

應用非劣性檢定評估非同步遠距教學的學習成效

盧宏益¹ 謝鑫能²

摘要

因應 COVID-19 的疫情，為讓學生不致受其影響而中斷學習，教師教學方式從實體教學調整為遠距教學模式。評估兩種不同教學方式的學生學習成效常透過差異性檢定進行評估，但若無顯著差異時，僅能說明兩種教學方式的學生學習成效是無差異。臨床醫學中的非劣性檢定是以新療法不劣於標準療法之概念進行評估，本研究應用臨床醫學中的非劣性檢定，以「數理統計」課程為例，利用廣義信賴區間之 95% 區間下界的方法，進行非同步遠距教學的學生學習表現相較於現場教學的學生學習表現之非劣性檢定，探究非同步遠距教學於數理科目的教學成效。研究結果發現，非同步遠距教學的學生平均成績與現場教學的學生平均成績雖未達顯著差異，但執行非劣性檢定顯示非同步遠距教學的學生學習表現並未低於現場教學的學生學習表現，探究原因可能是非同步遠距教學的學生可藉教學影片重複學習增強其數理推導能力，研究結果可做為教師進行遠距教學設計時的重要參考。

關鍵詞：非劣性檢定、非同步遠距教學、廣義信賴區間

1. 盧宏益，輔仁大學統計資訊學系教授

2. 謝鑫能，輔仁大學統計資訊學系助理教授

收件日期：2022.02.11；完成修改：2022.05.04；正式接受：2022.10.03

通訊作者：謝鑫能；Email：d89621202@ntu.edu.tw

地址：242062 新北市新莊區中正路 510 號 輔仁大學統計資訊學系

Applying Non-inferiority Test to Evaluate Student Learning Effectiveness of Asynchronous Distance Learning

Hung-Yi Lu¹ Hsin-Neng Hsieh²

Abstract

Due to the impact of the COVID-19 epidemic, teachers' teaching methods and student learning behavior models have from traditional teaching adjusted to a remote teaching model based on e-learning. Evaluating the learning effectiveness of students with two different teaching methods is often assessed through the difference test. But if there is no significant difference, it can only illustrate that there is no difference in the learning effectiveness of students in the two different teaching methods. In this study, we try to use the lower bound of the 95% generalized confidence interval to show that the learning effectiveness of students for digital teaching may not be better than traditional teaching methods, but at least non-inferior to the traditional teaching methods. The research results show that the learning performance of the digital teaching method is at least as good as the learning performance of the traditional teaching method, and it is not worse than the learning performance of the traditional teaching method. To explore the reasons of the research results, under the asynchronous distance teaching, students not only eliminate the limitations of learning time and space, but can watch teaching videos repeatedly according to their personal learning speed. When COVID-19 is still pandemic, we propose a non-inferiority test to evaluate students' learning effectiveness in digital teaching methods, which should have its application value.

Keywords: asynchronous distance learning, generalized confidence interval, non-inferiority test

1. Hung-Yi Lu, Professor, Department of Statistics and Information Science, Fu Jen Catholic University

2. Hsin-Neng Hsieh, Assistant Professor, Department of Statistics and Information Science, Fu Jen Catholic University

Received: 2022.02.11; Revised: 2022.05.04; Accepted: 2022.10.03

Corresponding Author: Hsin-Neng Hsieh; Email: d89621202@ntu.edu.tw

Address: No. 510, Zhongzheng Rd., Xinzhuang Dist., New Taipei City 242062, Taiwan
Department of Statistics and Information Science, Fu Jen Catholic University

壹、前言

從 2020 年起，全球受到 COVID-19（新冠肺炎）的疫情衝擊，教學場域面臨空前挑戰，維持正常教學運作並確保教學品質實為一重要且艱難的議題。為達停課不停學，讓學習不中斷，許多學校教師的教學方式與學生學習行為模式從傳統現場教學調整成遠距數位教學模式，透過遠距教學的安心就學方式，以因應疫情下的教學與學習（陳昭珍等，2021；陳瑜芬，2021）。相較於傳統現場教學，遠距教學突破時間與空間限制，讓學生可隨時透過網路進行線上學習。教師將不同屬性科目的學習概念統整於數位教學中，把關聯知識做連結並提供學習平臺，當學生自覺學習不足或成效不佳時，即可透過數位學習教材，藉由重複學習，以達適性化教學及補救學習之目的（盧宏益，2020）。

然而，某些課程設計並非採用遠距教學設計模式，雖因疫情影響而被迫改採遠距教學，但學生在遠距教學方式下的學習成效，是否至少能達到與傳統實體教學方式下的學習成效或是更優於傳統實體教學方式，則是值得深入探究（董力華，2021）。Piccoli 等（2001）的研究中對學習成效的定義，是指教學者在教學完成後評價學習者在知識、技能及態度上的改變。何希慧等（2010）則定義學習成效亦是指學習者在接受教育後，其知識、技能、態度及品格上的成長與變化。在高等教育中，學習成效可作為發展成績評比的依據，而無論是實體教學或是遠距教學，學習成果是最常被用來定義學習成效的指標，而學習成果一般包括測驗、期中、期末考試及學期成績（Boghikian-Whitby & Mortagy, 2008）。Noesgaard 與 Ørngreen（2015）的研究中說明了採用學習成果來定義學習成效是最具有有效性的方式，其研究亦指出教學者可以透過量化學習成果來達成其設定的學習目標，並藉此來評價學習成效的有效性。而評估知識領域的學習成果最常使用之指標則是學生的學業平均成績（Hiltz & Wellman, 1997），因而本研究以知識領域學習成果中的學業平均成績，作為評估採實體面授教學及遠距教學等兩種教學方式下的學生學習成效指標。

