**1102科學計算軟體期末報告**

**主題 : 棒球運動之投出球之轉速與其他變因之關係**

組別：第5組

組員：F64096180 姚敦翔、F64094031 張昊翰、F64091017 游崴鈞、

F64109519 林均霈、F64109527 蕭合亭、F64081088 李瑄哲

1. **簡介**

透過大聯盟數據，探討棒球運動中轉速的高低與其他變因之關係。

1. **材料介紹**

ff\_avg\_spin轉速:

投手投擲時球的旋轉速度，指標單位以每分鐘轉圈數(RPM)來衡量。

xslg被長打率 :

打擊者每一次擊球可以貢獻幾個壘包。

exit\_velocity\_avg擊球初速 :

打者把球擊出時，球離開球棒時之速度。

launch\_angle\_avg擊球仰角

打者把球擊出時，球離開球棒時與地面形成之角度。

hard\_hit\_percent強擊球率

只要擊球初超過95英里，及數強擊球，依此比例去計算強擊球率。

z\_swing\_miss\_percent揮空率 :

打者揮棒卻無觸擊球及數揮空，依比例去算揮空率。

groundballs\_percent滾地球比 :

擊球者擊出滾地球與飛球之比例。

flyballs\_percent飛球比

擊球者擊出飛球與滾地球之比例。

1. **結果與討論**

敘述統計

一張含有 文字 的圖片

自動產生的描述

盒型圖 : 被長打率與轉速之盒型圖

1. 在Excel中將被長打率由0.3始、每0.05作為組別間距，利用subset函數做分流的擷取分別擷取出八組
2. 繪製出每一組的被長打率與轉速間關係之盒形圖。

程式碼 :

xs1 <- subset(dataset, xslg < 0.3)

xs2 <- subset(dataset,xslg >= 0.3 & xslg < 0.35)

xs3 <- subset(dataset,xslg >= 0.35 & xslg < 0.4)

xs4 <- subset(dataset,xslg >= 0.4 & xslg < 0.45)

xs5 <- subset(dataset,xslg >= 0.45 & xslg < 0.5)

xs6 <- subset(dataset,xslg >= 0.5 & xslg < 0.55)

xs7 <- subset(dataset,xslg >= 0.55 & xslg < 0.6)

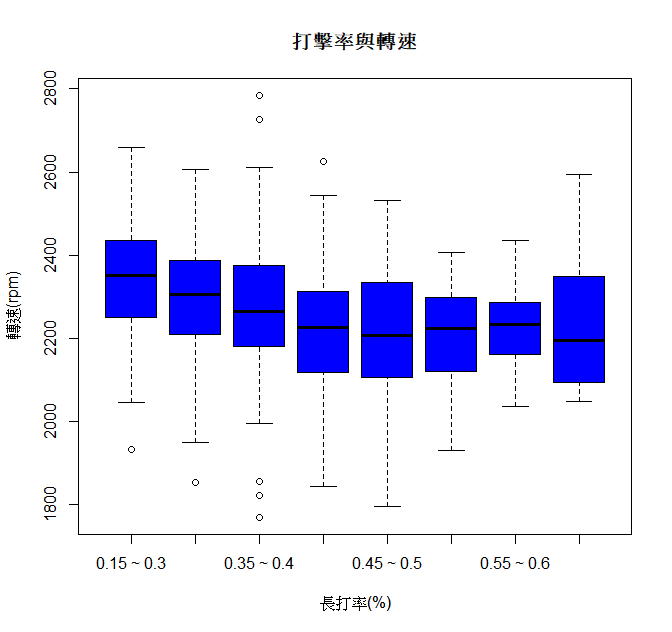
xs8 <- subset(dataset,xslg >= 0.6)

layout(matrix(c(1),1))

boxplot(xs1$ff\_avg\_spin, xs2$ff\_avg\_spin, xs3$ff\_avg\_spin, xs4$ff\_avg\_spin

, xs5$ff\_avg\_spin, xs6$ff\_avg\_spin, xs7$ff\_avg\_spin, xs8$ff\_avg\_spin, names =c(0.15~0.3,0.3~0.35,0.35~0.4,0.4~0.45,0.45~0.5,0.5~0.55,0.55~0.6,0.6~0.8)

