**一、研究背景與動機**

客戶性格分析是對公司理想客戶的詳細分析。它有助於企業更好地了解其客戶，並使他們更容易根據不同類型客戶的特定需求、行為和關注點修改產品。

客戶個性分析幫助企業根據來自不同類型客戶群的目標客戶修改其產品。例如，公司無需花錢向公司數據庫中的每個客戶推銷新產品，而是可以分析哪個客戶群最有可能購買該產品，然後僅在該特定客戶群上推銷該產品。

為了找到葡萄酒、肉製品、黃金的目標客戶，於是想先研究哪些因素會影響顧客的購買意願。

**二、研究目的**

分析顧客在葡萄酒、肉製品、黃金上的花費是否與顧客的教育水平、感情狀況、家庭年收入有關

**三、變數介紹**

ID : 顧客編號

Year\_Birth : 顧客出生年份

Education : 顧客教育水平(Basic : 國小或國中 Graduation : 高中 2n Cycle : 碩士 Master : 大學 PhD : 博士)

Marital\_Status : 感情狀況(Divorced : 離婚Married : 已婚 Single : 單身 Together : 交往中Widow : 寡婦)

Income : 顧客家庭年收入

Recency : 自顧客上次購買以來的天數

MntWines : 過去兩年在葡萄酒上的花費

MntMeatProducts : 過去兩年在肉製品上的花費

MntGoldProds : 過去兩年在黃金上的花費

wine : 把MntWines以<300、300~600、600~900、900~1200、>1200的金額分成五組

meat : 把MntMeatProducts以<150、150~300、300~500、450~700、>700的金額分成五組

gold : 把MntGoldProds以<50、100~150、150~200、50~100、>200的金額分成五組

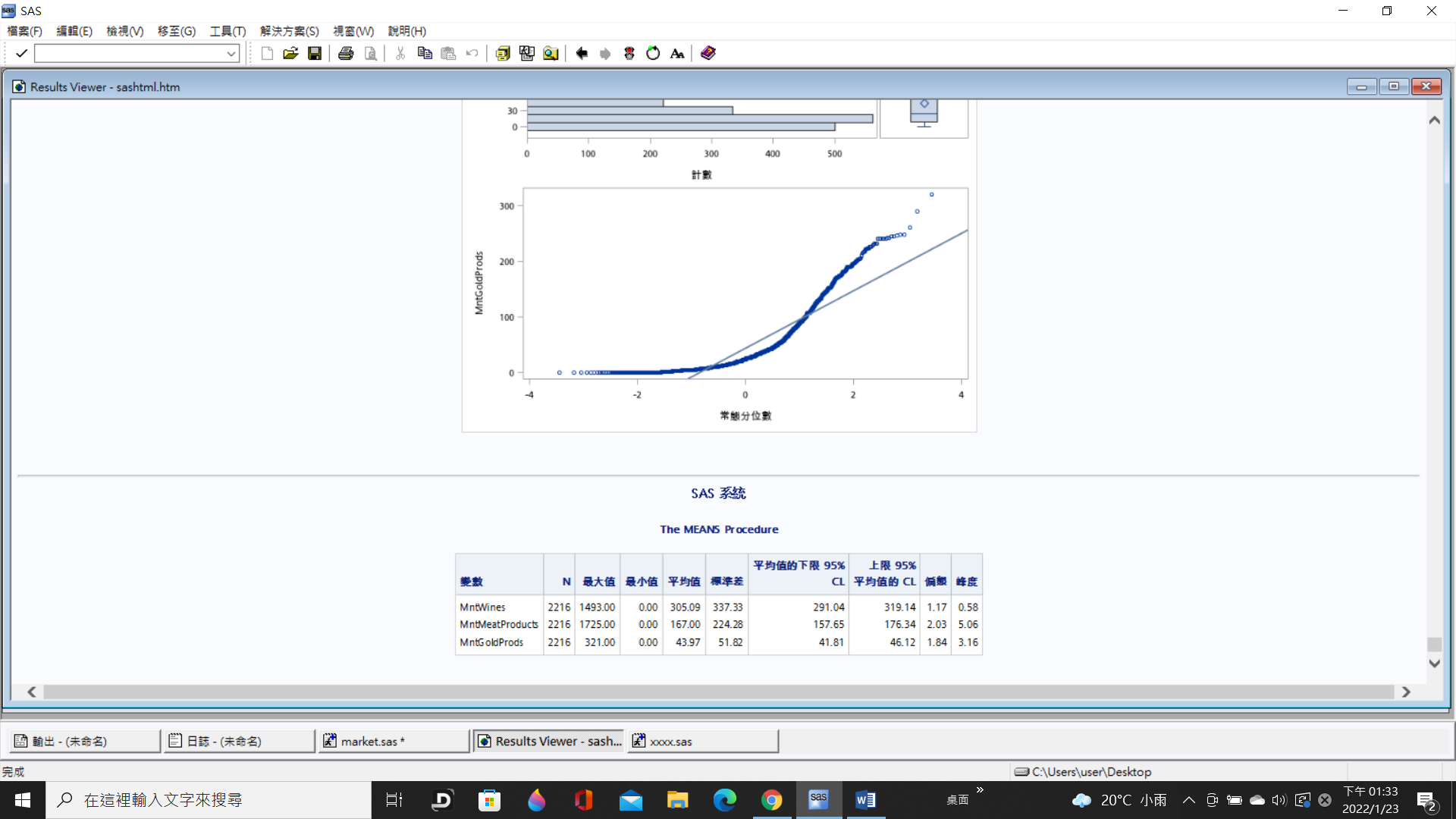
incomegp : 把Income以<30000、30000~50000、50000~70000、>70000的金額分成四組

**四、資料來源**

<https://www.kaggle.com/armanmostafazadeh/customer-personality-analysis-k-means>

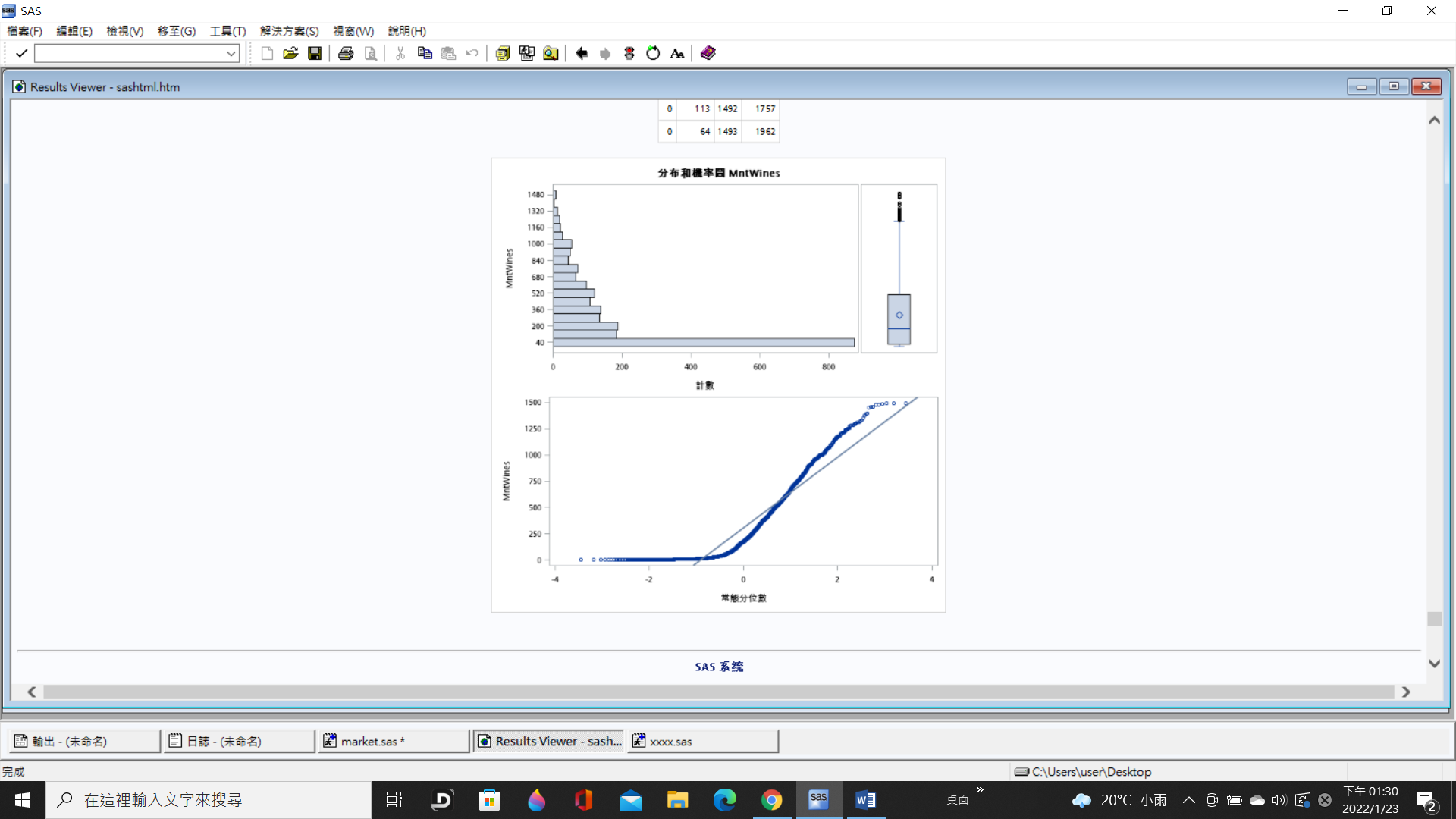
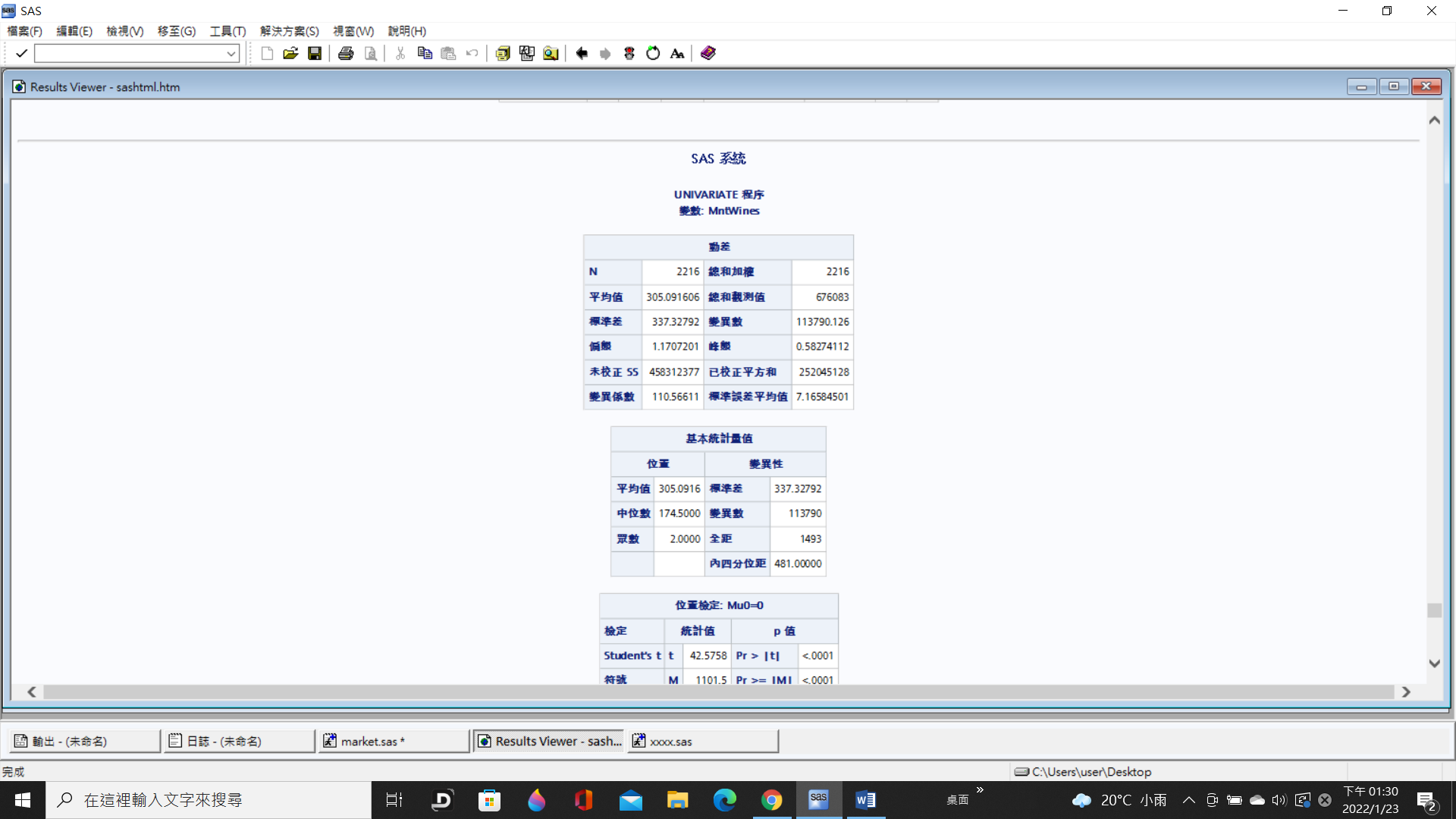
**五、敘述性統計**

