

# 中央研究院統計科學研究所

## 2016 年統計研習營 簡章

### 一、宗旨：

提供在學學生學習統計相關知識及參與學術研究機會，培育統計領域優秀人才。

### 二、研習時間：2016 年 7 月 27 日（三）至 8 月 12 日（五）

### 三、研習地點：中央研究院統計科學研究所

### 四、申請資格：對統計科學有興趣之國內外大學部在學學生（含應屆畢業生）。

### 五、申請辦法：[線上申請](#)，即日起至 2016 年 6 月 5 日止

### 六、申請文件：

1. 在學證明 (jpg 或 pdf)：學生證（正反面，含本學期註冊章）或其他可證明文件。
2. 歷年成績單 (jpg 或 pdf)：含 104 學年度第一學期成績
3. 推薦函 2 封：推薦人限大專院校教師或相關研究人員  
請推薦人於 2016 年 6 月 5 日前  
將推薦函逕寄本所承辦人林晏如小姐 [school@stat.sinica.edu.tw](mailto:school@stat.sinica.edu.tw)  
郵件主旨及推薦函檔名請設定為：統計研習營推薦函-申請人姓名  
推薦函參考格式如後附

### 七、課程簡介（各課程講師依中文姓氏筆畫排序）

**註：研習期間需自備筆電**

#### 預備課程（7 月 27 日~7 月 29 日）

- ◆機率統計課程入門（1.5 天）  
黃文璋 教授（國立高雄大學統計學研究所）
- ◆R 軟體基礎與實務應用（1.5 天）  
吳漢銘 教授（淡江大學數學系）

## 核心課程 (8月1日~8月5日、8月8日~8月12日)

### ◆社會網絡 (Social Networks) (2.5 天)

王泰期 博士 (國家實驗研究院國家高速網路與計算中心)  
林遠隆 博士 (中央研究院統計科學研究所)  
翁章譯 博士 (中央研究院統計科學研究所)  
楊振翔 副研究員 (中央研究院統計科學研究所)  
潘建興 副研究員 (中央研究院統計科學研究所)  
劉維中 副研究員 (中央研究院統計科學研究所)  
盧馬汀 博士 (中央研究院統計科學研究所)  
顏佐榕 助研究員 (中央研究院統計科學研究所)

### ◆生物資訊 (Bioinformatics) (2.5 天)

袁新盛 副研究員 (中央研究院統計科學研究所)  
陳璿宇 助研究員 (中央研究院統計科學研究所)  
黃耀廷 副教授 (國立中正大學資訊工程學系)  
劉佳鑫 先生 (Taiwan International Graduate Program 博士候選人)  
謝叔蓉 研究員 (中央研究院統計科學研究所)

### ◆資料科學 (Data Science) (2.5 天)

王偉仲 教授 (國立臺灣大學數學系)  
林宗男 教授 (國立臺灣大學電機工程學系)  
陳孟彰 研究員 (中央研究院資訊科學研究所)  
莊永裕 教授 (國立臺灣大學資訊工程學系)  
張源俊 研究員 (中央研究院統計科學研究所)

### ◆生物與醫學統計 (Biostatistics and Medical Statistics) (2.5 天)

王乃昕 教授 (美國密西根大學統計學系)  
沈志陽 研究員 (中央研究院生物醫學科學研究所)  
潘文涵 研究員 (中央研究院生物醫學科學研究所)  
楊欣洲 副研究員 (中央研究院統計科學研究所)  
劉玉麗 副研究員 (國家衛生研究院神經及精神醫學研究中心)

## 八、錄取結果：

本所將依據申請人所備之書面資料，由所方組成之審核小組進行審查並決定錄取人員。本年度預計招收 20 名學員，6 月底以 e-mail 個別通知審查結果，並於活動網頁公告相關資訊。

九、結業證書及研習費：

研習期間經評審成績優良之學員，將頒發結業證書及研習費（含部分住宿補助）。

十、住宿：

學員可申請於研習期間自費住宿（7/26晚~8/11晚，不含7/29、7/30、8/5、8/6四個晚上），由本所代為安排本院學術活動中心雙人房為原則（1,700元/晚，平均每人負擔850元/晚）。