, main = "打擊率與轉速", xlab = "長打率(%)", ylab = "轉速(rpm)", col ="blue")



直方圖

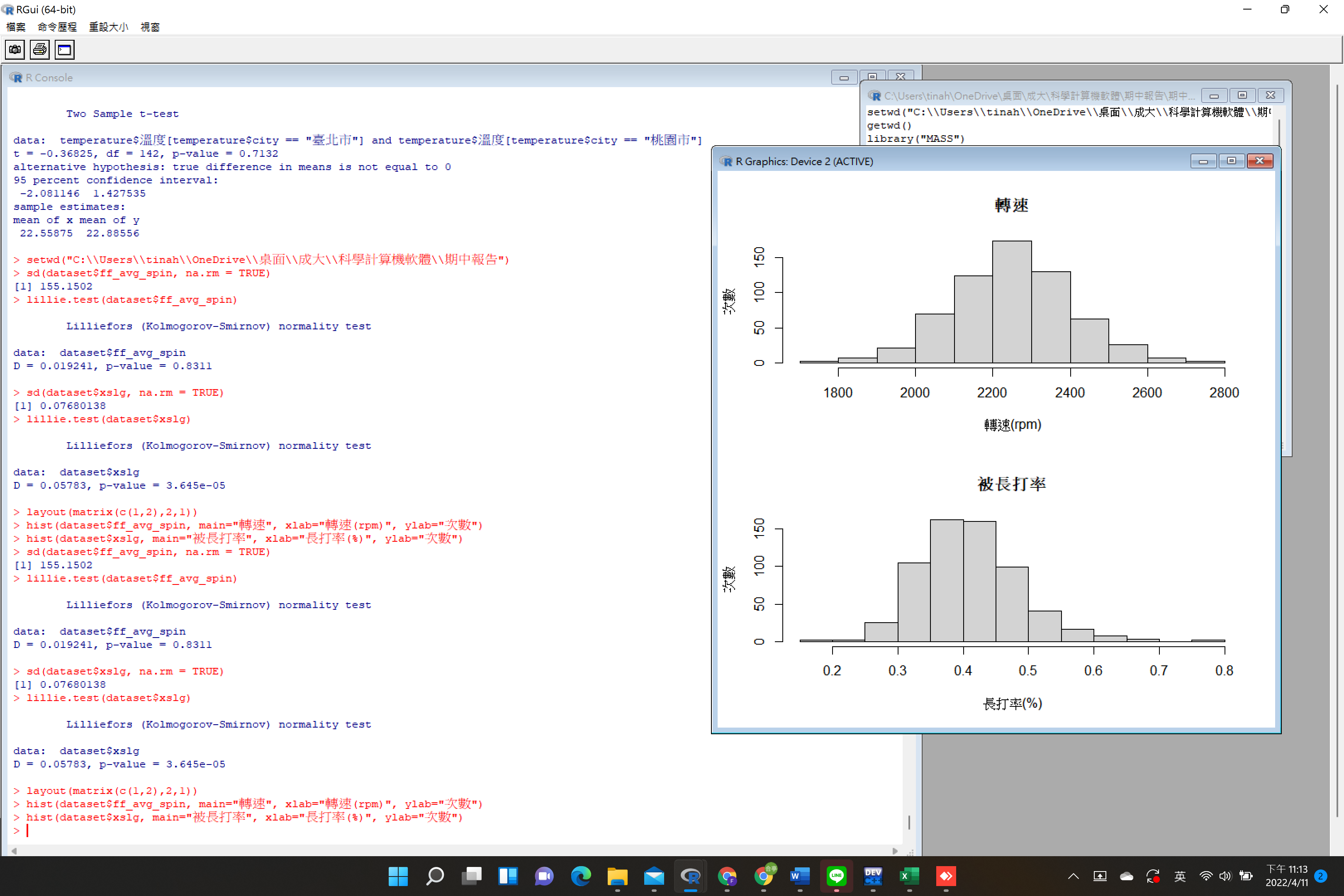
1. 算出轉速 (ff\_avg\_spin) 的標準差

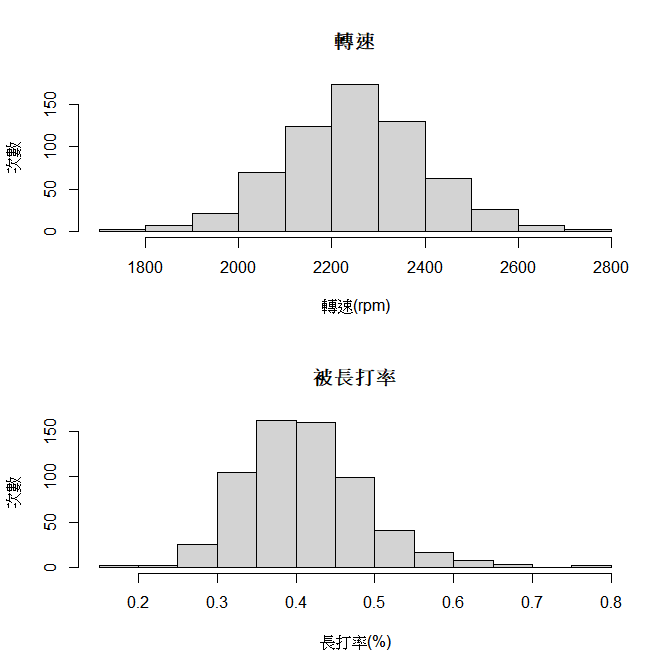
使用lillie檢定，p – value = 0.8311 > 0.05 ，屬常態分布。