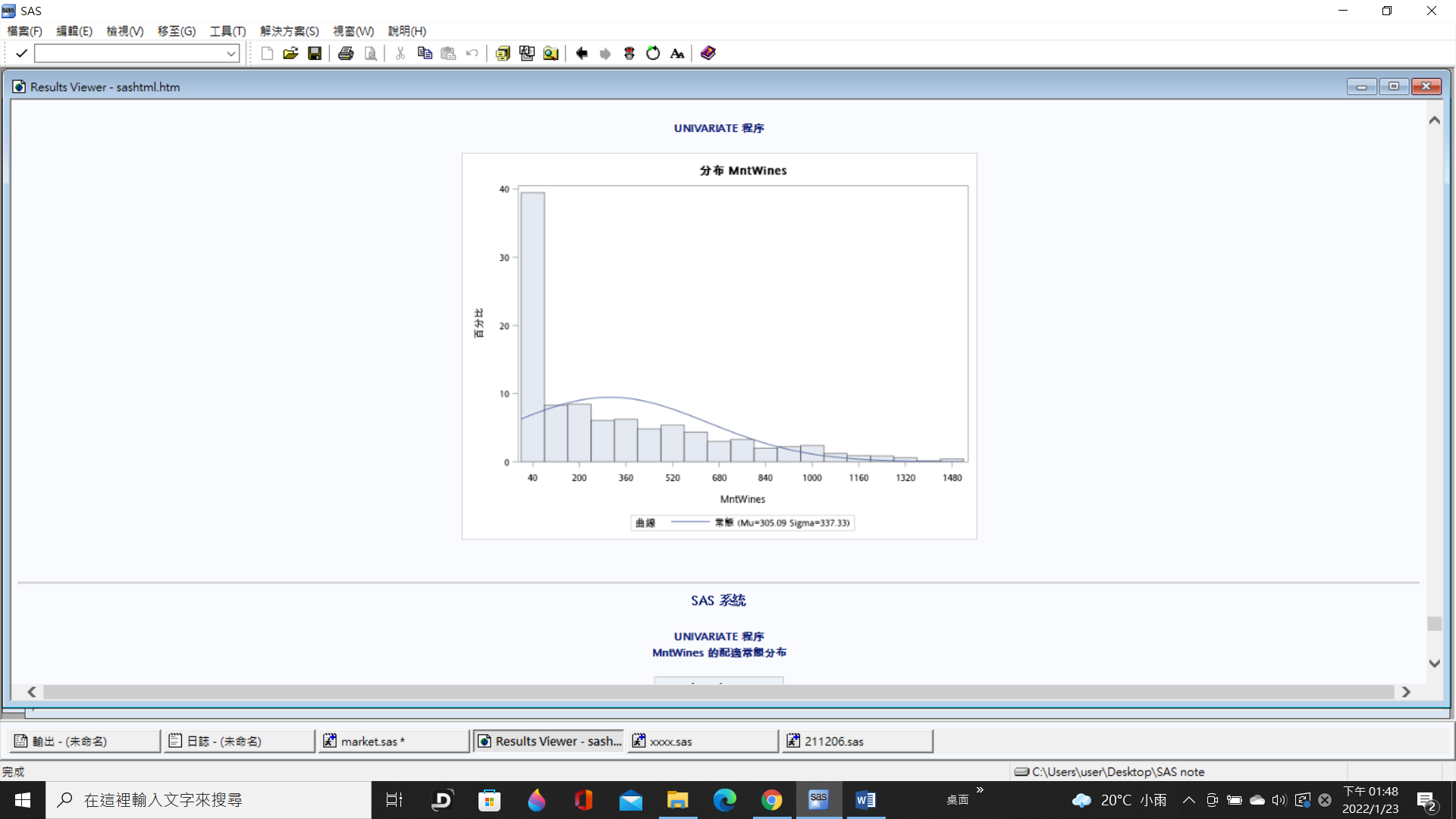
* 連續型變數：**MntWines、MntMeatProducts、MntGoldProds**的敘述性統計