本院學術活動中心客房 <http://gao.sinica.edu.tw/huo-dong/room.php>

十一、聯絡人：林晏如小姐、(02) 6614-5621、[school@stat.sinica.edu.tw](mailto:school@stat.sinica.edu.tw)

活動網址：<http://www3.stat.sinica.edu.tw/school2016/>

## 課程簡介

(各課程講師依中文姓氏筆畫排序)

### ◆機率統計課程入門 (1.5 天)

[黃文璋 教授](#) (國立高雄大學統計學研究所)

由於生活中充滿著隨機性，因此人們常需對現況做推測，或對未來做預測。這些推論如何產生？擲筊？藉助水晶球？雖可有各種推論方式，但科學的推論，便得仰賴數據，所謂數據會說話。利用統計方法，先收集資料，加以整理及分析，然後給出推論。

對於隨機現象，即使用的方法很好，推論仍不易次次準確，也難以保證預測的誤差必有多大。一般人受數學的薰陶較久，數學思維根深蒂固，因而初學統計，對很多概念便常感到迷惑。數學中常在證明，從給定前提下，推導出結果；而統計則是由結果，來判定前提是否可接受。運用統計，通常不是在證明那件事必然是對的。在隨機世界裡，一切都是假設，只是看那一假設可接受即使接受某一假設，真相為何，仍屬未知。另外，數學中所得結果往往為斬釘截鐵式的，而統計裡所給乃允許誤差下的機率式保證。隨機性的特質，使統計在理解層次上，並不易掌握。因而在實際應用時，由於未具備數據素養，常未能讓統計的功能充分發揮，甚至誤用，使數據說錯話。

在本課程中，將闡釋隨機性的概念，並介紹一些基本的統計思維。

### ◆R 軟體基礎與實務應用 (1.5 天)

[吳漢銘 教授](#) (淡江大學數學學系)

巨量資料(Big Data)的潮流帶動資料科學(Data Science)的興起。資料科學是綜合統計學、資訊科學及領域知識的一門新興學科，而 R 語言正是資料科學領域裡，最受歡迎的程式語言之一(見 [http://www.tiobe.com/tiobe\\_index](http://www.tiobe.com/tiobe_index) 及 <http://www.kdnuggets.com/2013/08/languages-for-analytics-data-mining-data-science.html>)。R 具有免費，開放源碼及學習資源豐富的優點，而其功能可透過套件增強。本課程開課之前，學員需自行先預習演練「R 語言統計計算環境介紹及物件基礎」和「資料的輸入與輸出」教材內容。課程三天主題為：(1) R 函式編寫，(2) 統計圖形及資料視覺化及(3) 資料分析應用主題。希望藉由 R 程式設計之技能訓練，奠定學員程式撰寫及資料分析的基礎。

#### ◆社會網絡 (Social Networks) (2.5 天)

[王泰期 博士](#) (國家實驗研究院國家高速網路與計算中心)

[林遠隆 博士](#) (中央研究院統計科學研究所)

[翁章譯 博士](#) (中央研究院統計科學研究所)

[楊振翔 副研究員](#) (中央研究院統計科學研究所)

[潘建興 副研究員](#) (中央研究院統計科學研究所)

[劉維中 副研究員](#) (中央研究院統計科學研究所)

[盧馬汀 博士](#) (中央研究院統計科學研究所)

[顏佐榕 助研究員](#) (中央研究院統計科學研究所)

當各位網民在臉書上玩遊戲交朋友時，你們知道臉書在後台分析著你們的交友網，以便將來為你們推薦新朋友嗎？當各位戰友在魔獸裡攻城掠地時，你們知道暴雪在後台分析著你們的表現，以便未來更新時為你們設定新的副本嗎？當各位網購族在奇摩商城裡瘋狂血拼時，你們知道雅虎在後台分析著你們的指標行為，以便下次特賣時為你們推薦喜歡的新品嗎？以上所有的分析都離不開當今最熱門的主題之一：社會網絡分析。

本研習將會以實際資料向學員完整地介紹這方面的知識，內容包括介紹網絡的基本定義、構造和特性，如何從點與點的關係量化點的位置以及其重要性，如何把點分群來簡化網絡和不同網絡生成的機制。藉由一些經典例子介紹社會網絡分析的用處和特性，例如人際間同質性現象、策略聯盟、資訊的擴散。接著引進一些基本的網絡概念、模型和統計方法，例如網絡結構、中心性、群聚係數、網絡基序和最小世界模型等。

