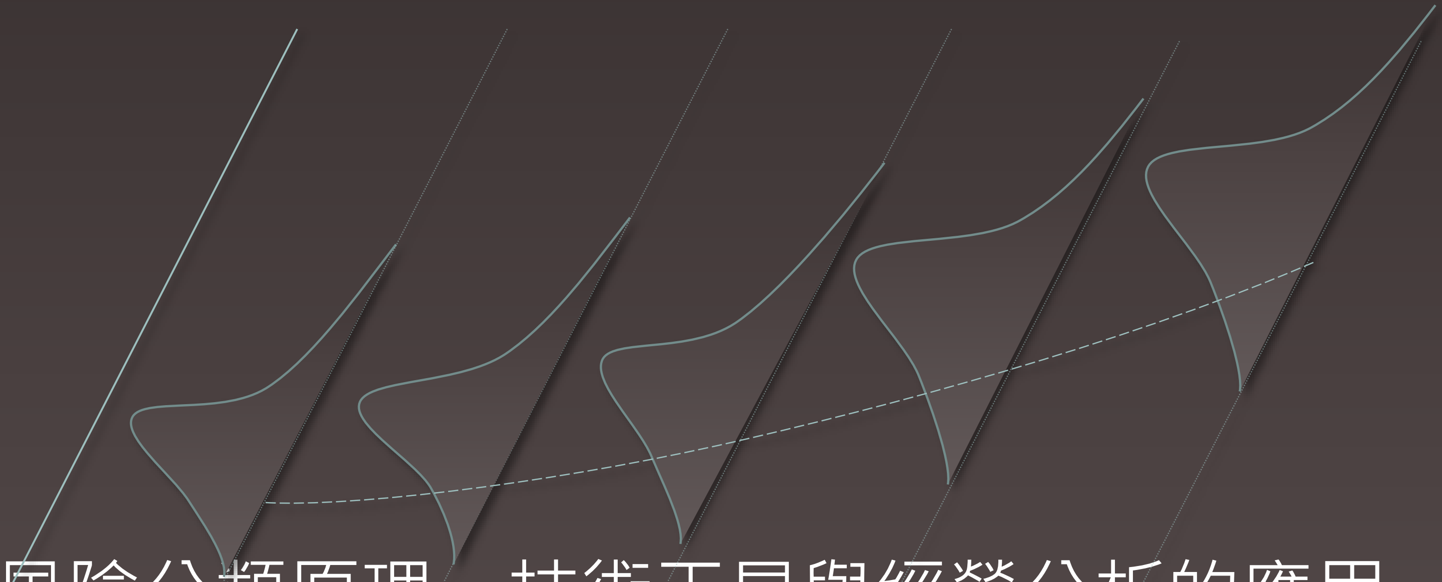


商品設計-風險分類原理、技術工具與經營分析的應用

廣義線性模型理論與R之應用

鄭弘偉、趙詩華

古典線性模型
- 定義

模型：

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon, \quad \varepsilon \sim \text{i.i.d.} \sim \text{Normal}(\,0\,,\ \sigma^2\,)$

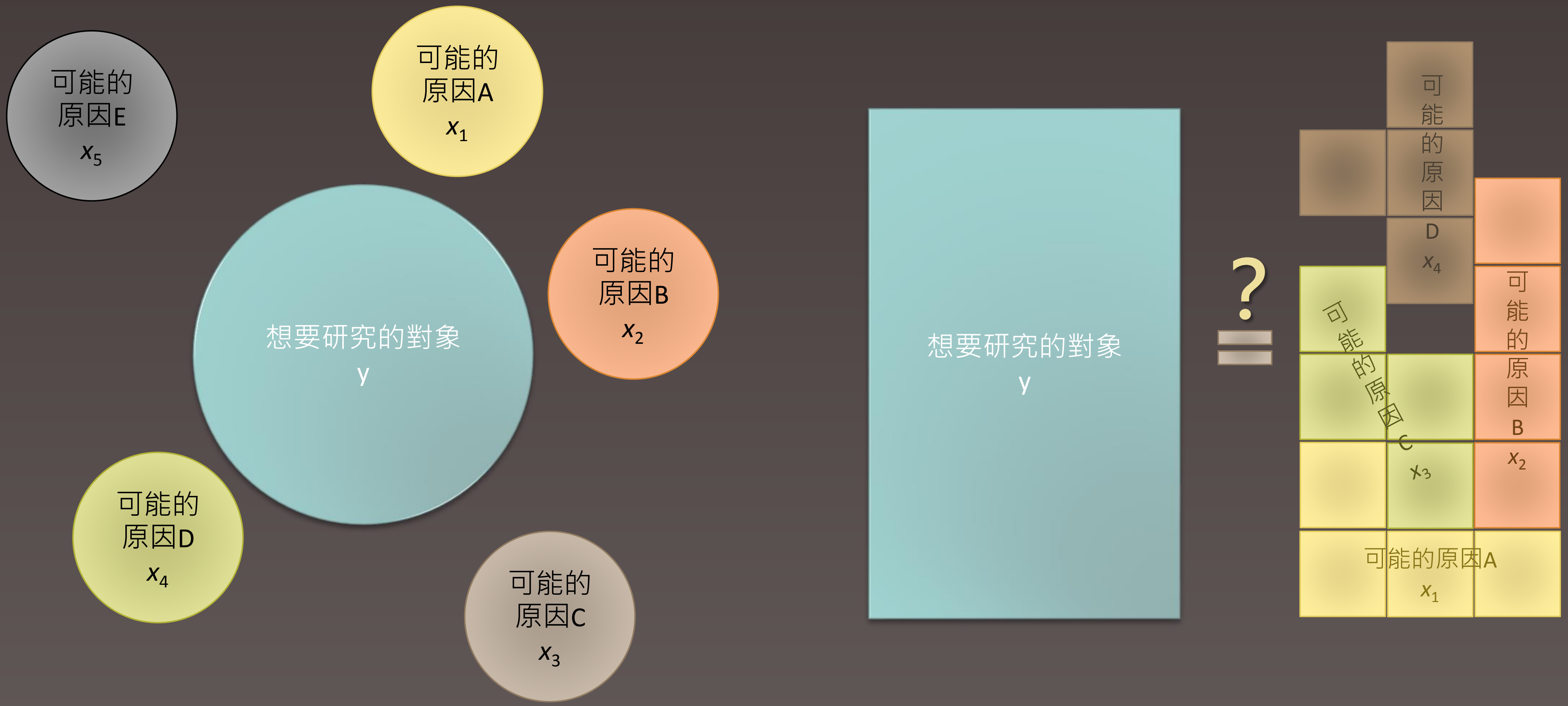
y：反應變數(response variable)

x_i：解釋變數(explanatory variable)

i.i.d.：各變數間相互獨立且來自同一個分配
(Independent and identically distributed random variables)

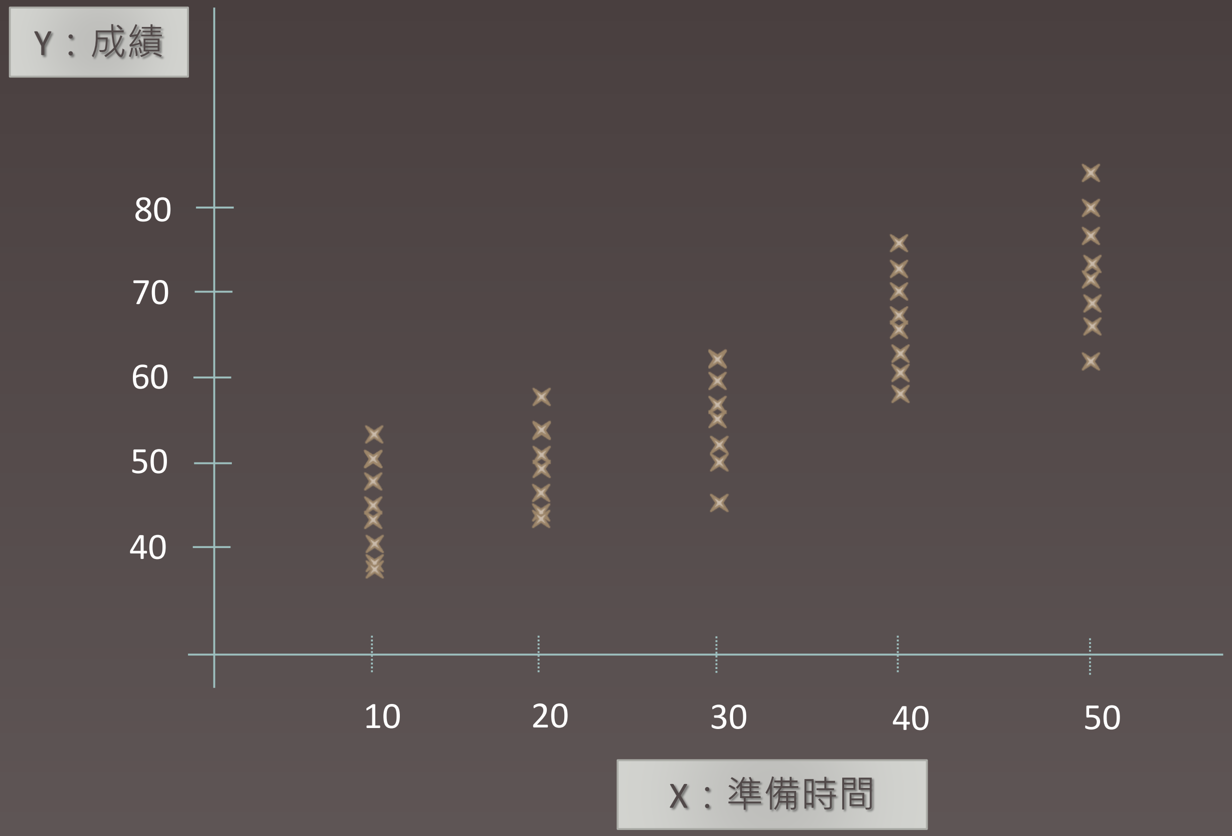
$\mathrm{E}(y) = \mu = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \cdots + \widehat{\beta_p} x_p$

古典線性模型
- 目標



可能的原因E x_5
可能的原因A x_1
想要研究的對象 y
可能的原因B x_2
可能的原因D x_4
可能的原因C x_3
想要研究的對象 y
?=
可能的原因D x_4
可能的原因B x_2
可能的原因C x_3
可能的原因A x_1