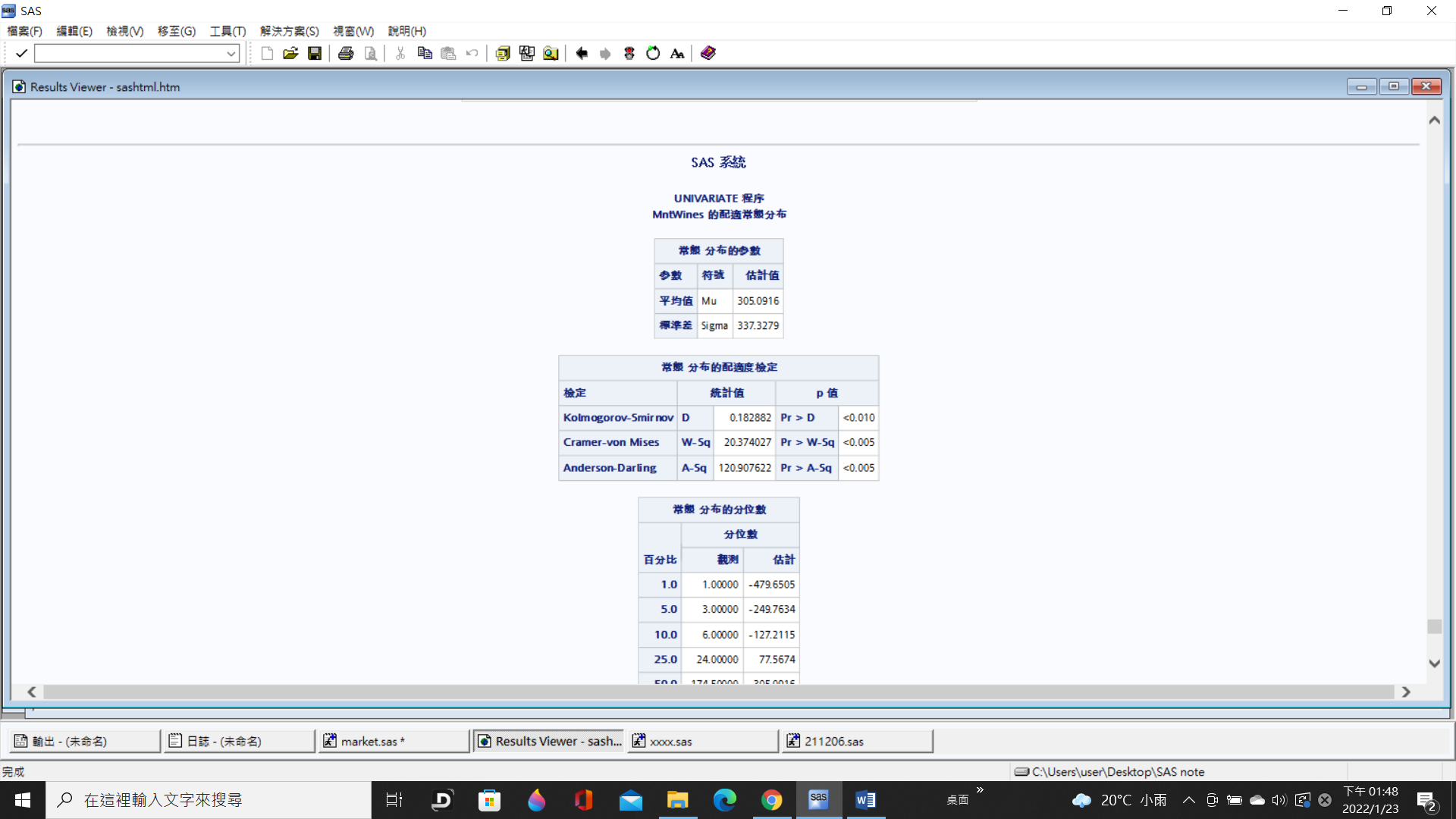


* **MntWines**此連續型資料的詳細資訊

並由紅色圈起來的部分可知，在各常態分佈的適配度檢定可知p-value<0.05，故知此變數不符合常態分配



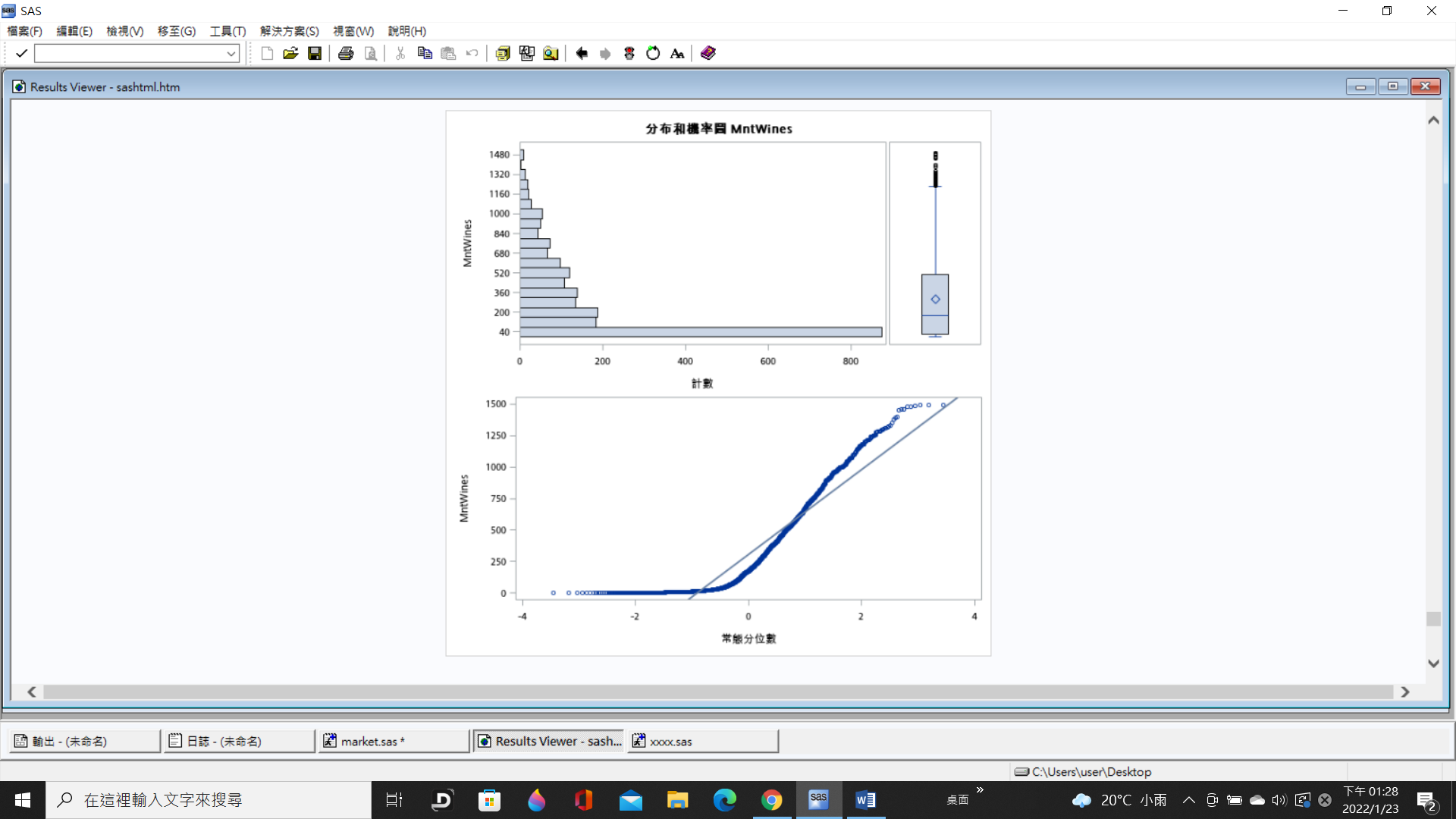


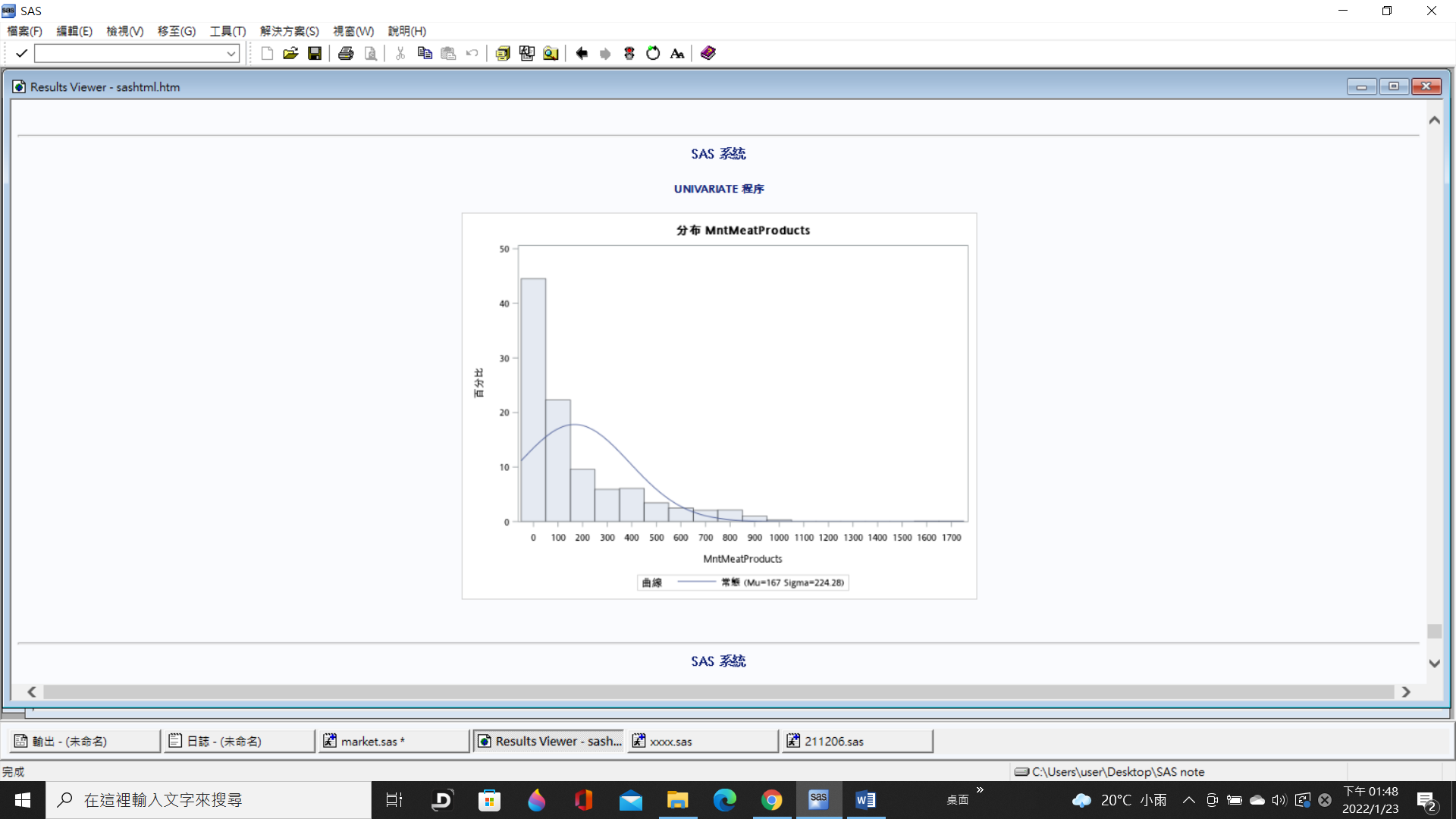


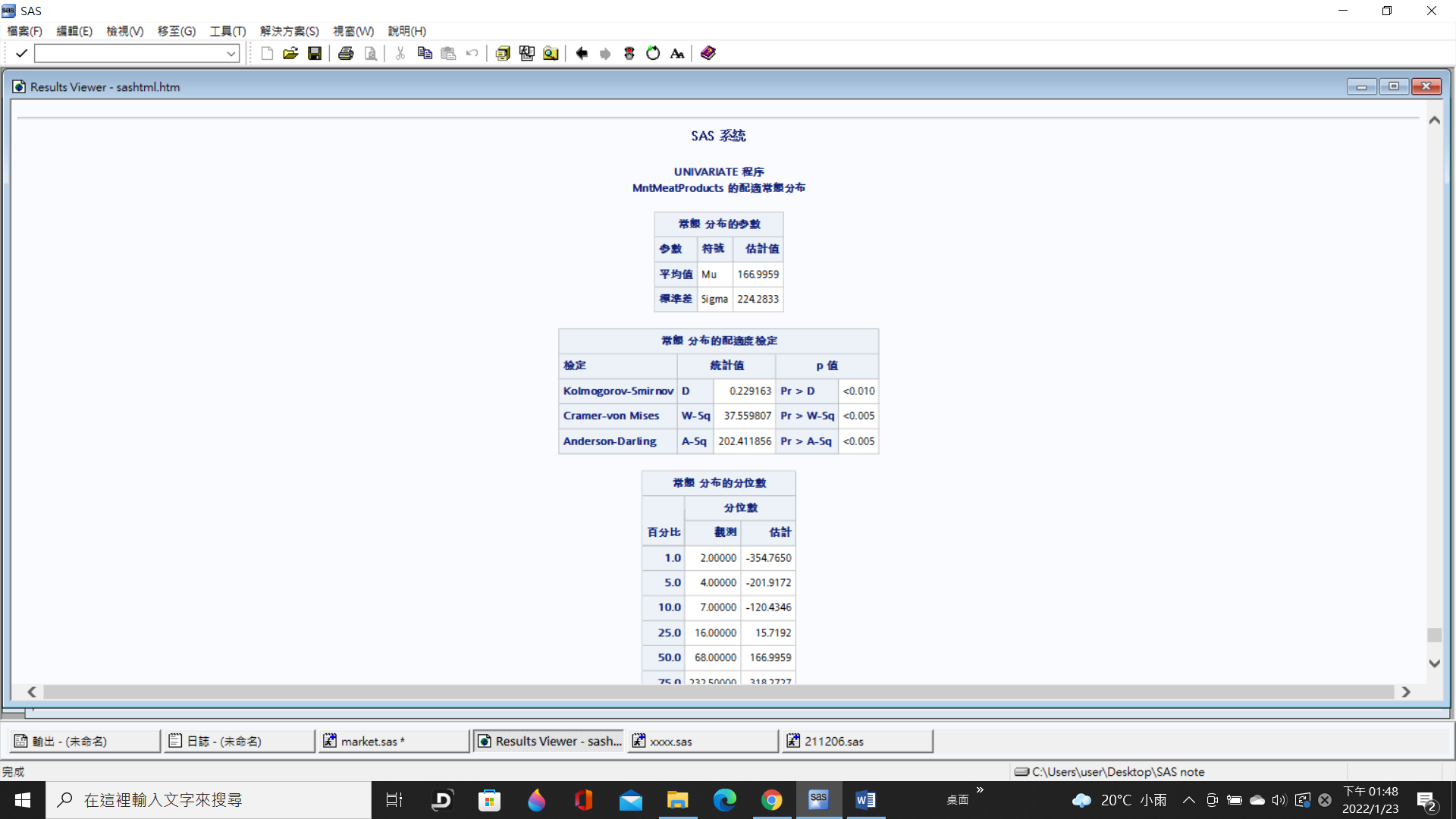
* **MntMeatProducts**此連續型資料的詳細資訊

並由紅色圈起來的部分可知，在各常態分佈的適配度檢定可知p-value<0.05，故知此變數不符合常態分配