1. 算出dataset之被長打率 (xslg) 的標準差

使用lillie檢定，p – value = 3.645e-05 < 0.05 ，屬非常態分布。

1. 使用hist做出兩個長方圖

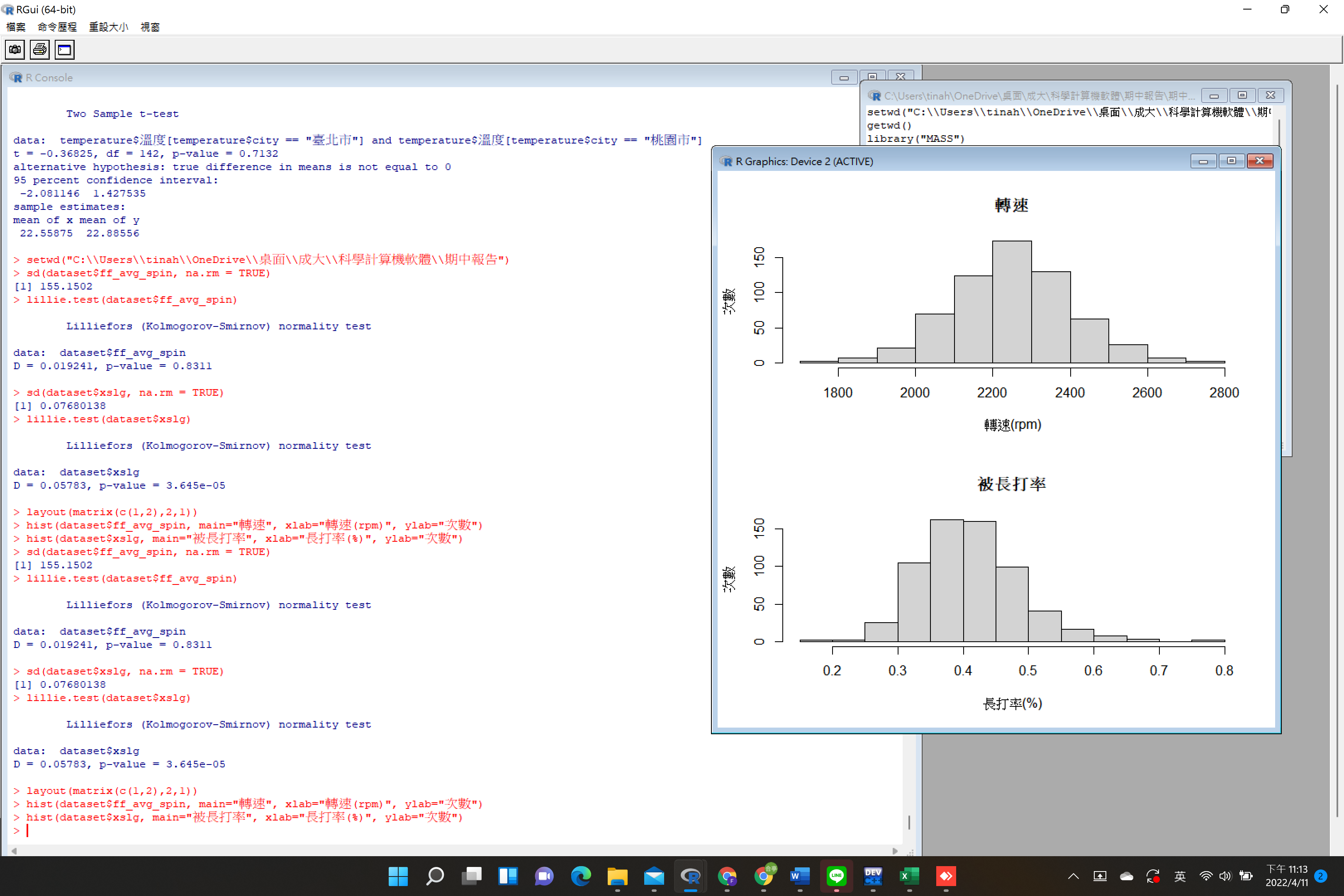


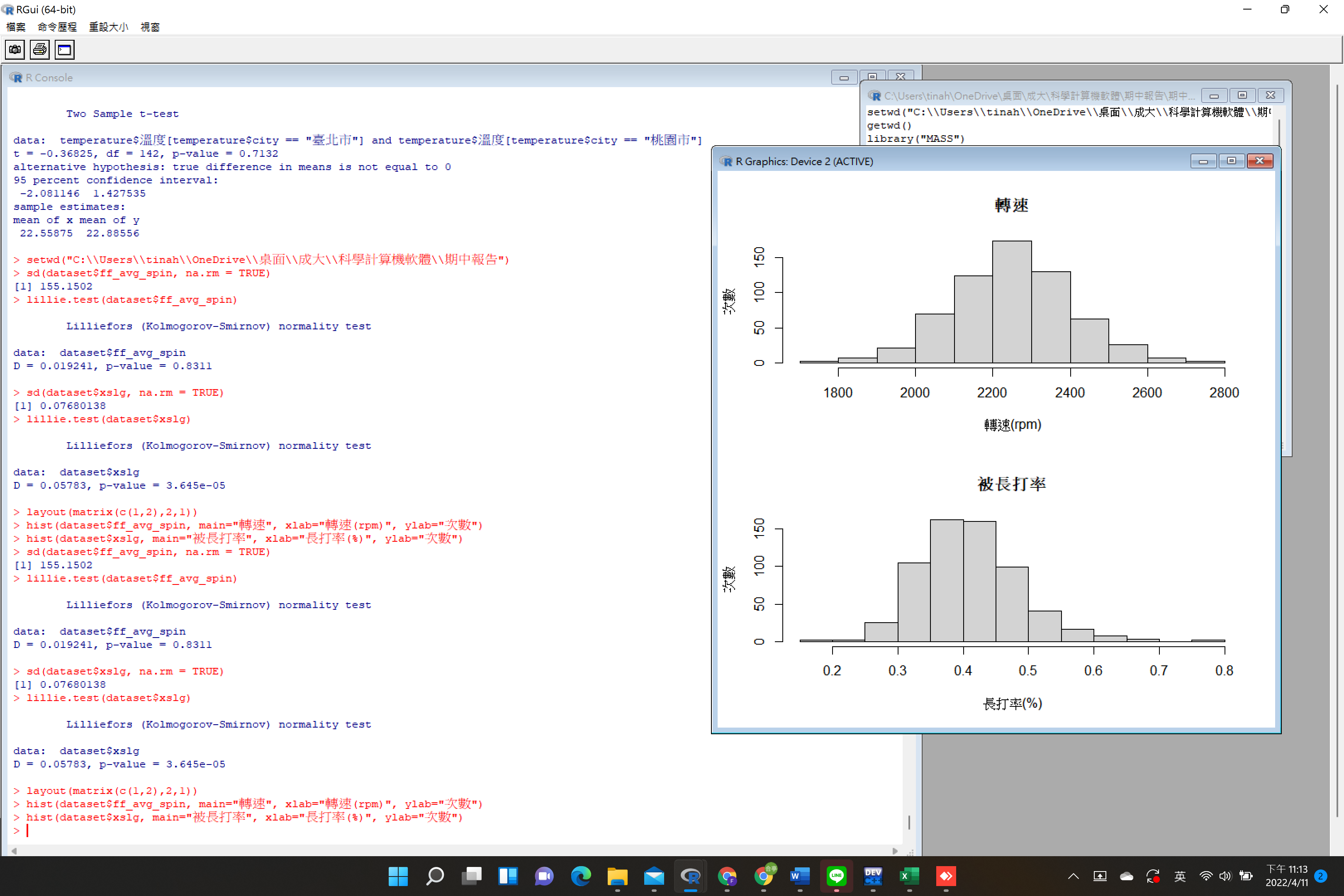


常態檢驗 : 使用lillie test進行常態分佈檢驗

轉速之p-value = 0.8311 > 0.05 ，屬常態分布。

被長打率之p-value = 3.645e-05 < 0.05 ，屬於非常態分佈。





t-test

轉速之資料利用subset函數做分流的擷取，分別擷取出高、低二組。

使用Levene Test 檢定，Pr 值 = 0.8717 > 0.05，變異數相同。

使用成對t檢定，因其變異數相同，故將var.equal設為TRUE

一張含有 文字 的圖片

自動產生的描述

ANOVA

使用一因子ANOVA檢定被長打率，變因為轉速區間2之資料，轉速區間以每400為一單位。

使用summary函數查看資料分析結果，可見p值 = 1.58e-05 < 0.05，可知均值間有差異，須進一步使用變異數檢定。