古典線性模型
- 想法



目標：
想要去了解每位學員準備考試的時間和成績的關係。

想要的結果？

1.如果學員花費多少時間準備，其預期成績為多少？

2.每多投入1小時，成績可以多幾分？

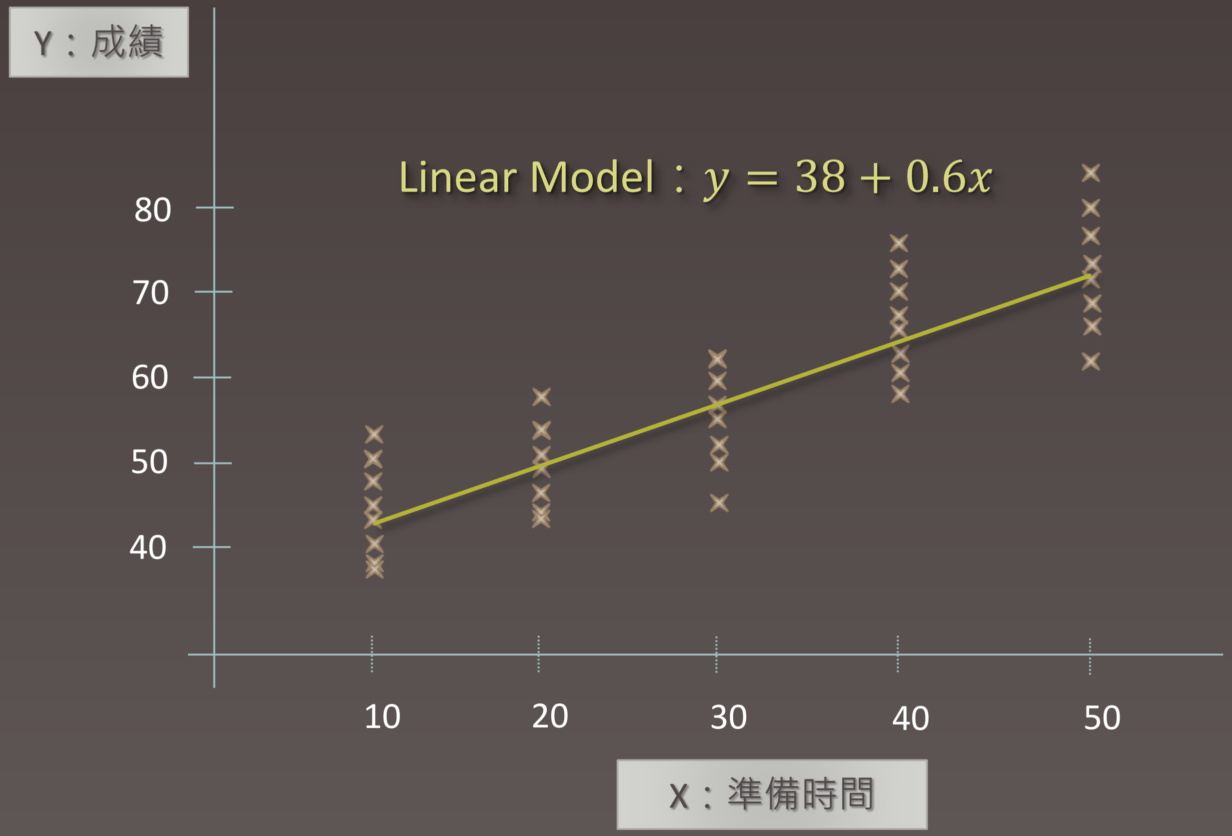
需要的前提假設？

1.每位學員程度皆相近。

2.在同樣準備時間下，每位學員的預期成績與實際成績差異僅來自於隨機波動。

3.全體學員的預期成績與實際成績的差異總合為0。

古典線性模型
- 模型配適



目標：
想要去了解每位學員準備考試的時間和成績的關係。

給定的假設

每個學員實際成績與預期成績之差值皆服從常態分配(0,σ²)，且各學員成績不相互影響。

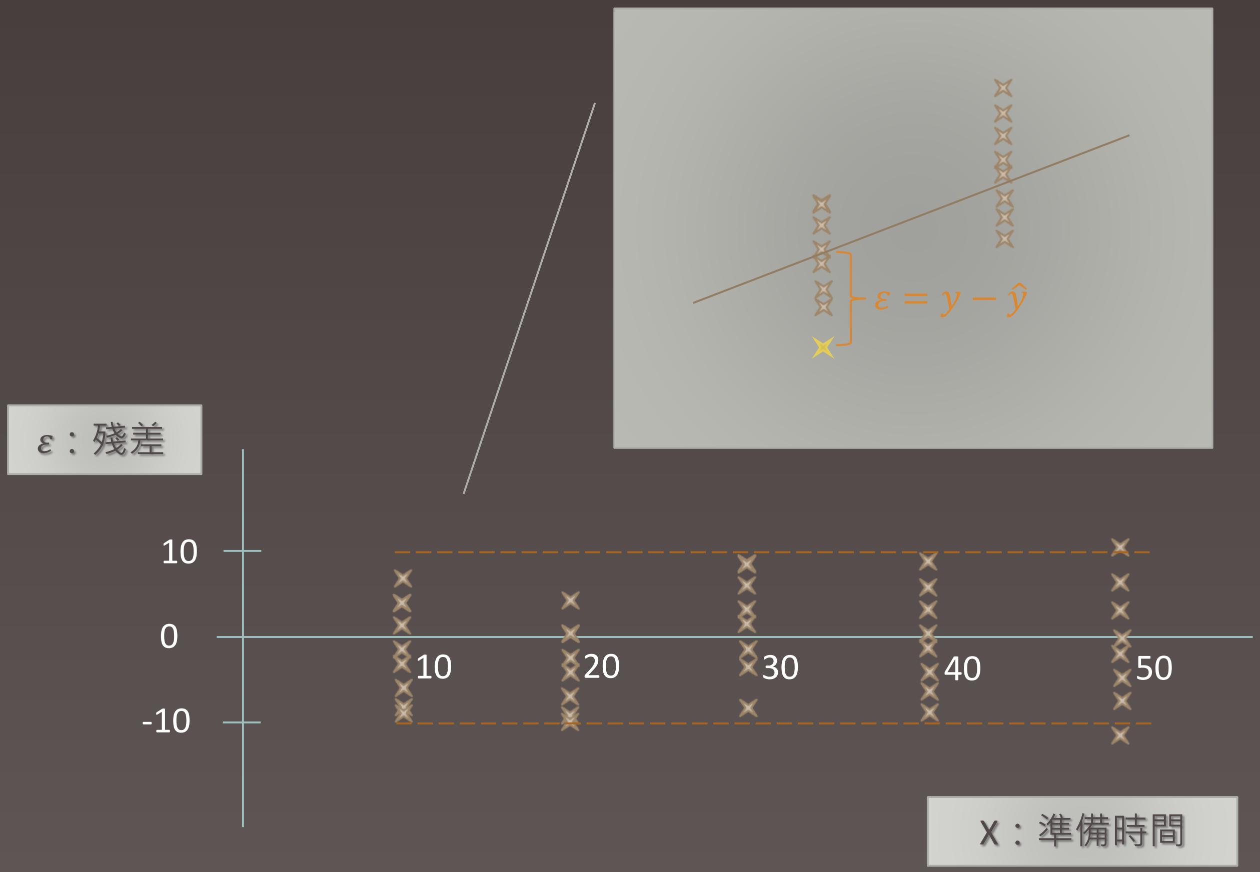
$\rightarrow \varepsilon \sim \text{i.i.d.} \sim \text{Normal}(0, \sigma^2)$

得到的結果

1.如果學員花費10小時準備，其預期成績為44分。

2.每多投入1小時，預期成績可增加6分。

古典線性模型
- 假設確認



目標：
確認配適結果是否顯著違反假設。

測試1：

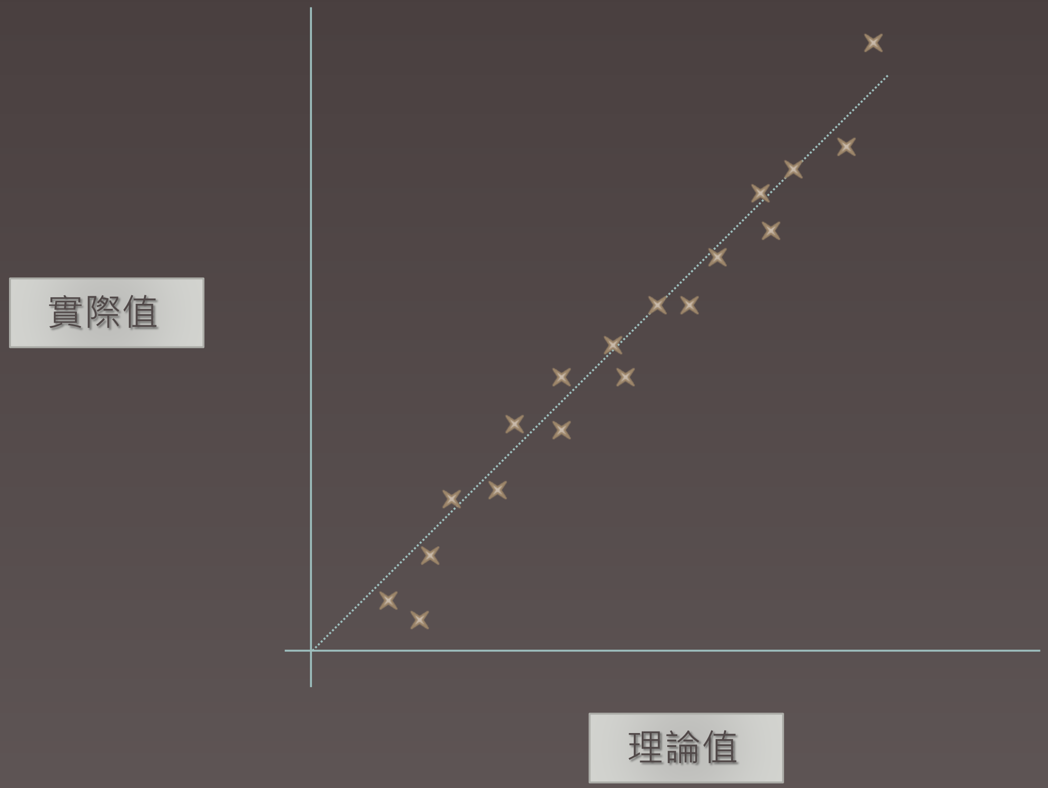
變異程度是否有顯著差異。
(Check for the constant variance)

工具：

1.殘差圖。

2.統計量檢定。

古典線性模型
- 假設確認



目標：
確認配適結果是否顯著違反假設。

測試2：

是否服從常態分配。
(Check for the normality)

工具：

1.常態機率圖(Q-Q Plot)。

2.直方圖。

古典線性模型
- 模型選擇

- 合適的線性模型不一定有一個。
- 解釋變數的增加會改善線性模型的配適能力，但會降低對於參數估計的精確度。
- 常用來判斷模型配適好壞的準則(Criteria)：
  - Akaike’s Information Criterion(AIC)

    $$\text{AIC} = -2l + 2p$$

  - Bayesian Information Criterion(BIC)

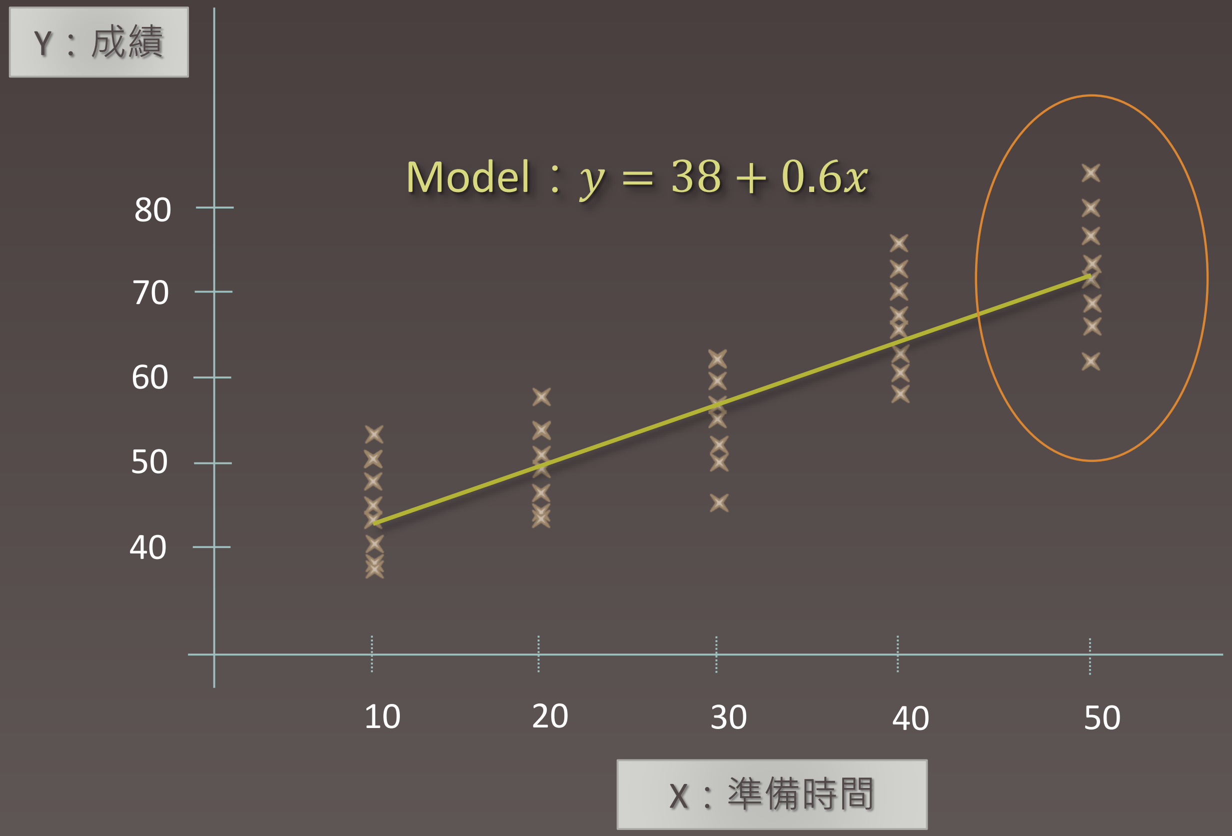
    $$\text{BIC} = -2l + p \cdot \ln n$$

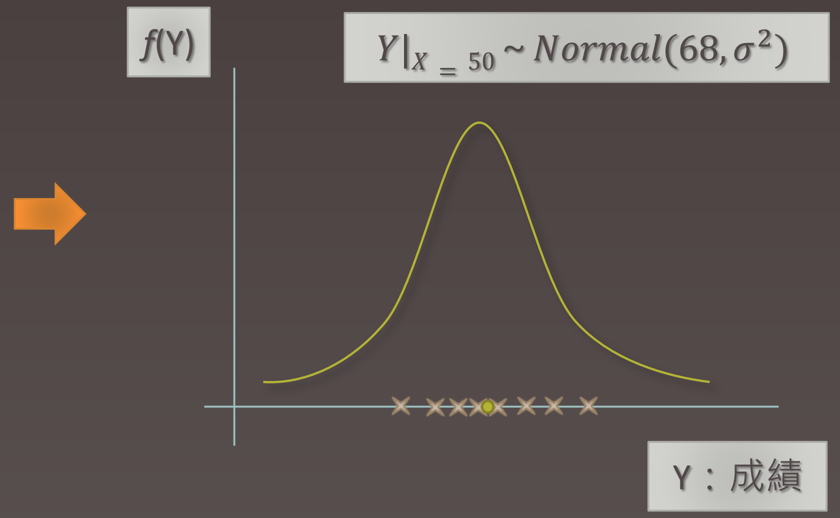
l：對數概似統計量(log-likelihood)；p：參數(β)個數；n：樣本數