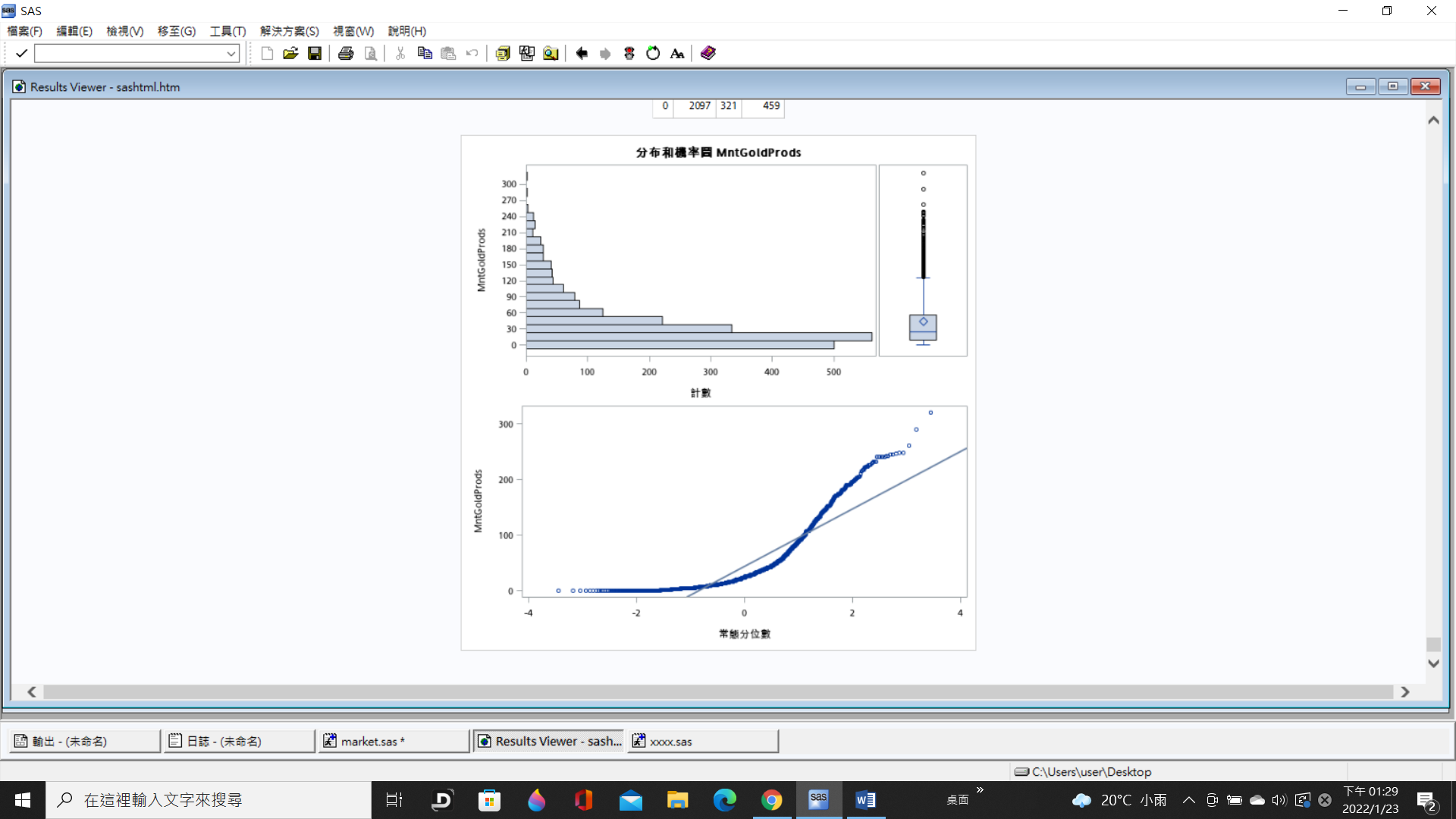


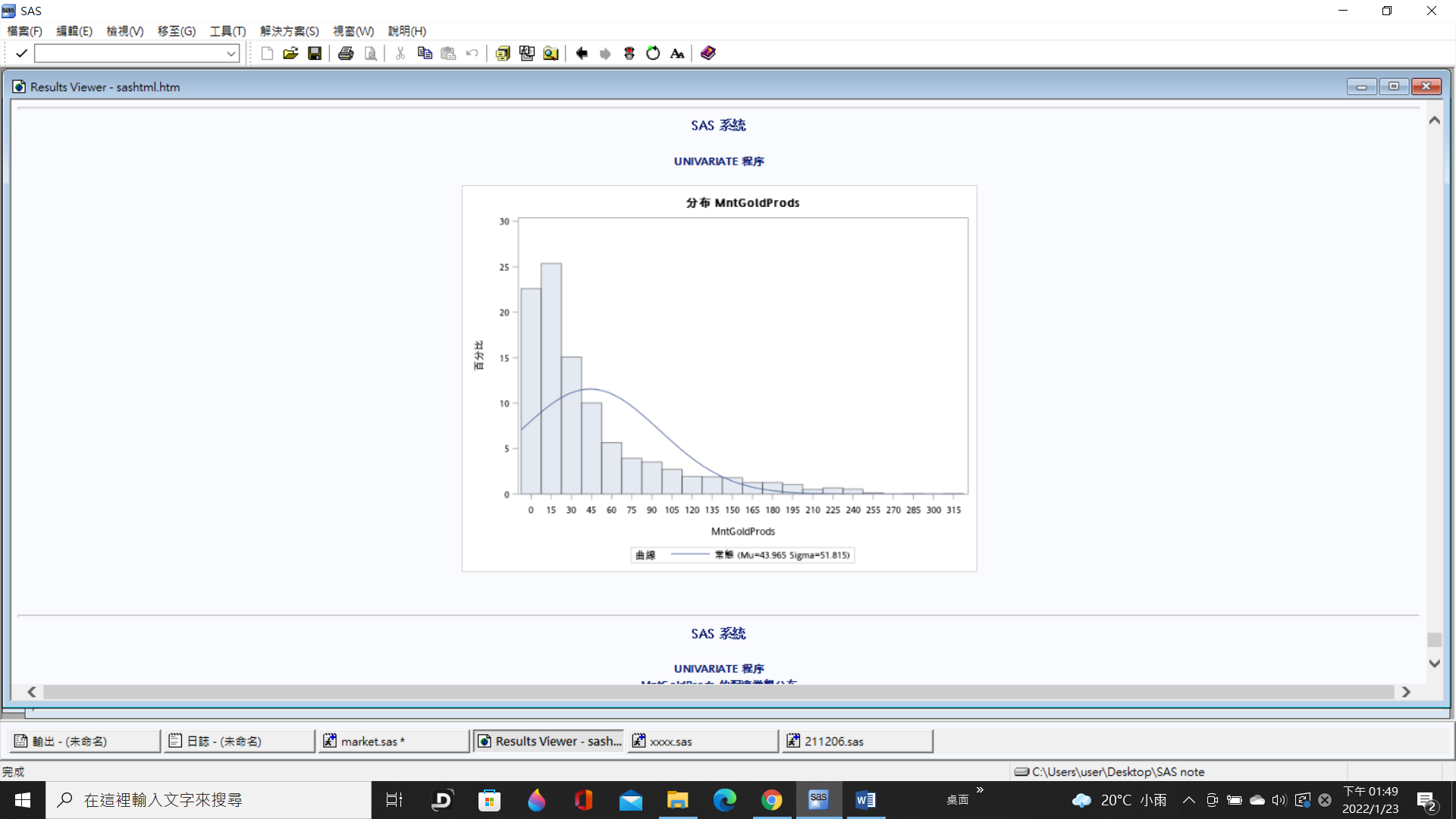


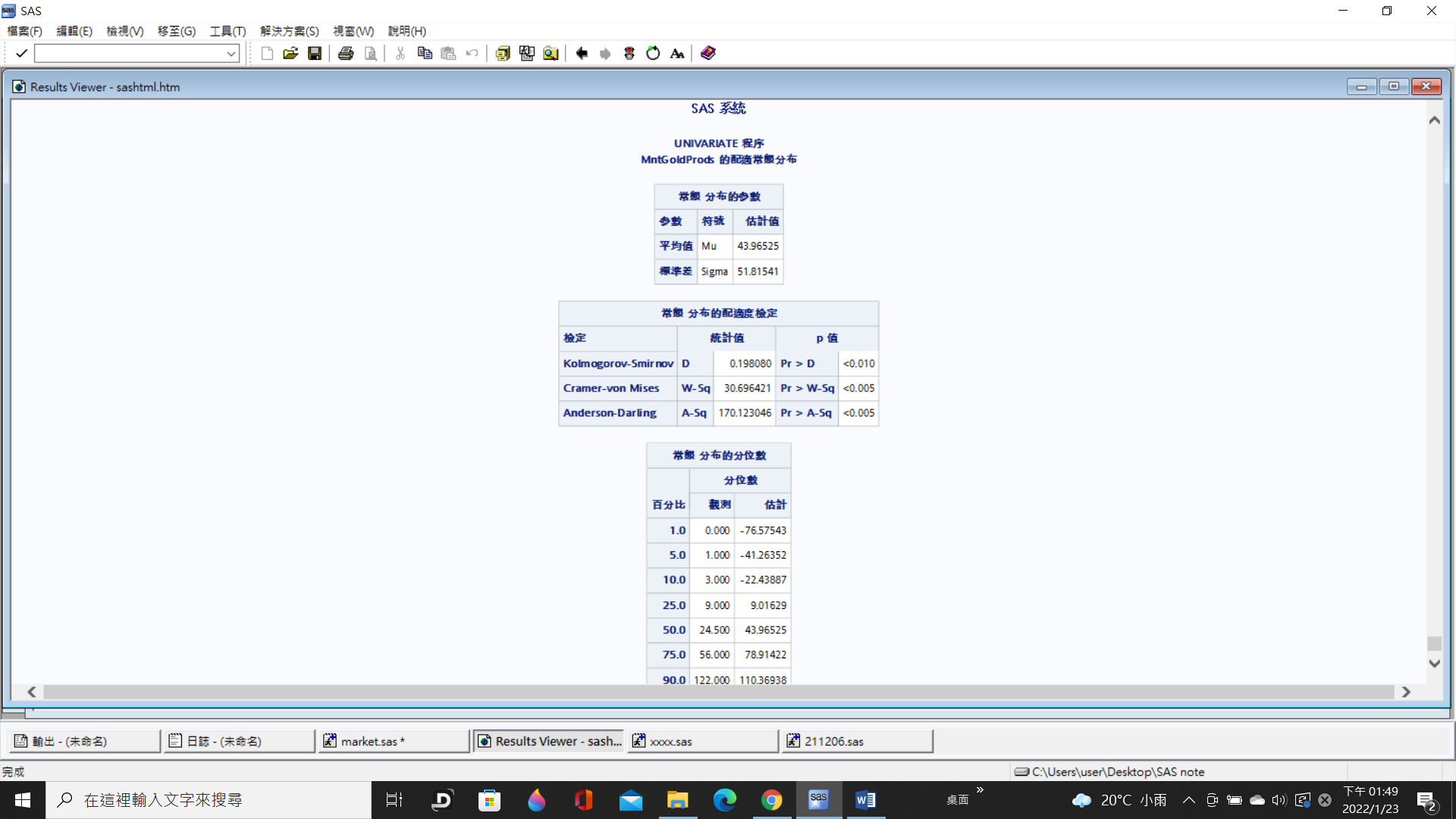
* **MntGoldProds**此連續型資料的詳細資訊

並由紅色圈起來的部分可知，在各常態分佈的適配度檢定可知p-value<0.05，故知此變數不符合常態分配



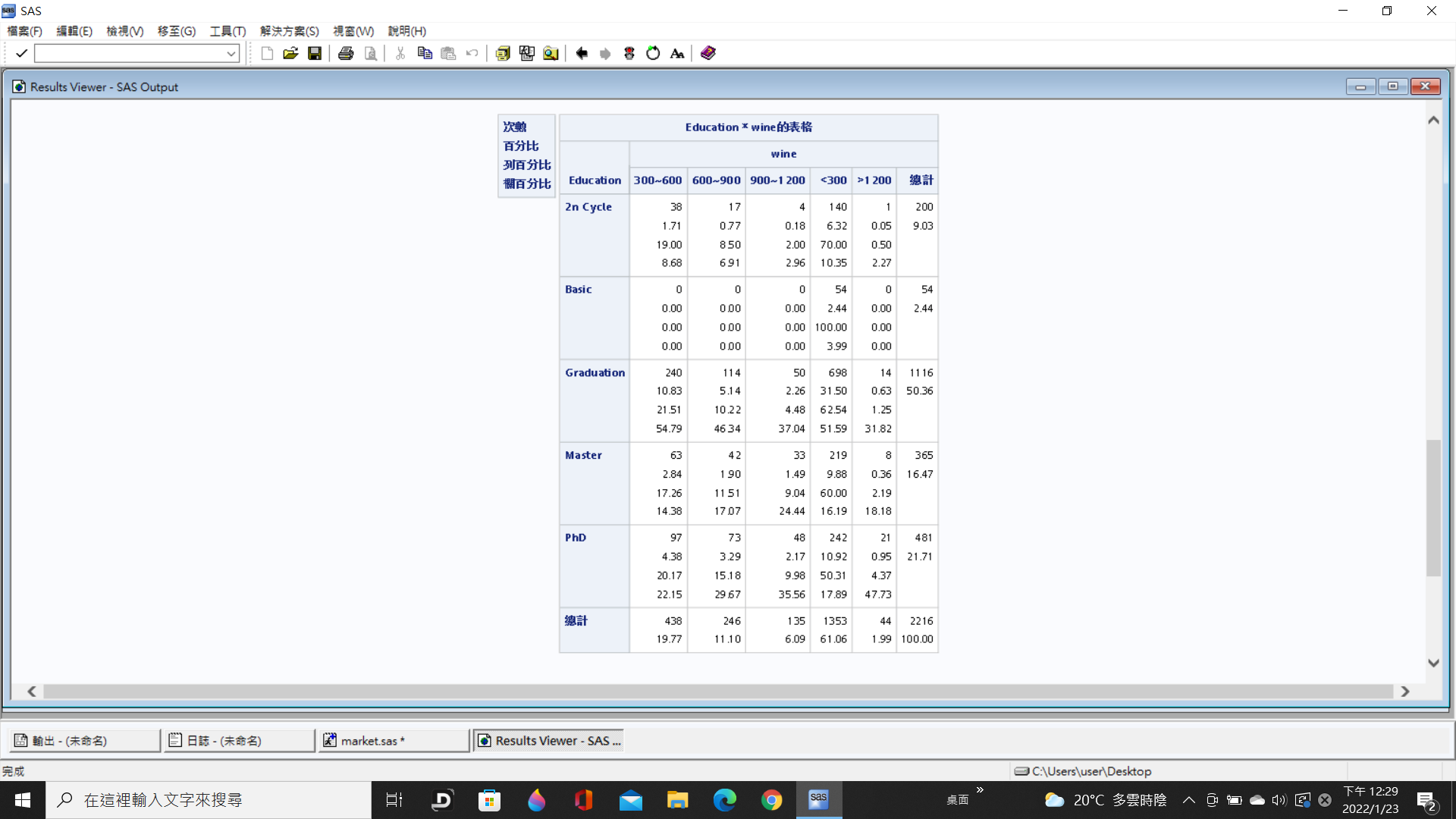




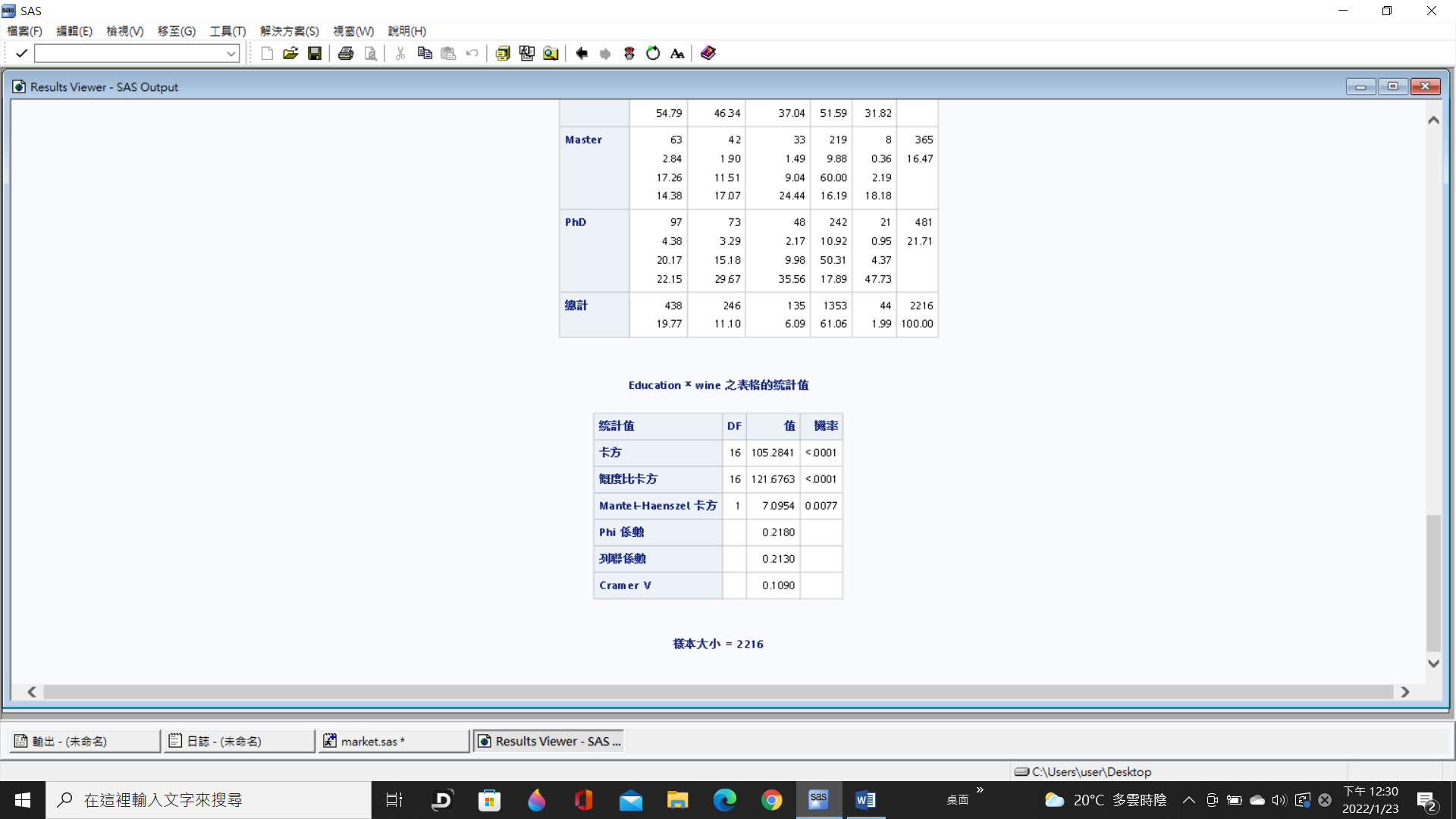


**六、獨立性檢定**

* 檢定**教育水平、感情狀況、家庭年收入**是否與顧客過去兩年在**葡萄酒**上的花費有關