受 COVID-19 疫情的肆虐，教學場域中教師所採用的教學方式從傳統面授（face to face teaching）教學，調整成以數位教學為主體的遠距教學（distance teaching）模式，以因應疫情影響而實施停課不停學之措施。本研究以

北部某私立綜合大學管理學院 109 學年度第一學期的「數理統計」課程為例，研究時點（2020 年底）雖屬防疫期間，然本土疫情尚在可控制中，未受疫情影響而調整成遠距教學模式，該校為執行防疫超前佈署，鼓勵教師先行實施遠距教學。本研究欲探究防疫期間同一課程採實體面授教學及非同步遠距教學等兩種教學方式下的學生學習表現，多數研究是以獨立樣本 t 檢定評估兩種教學方式的學生學習成效。舉例來說，楊建民（2010）探討採探究式教學法與講述式教學法等兩種教學法，進行國小五年級程式設計課程學習者的學習成效，透過獨立樣本 t 檢定對學習者測驗成績進行差異分析；研究結果顯示，兩種教學法對學習者的學習成效無顯著差異，而再進一步進行雙因數變異數分析（Two-Way Analysis of Variance, Two-Way ANOVA），發現不同數學學習成就者在不同教學法的學習成效則具顯著差異，且亦發現性別於不同教學法下的學習者學習成效達顯著差異。鄭韻如與高毓秀（2012）研究課室教學與數位教學方式於癌症照護課程的學習者學習表現，利用獨立樣本 t 檢定及共變數分析（Analysis of Covariance, ANCOVA）對學習者的學業成績進行分析，以探討不同教學方式的學習者學習成效；研究結果顯示，在癌症照護知識方面，採數位教學方式的學習者之學習成效較佳。羅芳吟與陳政煥（2021）探討因 COVID-19 疫情衝擊，利用成對樣本 t 檢定針對同時參與同步與非同步遠距教學的通識英語課程學習者於前後學習成就表現進行差異分析；研究結果發現，參與遠距教學通識英語課程學習者的學習成就表現與其參與遠距教學前的學習成就表現達顯著差異，該研究顯見疫情下的通識英語課程進行遠距教學，學習者的學業成就表現能達到其在實體課程時之學習成就表現。然而，欲進行疫情期間採遠距教學與實體教學的學生學習表現之獨立樣本 t 檢定，若檢定結果呈現未達統計上顯著差異，研究樣本往往僅能推論疫情期間採遠距教學與實體教學兩種教學方式下的學生學習成效並無統計上的顯著差異，這可能表示疫情期間採用遠距教學與實體教學兩種教學方式下的學生學習表現一樣好，抑或兩種教學方式的學生學習表現都不甚理想。遂此，在 COVID-19 疫情嚴峻的衝擊當下，本研究引用臨床醫學試驗中探討學名藥療效與安全性不劣於專利藥的非劣性檢定（non-inferiority test）評估角度，將其應用至教學場域中進行疫情期間採遠距教學模式與實體面授教學兩種教學方法下的學生學習表現之非劣性檢定，以評估疫情期間採遠距教學方式下的學生學習表現雖不一定能顯著優於實體面授教學下的學生學習表現，但透過非劣性檢定能夠說明疫情期間採遠距教學的學生學習表現不劣於

實體面授教學的學生學習表現，當然若反之則一樣可以檢驗之。Xiao 等（2017）應用非劣性檢定評估中國大陸全國性英語聽力測驗，使用影片與聲音混合方式進行英語聽力測驗的可信度與純粹使用聲音進行的英語聽力測驗之可信度；其研究結果發現，在英語聽力測驗中使用影片與聲音混合方式進行英語聽力測驗的可信度，不劣於純粹使用聲音所進行的英語聽力測驗之可信度。

過去已有諸多學者探討在不同學科採取非同步遠距教學方式的學生學習成效，例如：陳思光等（1997）對北部某大學醫學院牙醫系的牙科放射線學臨床實習課進行非同步遠距數位輔助教學，探究採非同步遠距教學下的學生修習牙科放射線學臨床實習課之學習成效；翁榮源與林宏樺（2013）以中部某大學的生活化學課程進行各種非同步遠距教學，探究採非同步遠距教學下的學生修習生活化學課程之學習成效；溫嘉榮與張建原（2014）以南部某科技大學的非同步遠距教學平臺，利用非同步遠距教學教授計算機概論課程，研析學生的學習成效等相關文獻。而探究數學科目採非同步遠距教學下的學生學習成效相關文獻則相對較少，又因國內 COVID-19 疫情初期的超前佈署下，因而催生探究數學科目課程運用遠距教學方式中的非同步遠距教學之學習者學習成效。本研究以北部某私立綜合大學管理學院 109 學年度第一學期的「數理統計」課程為例，利用廣義信賴區間（generalized confidence interval, GCI）95%區間下界（GCI 法）進行非劣性檢定，評估疫情期間教學者採用非同步遠距教學（asynchronous distance education）的教學模式時，學習者的學習成效（learning effectiveness）並不劣於實體面授教學模式下的學習者學習表現。

本研究共分為六個部分：第一部分為前言，說明研究背景及目的。第二部分針對研究相關主題進行文獻探討，說明廣義樞紐量（generalized pivotal quantity, GPQ）的定義、廣義信賴區間的構建方式，以及進行非劣性檢定的意義。第三部分為研究方法，透過 GPQ 構建疫情期間採非同步遠距教學與實體面授教學模式下的學習者學習成效差異 GCI 之 95% 下界（GCI 法），建立非劣性檢定，並說明其檢定程序。第四部分則是透過統計模擬研究，探討 GCI 法與 t 檢定統計量構建疫情期間採非同步遠距教學與實體面授教學模式下的學習者學習成效差異 95% 信賴區間下界（TCI 法）之型 I 誤差比率（type I error rate）及經驗檢定力（empirical power）等表現情形。第五部分以北部某私立綜合大學管理學院 109 學年度第一學期開設的「數理統計」課程作為