古典線性模型
- 步驟

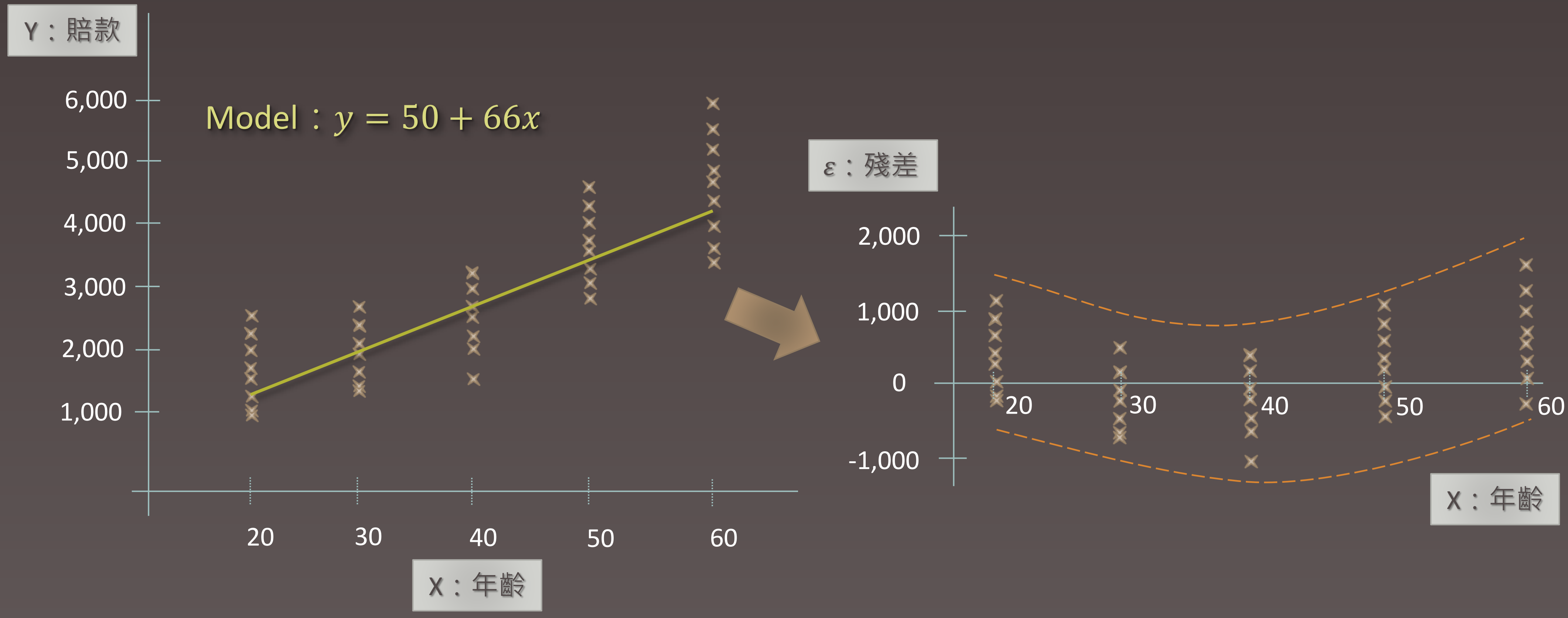
1. 尋找對反應變數有解釋能力之因子。
2. 配適模型(參數估計)。
3. 對給定之假設進行檢測。
4. 挑選適當之配適模型。

古典線性模型
- 統計上的解釋

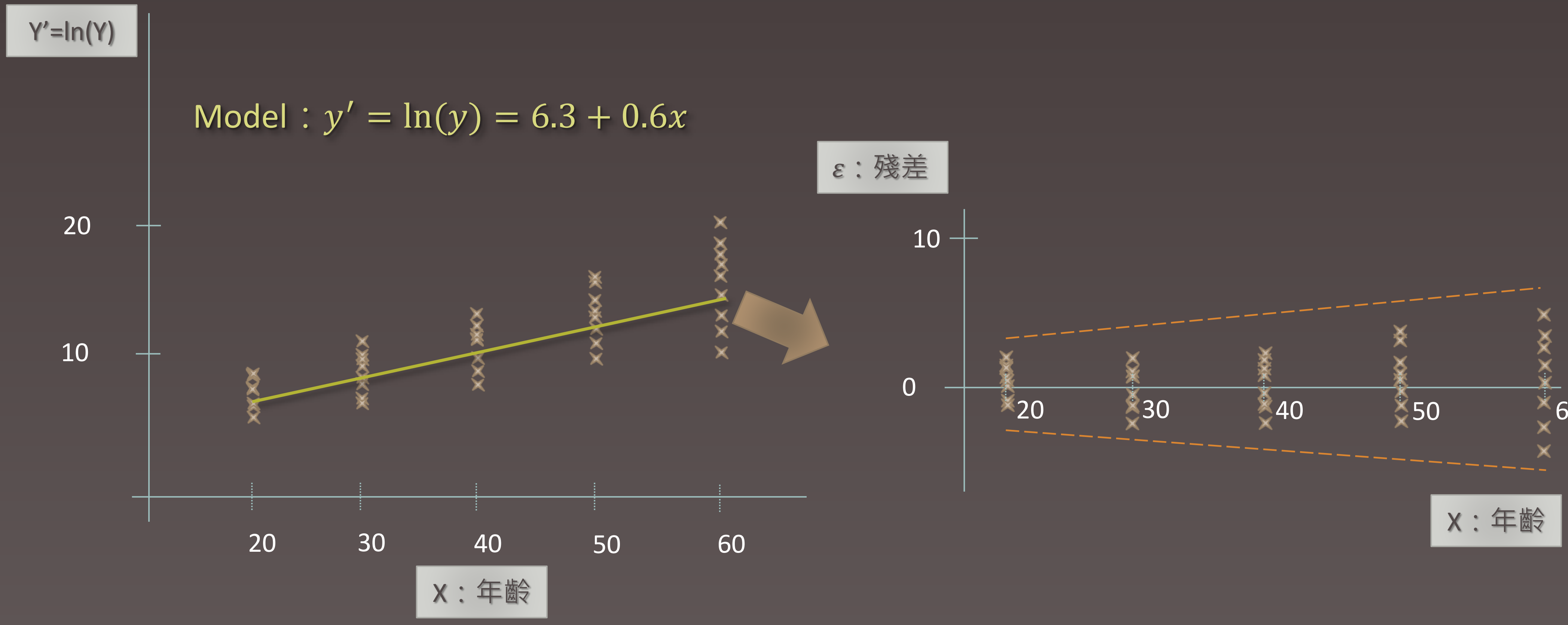




古典線性模型
- 處理保險資料



古典線性模型
- 處理保險資料 - 進行轉換



廣義線性模型
- 定義

模型：

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \quad y \sim \text{i.i.d.} \sim \text{指數簇}$$

y：反應變數(response variable)

x_i：解釋變數(explanatory variable)

i.i.d.：各變數間相互獨立且來自同一個分配
(Independent and identically distributed random variables)

$g(\mu)$：連結函數(link function)

指數簇： The Exponential Family ，包含Normal、Gamma、Inverse Gaussian、Poisson、Binomial及Negative Binomial等離散及連續型分配。

指數簇
- 定義

如果分配函數可改寫成下列模式：

$$f(y;\theta,\varphi) = exp\left\{c(y,\varphi) + \frac{y\cdot\theta - a(\theta)}{\varphi}\right\},$$

其中 θ 與 ϕ 為參數,
參數 θ 稱為標準參數(canonical parameter)且參數 ϕ稱為散度參數(dispersion parameter)。

$$E(y) = \dot{a}(\theta),\ Var(y) = \varphi \cdot \ddot{a}(\theta)\ ,$$

其中 $\dot{a}(\theta)$ 與 $\ddot{a}(\theta)$ 分別為 $a(\theta)$之一階與二階偏微分。

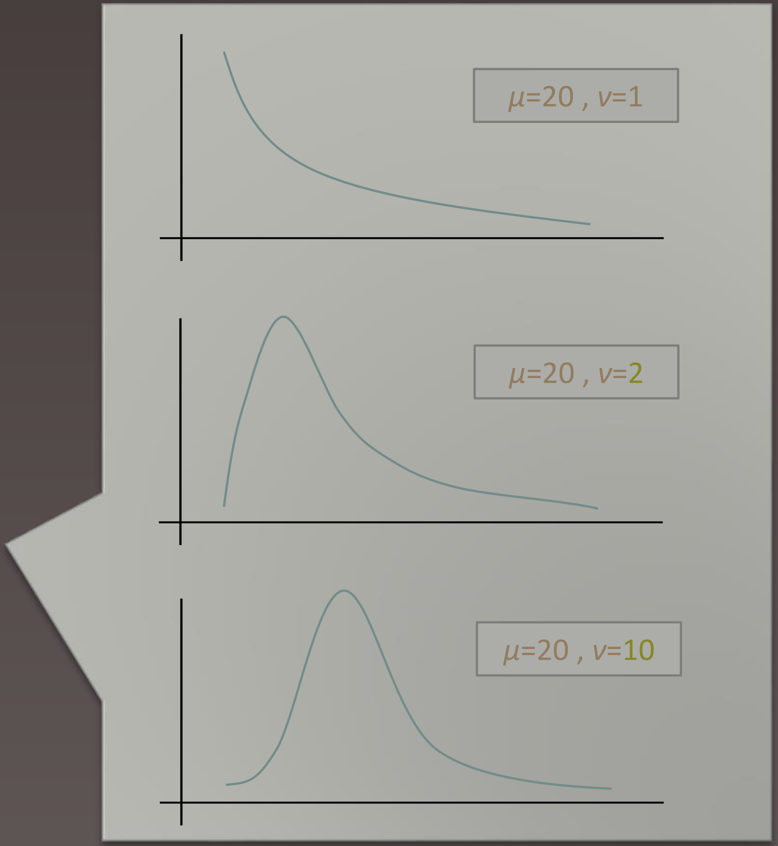
指數簇
- 以Gamma分配為例

$y \sim Gamma(\,\alpha\,,\ \beta\,)$

$$f(\,y\,;\alpha\,,\beta\,) = \frac{y^{\alpha-1} \cdot e^{-\beta y}}{\Gamma(\alpha)\cdot \beta^{-\alpha}}\ , y > 0 \text{ with } \mathrm{E}(y) = \frac{\alpha}{\beta}\ ,\ \mathrm{Var}(y) = \frac{\alpha}{\beta^2}$$

$$y \sim Gamma\left(\,\mu = \frac{\alpha}{\beta},\ \nu = \alpha\,\right)$$

$$f(\,y\,;\mu\,,\nu\,) = \frac{y^{\nu-1} \cdot e^{-\frac{\nu}{\mu} y}}{\Gamma(\nu)\cdot \left(\frac{\nu}{\mu}\right)^{-\nu}}\ , y > 0 \text{ with } \mathrm{E}(y) = \mu\ ,\ \mathrm{Var}(y) = \frac{\mu^2}{\nu}$$



指數簇
- 以Gamma分配為例

$y \sim Gamma(\mu, \nu)$

$$f(y;\mu,\nu) = \frac{y^{\nu-1} \cdot e^{-\frac{\nu}{\mu}y}}{\Gamma(\nu) \cdot \left(\frac{\nu}{\mu}\right)^{-\nu}},\, y > 0 \text{ with } \mathrm{E}(y) = \mu,\ \mathrm{Var}(y) = \frac{\mu^2}{\nu}$$

$$\ln\big(f(y;\mu,\nu)\big) = (\nu-1)\ln(y) - \frac{\nu}{\mu}y - \ln\big(\Gamma(\nu)\big) + \nu\ln(\nu) - \nu\ln(\mu)$$

$$= \big\{(\nu-1)\ln(y) - \ln\big(\Gamma(\nu)\big) + \nu\ln(\nu)\big\} + \frac{y\left(-\frac{1}{\mu}\right) - \ln(\mu)}{\frac{1}{\nu}},$$

$$\text{with } \theta = -\frac{1}{\mu}, a(\theta) = \ln(\mu) = -\ln(-\theta) \text{ and } \phi = \frac{1}{\nu}.$$

所以Gamma分配為指數簇，且

$$E(y) = \dot{a}(\theta) = -\frac{1}{\theta},\ Var(y) = \varphi \cdot \ddot{a}(\theta) = \frac{1}{\nu}\frac{1}{\theta^2} = \frac{\mu^2}{\nu}$$

指數簇

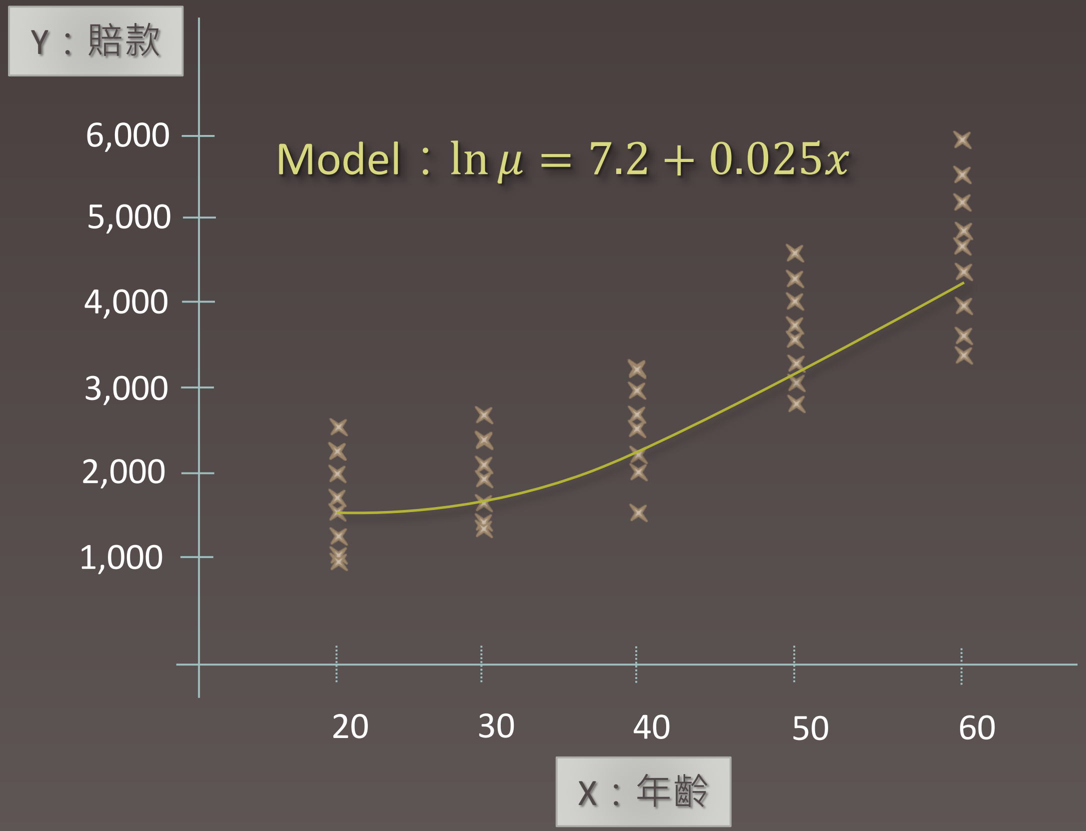
- 指數簇分配及其參數

分配	θ	$a(\theta)$	φ	$\dot{a}(\theta) = E(y)$	$\ddot{a}(\theta) = Var(y)/\varphi$
Binomial(n , π)	$\ln\frac{\pi}{1-\pi}$	$n\ln\left(1+e^{\theta}\right)$	1	$n\pi$	$n\pi(1-\pi)$
Poisson(μ)	$\ln\mu$	e^{θ}	1	μ	μ
Normal(μ , σ^2)	μ	$\frac{1}{2}\theta^2$	σ^2	μ	1
Gamma(μ , ν)	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$1/\nu$	μ	μ^2
Inverse Gaussian(μ , σ^2)	$-\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	σ^2	μ	μ^3
Negative Binomial(μ , κ)	$\ln\frac{\kappa\mu}{1+\kappa\mu}$	$-\frac{1}{\kappa}\pi\ln\left(1-\kappa e^{\theta}\right)$	1	μ	$\mu(1+\kappa\mu)$