Levene Test :

P值0.1161 > 0.05 ，可知xslg與轉速區間2之變異數相等。

Turkey :

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2-1 | 3-1 | 3-2 |
| p | 0.8106489(均值間無差異) | 0.0081192(均值間無差異) | 0.0000139(均值間無差異) |

一張含有 文字 的圖片

自動產生的描述

相關性檢定

使用spearman做相關性檢定

程式碼 : rcorr(as.matrix(data[ , 5:12]), type=c("spearman"))

|  |  |  |  |
| --- | --- | --- | --- |
| 相關性 | 被長打率 | 擊球初速 | 擊球仰角 |
| 轉速 | -0.27(低度負相關) | -0.06(低度負相關) | 0.26(低度正相關) |
| 相關性 | 強擊球率 | 揮空率 | 滾地球比 |
| 轉速 | -0.14(低度負相關) | 0.38(低度正相關) | -0.22(低度負相關) |
| 相關性 | 飛球比 |  | |
| 轉速 | 0.18(低度正相關) |

|  |  |  |  |
| --- | --- | --- | --- |
| 顯著性 | 被長打率 | 擊球初速 | 擊球仰角 |
| 轉速 | 0.000  (達統計上顯著性) | 0.1911  (未達統計上顯著性) | 0.000  (達統計上顯著性) |
| 顯著性 | 強擊球率 | 揮空率 | 滾地球比 |
| 轉速 | 0.0017  (達統計上顯著性) | 0.000  (達統計上顯著性) | 0.000  (達統計上顯著性) |
| 顯著性 | 飛球比 |  | |
| 轉速 | 0.000  (達統計上顯著性) |

一張含有 桌 的圖片

自動產生的描述

變數篩選(逐步)

排除不顯著變數，使用逐步法做變數篩選。

程式碼 :

model <- lm(ff\_avg\_spin轉速 ~ xslg被長打率 +

launch\_angle\_avg擊球仰角 + hard\_hit\_percent強擊球率 +

z\_swing\_miss\_percent揮空率 + groundballs\_percent滾地球比 +

flyballs\_percent飛球比, data= data)