實例，於疫情初期超前佈署，探究學習者在採非同步遠距教學與實體面授教學模式下的學習成效，利用 GCI 法與 TCI 法進行的非劣性檢定並說明其研究結果。最後，則是針對本研究進行總結及其研究限制。

貳、文獻探討

一、非同步遠距教學

「遠距教學」是一種新型態的教學模式，有別於傳統面授教學模式，遠距教學能不受時間與空間限制，結合資訊科技，教學者能藉由網路與電腦資訊科技，將教材傳遞給學習者，並與學習者進行即時或非即時的雙向互動教學方式（沈中偉，1998；鄒景平，2005）。非同步遠距教學屬於遠距教學方式之一，內涵為授課者與學生間無法直接對話討論，但可透過網頁平臺中的討論區功能進行互動。而實務上即使是現場面授教學，雖然授課者可直接與修課學生進行討論，但因課堂時間限制，也無法全然讓所有學生在課堂內發表意見。因此，透過網頁平臺的討論區功能進行討論，可以讓更多學生進行互動及討論（楊奕農、柴蕙質，2002）。

過去已有學者針對非同步遠距教學與傳統教學進行學生學習成效的比較研究，然而不同研究所採用的學習成效評估指標與分析方法則有所差異。舉例來說，Allgood 與 Walstad（1999）及 Chizmar 與 Walbert（1999）利用測驗成績作為學習成效指標，以平均數 t 檢定評估非同步遠距教學與傳統教學方式下的學生測驗成績之差異分析；Gregor 與 Cuskelly（1994）、田耐青與洪明洲（1998），以及李忠謀與邱瓊芳（1999）透過學習滿意度問卷調查作為非同步遠距教學的學習成效指標，並利用敘述統計方法進行評估；Webster 與 Hackley（1997）亦是透過學習滿意度問卷調查作為學習成效指標，但他們使用多元迴歸分析法進行非同步遠距教學與傳統教學方式下的學生學習成效評估；Davies 與 Berrow（1998）、Motiwalla 與 Tello（2000），以及 Chou（2001）透過學生學習滿意度及學生測驗成績的兩種學習成效指標，衡量非同步遠距教學與傳統教學方式的學生學習成效，Motiwalla 與 Tello 使用敘述統計法、Chou 利用平均數 t 檢定法，而 Davies 與 Berrow 則是引用相關係數進行非同步遠距教學與傳統教學方式的學生學習成效之評估。此外，楊奕農與柴蕙質（2002）在研究中假設影響學習成效的各類因素與學習成效間呈現

線性關係，利用多元迴歸模型探討影響非同步遠距學習成效的可能因素；劉通國（2002）針對學習風格與電腦自我效能對非同步遠距學習成效及滿意度的研究，研究結果提出非同步遠距教學下的學生學習風格與學習成效並無顯著差異。黃秀仍（2007）探討進修部學生對非同步遠距教學的看法，研究發現進修部學生對非同步遠距教學的需要度相當高，推測原因可能是可以彈性學習；此外，研究中亦發現，學生非常重視非同步遠距教學中的電子化教材品質，因此授課時間安排及電子化教材製作則是採非同步遠距教學下的學生較重視之項目。

2020 年，因著 COVID-19 的疫情影響，為維持學習不中斷，諸多教師從傳統實體教學調整成遠距數位教學模式。羅芳吟與陳政煥（2021）針對中部某私立綜合大學採同步與非同步遠距教學的大一通識英語課程學生為例，利用成對樣本 t 檢定探究學習者參與遠距教學前後學習成就表現的差異分析；研究結果發現，學習者參與遠距教學通識英語課程的學習成就表現顯著較參與遠距教學前的學習成就表現進步，該研究顯示疫情下的通識英語課程進行遠距教學，學習者的學業成就表現能達到其在實體課程時的學習成就表現。馬國華等（2021）於疫情期間探討南部某私立大學醫學院物理治療系在實習課中運用雲端教學影片能否提升物理治療實習生的學習成效，雲端教學影片內容包括實習評估、治療及衛教等部分，實習科別分別在骨科、神經科、兒童科等科別與各科臨床實務進行雲端教學影片教學；研究結果發現，物理治療實習生在臨床演練評量（mini-clinical evaluation exercise, Mini-CEX）及直接觀察臨床紀錄評估（direct observation of procedural skills, DOPS）的後測實習表現顯著較其前測實習表現佳。

二、非劣性檢定

在醫學倫理考量下，對於如腫瘤會危及生命的疾病之臨床試驗來說，如果市面上已有行之多年、具有療效（treatment effect）且安全（safety）的標準藥（即活性對照藥），則單純的安慰劑對照試驗是不太可能被採用進行臨床試驗。實務上，除非在腫瘤臨床試驗中的新藥療效確實優於標準藥療效，或者使用標準藥治療再加上新藥的合併療效優於單純使用標準藥療效，而如果新藥療效優於標準藥的可能性較小，但是新藥保留了標準藥大部分的療效，且新藥比標準藥更具有安全性、耐受性、方便性等方面的優勢，此時在臨床試驗中常會使用非劣性檢定進行這類新藥療效的評估。