連結函數
- 連結函數形式及各分配主要連結函數

連結函數	函數形式	主要適用之分配
Identity	μ	Normal
Log	$\ln \mu$	Poisson
Power	μ^p	Gamma(p=-1)、Inverse Gaussian(p=-2)
Square root	$\sqrt{\mu}$	
Logit	$\ln \frac{\mu}{1-\mu}$	Binomial

廣義線性模型
- 模型配適



目標：
想要去了解賠款和年齡間的關係。

選定的條件

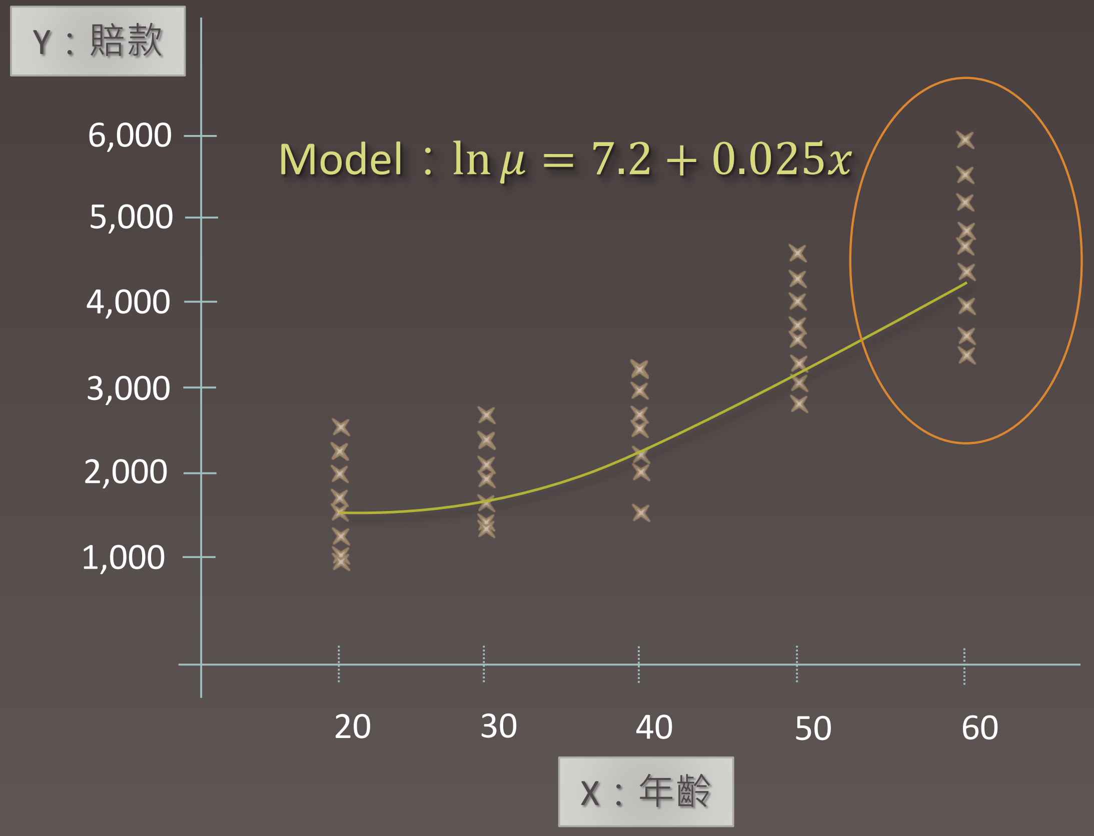
$$\begin{cases} y \sim \text{Gamma}(\mu, \nu) \\ \text{Link Function : Log−Link} \\ g(\mu) = \beta_0 + \beta_1 x \end{cases}$$

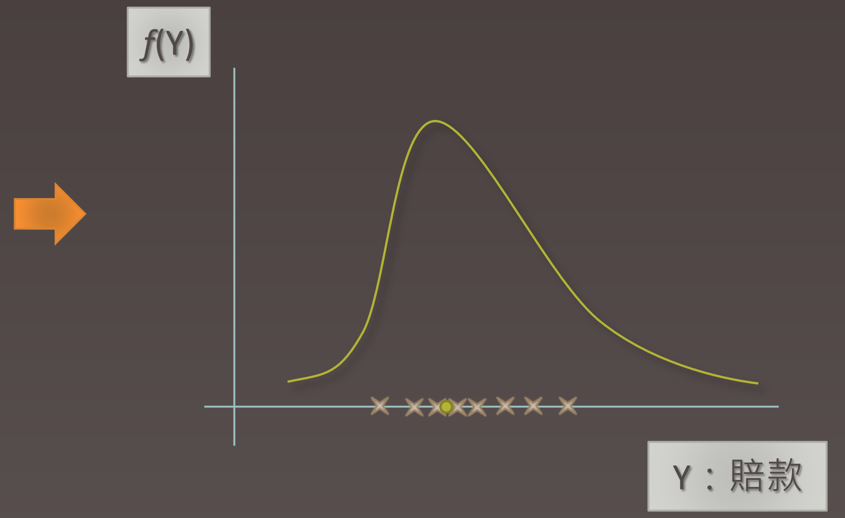
得到的結果

1.如果被保險人年齡為30歲者，其賠款預期將服從$\text{Gamma}(\mu = exp\{7.2 + 0.025 \times 30\}, \nu)$ 。

2.且預期被保險人年齡每增加10歲，其平均預期賠款將增加$exp\{0.025 \times 10\}$倍。

廣義線性模型
- 配適結果

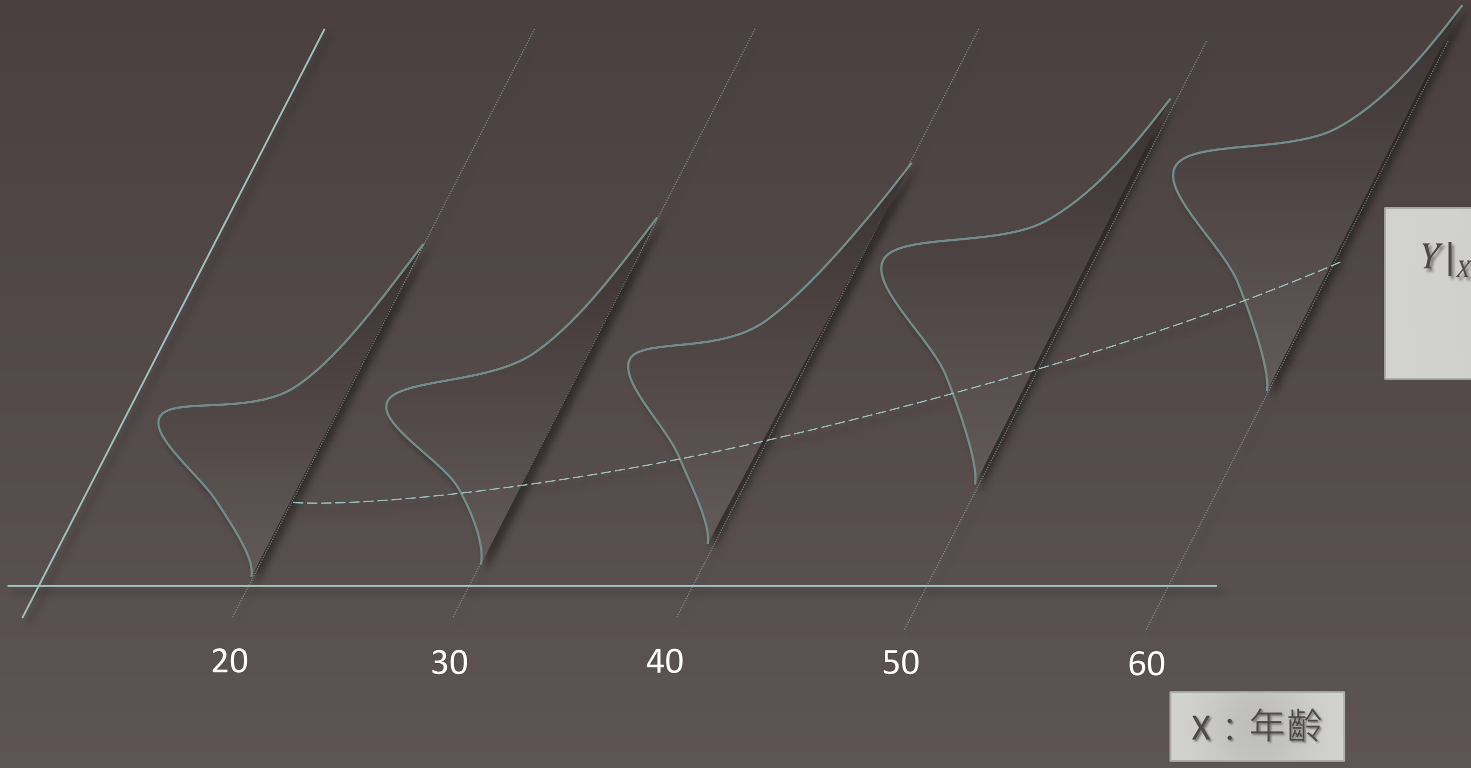




$$Y|_{X=60} \sim Gamma(\mu = e^{7.2+0.025\times 60}, \nu)$$

$$\mathrm{E}(Y|_{X=60}) = \mu \,, \mathrm{Var}(Y|_{X=60}) = \nu \cdot \mu^2$$

廣義線性模型
- 統計上的解釋



$$Y|_{X=60} \sim Gamma(\hat{\mu} = e^{7.2+0.025\times 60}, \hat{\nu} = 0.008)$$

$$\mathrm{E}(Y|_{X=60}) = \hat{\mu}\,, \mathrm{Var}(Y|_{X=60}) = \hat{\nu}\cdot\hat{\mu}^2$$

廣義線性模型
- 模型選擇

- 合適的線性模型不一定有一個。
- 解釋變數的增加會改善線性模型的配適能力，但會降低對於參數估計的精確度。
- 使用AIC或BIC來判斷模型配適好壞的準則(Criteria)。

廣義線性模型
- 步驟

1. 尋找對反應變數有解釋能力之因子。
2. 配適模型(參數估計)。
3. 對給定之假設進行檢測。
4. 挑選適當之配適模型。

結論

古典線性模型：

1. 反應變數Y必須服從常態分配，且變異數均相同。
2. 反應變數Y與解釋變數X間之關係方程式僅允許“直線性”相關。

廣義線性模型：

1. 反應變數Y服從之分配為指數簇之一員，且變異數可不同。
(若選擇之分配為常態分配，則變異數仍均相同，同古典線性模型。)
2. 反應變數Y與解釋變數X間之關係方程式為“線性”相關。
(古典線性模型僅可選擇連結函數中的Identity - Link形式。)

Generalized Linear Models in R

工欲善其事，必先利其器。

《論語‧衛靈公》

R 是 ...

