應用迴歸分析期末報告

主題：臺灣地區果品生產概況

組員：05170463 蔡宜諠

05170561 張芳瑜

工作分配表

|  |  |
| --- | --- |
| 姓名 | 負責部分 |
| 05170463 蔡宜諠 | 迴歸模型分析、結論與建議 |
| 05170561 張芳瑜 | 緒論、台灣果品生產分析 |

目錄

第一章 緒論

第一節 研究背景與動機

第二節 研究目的

第三節 研究範圍與資料來源

第四節 變數介紹

第五節 研究流程

第二章 台灣果品生產分析

第一節 以年度區分分析果品生產

第二節 以地區區分分析果品生產

第三章 迴歸模型分析

第一節 建立簡單線性迴歸模型

第二節 建立多元線性迴歸模型

第四章 結論與建議

第一節 結論

第二節 建議

**第一章 緒論**

**第一節 研究背景與動機**

台灣位處亞熱帶地區，但因為海拔高度不同的環境條件豐富，以致熱帶至寒帶的氣候條件都有，所以可供栽植的水果種類繁多，像是來自中國的梅、李、桃以及源自中亞的蘋果、葡萄等溫帶水果，或是源自東南亞的楊桃、芒果、蓮霧、香蕉以及中南美洲的番荔枝(俗稱釋迦)、百香果、鳳梨、三角柱仙人掌(俗稱火龍果)、酪梨、木瓜、番石榴等熱帶水果，都先後來到台灣寶島，同時我們也研發出許多優異的品種，讓大家的飲食生活更形增色不少。

今天的台灣被譽為水果王國，水果早已成為我們日常生活的一部份，街上隨處都可見玲瑯滿目的水果攤。台灣水果百百款，而每樣水果盛產時期皆不盡相同，因現今科技的農業技術發達，許多水果已可一年四季皆可生產，水果好壞仍需視當時氣候、栽種地區等因素，與食用者個人口味不同也有很大關係。

最近台灣香蕉產量過剩，導致造成價格崩跌，不論香蕉等級好壞，全送到收購點，造成農會集貨場香蕉堆積如山，看著農政官員以刀片破壞外觀（去商品化），送去當堆肥或餵豬，一年心血就這樣丟棄，農民看了心都淌血卻很無奈。

因為這新聞，讓我們想了解台灣水果的生產概況，很多水果是從哪裡來的我們都不清楚，我們可以藉由這主題來研究，而且我們還能比較各縣市每項水果的生產數量和地區之間的差異。

**第二節 研究目的**

1.了解台灣每年產量的變化

2.了解台灣地區變化

3.求出簡單線性迴歸線

4.求出多元線性迴歸線

**第三節 研究範圍與資料來源**

1. 研究範圍：2013-2017年度臺灣地區17種果品

二、資料來源：政府資料開放平台

https://data.gov.tw/dataset/9680

**第四節 變數介紹**

|  |  |  |
| --- | --- | --- |
| 變數名稱 | 變數中文 | 介紹 |
| year | 年度 | 西元年為主 |
| city | 地區別 | 全台的縣市 |
| kinds | 果品類別 | 共有17種水果和一個其他水果 |
| pantingmany | 種植株數 |  |
| rewardmany | 收穫株數 |  |
| pantingarea | 種植面積 |  |
| rewardarea | 收穫面積 |  |
| manygoods | 每株平均產量 |  |
| areagoods | 每公頃平均產量 |  |
| goods | 產量 |  |

**第五節 研究流程**

**第二章 台灣果品生產分析**

**第一節 以年度區分分析果品生產**

1.我把各年的平均數做繪圖

****

1. 由折線圖可以明顯發現，2014年產量最高，2016年產量最低
2. 2014年台灣水果，是主導台灣經濟發展的重要支柱，大量水果外銷所創造的外匯收入，已成為台灣產業發展的隱形力量，所以2014年產量極高。