在醫學領域中，非劣性檢定主要在評估研發新藥與標準藥（已上市藥）的相對療效（吳雅琪，2012）。近年來，學者在非劣性檢定研究如當黃金標準試驗（gold standard test）存在時，Li 等（2008a）以接受者操作特徵（receiver operating characteristics, ROC）曲線下的部分面積（partial area under the ROC curve, PAUROC）作為評估新診斷方法的準確性指標，採用廣義檢定變數（generalized test variable, GTV）構建廣義 p 值（generalized p value, GPV），進行成對部分 ROC 曲線下面積差的非劣性檢定。模擬結果顯示，GPV 不僅適切地控制非劣性檢定的檢定大小（size），且其檢定力（power）優於無母數方法（nonparametric method）及最大概似法（maximum likelihood method）。而當黃金標準試驗不存在時，Shih 等（2016）利用 EM 演算法（expectation-maximization algorithm）結合拔靴法（bootstrap method），進行成對部分 ROC 曲線下面積差的非劣性檢定，經統計模擬研究發現所提方法的檢定大小能充分控制在宣稱的顯著水準（significance level）之外，同時檢定力表現與黃金標準試驗存在時一樣好；范又方（2019）針對成對二元資料，在成對設計中利用強韌概似函數方法（robust likelihood function method）建立強韌分數檢定統計量（robust score test statistic），進行等效性（equivalence）與非劣性檢定，以評估新藥療效或醫療診斷醫學方法是否與標準藥或現行診斷方法具相同療效或是不劣於標準方法，透過統計模擬研究發現，強韌分數檢定統計量所構建等效性或非劣性檢定的檢定大小接近給定的顯著水準，其檢定力亦有不錯表現；Hsieh 與 Lu（2020）探究新藥能保持標準藥多大比例的療效問題，利用廣義 p 值法（GPV-based 法）構建新藥維持標準藥療效比例的非劣性檢定，透過大規模統計模擬研究，驗證 GPV-based 法能控制檢定大小接近宣稱的顯著水準且亦具高檢定力表現等。

然而，非劣性檢定應用在教育領域中的研究相對較少，李秋璿（2010）探討中部某醫學大學護理學系的精神科暴力處置課程，在研究中先以獨立樣本 t 檢定分析網路教學與傳統教學方式下的學生知識面學習表現之差異， t 檢定結果顯示兩種不同教學方式下的學生知識面向之學習表現並無顯著差異。進一步以 t 檢定統計量進行兩種不同教學方式下的學生知識面學習表現之非劣性檢定，研究發現採網路教學方式的學生知識學習表現並不劣於傳統教學方式的學生學習表現。Xiao 等（2017）在中國大陸全國性的英語聽力測驗中，應用非劣性檢定評估使用影片與聲音混合方式進行英語聽力測驗的可信度與純粹使用聲音所進行英語聽力測驗的可信度，研究結果發現在英語聽力

測驗中，使用影片與聲音混合方式進行英語聽力測驗的可信度不劣於純粹使用聲音所進行英語聽力測驗的可信度。

三、廣義信賴區間

參數 (parameter) 的區間估計是以統計量之抽樣分配來進行構建。當統計量的抽樣分配太過複雜且存在干擾參數 (nuisance parameter) 時，較不易推導出參數的確切信賴區間 (exact confidence interval)。為克服此問題，Weerahandi (1993) 定義了 GPQ，透過 GPQ 來構建出感興趣參數的 GCI，因此 GCI 是構建於包含干擾參數的確切信賴區間。

假設 $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ 的機率分配為 $f(\mathbf{x}; \xi)$ ， $\xi = (\theta, \boldsymbol{\eta})$ 為未知參數向量，其中 θ 為感興趣參數， $\boldsymbol{\eta}$ 為干擾參數。令 $\mathbf{x} = (x_1, \dots, x_n)$ 為 \mathbf{X} 的樣本觀察值。若 $R = R(\mathbf{X}; \mathbf{x}, \xi)$ 滿足下列二個性質：

性質 A：R 的機率分配與任何未知參數無關。

性質 B： $r_{obs} = R(\mathbf{x}; \mathbf{x}, \xi)$ 與干擾參數無關，即 r_{obs} 是 \mathbf{x}, θ 的函數。

則稱 $R = R(\mathbf{X}; \mathbf{x}, \xi)$ 為 θ 的 GPQ。令 r_1, r_2 使得 (1) 式滿足：

$$P(r_1 \leq R(\mathbf{X}; \mathbf{x}, \xi) \leq r_2) = 1 - \alpha \quad (1)$$

透過蒙地卡羅演算法 (Monte-Carlo algorithm)，求得 $r_1 = R_{\alpha/2}$ ，為 R 的第 $100 \times \alpha/2$ 百分位數； $r_2 = R_{1-\alpha/2}$ ，為 R 的第 $100 \times (1 - \alpha/2)$ 百分位數，因此 $(R_{\alpha/2}, R_{1-\alpha/2})$ 為 θ 之 $100 \times (1 - \alpha)\%$ 雙尾的 GCI。

近年來，已有許多學者將廣義推論應用在各領域中，例如：Lin 與 Liao (2006) 探討在一般平衡線性隨機模型 (general balanced linear random models) 中利用 GPQ 構建廣義容許區間 (tolerance interval)，經統計模擬研究發現，以 GPQ 為基準所構建的容許區間具有較佳的覆蓋率 (coverage probability) 及較短的區間期望長度 (expected length)。Lin 與 Liao (2008) 在平衡混合線性模型 (balanced mixed linear models) 下提出利用 GPQ 來構建預測區間 (prediction interval)，透過模擬研究結果發現，由 GPQ 所構建的預測區間，其覆蓋率接近研究設定的信賴水準 (confidence level)，其區間期望長度亦較短；當黃金標準試驗存在時，Li 等 (2008b) 以 ROC 曲線下面積 (area under the ROC curve, AUROC) 作為醫學診斷工具的評估指標，透過 GPQ 構建成對 ROC 曲線下面積差的 GCI。由統計模擬研究發現，相較無母數

