

R 語言教學

鍾旻錡, 陳柏瑜

Statistics with Recitation

NTU Econ

2020.09.29

Outline

- 1 簡介
- 2 安裝 R
- 3 資料型態
- 4 R as a calculator
- 5 Function
- 6 Packages

R 是什麼

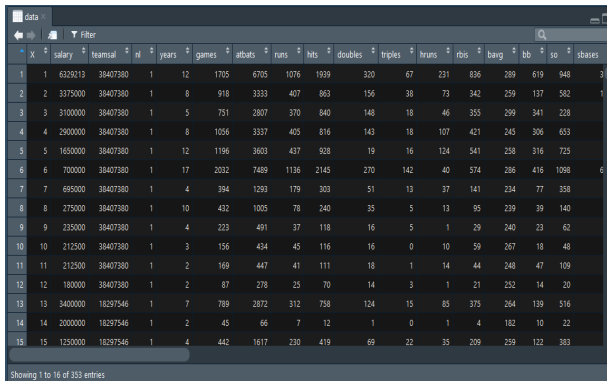
- 一種程式語言



- 為什麼選擇 R

R 做資料分析

● 讀取資料



The screenshot shows an R data viewer window with a table of 15 rows and 18 columns. The columns represent various statistics for baseball players, including salary, teamsal, nl, years, games, atbats, runs, hits, doubles, triples, hruns, rbs, bavg, bb, so, and sbases. The data is sorted by salary in descending order.

X	salary	teamsal	nl	years	games	atbats	runs	hits	doubles	triples	hruns	rbs	bavg	bb	so	sbases
1	6329213	38407380	1	12	1705	6705	1076	1939	320	67	231	836	289	619	948	3
2	3375000	38407380	1	8	918	3333	407	863	156	38	73	342	259	137	582	1
3	3100000	38407380	1	5	751	2807	370	840	148	18	46	355	299	341	228	
4	2900000	38407380	1	8	1056	3337	405	816	143	18	107	421	245	306	653	
5	1650000	38407380	1	12	1196	3603	437	928	19	16	124	541	258	316	725	
6	700000	38407380	1	17	2032	7489	1136	2145	270	142	40	574	286	416	1098	6
7	695000	38407380	1	4	394	1293	179	303	51	13	37	141	234	77	358	
8	275000	38407380	1	10	432	1005	78	240	35	5	13	95	239	39	140	
9	235000	38407380	1	4	223	491	37	118	16	5	1	29	240	23	62	
10	212500	38407380	1	3	156	434	45	116	16	0	10	59	267	18	48	
11	212500	38407380	1	2	169	447	41	111	18	1	14	44	248	47	109	
12	180000	38407380	1	2	87	278	25	70	14	3	1	21	252	14	20	
13	3400000	18297546	1	7	789	2872	312	758	124	15	65	375	264	139	516	
14	2000000	18297546	1	2	45	66	7	12	1	0	1	4	182	10	22	
15	1250000	18297546	1	4	442	1617	230	419	69	22	35	209	259	122	383	

Showing 1 to 16 of 353 entries

R 做資料分析

● 統計分析

```
Call:
lm(formula = lsalary ~ hruns + bavg + bb, data = data)

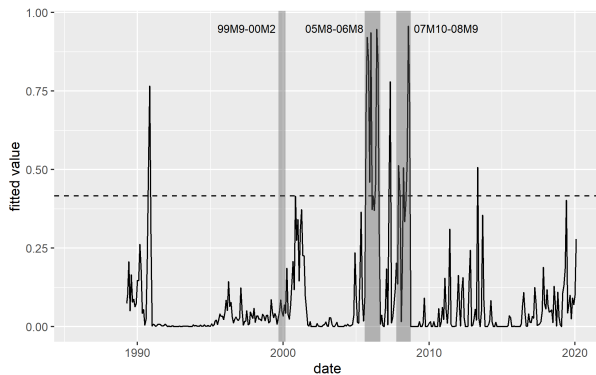
Residuals:
    Min       1Q   Median       3Q      Max
-4.0912 -0.7000 -0.0878  0.7512  1.9351

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.182e+01  3.350e-01  35.290 < 2e-16 ***
hruns        4.168e-03  1.059e-03   3.936 9.99e-05 ***
bavg         4.005e-03  1.318e-03   3.040 0.00255 **
bb           1.906e-03  3.498e-04   5.448 9.65e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9128 on 349 degrees of freedom
Multiple R-squared:  0.4091,    Adjusted R-squared:  0.404
F-statistic: 80.55 on 3 and 349 DF,  p-value: < 2.2e-16
```