3.2016年水果產量大幅下降，是這五年來產量最低的，原因是2016年寒流、颱風、暖冬，重創農業。「水果王國」怎麼了？農業學者、氣象專家與農民，將禍首指向極端氣候。農委會副主委、農經學者陳吉仲說，「農業部門受氣候變遷影響最嚴重，」尤其，當極端天氣發生在農作成長關鍵期，如抽芽、開花當週遇到寒流、暴雨，「結果率都會大大降低，影響收成，」台大農藝系退休教授郭華仁憂心。2016年台灣遭逢一個霸王寒流、兩個強颱，加上68年來最熱的冬至；極端氣候，對需要穩定的農業造成極端破壞，推升出極端價格。

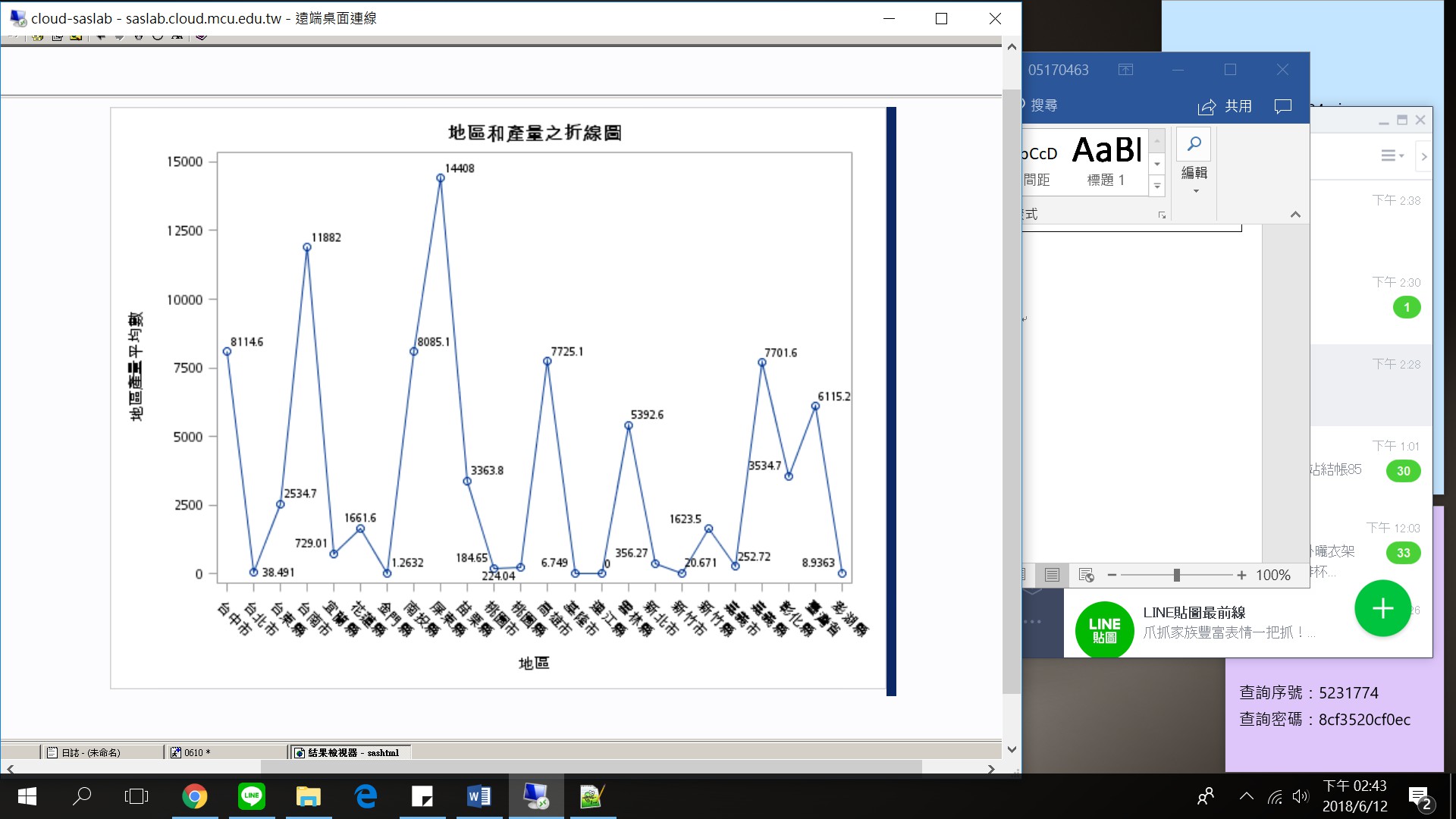
2.我分年做產量比較

|  |  |
| --- | --- |
|  | 2013年總和前三高的水果，第一鳳梨，第二香蕉，第三芒果，最低分別是，溫州蜜柑、斗柚、和蘋果。 |
|  | 2014年總和前三高的水果，第一鳳梨，第二香蕉，第三柳橙，最低分別是，溫州蜜柑、斗柚、和蘋果。 |
|  | 2015年總和前三高的水果，第一鳳梨，第二香蕉，第三芒果，最低分別是，溫州蜜柑、斗柚、和晚崙西亞橙。 |
|  | 2016年總和前三高的水果，第一鳳梨，第二香蕉，第三番石柳，最低分別是，溫州蜜柑、斗柚、和晚崙西亞橙。 |
|  | 2017年總和前三高的水果，第一鳳梨，第二香蕉，第三番石柳，最低分別是，溫州蜜柑、斗柚、和晚崙西亞橙。 |

很明顯的這五張圖，水果總和最高的都是鳳梨和香蕉。鳳梨在台灣不只是豐富的水果來源植物，也是重要的經濟作物，除大量供作水果食用外，加工製造鳳梨罐頭及飲料、果醬銷售國內外，製作蜜餞、鳳梨酥等甜點亦頗負盛名，尚可利用殘渣製造酒、醋。目前台灣鳳梨外銷地區以日本、香港、新加坡、加拿大為主。台灣的香蕉不只國內暢銷，還有以外銷到日本為主。

**第二節 以地區區分分析果品生產**

1.以地區區分畫出的平均數圖



1.屏東位於台灣最南端，所有的水果節氣都比台灣任何一個地方早，遭遇的寒流不會比中部強，凍傷等寒害問題較少。屏東的香蕉面積為全臺第1大產區。全年都可生產，屏東縣種植的愛文芒果面積是全臺第2大，。屏東位屬臺灣最南端，加上地理環境臨海，所以有充足的陽光照射及海風吹拂，加上土壤等條件適合種植及農民專業技術不斷精進，所生產的果品品質極佳。