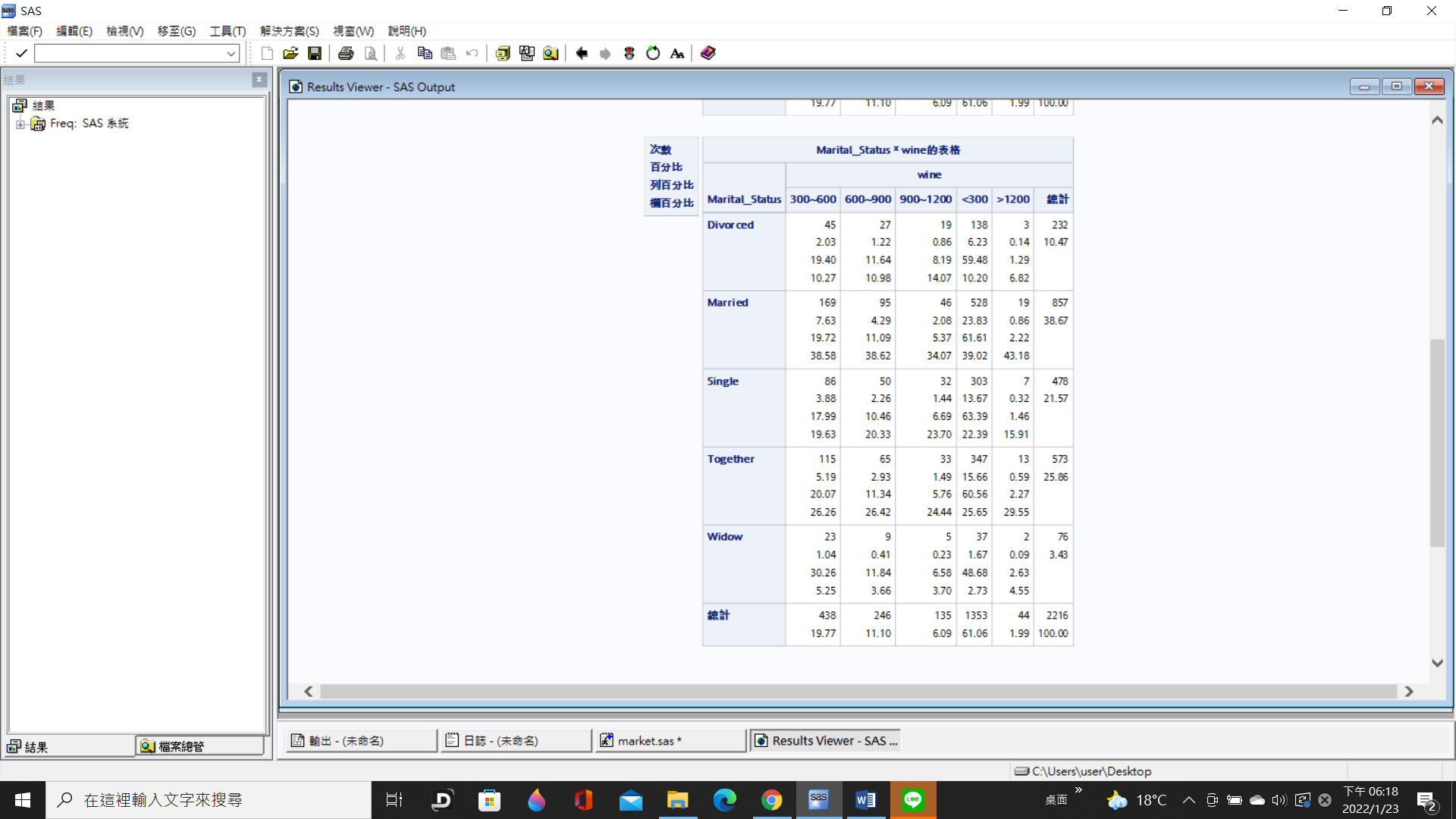


圖一 教育水平跟過去兩年在葡萄酒上的花費的列聯表

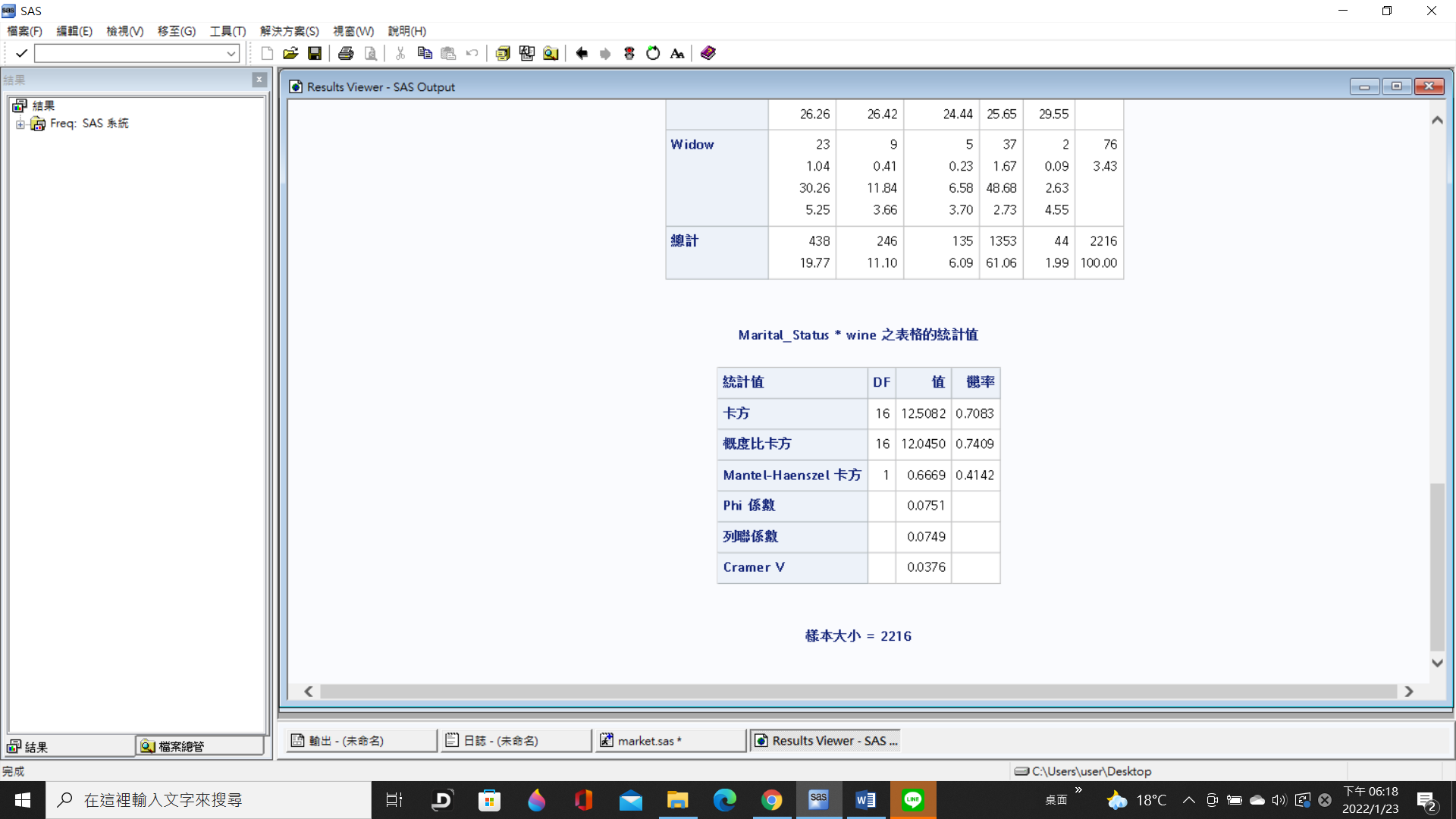


圖二 教育水平跟過去兩年在葡萄酒上的花費的卡方檢定(獨立性檢定)

由圖二可知，卡方的p-value小於0.05，所以可rejH0，也就是教育程度與過去兩年在葡萄酒上的花費不相互獨立，他們之間會互相影響。

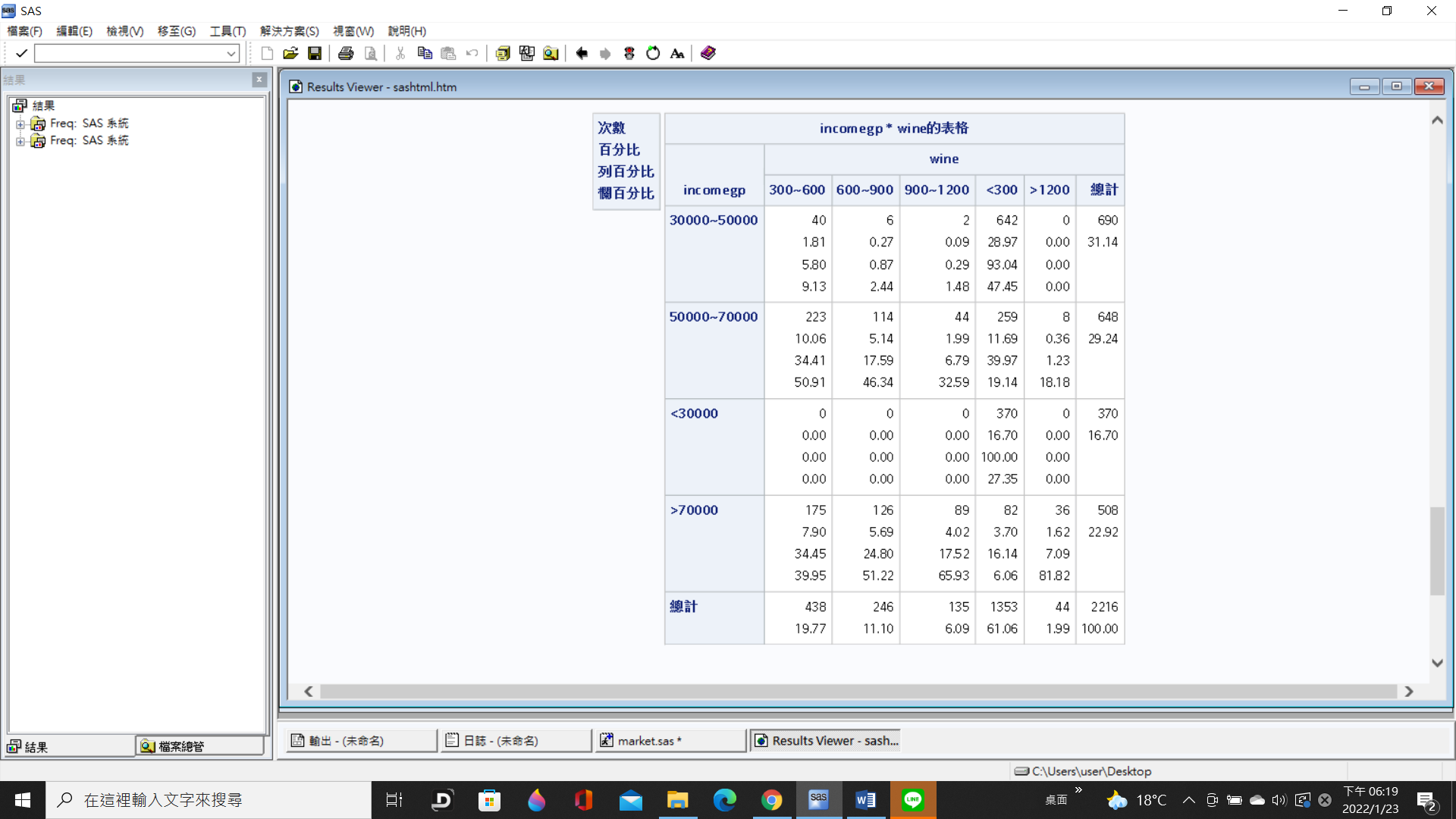


圖三 感情狀況跟過去兩年在葡萄酒上的花費的列聯表

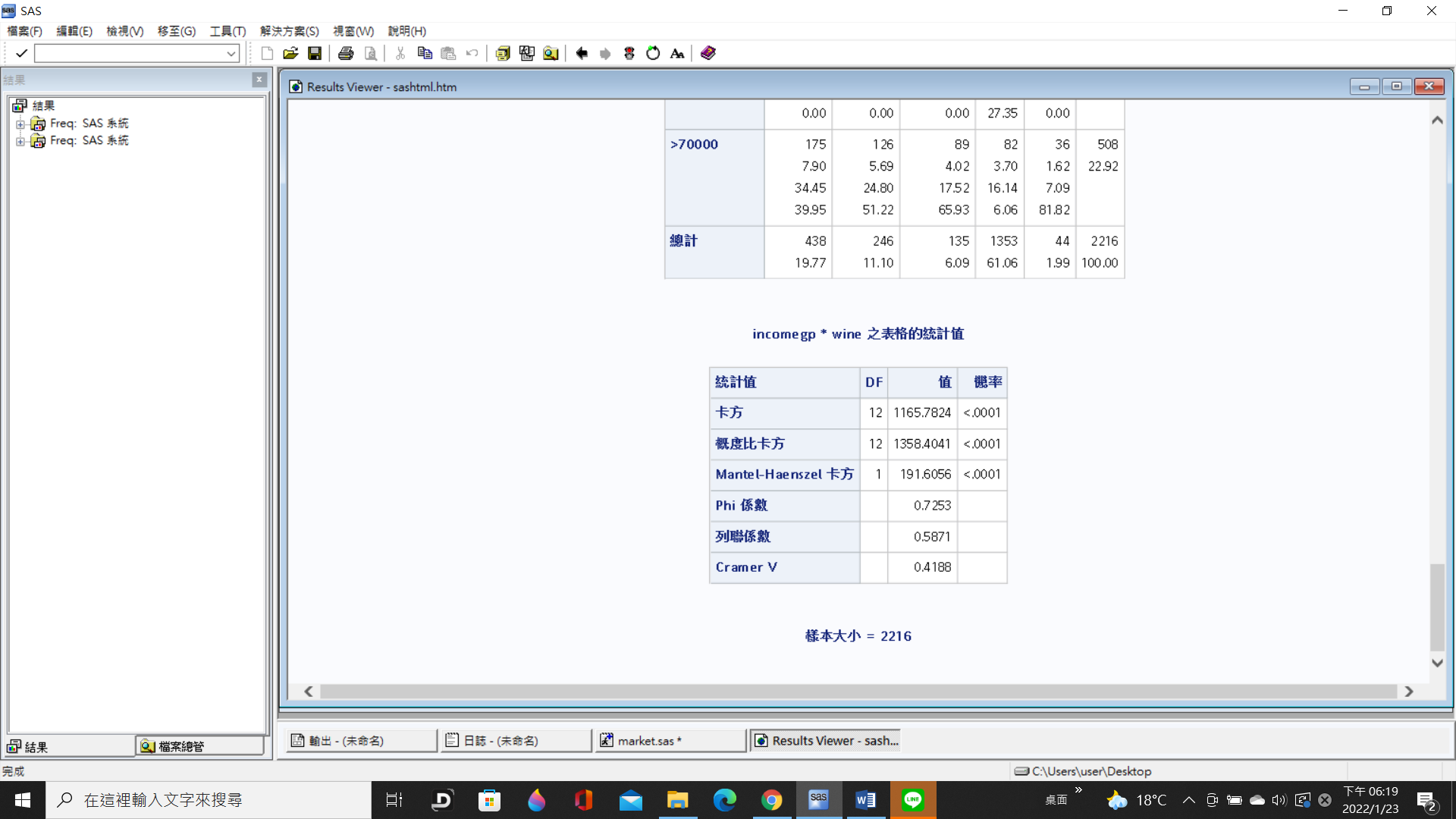


圖四 感情狀況跟過去兩年在葡萄酒上的花費的卡方檢定(獨立性檢定)