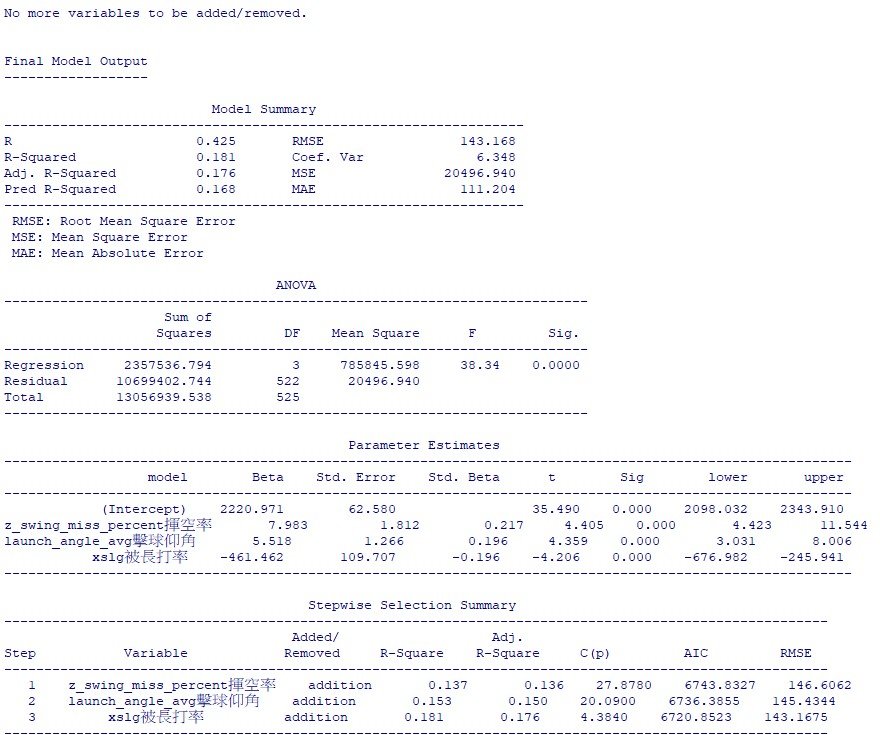
ols\_step\_both\_p(model, penter = 0.05, prem = 0.1 , details = TRUE)

分析 :

揮空率、擊球仰角、被長打率加入，淘汰掉p值大於0.1之變數及球初速、強擊球率、滾地球比、飛球比。

Sig :

由sig可看見，剩餘項目皆達統計上之顯著性。



共線性

將逐步變數篩選出之變數做共線性檢定。

程式碼 :

mode2 <- lm(ff\_avg\_spin轉速 ~ xslg被長打率 +

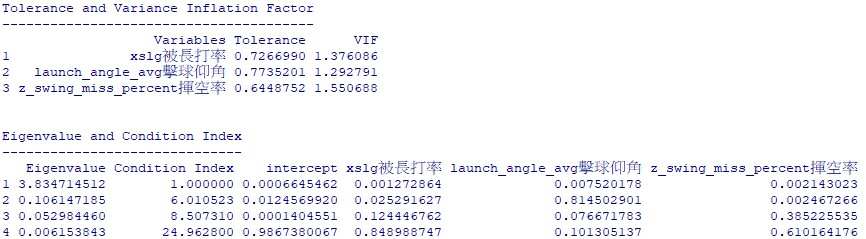
launch\_angle\_avg擊球仰角 + z\_swing\_miss\_percent揮空率, data= data)

ols\_coll\_diag(mode2)

分析 :

VIF 值皆未大於3，表示其共線性數於不嚴重的，無須剔除。

|  |  |  |  |
| --- | --- | --- | --- |
| 與轉速之共線性 | 被長打率 | 擊球仰角 | 揮空率 |
| VIF | 1.376 | 1.293 | 1.551 |



線性回歸

程式碼 :

summary(mode2)

分析 :

|  |  |  |  |
| --- | --- | --- | --- |
| 截距 | 被長打率 | 擊球仰角 | 揮空率 |
| 2220.971 | -461.462 | 5.518 | 7.983 |

模型公式 :

轉速 = 2220.971 – 461.462 \* 被長打率 + 5.518 \* 擊球仰角 +

7.983 \* 揮空率

R-square = 0.181

Adjust R-square = 0.176

一張含有 文字 的圖片

自動產生的描述

1. **結論**

資料數連續型資料，我們做了變數挑選、共線性檢定、相關分析、線性回歸。

˙透過相關性檢定剔除掉擊球初速。

˙由逐步變數篩選留下轉速與被長打率、擊球仰角、揮空率，可看到剩餘變數皆達統計上的顯著性。

˙利用共線性檢定，檢查出未有嚴重共線性，無須剔除任一碧數。

˙最後，做出線性迴規模型

1. **資料來源**

參考資料 : <https://baseballsavant.mlb.com/statcast_search>