- ✓ 自由軟體(Free-Software, GNU協定)
- ✓ 開放原始碼的統計、繪圖軟體
- ✓ 建構在貝爾實驗室S語言基礎的軟體
- ✓ 『免付費』的公開軟體

窈窕淑女，君子好逑。

《詩經◦關雎》

R 安裝步驟—

https://www.r-project.org/ or

R 安裝步驟二

[Home]

Download

CRAN

R Project

About R
Contributors
What's New?
Mailing Lists
Bug Tracking
Conferences
Search

R Foundation

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

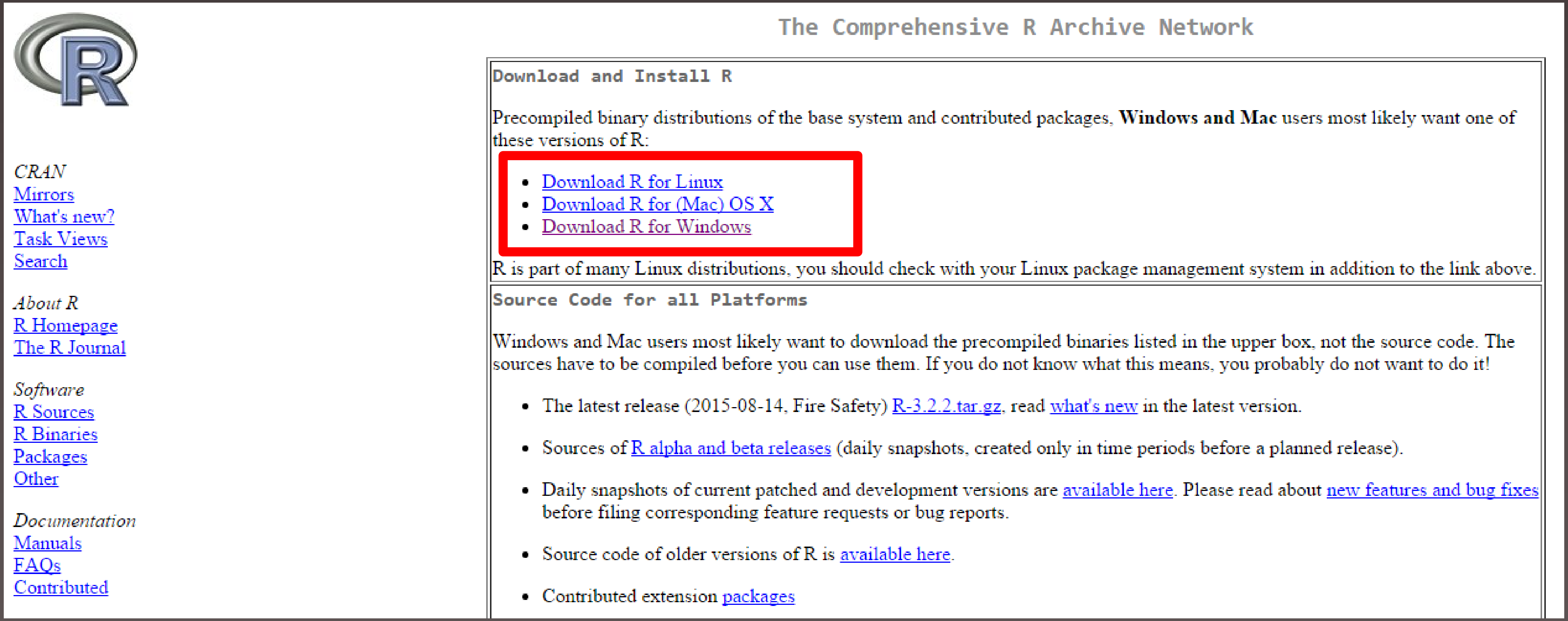
News

- **R version 3.2.2 (Fire Safety)** has been released on 2015-08-14.
- **The R Journal Volume 7/1** is available.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.
- **useR! 2015**, will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.
- **useR! 2014**, took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.

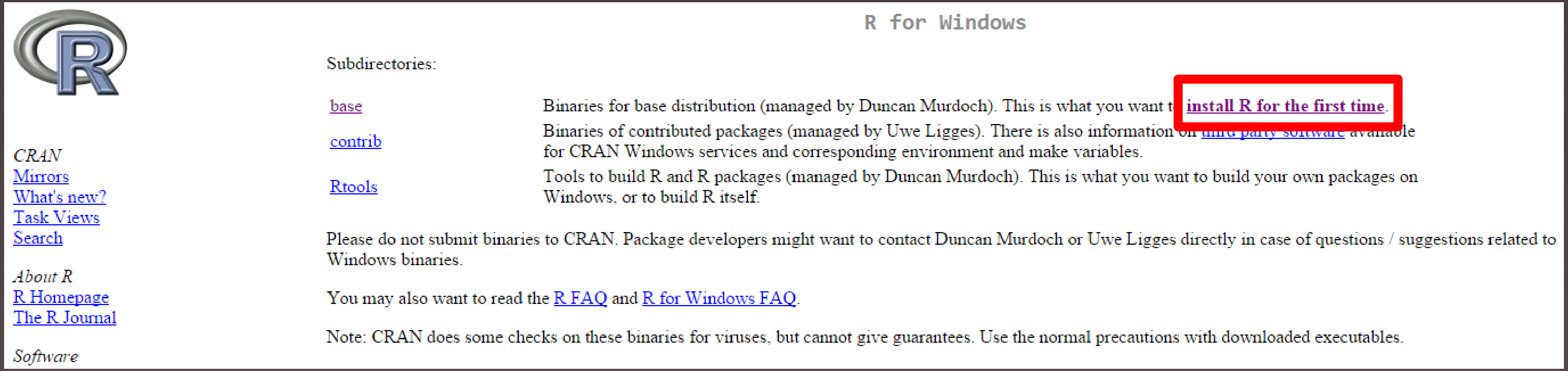
R 安裝步驟三

Switzerland		
https://stat.ethz.ch/CRAN/		ETH Zürich
http://stat.ethz.ch/CRAN/		ETH Zürich
Taiwan		
http://ftp.yzu.edu.tw/CRAN/	→ 元智大學	Department of Computer Science and Engineering, Yuan Ze University
http://cran.csie.ntu.edu.tw/	→ 台灣大學	National Taiwan University, Taipei
Thailand		
http://mirrors.psu.ac.th/pub/cran/		Prince of Songkla University, Hatyai
Turkey		
http://cran.pau.edu.tr/		Pamukkale University, Denizli
http://cran.ncc.metu.edu.tr/		Middle East Technical University Northern Cyprus Campus, Mersin
UK		
https://www.stats.bris.ac.uk/R/		University of Bristol
http://www.stats.bris.ac.uk/R/		University of Bristol

R 安裝步驟四


The Comprehensive R Archive Network
CRAN
Mirrors
What's new?
Task Views
Search
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
Packages
Other
Documentation
Manuals
FAQs
Contributed
Download and Install R
Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:
Download R for Linux
Download R for (Mac) OS X
Download R for Windows
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.
Source Code for all Platforms
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!
The latest release (2015-08-14, Fire Safety) R-3.2.2.tar.gz, read what's new in the latest version.
Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
Source code of older versions of R is available here.
Contributed extension packages

R 安裝步驟五



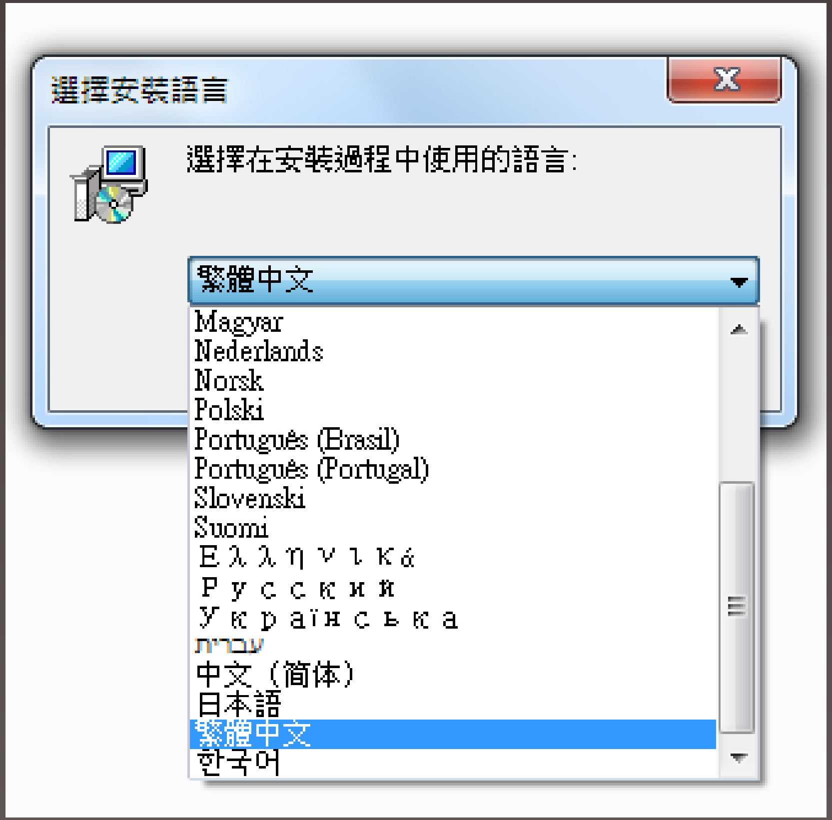
R for Windows
CRAN
Mirrors
What's new?
Task Views
Search
About R
R Homepage
The R Journal
Software
Subdirectories:
base
Binaries for base distribution (managed by Duncan Murdoch). This is what you want t install R for the first time.
contrib
Binaries of contributed packages (managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
Rtools
Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.
Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.
You may also want to read the R FAQ and R for Windows FAQ.
Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

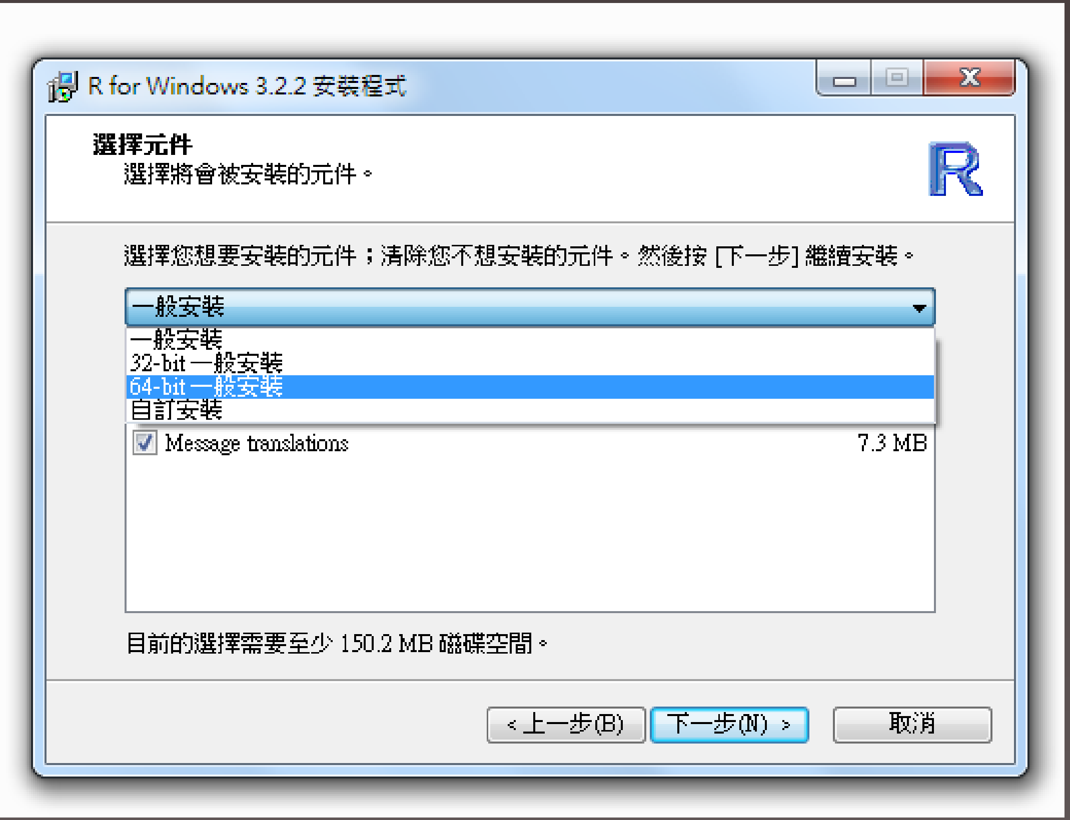
R 安裝步驟六


R-3.2.2 for Windows (32/64 bit)
Download R 3.2.2 for Windows (62 megabytes, 32/64 bit)
New features in this version
If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the md5sum of the .exe to the true fingerprint. You will need a version of md5sum for windows: both graphical and command line versions are available.
Frequently asked questions
Does R run under my version of Windows?
How do I update packages in my previous version of R?
Should I run 32-bit or 64-bit R?
CRAN
Mirrors
What's new?
Task Views
Search
About R
R Homepage
The R Journal