法及最大概似法所構建的成對 ROC 曲線下面積差之區間估計，以 GPQ 為基準構建的成對 ROC 曲線下面積差之 GCI 具較佳覆蓋率及較短的區間期望長度。Chang 等（2013）探討當黃金標準試驗不存在時，提出結合 EM 演算法、拔靴法及 GPQ 的啟發式方法（heuristic method），構建成對 ROC 曲線下面積差的 GCI；透過統計模擬研究發現，啟發式方法構建的 GCI，其覆蓋率接近宣稱的信賴水準並具較短的區間期望長度。Hsieh 等（2009）在體外溶離率試驗中，利用相似因子（similarity factor）進行變化前後的藥品相似性評估，因為相似因子的機率分配複雜，因而進行溶離率剖析曲線相似性（similarity of dissolution profiles）檢定就相對困難；他們利用 GPV 法進行溶離率剖析曲線相似性評估，在不同溶解度平均數差、變異數及藥品樣本數等參數組合進行統計模擬研究，以探討 GPV 法與拔靴法在偏誤、檢定大小與檢定力等性質的比較；模擬研究結果發現，隨著藥品樣本數增加，利用 GPV 法與拔靴法評估溶離率剖析曲線的相似性表現一樣好。Tsai（2010）研究線性測量誤差模型（linear measurement error model）中斜率參數的區間估計，利用 GPQ 進行線性測量誤差模型中斜率參數的 GCI 構建；透過蒙地卡羅模擬（Monte Carlo simulation）研究發現，由 GPQ 所構建的測量誤差模型中斜率參數的 GCI，其覆蓋率接近宣稱的信賴水準，亦具較短的區間期望長度。Wu 與 Hsieh（2014）針對具高度偏態的漁業研究資料，利用 GPQ 構建出 delta 對數常態分配（delta log-normal distribution）的平均數之 GCI，模擬研究結果顯示以 GPQ 構建的 GCI 不但提供足夠的覆蓋率，而且提供合理的相對偏誤（relative bias）等。

上述所提及的前人研究，大部分是將廣義推論應用在生物統計或工業統計領域，Yu 與 Tsai（2012）則是將 GCI 應用在教育領域的研究中。Yu 與 Tsai 在線性迴歸模型中利用 GPQ 構建中介效果（mediated effects）的 GCI，經統計模擬研究發現，在小樣本及非常態樣本時，中介效果的 GCI 表現更具穩健性。

參、研究方法

假設 X_E, X_C 分別表示採用非同步遠距教學法〔實驗組（E）〕和現場教學法〔控制組（C）〕學生的成績。假設在兩種教學法下修讀學生成績相互獨立且分別服從平均數為 μ_E, μ_C 及變異數為 σ_E^2, σ_C^2 的常態分配， n_E, n_C 分別表

示實驗組及控制組的修讀學生數，兩組學生數並不需相等。不失一般性，除了教學法外，控制其他影響實驗組（E）和控制組（C）學習成效的因素，探究兩種不同教學法下的學生成績表現，其成績愈高代表該教學方法對學習表現愈佳。欲比較 COVID-19 的疫情衝擊下，採非同步遠距教學法的學生學習表現至少不劣於現場教學的學生學習表現，同時亦可防止疫情在教室內發生群聚感染，因此本研究透過構建 $\mu_E - \mu_C$ 之 GCI 下界進行（2）式的非劣性檢定。

$$H_0 : \mu_E - \mu_C \leq -M \quad \text{V.S.} \quad H_1 : \mu_E - \mu_C > -M \quad (M > 0) \quad (2)$$

其中， $-M$ 為非劣性檢定臨界值（margin）。

假設 X_1, \dots, X_{n_E} 與 Y_1, \dots, Y_{n_C} 為兩組來自 $N(\mu_E, \sigma_E^2)$ 和 $N(\mu_C, \sigma_C^2)$ 獨立的隨機樣本， \bar{X} 與 \bar{Y} 分別為 μ_E 和 μ_C 的充分統計量（sufficient statistic）， S_E^2 與 S_C^2 分別是 σ_E^2 和 σ_C^2 的充分統計量，而 \bar{x} 與 \bar{y} 分別為 \bar{X} 與 \bar{Y} 的觀測值， s_E^2 與 s_C^2 分別為 S_E^2 與 S_C^2 的觀測值。

Case 1： σ_E^2 與 σ_C^2 未知且不相等

$$\text{令} \quad Z_1 = \frac{(\bar{X} - \bar{Y}) - (\mu_E - \mu_C)}{\sqrt{\text{Var}(\bar{X} - \bar{Y})}} = \frac{(\bar{X} - \bar{Y}) - (\mu_E - \mu_C)}{\sqrt{\frac{\sigma_E^2}{n_E} + \frac{\sigma_C^2}{n_C}}}, \quad U_1 = \frac{(n_E - 1)S_E^2}{\sigma_E^2},$$

$U_2 = \frac{(n_C - 1)S_C^2}{\sigma_C^2}$ ，則 $\mu_E - \mu_C$ 之 GPQ 定義為：

$$R_{\mu_E - \mu_C} = (\bar{x} - \bar{y}) - Z_1 \sqrt{\frac{(n_E - 1)s_E^2}{n_E U_1} + \frac{(n_C - 1)s_C^2}{n_C U_2}} \quad (3)$$

$$= (\bar{x} - \bar{y}) - \frac{(\bar{X} - \bar{Y}) - (\mu_E - \mu_C)}{\sqrt{\frac{\sigma_E^2}{n_E} + \frac{\sigma_C^2}{n_C}}} \sqrt{\frac{\sigma_E^2 s_E^2}{n_E S_E^2} + \frac{\sigma_C^2 s_C^2}{n_C S_C^2}} \quad (4)$$