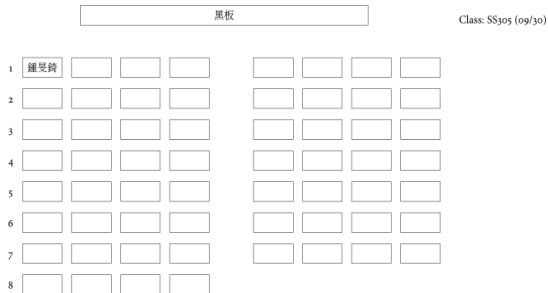
R 做資料分析

● 視覺化



其他範例

- 自動排座位



- 參考網址: <https://github.com/MinChi-o314/seat>

安裝

- R

<https://cran.csie.ntu.edu.tw/>

- R studio

<https://rstudio.com/products/rstudio/download/#download>

資料種類

- numeric, integer, character
- logic
- vector, matrix, data.frame
- factor, list, array

資料種類

- 查詢:

ex: `is.numeric`, `class`

- 設定:

ex: `as.numeric`

R as a calculator

常見的運算元(operator) 有:

- 四則運算: $+$, $-$, $*$, $/$
- 次方: $^$ 或 $**$
- 餘數: $\%\%$
- $>$, $<$, $==$, $!=$
- $\%*\%$: 矩陣乘法

R as a calculator

常見的保留字有:

- 圓周率: pi
- 無窮大: Inf
- Not Available: NA
- Not a Number: NaN

Practice

- 2的30次方除以63的餘數是多少?

R as a calculator

- Practice



WIFI

南航饮食服务中心

账号：馨园民族餐厅

密码：结果取前8位 ↓

$$\int_{-2}^2 \left(x^3 \cos \frac{x}{2} + \frac{1}{2} \right) \sqrt{4-x^2} dx$$

Function 函式

函式可以表達純量, 當作一個純數值

- `sqrt(100)`
- `log(1)`
- `factorial(3)`
- `exp(1)`
- `abs(-5)`

Practice

- `X Binomial(n=4,p=0.6)`
- 試用R 計算 $\Pr(X=2)$

內建函式: 隨機抽樣

函式通常由名稱及引數組成, 如: 隨機抽樣的`sample()` 函式

`sample(x, size, replace = FALSE, prob = NULL)`

刮號內則為函式的輸入(input), 以逗號分隔

- `x`: 抽樣的pool
- `size`: 抽幾個樣本出來
- `replace`: 取後是否放回
- `prob`: `x` vector 當中每個元素被抽取到的機率值

內建函式: 柱狀圖

`barplot(height, ...)`

- `height`: 柱高
- `names.arg`: 每根柱子的標籤

Demo

- 假設今有一枚公正硬幣, 擲10次後用長條圖畫出出現正反面的次數

內建函式: 散布圖

`plot(x, y, ...)`

- `x`: x 軸座標值
- `y`: y 軸座標值
- `type`: 圖的類型, 有 'l', 'p', 'h', 's', 'b'

自定義函式

自定義函式, 如 `triple()`:

- `triple <- function(n){3*n}`

Demo

- 找出 `Binomial(n, p)` 的 pmf 在 $X=x$ 時的高度
- 擲三次公正骰子, 點數 3 出現恰好 2 次的機率為?

Practice

- 承上, 驗證點數 3 出現恰好 0 次的機率加上恰好 1 到 3 次的機率, 總和為一

自定義函式

Monty Hall Problem Demo

Packages 套件

指令

- `install.packages("套件名稱")`
- `library(套件名稱)`

感謝大家聆聽