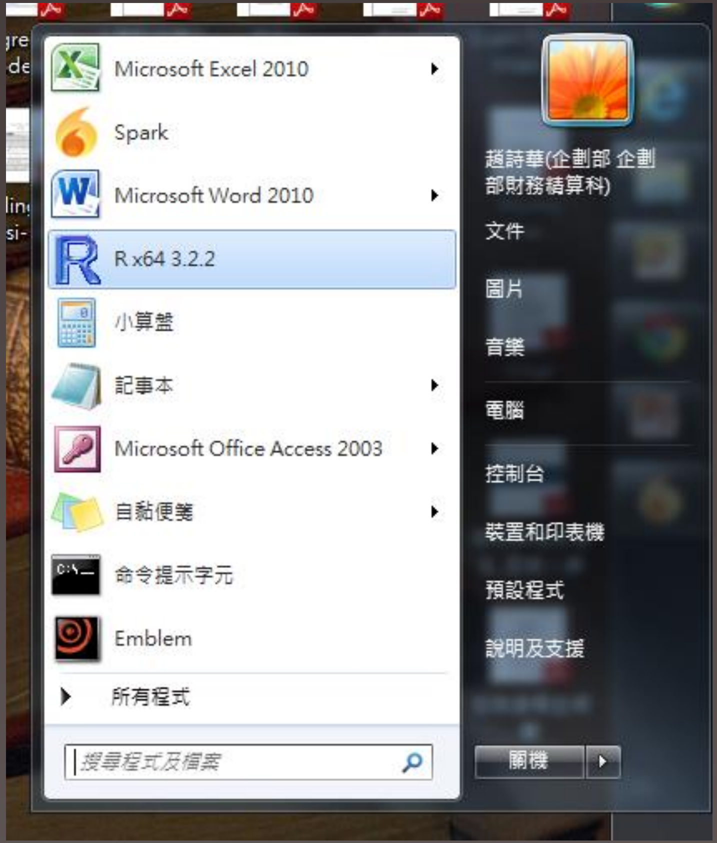
R 安裝步驟七

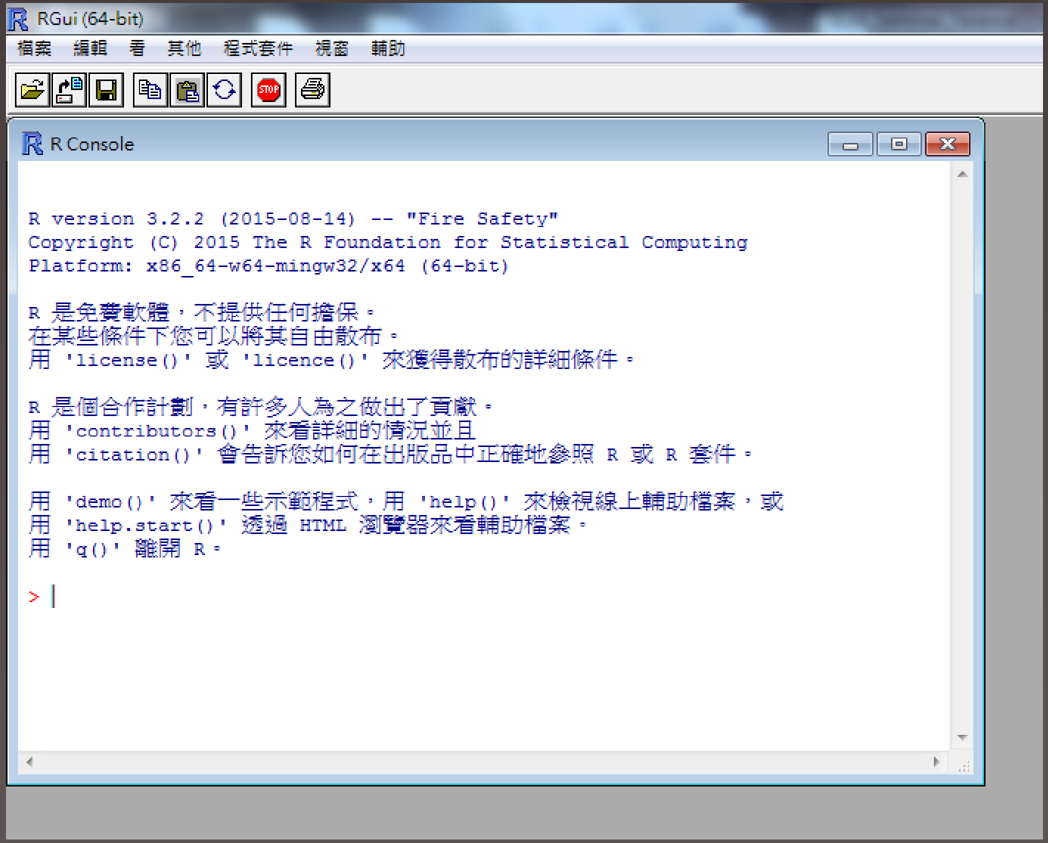
R-3.2.2-win (as of 2015.10.01)



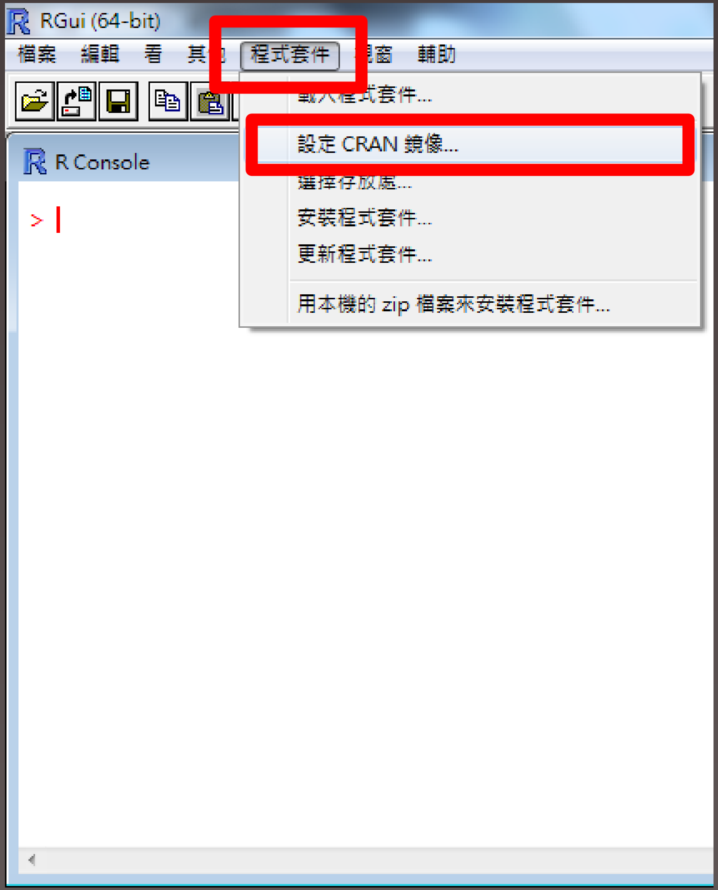


R 設定步驟一

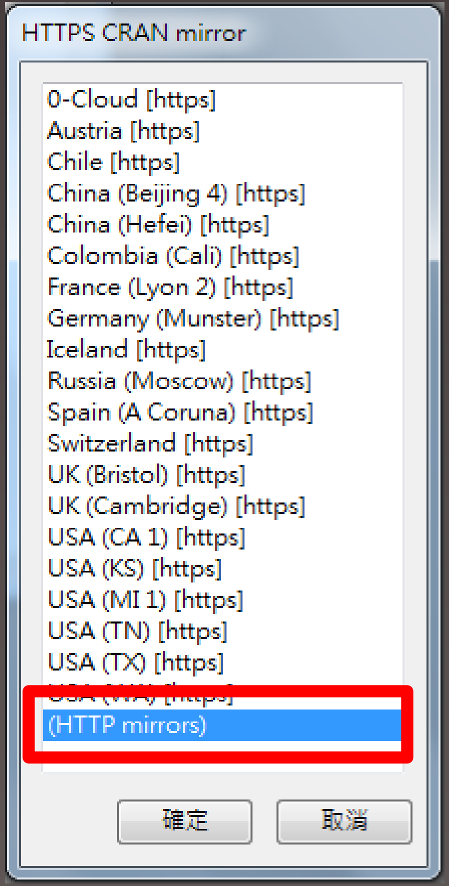

Microsoft Excel 2010
Spark
Microsoft Word 2010
R x64 3.2.2
小算盤
記事本
Microsoft Office Access 2003
自黏便箋
命令提示字元
Emblem
所有程式
搜尋程式及檔案
趙詩華(企劃部 企劃部財務精算科)
文件
圖片
音樂
電腦
控制台
裝置和印表機
預設程式
說明及支援
關機


RGui (64-bit)
檔案 編輯 看 其他 程式套件 視窗 輔助
R Console
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條件。
R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。
用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。
>

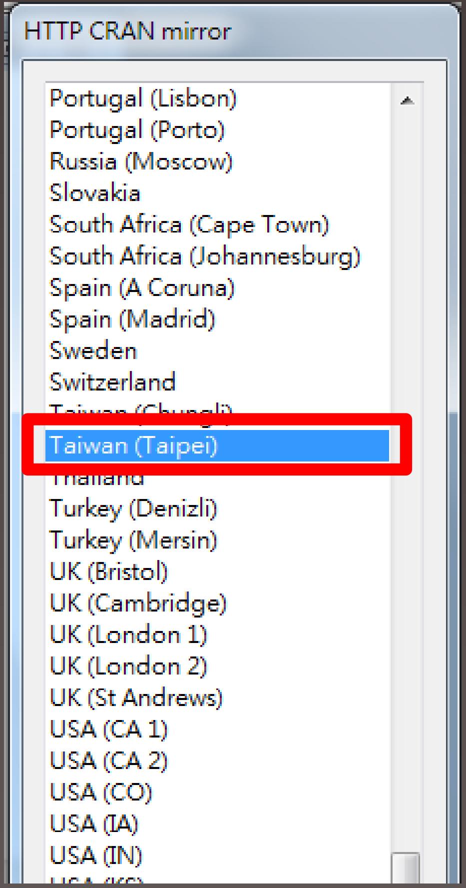
R 設定步驟二



RGui (64-bit)
檔案
編輯
看
程式套件
輔助
R Console
設定 CRAN 鏡像...
安裝程式套件...
更新程式套件...
用本機的 zip 檔案來安裝程式套件...

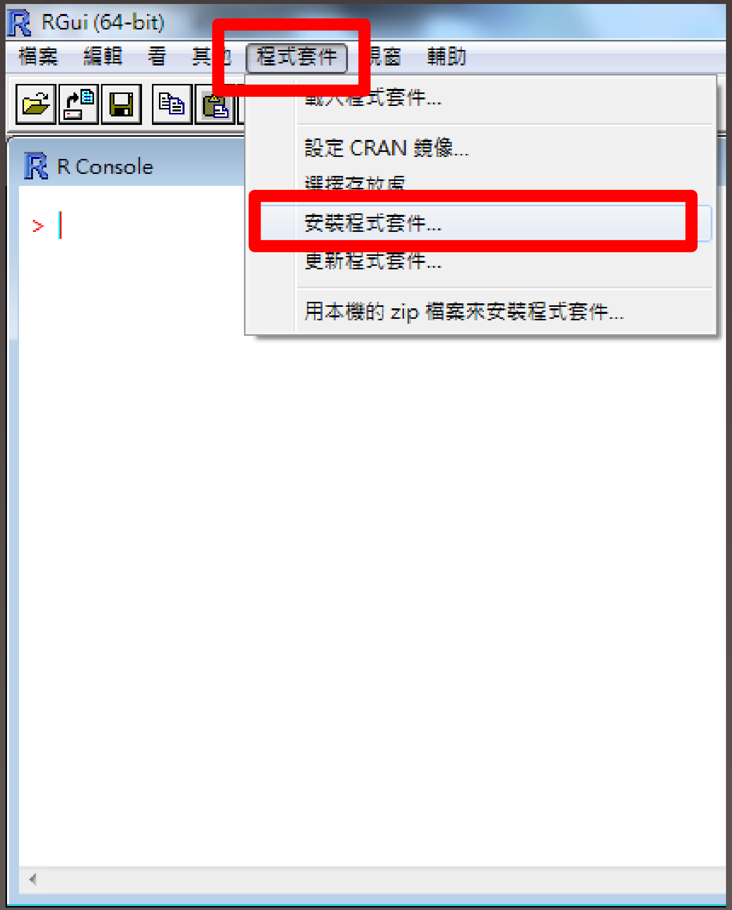


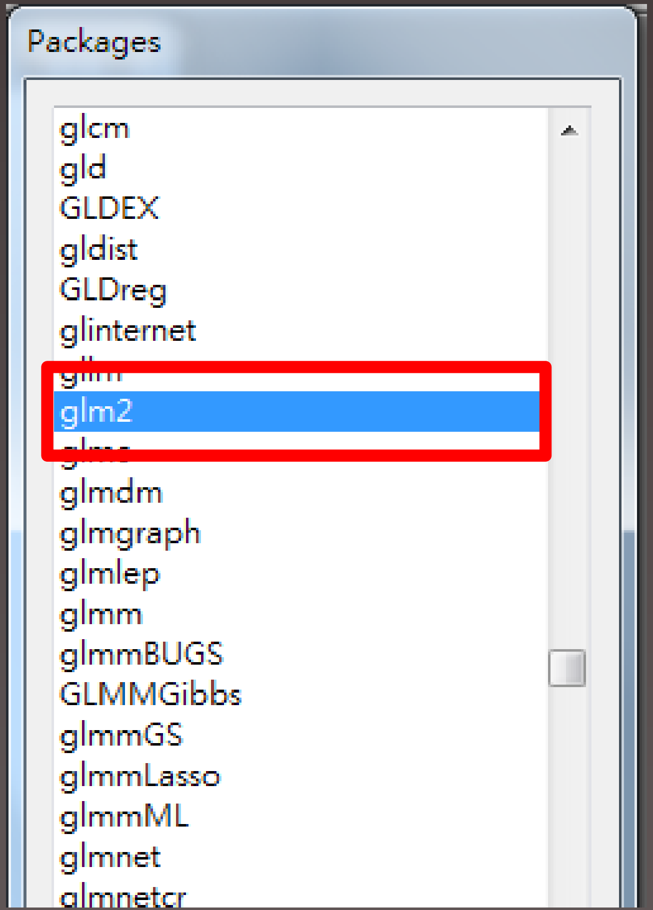
HTTPS CRAN mirror
0-Cloud [https]
Austria [https]
Chile [https]
China (Beijing 4) [https]
China (Hefei) [https]
Colombia (Cali) [https]
France (Lyon 2) [https]
Germany (Munster) [https]
Iceland [https]
Russia (Moscow) [https]
Spain (A Coruna) [https]
Switzerland [https]
UK (Bristol) [https]
UK (Cambridge) [https]
USA (CA 1) [https]
USA (KS) [https]
USA (MI 1) [https]
USA (TN) [https]
USA (TX) [https]
(HTTP mirrors)
確定
取消