由圖四可知，卡方的p-value大於0.05，所以不可rejH0，也就是感情狀況與過去兩年在葡萄酒上的花費相互獨立，他們之間不會互相影響。



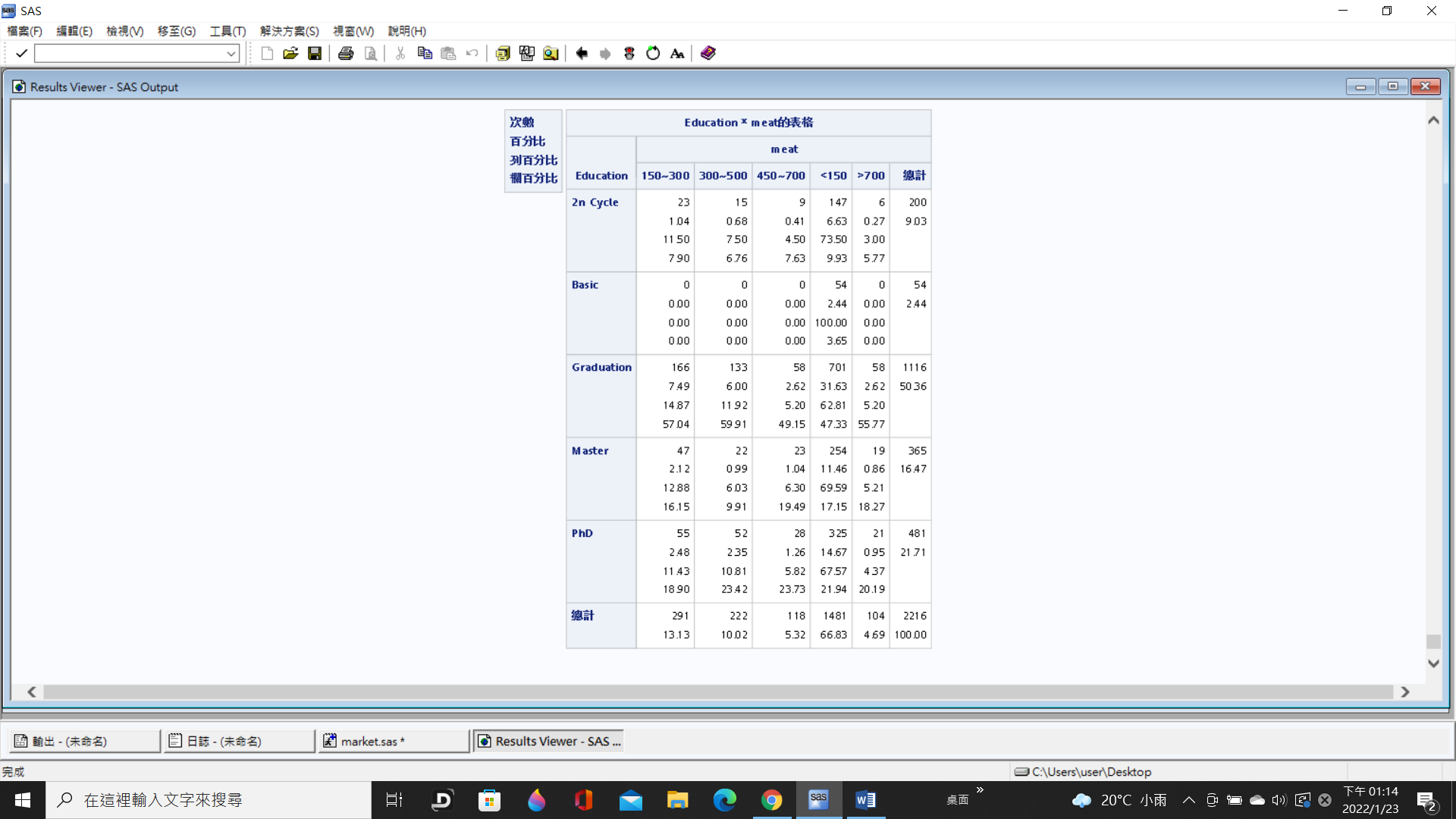
圖五 家庭年收入跟過去兩年在葡萄酒上的花費的列聯表



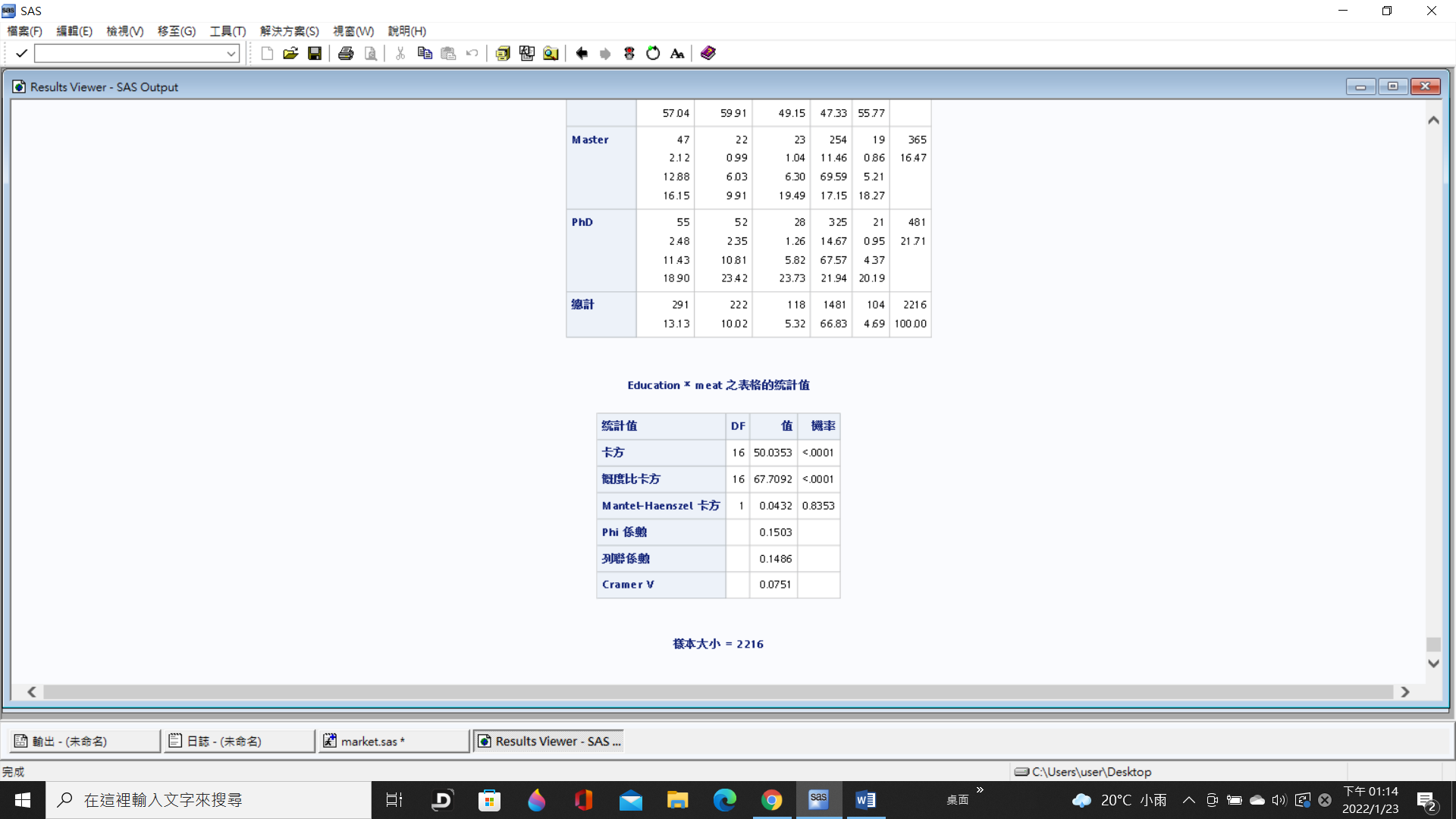
圖六 家庭年收入跟過去兩年在葡萄酒上的花費的卡方檢定(獨立性檢定)

由圖六可知，卡方的p-value小於0.05，所以可rejH0，也就是家庭年收入與過去兩年在葡萄酒上的花費不相互獨立，他們之間會互相影響。

* 檢定**教育水平、感情狀況、家庭年收入**是否與顧客過去兩年在**肉製品**上的花費有關

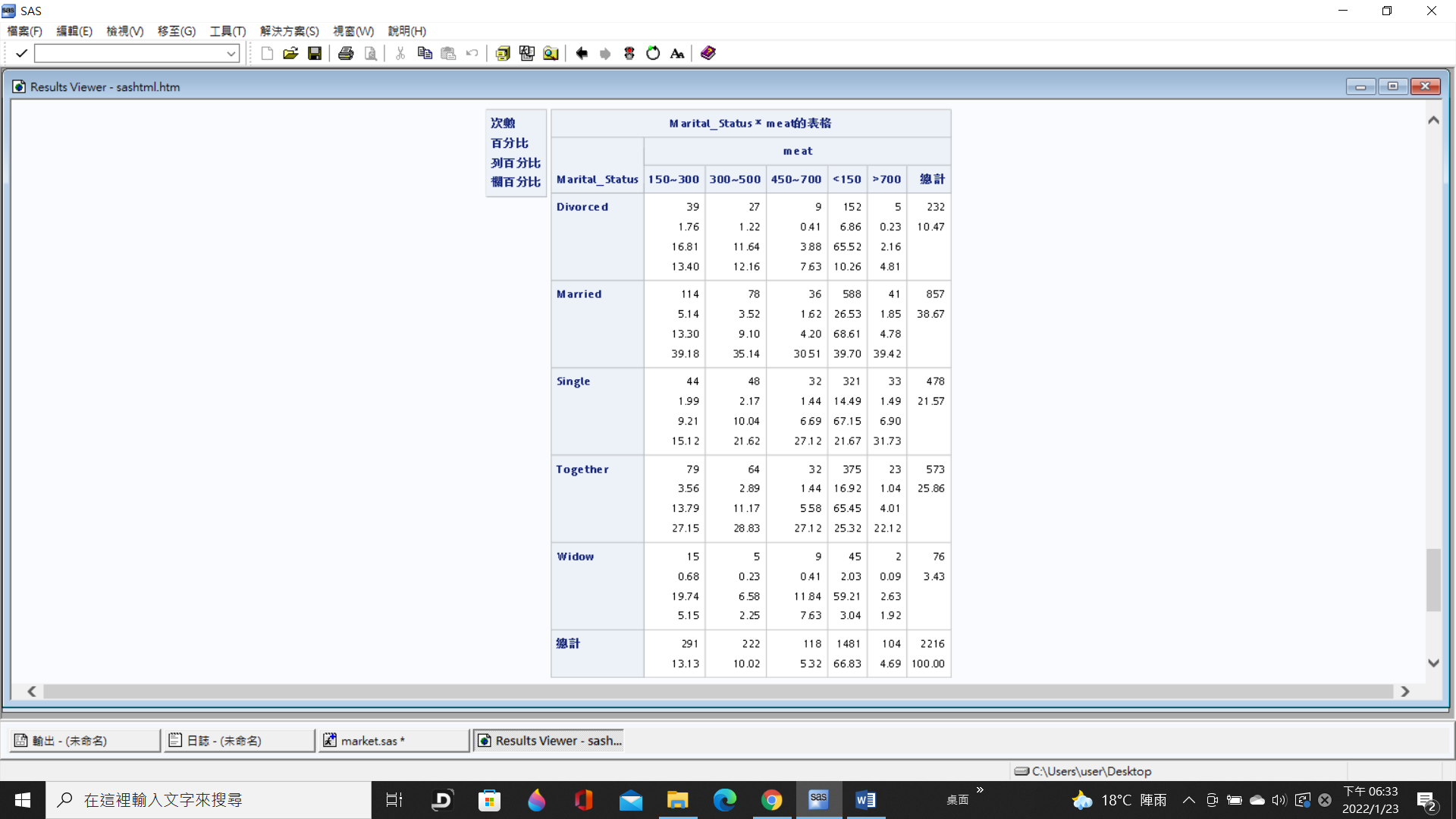


圖七 教育水平跟過去兩年在肉製品上的花費的列聯表



圖八 教育水平跟過去兩年在肉製品上的花費的卡方檢定(獨立性檢定)