2.台南文旦、鳳梨、芒果，因為台南比較氣候也不定熱的時候很熱冷的時候很冷，而且比較潮濕，所以在當地也研究過其他農業，不過都沒進展到後來才知道當地可以種這些東西鳳梨和芒果文旦都出產到日本、大陸和印尼。

2.下面是以產品區分的地區分區圖：

|  |  |
| --- | --- |
|  | 這16種水果中，產量最多得是鳳梨，鳳梨大多生產在高屏、其次是雲嘉南、再來是中彰投。 |
|  | 這16種水果中，產量最多得是香蕉，香蕉大多生產在高屏、其次是雲嘉南、還有中彰投和東部。和前一張鳳梨圖的生產地區是差不多的。 |
|  | 這三種水果是橙柑的產量較高，橙柑大多生產在雲嘉南、中彰投和桃竹苗，和前兩張圖不太一樣。 |

**第三章 建立線性迴歸模型**

**第一節 建立簡單線性迴歸模型**

1.自變數為種植面積：

|  |  |
| --- | --- |
|  | 1. 從綠色框框可以看得出來，=0.3787，這代表種植面積解釋了37.87%的產量，算是不太好的。 2. 從紅色框框可以看出p-value < 0.001，因此有足夠證據顯示依變數(產量)得以被自變數(種植面積)所解釋。 |
|  | 1.因為我建立的是簡單迴歸模型，所以可以由藍色的框框可以看出Pr>|t|小於 0.05，達顯著水準，表示自變項(種植面積)能讓模式更良好  2.由黃色框框可以求得簡單線性迴歸線：產量=166.772+種植面積0.164 |
|  | 此為殘差對種植面積作的散佈圖，目的是為了觀察殘差是否與自變數之間存在某種關係。  圖中殘差分布的情形無顯著存在某趨勢性，我們可視殘差與自變數種植面積無關。 |
|  | 此為為殘差對依變數產量作的各種圖形分析中的截取重要部分。  上面的圖，我們用來同樣觀察殘差是否均勻散佈在此圖當中，看起來是有的。  下面的圖中的常態機率圖，可以看得出來符合直線，所以為常態分佈。 |
|  | 此為種植面積和產量之散佈圖，並繪製估計出來的簡單線性迴歸線與其 95％期望值估計信賴值區間與 95％預測值信賴區間。 |