其中， Z_1 為標準常態隨機變數， U_1 與 U_2 分別為服從自由度為 $n_E - 1$ 和 $n_C - 1$ 的卡方隨機變數，且 Z_1 、 U_1 與 U_2 相互獨立。由（3）式可知， $R_{\mu_E - \mu_C}$ 的機率分配與未知參數 $\mu_E - \mu_C$ 無關。將樣本平均數觀測值 \bar{x} 及 \bar{y} 與樣本變異數觀測值 s_E^2 及 s_C^2 分別代入 \bar{X} 、 \bar{Y} 、 S_E^2 與 S_C^2 後，可將（4）式還原成有興趣的參數 $\mu_E - \mu_C$ 。因此， $R_{\mu_E - \mu_C}$ 符合 GPQ 定義中的性質 A 及性質 B，因此 $R_{\mu_E - \mu_C}$

為 $\mu_E - \mu_C$ 的 GPQ。

Case 2： σ_E^2 與 σ_C^2 未知且相等為 σ^2

$$\text{令 } Z_2 = \frac{(\bar{X} - \bar{Y}) - (\mu_E - \mu_C)}{\sqrt{\text{Var}(\bar{X} - \bar{Y})}} = \frac{(\bar{X} - \bar{Y}) - (\mu_E - \mu_C)}{\sqrt{\frac{\sigma^2}{n_E} + \frac{\sigma^2}{n_C}}}, U_3 = \frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{\sigma^2},$$

則 $\mu_E - \mu_C$ 之 GPQ 定義為：

$$R_{\mu_E - \mu_C} = (\bar{x} - \bar{y}) - Z_2 \sqrt{\frac{(n_E + n_C - 2)S_{pooled}^2}{n_E U_3} + \frac{(n_E + n_C - 2)S_{pooled}^2}{n_C U_3}} \quad (5)$$

$$= (\bar{x} - \bar{y}) - \frac{(\bar{X} - \bar{Y}) - (\mu_E - \mu_C)}{\sqrt{\frac{\sigma^2}{n_E} + \frac{\sigma^2}{n_C}}} \sqrt{\frac{\sigma^2 S_{pooled}^2}{n_E S_{pooled}^2} + \frac{\sigma^2 S_{pooled}^2}{n_C S_{pooled}^2}} \quad (6)$$

其中， S_{pooled}^2 為合併樣本變異數（pooled sample variance），是 σ^2 的充分統計量，而 s_{pooled}^2 為 S_{pooled}^2 的觀測值。 Z_2 為標準常態隨機變數， U_3 是服從自由度為 $n_E + n_C - 2$ 的卡方隨機變數，且 Z_2 與 U_3 相互獨立。由（5）式可知， $R_{\mu_E - \mu_C}$ 的機率分配與未知參數 $\mu_E - \mu_C$ 無關。將樣本平均數觀測值 \bar{x} 及 \bar{y} 與合併樣本變異數觀測值 s_{pooled}^2 分別代入 \bar{X} 、 \bar{Y} 與 S_{pooled}^2 後，（6）式將還原成有興趣參數 $\mu_E - \mu_C$ 。因此， $R_{\mu_E - \mu_C}$ 符合 GPQ 定義中的性質 A 及性質 B，故 $R_{\mu_E - \mu_C}$ 為 $\mu_E - \mu_C$ 的 GPQ。

針對（2）式的非劣性檢定形式，經由蒙地卡羅演算法構建 $\mu_E - \mu_C$ 的 $100 \times (1 - \alpha)\%$ 之 GCI 下界進行非劣性檢定程序步驟如下：

Case 1： σ_E^2 與 σ_C^2 未知且不相等

步驟 1：選擇模擬的次數 H ，對 $1 \leq h \leq H$ ，針對每一次模擬選擇 $M = 10000$ 個蒙地卡羅隨機樣本。對任意的 $1 \leq m \leq M$ ，產生相互獨立的標準常態隨機變數值 $Z_{1,m,h}$ 及自由度分別為 $n_E - 1$ 與 $n_C - 1$ 之 $U_{1,m,h}$ 與 $U_{2,m,h}$ 卡方隨機變數值。

步驟 2：利用（3）式計算 $R_{\mu_E - \mu_C, m, h}$ 。

步驟 3：對第 h 次所計算出來的 m 個 GPQ 經過排序後，得到第 $M \times (\alpha)$ 位置對應的 GPQ， $R_{\mu_E - \mu_C, (l), h}$ ，即為 95%GCI 的區間下界。

步驟 4：計算 $\bar{R}_{\mu_E - \mu_C, (l)} = \frac{1}{H} \sum_{h=1}^H R_{\mu_E - \mu_C, (l), h}$ 。

若 $\bar{R}_{\mu_E - \mu_C, (l)} > -M$ ，則拒絕 (2) 式的 $H_0: \mu_E - \mu_C \leq -M$ ，表示在 COVID-19 的疫情衝擊下，統計上有充分證據說明採取非同步遠距教學法的學生學習表現至少不會較現場教學的學生學習表現差。

Case 2： σ_E^2 與 σ_C^2 未知且相等為 σ^2

步驟 1：選擇模擬的次數 H ，對 $1 \leq h \leq H$ ，針對每一次模擬選擇 $M=10000$ 個蒙地卡羅隨機樣本。對任意的 $1 \leq m \leq M$ ，產生相互獨立的標準常態隨機變數值 $Z_{2,m,h}$ 及自由度分別為 $n_E + n_C - 2$ 的 $U_{3,m,h}$ 卡方隨機變數值。

步驟 2：利用 (5) 式計算 $R_{\mu_E - \mu_C, m, h}$ 。

步驟 3：對第 h 次所計算出來的 m 個 GPQ 經過排序後，得到第 $M \times (\alpha)$ 位置對應的 GPQ， $R_{\mu_E - \mu_C, (l), h}$ ，即為 95%GCI 的區間下界。

步驟 4：計算 $\bar{R}_{\mu_E - \mu_C, (l)} = \frac{1}{H} \sum_{h=1}^H R_{\mu_E - \mu_C, (l), h}$ 。