由圖八可知，卡方的p-value小於0.05，所以可rejH0，也就是教育程度與過去兩年在肉製品上的花費不相互獨立，他們之間會互相影響。

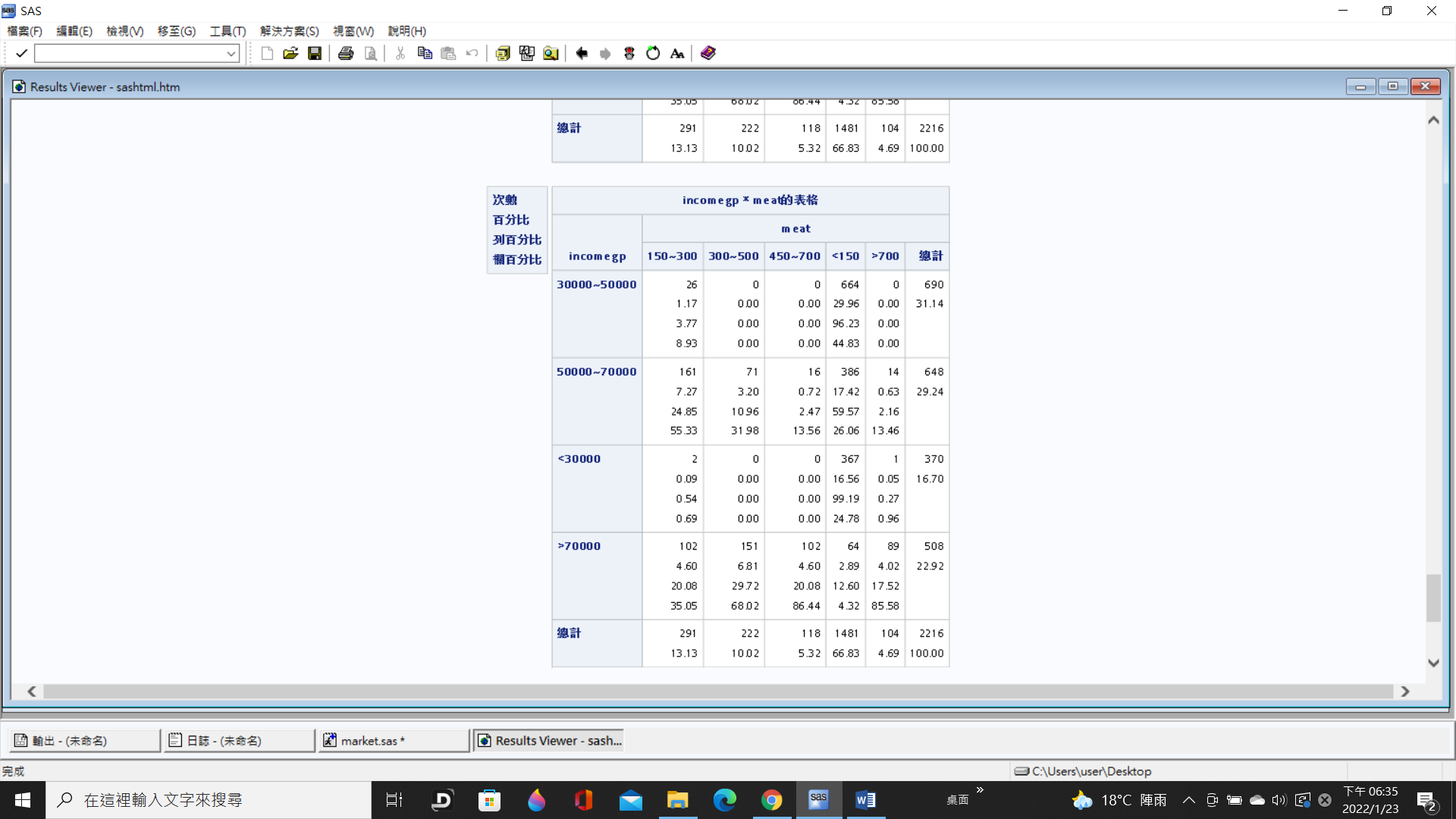


圖九 感情狀況跟過去兩年在肉製品上的花費的列聯表

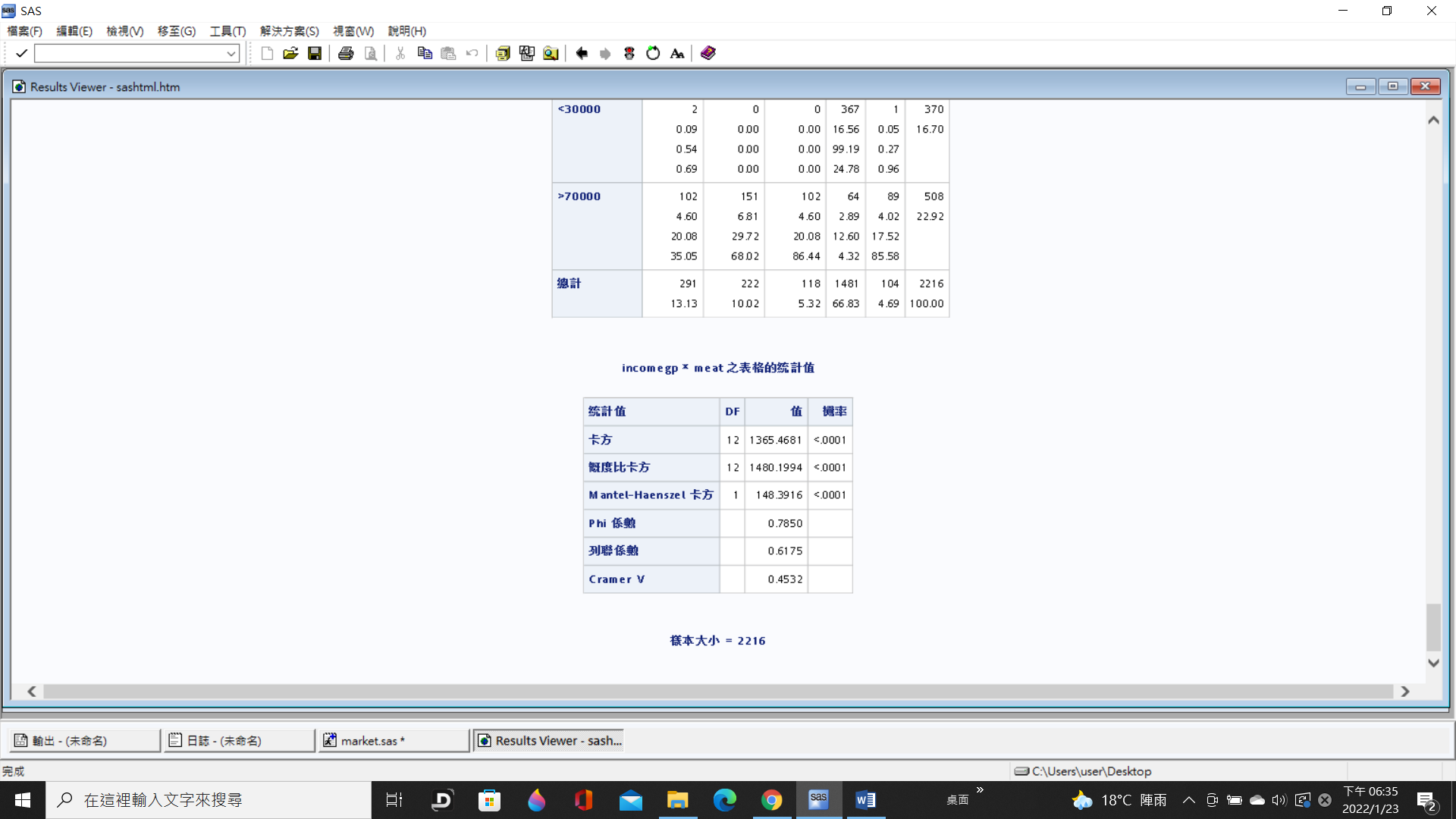


圖十 感情狀況跟過去兩年在肉製品上的花費的卡方檢定(獨立性檢定)

由圖十可知，卡方的p-value小於0.05，所以可rejH0，也就是感情狀況與過去兩年在肉製品上的花費不相互獨立，他們之間會互相影響。



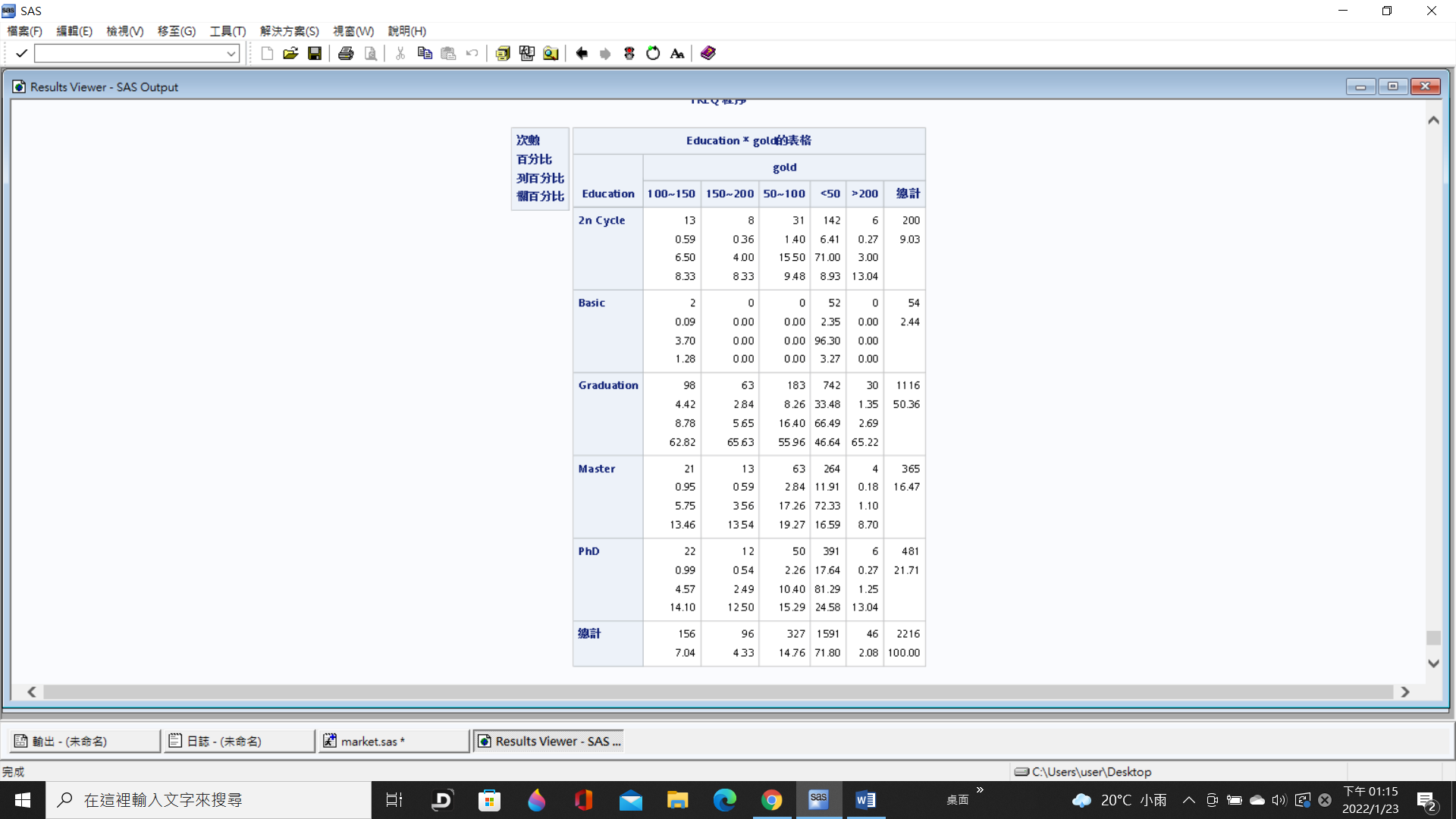
圖十一 家庭年收入跟過去兩年在肉製品上的花費的列聯表



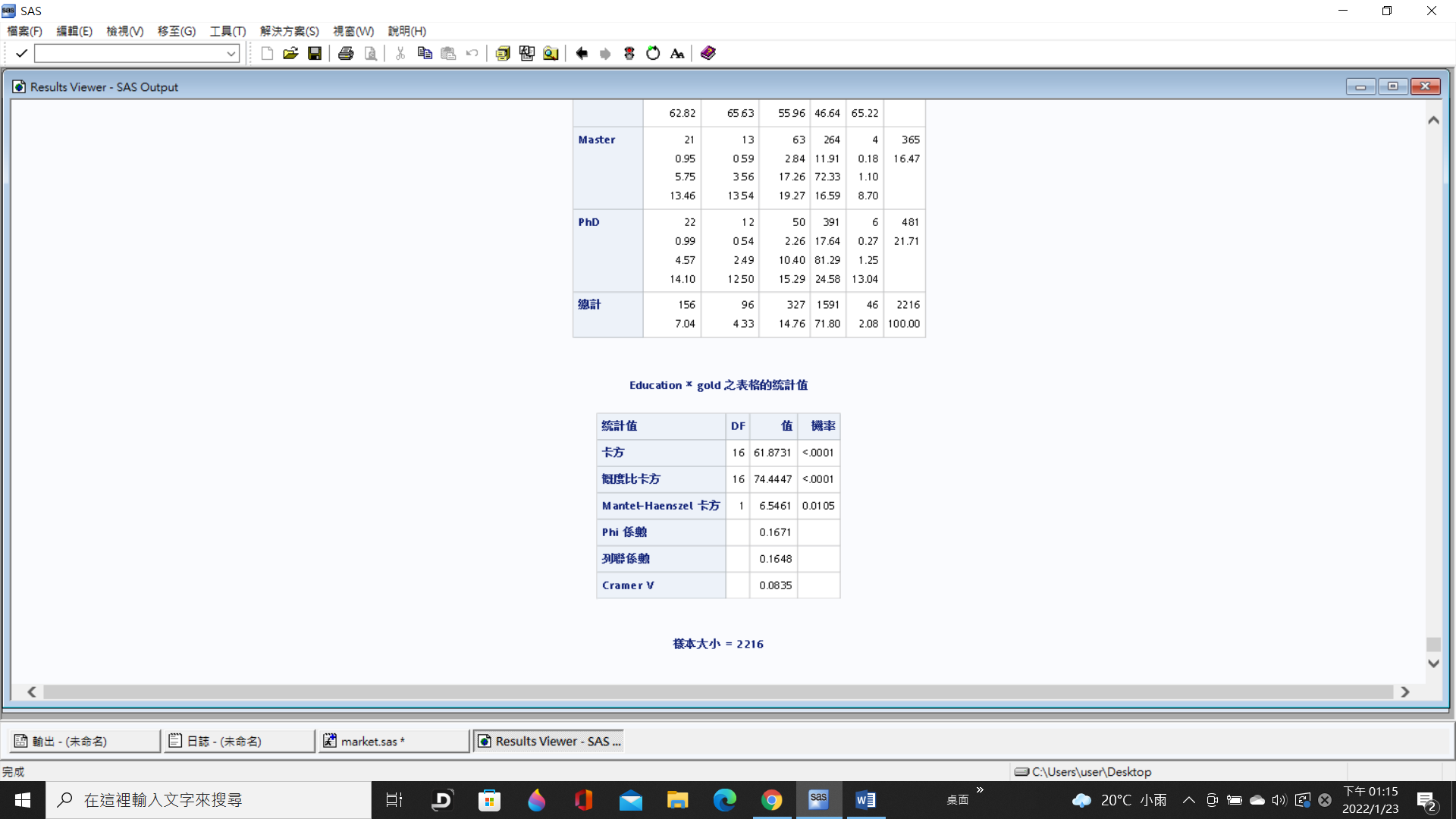
圖十二 家庭年收入跟過去兩年在肉製品上的花費的卡方檢定(獨立性檢定)