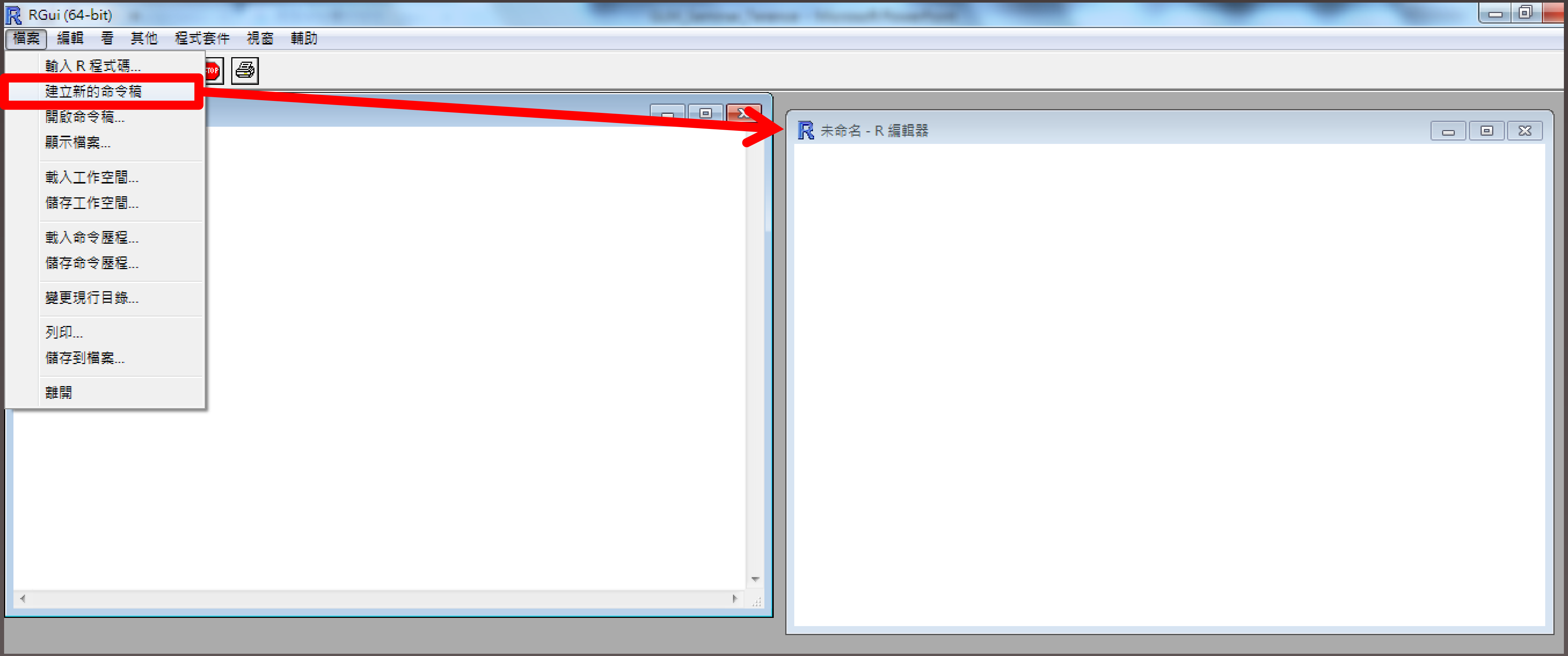
HTTP CRAN mirror
Portugal (Lisbon)
Portugal (Porto)
Russia (Moscow)
Slovakia
South Africa (Cape Town)
South Africa (Johannesburg)
Spain (A Coruna)
Spain (Madrid)
Sweden
Switzerland
Taiwan (Taipei)
Turkey (Denizli)
Turkey (Mersin)
UK (Bristol)
UK (Cambridge)
UK (London 1)
UK (London 2)
UK (St Andrews)
USA (CA 1)
USA (CA 2)
USA (CO)
USA (IA)
USA (IN)

R 設定步驟三


RGui (64-bit)
檔案 編輯 看 其他 程式套件 視窗 輔助
R Console
設定 CRAN 鏡像...
安裝程式套件...
更新程式套件...
用本機的 zip 檔案來安裝程式套件...


Packages
glcm
gld
GLDEX
gldist
GLDreg
glinternet
glm2
glmdm
glmgraph
glmlep
glmm
glmmBUGS
GLMMGibbs
glmmGS
glmmLasso
glmmML
glmnet

R 基本操作


RGui (64-bit)
檔案 編輯 看 其他 程式套件 視窗 輔助
輸入 R 程式碼...
建立新的命令稿
開啟命令稿...
顯示檔案...
載入工作空間...
儲存工作空間...
載入命令歷程...
儲存命令歷程...
變更現行目錄...
列印...
儲存到檔案...
離開
未命名 - R 編輯器

R 基本操作

- library(套件名稱) -> 載入套件

  (常使用的套件如：MASS、ggplot2、glm2)

- Ctrl+F5 -> 執行選取之程式碼

- ?+指令 -> 在CRAN中查詢

  (如：?glm -> http://127.0.0.1:17786/library/stats/html/glm.html)

R 套件

- 統計至 2015.6.18 約有 6,000 多個套件免費使用

  (http://blog.revolutionanalytics.com/2015/06/fishing-for-packages-in-cran.html)

- glm 功能內建於 stat 套件中
- glm2 為 Ian Marschner 所開發，增加模型配適收斂的穩定性

  (https://cran.r-project.org/web/packages/glm2/glm2.pdf)

R 套件(續)

- glm 與 glm2 於 R 中可使用之分配及連結函數

分配	預設連結函數
binomial	logit
gaussian	identity
Gamma	inverse
poisson	log
quasibinomial	logit
quasipoisson	log

R 套件(續)

- 負二項分配配適建置於 MASS 套件中

分配	預設連結函數
glm.nb	log

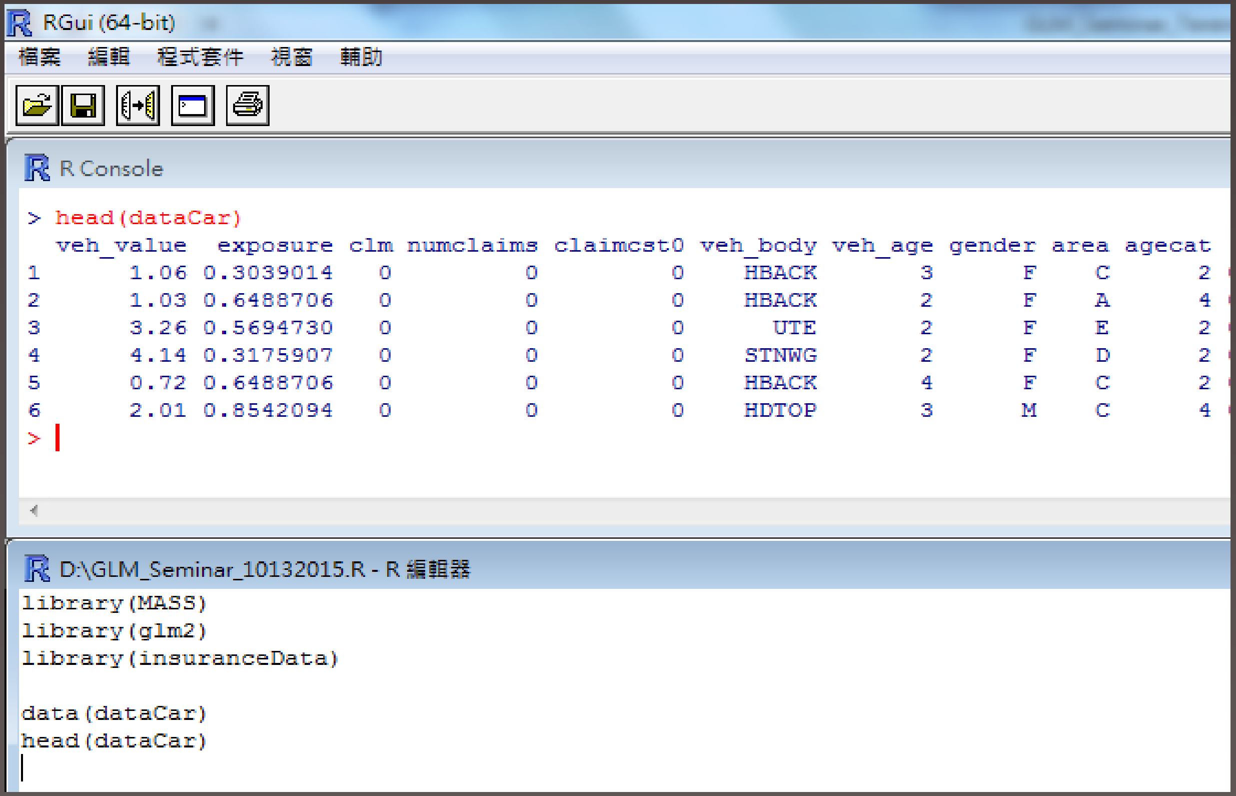
視其所以，觀其所由，察其所安。

《論語‧為政》

範例一

- 採用 insuranceData 套件中的dataCar資料

```
library(insuranceData)
data(dataCar)
head(dataCar)
```

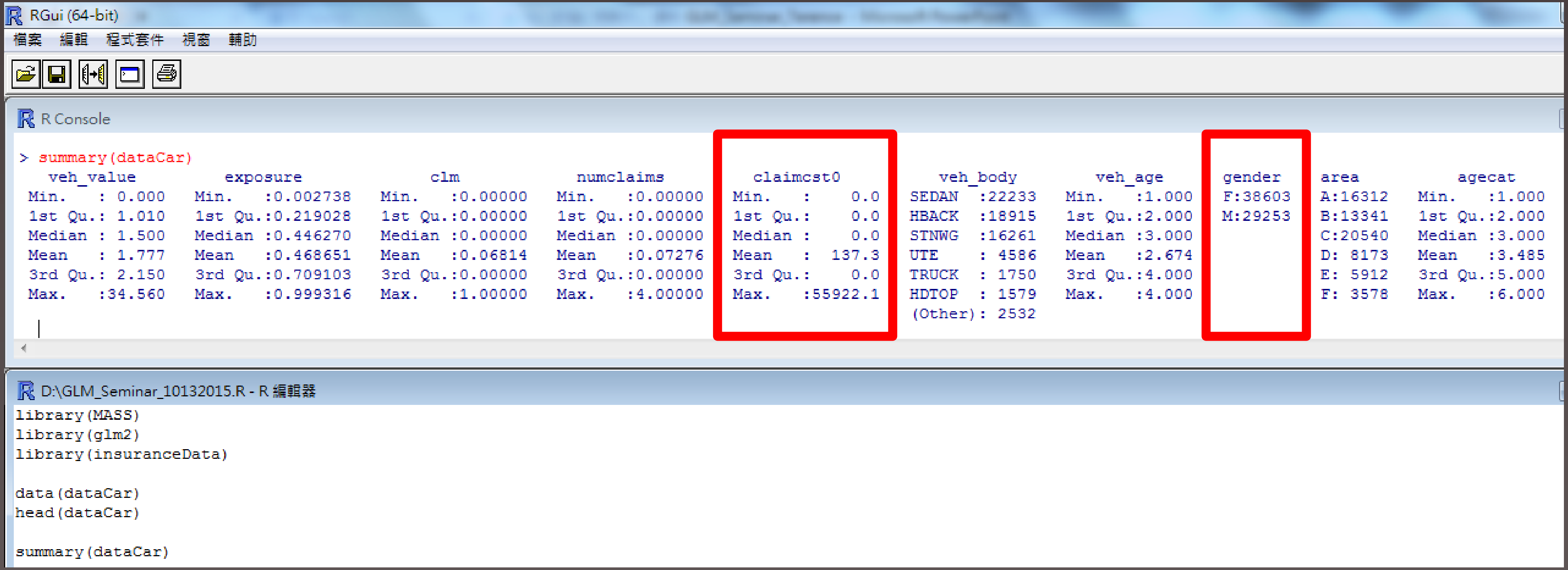


範例一(續)

資料欄位	解釋
veh_value	車輛價值 (萬元)
exposure	Exposure
clm	是否發生賠案 (否 = 0 , 是 = 1)
numclaims	理賠件數
claimcst0	理賠金額 (0 = 無理賠)
veh_body	車輛種類
veh_age	車齡分類 (1 – 4 , 新 – 舊)
gender	性別 (女性 = F , 男性 = M)
area	地區別 (A – F)
agecat	年齡分類 (1 – 6 , 小 – 大)

範例一(續)

- 基本敘述統計 summary(dataCar)



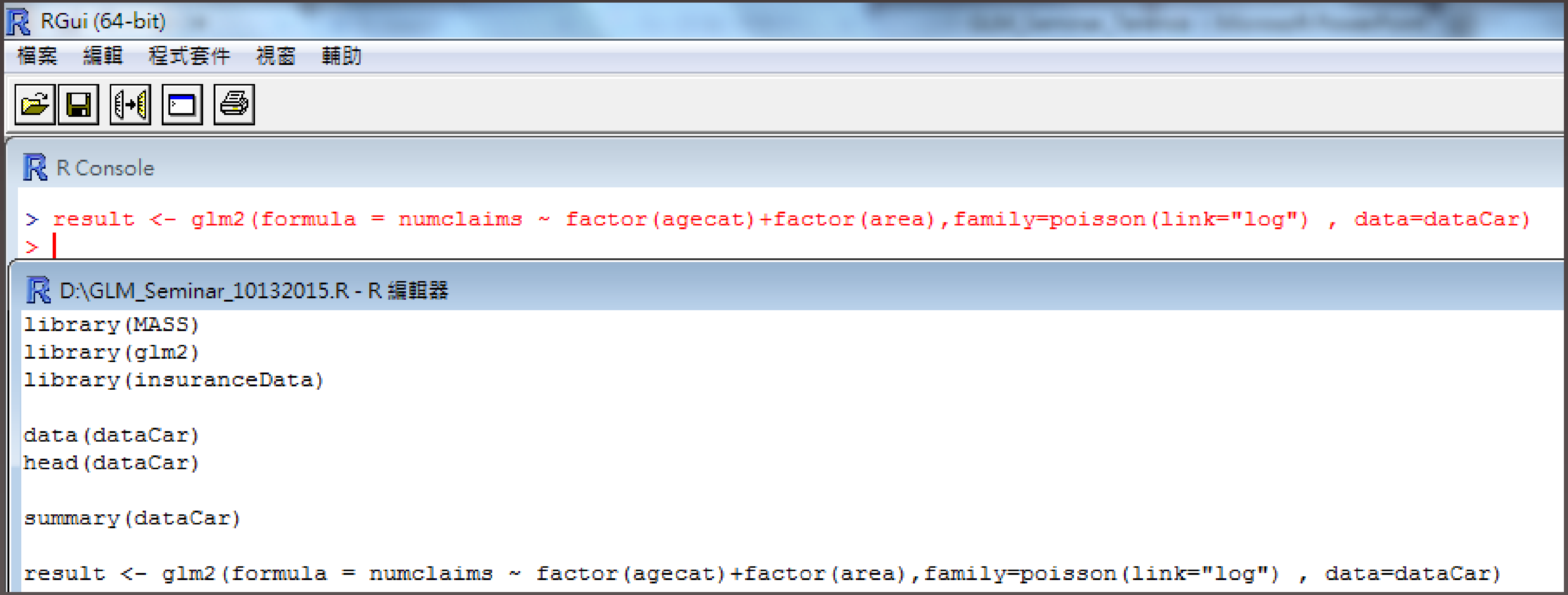
範例一(續)

