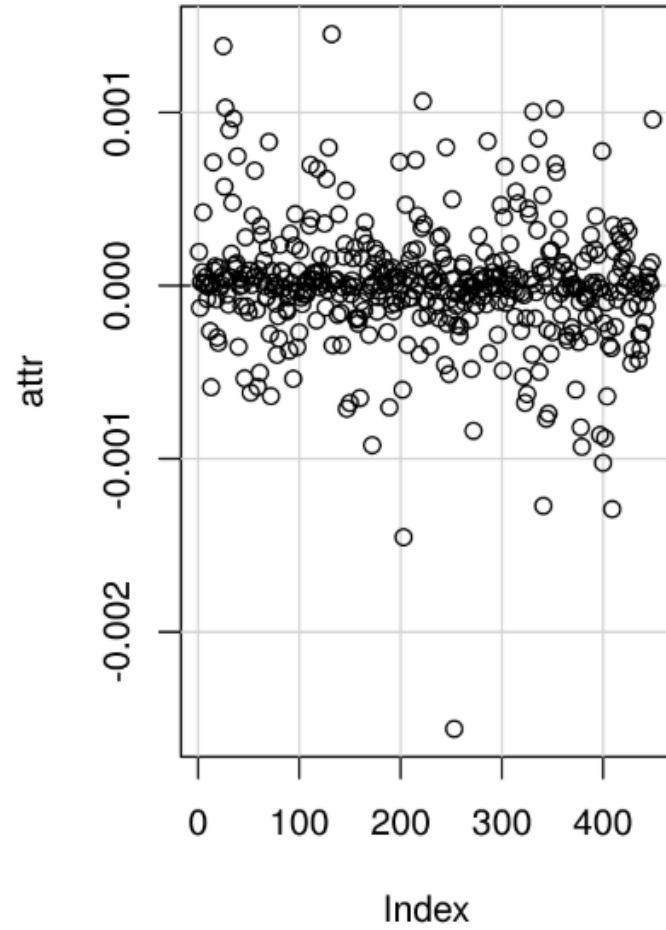


IV 的獨立性

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.1196148 1.755918 0.02
## Alternative hypothesis: rho != 0
```

- 統計值為1.76，p值<0.05
(不通過)

dfbeta Plots



總結

- 整體模型表明, 魅力是解釋積極反應的最重要的變數, 雖然有趣和分享也很重要。
- 這些見解可以用在讓一個速配約會的機構中了解如何進行配對, 或是對現代社會中的約會進行研究。