由圖十二可知，卡方的p-value小於0.05，所以可rejH0，也就是家庭年收入與過去兩年在肉製品上的花費不相互獨立，他們之間會互相影響。

* 檢定**教育水平、感情狀況、家庭年收入**是否與顧客過去兩年在**黃金**上的花費有關

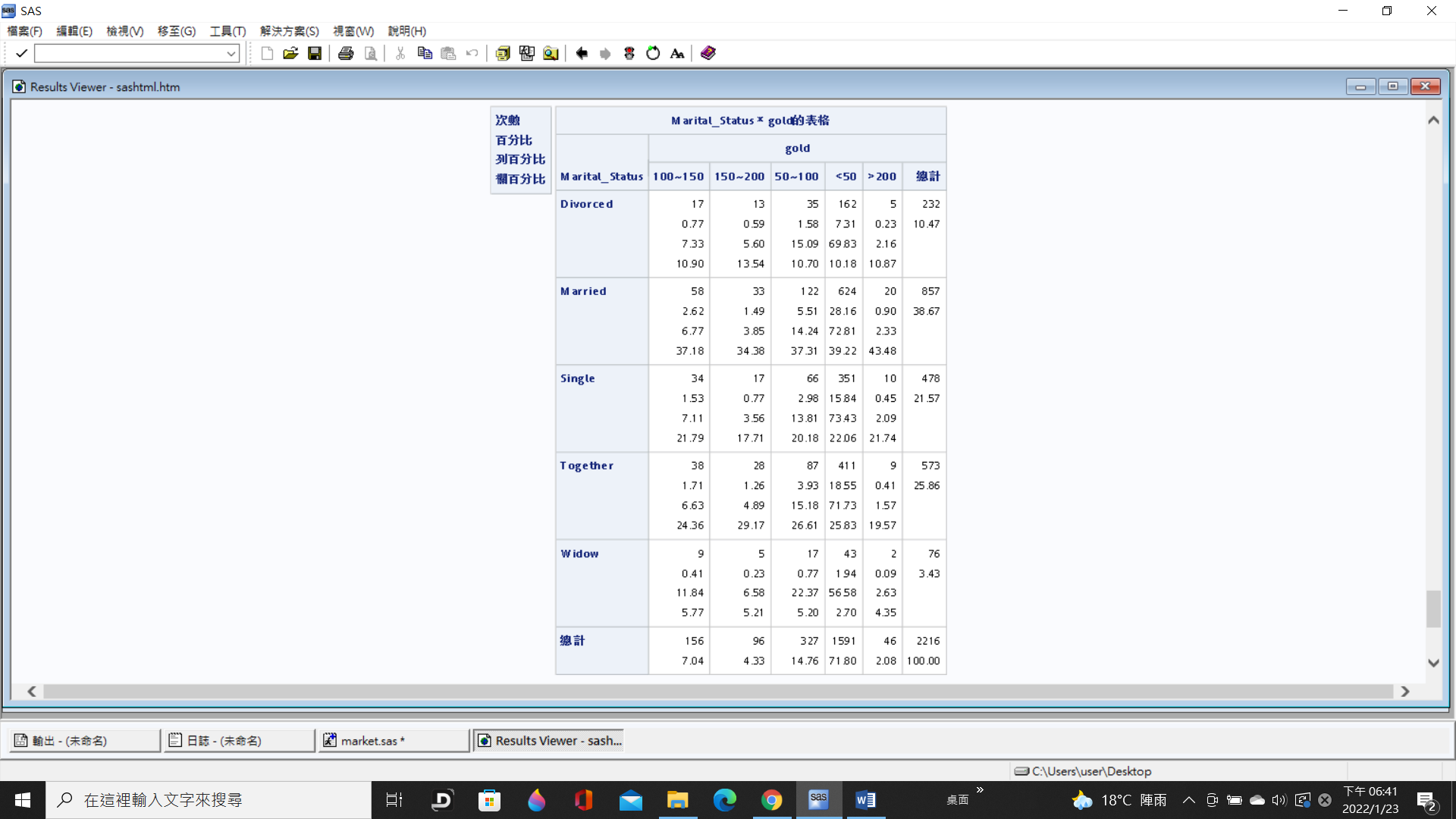


圖十三 教育水平跟過去兩年在黃金上的花費的列聯表

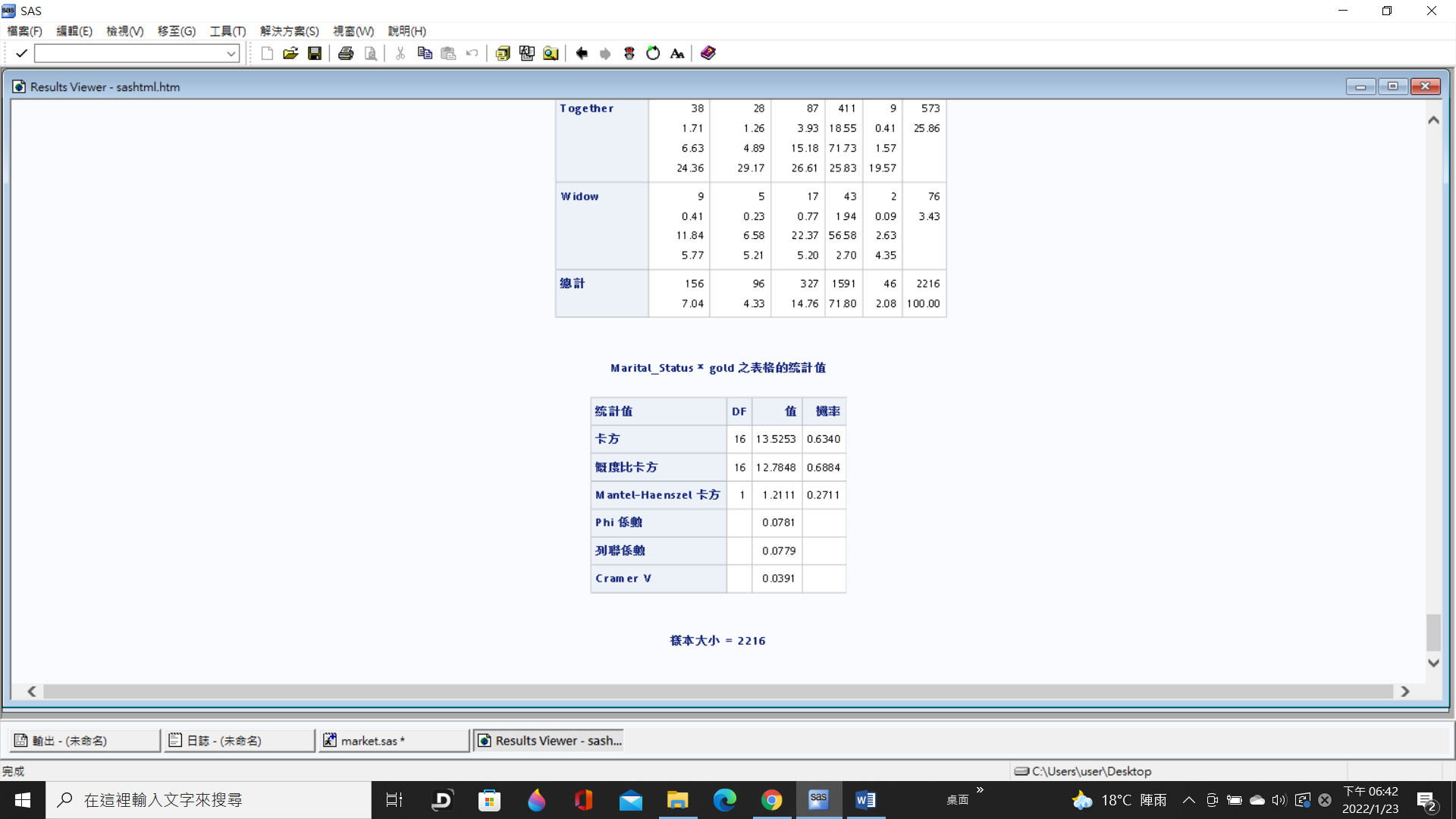


圖十四 教育水平跟過去兩年在黃金上的花費的卡方檢定(獨立性檢定)

由圖十四可知，卡方的p-value小於0.05，所以可rejH0，也就是教育水平與過去兩年在黃金上的花費不相互獨立，他們之間會互相影響。

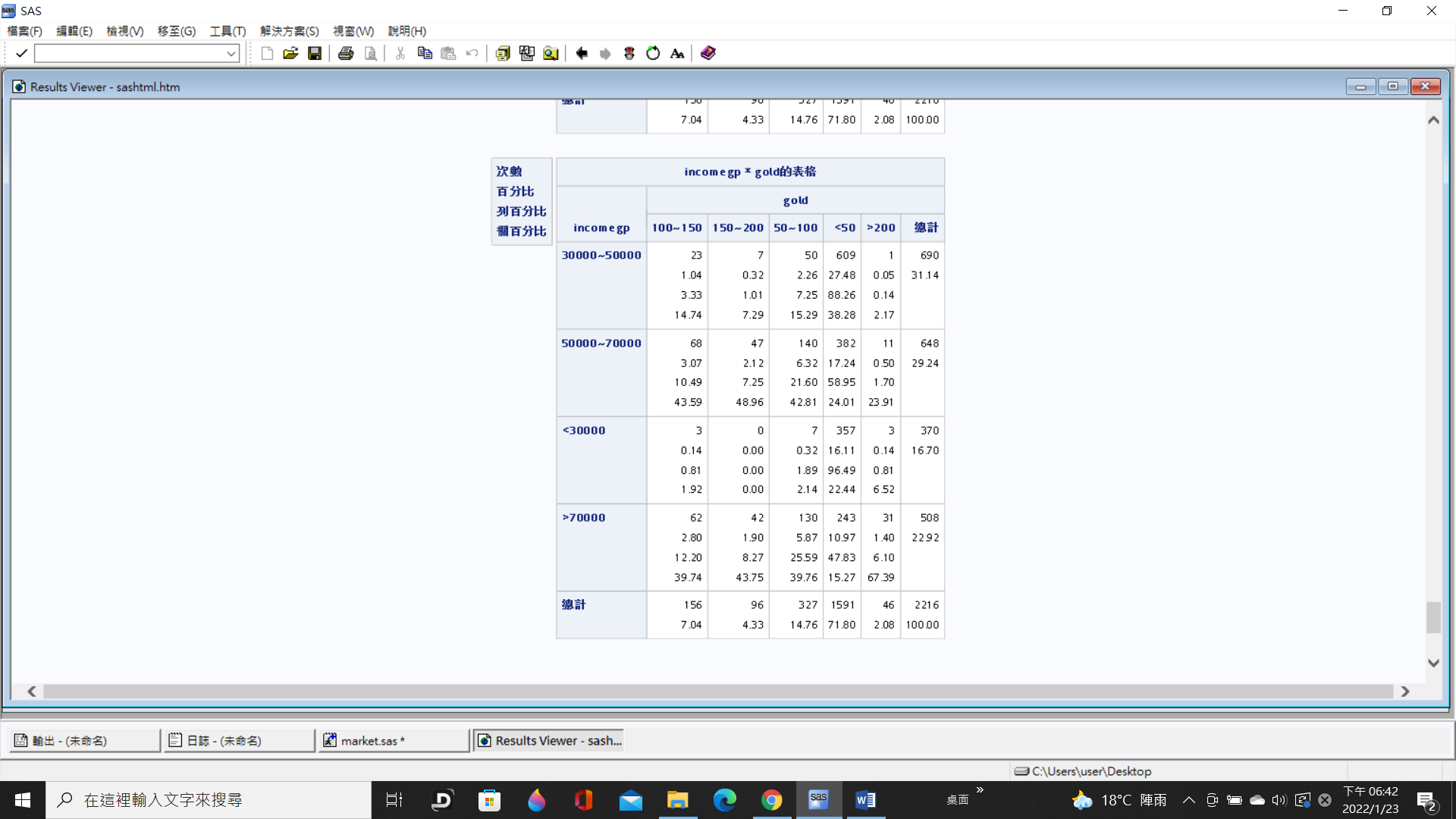


圖十五 感情狀況跟過去兩年在黃金上的花費的列聯表

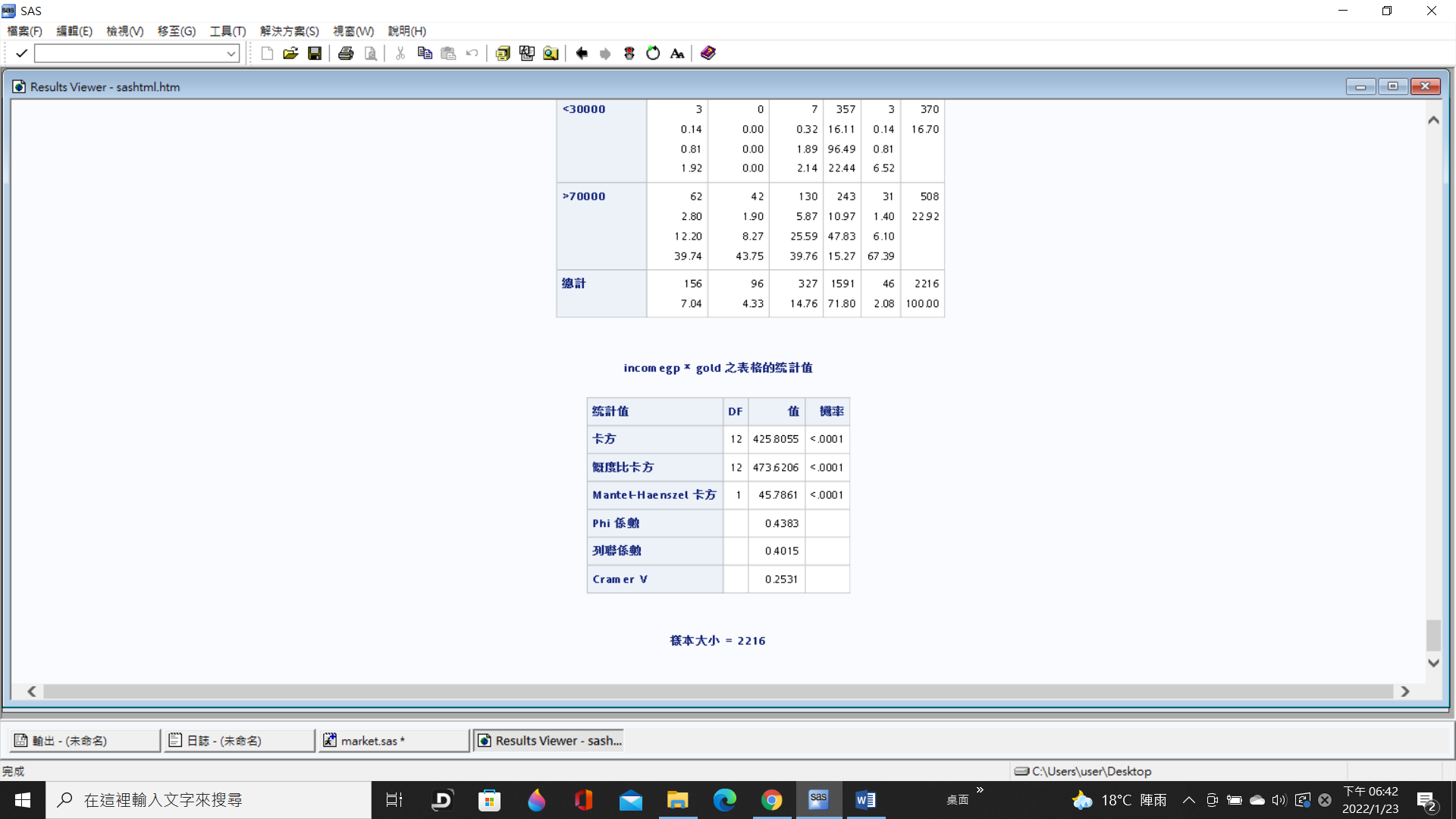


圖十六 感情狀況跟過去兩年在黃金上的花費的卡方檢定(獨立性檢定)

由圖十六可知，卡方的p-value大於0.05，所以不可rejH0，也就是感情狀況與過去兩年在黃金上的花費相互獨立，他們之間不會互相影響。



圖十七 家庭年收入跟過去兩年在黃金上的花費的列聯表



圖十八 家庭年收入跟過去兩年在黃金上的花費的卡方檢定(獨立性檢定)

由圖十八可知，卡方的p-value小於0.05，所以可rejH0，也就是家庭年收入與過去兩年在黃金上的花費不相互獨立，他們之間會互相影響。

**七、結論**

綜合圖一到圖十八，可得到以下結論 :

|  |  |  |  |
| --- | --- | --- | --- |
|  | 過去兩年在  葡萄酒上的花費 | 過去兩年在  肉製品上的花費 | 過去兩年在  黃金上的花費 |
| 顧客教育水平 | 有相關 | 有相關 | 有相關 |
| 顧客感情狀況 | 無相關 | 有相關 | 無相關 |
| 顧客家庭年收入 | 有相關 | 有相關 | 有相關 |

不論是在葡萄酒、肉製品還是黃金上的花費，都跟顧客的教育水平、家庭年收入有關;但顧客的感情狀況，只跟肉製品上的花費有關，跟葡萄酒、黃金的花費則無關。

也就是說，若是廠商想要販售葡萄酒、肉製品，或是黃金的商品，可針對不同教育水平、家庭年收入的顧客，分別制定合適的價格與促銷活動。但是，如果廠商想在情人節這類的節日針對顧客的感情狀況去制定銷售價格等等的，只有對於肉製品的廠商會比較適合。