- 模型配適指令

result <- glm2(formula , family , data)

- result → 儲存位置
- formula → 迴歸式
- family → 配適分配 & 連結函數
- data → 資料位置

範例一(續)


RGui (64-bit)
檔案 編輯 程式套件 視窗 輔助
R Console
> result <- glm2(formula = numclaims ~ factor(agecat)+factor(area),family=poisson(link="log") , data=dataCar)
>
D:\GLM_Seminar_10132015.R - R 編輯器
library(MASS)
library(glm2)
library(insuranceData)
data(dataCar)
head(dataCar)
summary(dataCar)
result <- glm2(formula = numclaims ~ factor(agecat)+factor(area),family=poisson(link="log") , data=dataCar)

範例一(續)

- GLM分析報表 summary(result)

RGui (64-bit)

檔案 編輯 程式套件 視窗 輔助

R Console

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4532  -0.3929  -0.3827  -0.3479   5.0926

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -2.396574   0.050483 -47.473  < 2e-16 ***
factor(agecat)2   -0.166871   0.053912  -3.095 0.001966 **
factor(agecat)3   -0.191743   0.052414  -3.658 0.000254 ***
factor(agecat)4   -0.219510   0.052455  -4.185 2.86e-05 ***
factor(agecat)5   -0.409952   0.058786  -6.974 3.09e-12 ***
factor(agecat)6   -0.417655   0.066983  -6.235 4.51e-10 ***
factor(area)B      0.055743   0.042743   1.304 0.192183
factor(area)C      0.001689   0.038946   0.043 0.965413
factor(area)D     -0.116697   0.052498  -2.223 0.026223 *
factor(area)E     -0.034412   0.057175  -0.602 0.547261
factor(area)F      0.120794   0.064542   1.872 0.061268 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138

Number of Fisher Scoring iterations: 6
```

RGui (64-bit)

檔案 編輯 程式套件 視窗 輔助

R Console

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4532  -0.3929  -0.3827  -0.3479   5.0926

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.396574   0.050483 -47.473  < 2e-16 ***
factor(agecat)2 -0.166871   0.053912  -3.095 0.001966 **
factor(agecat)3 -0.191743   0.052414  -3.658 0.000254 ***
factor(agecat)4 -0.219510   0.052455  -4.185 2.86e-05 ***
factor(agecat)5 -0.409952   0.058786  -6.974 3.09e-12 ***
factor(agecat)6 -0.417655   0.066983  -6.235 4.51e-10 ***
factor(area)B    0.055743   0.042743   1.304 0.192183
factor(area)C    0.001689   0.038946   0.043 0.965413
factor(area)D   -0.116697   0.052498  -2.223 0.026223 *
factor(area)E   -0.034412   0.057175  -0.602 0.547261
factor(area)F    0.120794   0.064542   1.872 0.061268 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138

Number of Fisher Scoring iterations: 6
```

RGui (64-bit)

檔案 編輯 程式套件 視窗 輔助

R Console

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4532  -0.3929  -0.3827  -0.3479   5.0926

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.396574   0.050483 -47.473  < 2e-16 ***
factor(agecat)2 -0.166871   0.053912  -3.095 0.001966 **
factor(agecat)3 -0.191743   0.052414  -3.658 0.000254 ***
factor(agecat)4 -0.219510   0.052455  -4.185 2.86e-05 ***
factor(agecat)5 -0.409952   0.058786  -6.974 3.09e-12 ***
factor(agecat)6 -0.417655   0.066983  -6.235 4.51e-10 ***
factor(area)B    0.055743   0.042743   1.304 0.192183
factor(area)C    0.001689   0.038946   0.043 0.965413
factor(area)D   -0.116697   0.052498  -2.223 0.026223 *
factor(area)E   -0.034412   0.057175  -0.602 0.547261
factor(area)F    0.120794   0.064542   1.872 0.061268 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138

Number of Fisher Scoring iterations: 6
```

RGui (64-bit)

檔案 編輯 程式套件 視窗 輔助

R Console

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4532  -0.3929  -0.3827  -0.3479   5.0926

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.396574   0.050483 -47.473  < 2e-16 ***
factor(agecat)2 -0.166871   0.053912  -3.095 0.001966 **
factor(agecat)3 -0.191743   0.052414  -3.658 0.000254 ***
factor(agecat)4 -0.219510   0.052455  -4.185 2.86e-05 ***
factor(agecat)5 -0.409952   0.058786  -6.974 3.09e-12 ***
factor(agecat)6 -0.417655   0.066983  -6.235 4.51e-10 ***
factor(area)B    0.055743   0.042743   1.304 0.192183
factor(area)C    0.001689   0.038946   0.043 0.965413
factor(area)D   -0.116697   0.052498  -2.223 0.026223 *
factor(area)E   -0.034412   0.057175  -0.602 0.547261
factor(area)F    0.120794   0.064542   1.872 0.061268 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138

Number of Fisher Scoring iterations: 6
```

RGui (64-bit)

檔案 編輯 程式套件 視窗 輔助

R Console

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4532  -0.3929  -0.3827  -0.3479   5.0926

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.396574   0.050483 -47.473  < 2e-16 ***
factor(agecat)2 -0.166871   0.053912  -3.095 0.001966 **
factor(agecat)3 -0.191743   0.052414  -3.658 0.000254 ***
factor(agecat)4 -0.219510   0.052455  -4.185 2.86e-05 ***
factor(agecat)5 -0.409952   0.058786  -6.974 3.09e-12 ***
factor(agecat)6 -0.417655   0.066983  -6.235 4.51e-10 ***
factor(area)B    0.055743   0.042743   1.304 0.192183
factor(area)C    0.001689   0.038946   0.043 0.965413
factor(area)D   -0.116697   0.052498  -2.223 0.026223 *
factor(area)E   -0.034412   0.057175  -0.602 0.547261
factor(area)F    0.120794   0.064542   1.872 0.061268 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138

Number of Fisher Scoring iterations: 6
```

RGui (64-bit)

檔案 編輯 程式套件 視窗 輔助

R Console

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4532  -0.3929  -0.3827  -0.3479   5.0926

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.396574   0.050483 -47.473  < 2e-16 ***
factor(agecat)2 -0.166871   0.053912  -3.095 0.001966 **
factor(agecat)3 -0.191743   0.052414  -3.658 0.000254 ***
factor(agecat)4 -0.219510   0.052455  -4.185 2.86e-05 ***
factor(agecat)5 -0.409952   0.058786  -6.974 3.09e-12 ***
factor(agecat)6 -0.417655   0.066983  -6.235 4.51e-10 ***
factor(area)B    0.055743   0.042743   1.304 0.192183
factor(area)C    0.001689   0.038946   0.043 0.965413
factor(area)D   -0.116697   0.052498  -2.223 0.026223 *
factor(area)E   -0.034412   0.057175  -0.602 0.547261
factor(area)F    0.120794   0.064542   1.872 0.061268 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138

Number of Fisher Scoring iterations: 6
```

範例一(續)

- 估計值之轉換依連結函數而定

```
> cbind(Relativity=exp(coef(result)),Estimate=coef(result))
                  Relativity     Estimate
(Intercept)       0.09102932 -2.396573633
factor(agecat)2   0.84630889 -0.166870873
factor(agecat)3   0.82551905 -0.191742944
factor(agecat)4   0.80291205 -0.219510094
factor(agecat)5   0.66368215 -0.409951941
factor(agecat)6   0.65858953 -0.417654799
factor(area)B     1.05732572  0.055742819
factor(area)C     1.00169022  0.001688795
factor(area)D     0.88985490 -0.116696861
factor(area)E     0.96617352 -0.034411837
factor(area)F     1.12839255  0.120794096
>
```

範例一(續)

- 模型選擇考參考AIC，AIC值較小之模型較佳。

AIC = 36,138

```
> summary(result)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = poisson(link = "log"),
    data = dataCar)

    Null deviance: 26768  on 67855  degrees of freedom
Residual deviance: 26681  on 67845  degrees of freedom
AIC: 36138
```

AIC = 18,878

```
> result_2 <- glm2(formula = numclaims ~ factor(agecat)+factor(area),family=gaussian(link="ident
> summary(result_2)

Call:
glm2(formula = numclaims ~ factor(agecat) + factor(area), family = gaussian(link = "identity"),
    data = dataCar)

    Null deviance: 5251.8  on 67855  degrees of freedom
Residual deviance: 5245.5  on 67845  degrees of freedom
AIC: 18878
```

知是行之始，行是知之成。

《傳習錄》

Easy GLM

https://terencechaoapplication.shinyapps.io/EasyGLM

Easy GLM

- ✓ 透過 Rstudio 公司開發的 Shiny 套件製作
- ✓ 資料須為 .csv 檔或 .txt檔
- ✓ 欄位名稱須以英文標示
- ✓ 每月使用時數為 250 個小時

shinyapps.io Powered by RStudio

Easy GLM ! | 資料選擇 | 敘述統計 | GLM分析 | More

請選擇檔案 (csv檔 / txt檔)

選擇檔案 未選擇任何檔案

☑ 第一列是否為表頭(勾選為是)

資料分隔方式

- ◉ 逗點
- ○ 分號
- ○ Tab

需呈現資料筆數

1 6 100

1 11 21 31 41 51 61 71 81 91 100

shinyapps.io Powered by RStudio

Easy GLM ! 資料選擇 敘述統計 GLM分析 More

請選擇檔案 (csv檔 / txt檔)

選擇檔案 AutoBI.csv

Upload complete

☑ 第一列是否為表頭(勾選為是)

資料分隔方式

- ◉ 逗點
- ○ 分號
- ○ Tab

需呈現資料筆數

1 6 100

1 11 21 31 41 51 61 71 81 91 100

	Loss	Age	Territory	Gender	Marri	ClaimCount
1	34.94	50	1	1		5
2	10.89	28	2	2	2	13
3	0.33	5	2	1	2	66
4	11.04	32	1	1	1	71
5	0.14	30	2	1	4	96
6	0.31	35	1	2	1	97

shinyapps.io Powered by RStudio

Easy GLM ! 資料選擇 敘述統計 GLM分析 More

```
      Loss                Age          Territory        Gender     
 Min.   :   0.0050   Min.   : 0.00   Min.   :1.00   Min.   :1.000  
 1st Qu.:   0.6825   1st Qu.:19.75   1st Qu.:1.00   1st Qu.:1.000  
 Median :   2.3325   Median :31.00   Median :1.00   Median :2.000  
 Mean   :   6.4368   Mean   :32.70   Mean   :1.49   Mean   :1.554  
 3rd Qu.:   3.9978   3rd Qu.:43.00   3rd Qu.:2.00   3rd Qu.:2.000  
 Max.   :1067.6970   Max.   :95.00   Max.   :2.00   Max.   :2.000  
                     NA's   :152                    NA's   :8      
     Marri         ClaimCount   
 Min.   :1.000   Min.   :    5  
 1st Qu.:1.000   1st Qu.: 7141  
 Median :2.000   Median :14268  
 Mean   :1.583   Mean   :14125  
 3rd Qu.:2.000   3rd Qu.:21377  
 Max.   :4.000   Max.   :28294  
 NA's   :14                     
```

shinyapps.io Powered by RStudio

Easy GLM ! 資料選擇 敘述統計 GLM分析 More ▾

請選擇反應變數

- (●) Loss
- ( ) Age
- ( ) Territory
- ( ) Gender
- ( ) Marri
- ( ) ClaimCount

請選擇解釋變數

- [ ] Loss
- [x] Age
- [ ] Territory
- [ ] Gender
- [ ] Marri
- [ ] ClaimCount

是否分析交互影響

- ( ) 是
- (●) 否

請選擇分配函數

- (●) Normal
- ( ) Gamma
- ( ) Poisson
- ( ) Quasi-Posiion

請選擇連結函數

- (●) Identity
- ( ) Log
- ( ) Inverse

請選擇 Prior Weights

不使用P.W.

請選擇Offset

不使用Offset

目前版本僅可選擇單一變數

點我進行模型配適

GLM分析報表

shinyapps.io   Powered by RStudio

Easy GLM !   資料選擇   敘述統計   GLM分析   More

請選擇反應變數

- (•) Loss
- ( ) Age
- ( ) Territory
- ( ) Gender
- ( ) Marri
- ( ) ClaimCount

請選擇解釋變數

- [ ] Loss
- [ ] Age
- [ ] Territory
- [x] Gender
- [ ] Marri
- [ ] ClaimCount

是否分析交互影響

- ( ) 是
- (•) 否

請選擇分配函數

- (•) Normal
- ( ) Gamma
- ( ) Poisson
- ( ) Quasi-Posiion

請選擇連結函數

- (•) Identity
- ( ) Log
- ( ) Inverse

請選擇 Prior Weights

不使用P.W.

請選擇Offset

不使用Offset

目前版本僅可選擇單一變數

點我進行模型配適

GLM分析報表

```
Call:
glm2(formula = formula_1, family = family_dist, data = userdata(),
    maxit = 50)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
  -6.75    -5.67    -4.09    -2.39  1060.94

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       6.0403     1.6580   3.643 0.000282 ***
factor(Gender)2   0.7176     2.2275   0.322 0.747415
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1338.756)

    Null deviance: 1459383  on 1091  degrees of freedom
Residual deviance: 1459244  on 1090  degrees of freedom
  (8 observations deleted due to missingness)
AIC: 10965

Number of Fisher Scoring iterations: 2
```


Q & A