本研習除了導師們課堂知識的分享外，學員們會以有獎比賽的方式來了解社會網絡形成和變化，務求學員在這 2.5 天的課程中，能夠以最輕鬆的互動來了解如何分析這個當今人類不可或缺的網絡世界。

#### ◆生物資訊 (Bioinformatics) (2.5 天)

[袁新盛 副研究員](#) (中央研究院統計科學研究所)

[陳璿宇 助研究員](#) (中央研究院統計科學研究所)

[黃耀廷 副教授](#) (國立中正大學資訊工程學系)

劉佳鑫 先生 (Taiwan International Graduate Program 博士候選人)

[謝叔蓉 研究員](#) (中央研究院統計科學研究所)

隨著分子生物技術的快速發展，新穎的生物技術產生的大量數據，衍伸出複雜的資料分析問題，生物資訊學則為解決此複雜大量數據所衍生出的方法。在本課程中將邀請各領域專家介紹不同分子層次包含基因體、轉錄體以及蛋白質體的資料分析方法，以及其相關應用，期望經由修習本課程後，能對生物資訊學有初步的認識。

#### ◆資料科學 (Data Science) (2.5 天)

[王偉仲 教授](#) (國立臺灣大學數學系)

[林宗男 教授](#) (國立臺灣大學電機工程學系)

[陳孟彰 研究員](#) (中央研究院資訊科學研究所)

[莊永裕 教授](#) (國立臺灣大學資訊工程學系)

[張源俊 研究員](#) (中央研究院統計科學研究所)

隨著資訊與生醫等科技產業的蓬勃發展，各種形態的資料正快速地產生與累積，如何了解各種大量資料特性，並釐清數據間錯綜複雜的關聯與因果，以開發新的資料分析方法，發掘資料中之價值與知識，並為相關專業領域及決策單位提供決策依據，這是資料科學的發展重點。

資料科學是一門跨領域的學問，基本上由三大面向所構成：資訊科學、數學與統計學以及領域知識。本課程將由以上臺灣大學與中央研究院在相關資料科學領域的教學與研究同仁帶領學員們一窺資料科學之究竟。

#### ◆生物與醫學統計 (Biostatistics and Medical Statistics) (2.5 天)

課程主軸：統計科學與營養、藥物、毒品、基因體研究的對話 (A crosstalk of statistics with nutrition, medicine, drug, and genomics)

[王乃昕 教授](#) (美國密西根大學統計學系)

[沈志陽 研究員](#) (中央研究院生物醫學科學研究所)

[潘文涵 研究員](#) (中央研究院生物醫學科學研究所)

[楊欣洲 副研究員](#) (中央研究院統計科學研究所)

[劉玉麗 副研究員](#) (國家衛生研究院神經及精神醫學研究中心)

健康好生活，營養要均衡，藥物莫亂用，毒品不相逢。在 2.5 天的課程中，傑出講員們將透過深入淺出的方式，帶領學員們進入「營養」、「藥物」、「毒品」研究領域，並瞭解其與「基因體」生物科技的連結。大家將體驗「統計」科學在這場生活與科技的對話中所扮演的重要角色，並習得從事這類型研究必備的統計基礎、分析方法與程式操作。

## 中央研究院統計科學研究所 2016 年統計研習營 推薦函

申請人姓名 \_\_\_\_\_

推薦人姓名 \_\_\_\_\_

與申請人之關係

任課老師（科目 \_\_\_\_\_，授課年度 \_\_\_\_\_）

其它： \_\_\_\_\_

評鑑申請人：						
	傑出 (前 5%)	優秀 (前 15%)	佳 (前 30%)	尚可 (前 50%)	差	不清楚
學習能力						
分析能力						
表達能力						
獨立思考能力						
從事研究之潛力						
評語：						

推薦人簽章 \_\_\_\_\_

日期 \_\_\_\_\_

任職單位 \_\_\_\_\_

職稱 \_\_\_\_\_

請於 2016 年 6 月 5 日 前將推薦函逕寄中研院統計所承辦人林晏如小姐

[school@stat.sinica.edu.tw](mailto:school@stat.sinica.edu.tw)

郵件主旨及推薦函檔名請設定為：**統計研習營推薦函-申請人姓名**。