若 $\bar{R}_{\mu_E - \mu_C, (l)} > -M$ ，則拒絕 (2) 式的 $H_0: \mu_E - \mu_C \leq -M$ ，表示雖面對 COVID-19 的疫情衝擊，統計上具充分證據說明在教學場域中採取非同步遠距教學法下的學生學習表現至少不會比採現場面授教學的學生學習表現差。

肆、統計模擬研究

研究者利用廣義樞紐量所構建廣義信賴區間的下界 (GCI 法) 進行 (2) 式之非劣性檢定，透過蒙地卡羅模擬 GCI 法的型 I 誤差比率及經驗檢定力，並與 t 統計量所構建的 t 信賴區間之下界 (TCI 法) 進行比較。

一、模擬參數設定

假設採用非同步遠距教學法〔實驗組 (E)〕下的學生成績服從平均數為 μ_E 、變異數為 σ_E^2 的常態分配；而採用現場教學法〔控制組 (C)〕的學生成績則是服從平均數為 μ_C 、變異數為 σ_C^2 的常態分配。由於本研究實例中的非同步遠距教學與現場教學之學生，在機率模型單元成績的變異數經 Levene F 檢定是呈現變異數相等的結果 ($F=0.01, p=0.93$)，因此本研究僅考慮 $\sigma_E^2 = \sigma_C^2 = \sigma^2$ 下的統計模擬研究情形，其中 σ^2 表示採兩種教學法下的學生成績

之母體共同變異數； n_E, n_C 分別表示實驗組及控制組的修讀學生數，兩組學生數並不需相等。不失一般性，其成績愈高代表該教學方法對學習表現愈佳。

統計模擬研究參數的設定為：(1)非同步遠距教學法〔實驗組（E）〕與現場教學法〔控制組（C）〕的學生平均成績差異考慮下列三種情形：(i) $\mu_E - \mu_C = 40 - 45$ ；(ii) $\mu_E - \mu_C = 50 - 55$ ；(iii) $\mu_E - \mu_C = 60 - 65$ ；(2)非同步遠距教學法〔實驗組（E）〕與現場教學法〔控制組（C）〕的學生成績變異數均設定為 10（即 $\sigma_E^2 = \sigma_C^2 = \sigma^2 = 10$ ）；(3)採非同步遠距教學法〔實驗組（E）〕與現場教學法〔控制組（C）〕的模擬樣本數 n_E 和 n_C 設定為 20、50、70、100、200 且 n_E 小於等於 n_C 。針對每種參數組合，獨立生成 10000 組樣本資料。從每組所模擬出的樣本資料，透過 t 統計量建立 t 信賴區間的下界。此外，針對每組模擬樣本資料，透過蒙地卡羅演算法產生 10000 個 GPQs，構建廣義信賴區間的下界，分別利用 t 信賴區間與廣義信賴區間的下界進行（2）式之非劣性統計檢定程序。評價 GCI 法與 TCI 法執行（2）式的非劣性檢定是以型 I 誤差比率及經驗檢定力進行評估。型 I 誤差比率是指（2）式的虛無假設（null hypothesis）為真時，所建立 10000 個 GCI 下界與 TCI 下界大於（2）式中 $-M$ 的比例。當型 I 誤差比率愈接近宣稱的顯著水準，表示檢定方法於進行非劣性檢定時可適當地控制型 I 誤差。除進行非劣性檢定時可適當的控制型 I 誤差外，檢定方法的經驗檢定力亦是評估檢定方法之指標。當宣稱的顯著水準固定時，檢定方法的經驗檢定力愈高，表示檢定方法在進行非劣性檢定時的檢定準確度愈高。在不同統計模擬研究參數的組合下，GCI 法與 TCI 法進行（2）式非劣性檢定的型 I 誤差比率模擬結果列於表 1，而檢定力的模擬結果請見表 2 及圖 1。

二、模擬結果分析

利用 GCI 法與 TCI 法針對（2）式進行非劣性檢定，由表 1 可知， $\mu_E - \mu_C$ 在三種不同的情況下，GCI 法與 TCI 法進行非劣性檢定所模擬的型 I 誤差比率大都很接近給定的顯著水準 0.05，且隨著樣本數增加，GCI 法與 TCI 法的型 I 誤差比率愈接近 0.05。此外，模擬型 I 誤差比率均落在 0.0457 至 0.0543 間的容許誤差範圍內，兩種檢定法的模擬型 I 誤差比率差異不大，顯見 GCI 法與 TCI 法進行非劣性檢定均嚴謹且適切控制執行（2）式非劣性檢定時的型 I 誤差。