1.自變數為收穫面積：

|  |  |
| --- | --- |
|  | 1. 從綠色框框可以看得出來，=0.3687，這代表收穫面積解釋了36.87%的產量，算是不太好的。 2. 從紅色框框可以看出p-value < 0.001，因此有足夠證據顯示依變數(產量)得以被自變數(收穫面積)所解釋。 |
|  | 1.因為我建立的是簡單迴歸模型，所以可以由藍色的框框可以看出Pr>|t|小於 0.05，  達顯著水準，表示自變項(收穫面積)能讓模式更良好  2.由黃色框框可以求得簡單線性迴歸線：產量=168.05672+收穫面積0.168 |
|  | 此為殘差對種植面積作的散佈圖，目的是為了觀察殘差是否與自變數之間存在某種關係。  圖中殘差分布的情形無顯著存在某趨勢性，我們可視殘差與自變數收穫面積無關。 |
|  | 此為為殘差對依變數產量作的各種圖形分析中的截取重要部分。  上面的圖，我們用來同樣觀察殘差是否均勻散佈在此圖當中，看起來是有的。  下面的圖中的常態機率圖，可以看得出來符合直線，所以為常態分佈。 |
|  | 此為收穫面積和產量之散佈圖，並繪製估計出來的簡單線性迴歸線與其 95％期望值估計信賴值區間與 95％預測值信賴區間。 |

**第二節 建立多元線性迴歸模型**

自變數為種植多寡、收穫多寡、種植面積和收穫面積：

|  |  |
| --- | --- |
|  | 這次第四個字變數加入後的資料  1.從綠色框框可以看得出來，=0.6267，這代表收穫面積解釋了62.67%的產量，算是不錯的。  2.從紅色框框可以看出p-value < 0.001，因此有足夠證據顯示依變數(產量)得以被自變數(收穫面積)所解釋。 |
|  | 1.因為我建立的是簡單迴歸模型，所以可以由藍色的框框可以看出Pr>|t|小於 0.05，達顯著水準，表示自變項(收穫面積)能讓模式更良好  2.由黃色框框可以求得簡單線性迴歸線：產量=129.43+種植株數0.0002878+收穫株數0.0003091+種植面積5.70749+收穫面積5.78418 |
|  | 此為殘差對四個字變數作的散佈圖，目的是為了觀察殘差是否與自變數之間存在某種關係。  圖中殘差分布的情形無顯著存在某趨勢性，我們可視殘差與自變數無關。 |
|  | 此為為殘差對依變數產量作的各種圖形分析中的截取重要部分。  上面的圖，我們用來同樣觀察殘差是否均勻散佈在此圖當中，看起來是有的。  下面的圖中的常態機率圖，可以看得出來符合直線，所以為常態分佈。 |

**第四章 結論及建議**

**第一節 結論**

1. 2014年產量最高，2016年產量最低，而其中產量最低是因為氣候的影響。

2..生產地區大多集中在高屏及雲嘉南

3.鳳梨及香蕉歷年來都是台灣最多產量的水果。

4. 獲得兩個簡單線性迴歸線：產量=166.772+種植面積×0.164

產量=168.05672+收穫面積×0.168。

1. 獲得一個多元線性迴歸線：產量=129.43+種植株數0.0002878+收穫株數0.0003091+種植面積5.70749+收穫面積5.78418。

**第二節 建議**

1. 我們時常在台灣吃的到蘋果，但台灣種植蘋果的數量相較之下比較少，跟香蕉相比，它的價位也是比較高的，所以我們應該在台灣多種一點蘋果，更能降低蘋果的價錢。
2. 東部是全台灣地廣人稀的地方，但種植相對來說卻沒有很多，我覺得可以多多利用東部的土地去種植更多高價位的東西。