表 1 GCI 法與 TCI 法進行非劣性檢定的型 I 誤差比率

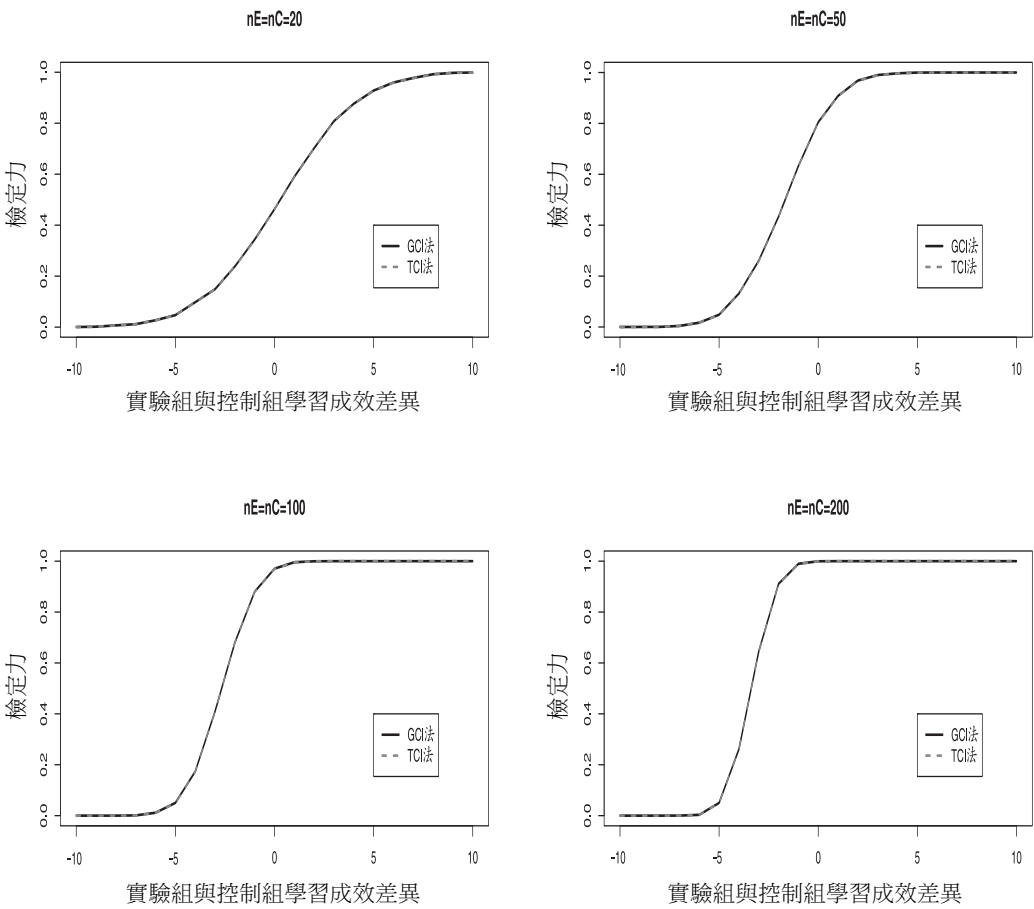
n_E	n_C	$\mu_E - \mu_C$					
		40-45		50-55		60-65	
		GCI 法	TCI 法	GCI 法	TCI 法	GCI 法	TCI 法
20	20	0.0471	0.0472	0.0476	0.0473	0.0483	0.0481
	50	0.0492	0.0485	0.0493	0.0490	0.0491	0.0488
	70	0.0494	0.0490	0.0499	0.0498	0.0501	0.0501
	100	0.0492	0.0489	0.0512	0.0509	0.0509	0.0507
	200	0.0511	0.0517	0.0518	0.0512	0.0514	0.0511
50	50	0.0484	0.0482	0.0480	0.0474	0.0494	0.0484
	70	0.0506	0.0508	0.0481	0.0478	0.0501	0.0502
	100	0.0488	0.0483	0.0502	0.0500	0.0502	0.0505
	200	0.0505	0.0497	0.0511	0.0505	0.0510	0.0507
70	70	0.0485	0.0476	0.0491	0.0492	0.0499	0.0490
	100	0.0493	0.0491	0.0502	0.0502	0.0500	0.0494
	200	0.0505	0.0508	0.0503	0.0498	0.0509	0.0502
100	100	0.0501	0.0494	0.0491	0.0489	0.0492	0.0490
	200	0.0509	0.0507	0.0496	0.0490	0.0502	0.0503
200	200	0.0502	0.0501	0.0503	0.0504	0.0513	0.0512

再者，研究者考慮 $\mu_E - \mu_C = 40 - 45$ 與 $n_E = n_C$ 為 20、50、100 及 200 的情況下，模擬 GCI 法與 TCI 法進行非劣性檢定時的檢定力。由表 2 可知，GCI 法的檢定力（當 $-M > -5$ 時）均略高於 TCI 法，尤其在小樣本時，GCI 法的檢定力表現則是略優於 TCI 法。此外，隨著樣本數增加，GCI 法與 TCI 法的檢定力愈為接近。GCI 法與 TCI 法均能嚴謹且適切控制執行非劣性檢定時的型 I 誤差，在小樣本時，GCI 法的檢定力表現略優於 TCI 法，而在大樣本時，兩種方法進行非劣性檢定的檢定力表現則是一樣好。從圖 1 的檢定力函數圖形亦可知，在不同的樣本數組合下，GCI 法與 TCI 法的檢定力表現差異不大，且隨著樣本數增加，兩種檢定法的檢定力亦呈現增加之趨勢。而在小樣本時，GCI 法的檢定力表現則略優於 TCI 法。

表 2 GCI 法與 TCI 法進行非劣性檢定的經驗檢定力

$-M$	$\mu_E - \mu_C$							
	40-45							
	$n_E = n_C = 20$		$n_E = n_C = 50$		$n_E = n_C = 100$		$n_E = n_C = 200$	
	GCI 法	TCI 法	GCI 法	TCI 法	GCI 法	TCI 法	GCI 法	TCI 法
-10	0.0005	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-9	0.0016	0.0014	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000
-8	0.0068	0.0066	0.0004	0.0004	0.0000	0.0000	0.0000	0.0000
-7	0.0117	0.0110	0.0042	0.0042	0.0010	0.0010	0.0002	0.0002
-6	0.0269	0.0266	0.0170	0.0167	0.0114	0.0110	0.0032	0.0032
-5	0.0471	0.0472	0.0484	0.0482	0.0501	0.0494	0.0502	0.0501
-4	0.0977	0.0969	0.1316	0.1306	0.1732	0.1728	0.2612	0.2609
-3	0.1487	0.1484	0.2607	0.2583	0.4093	0.4086	0.6456	0.6451
-2	0.2378	0.2372	0.4339	0.4335	0.6797	0.6789	0.9108	0.9108
-1	0.3436	0.3418	0.6329	0.6327	0.8799	0.8795	0.9896	0.9892
0	0.4627	0.4622	0.8046	0.8039	0.9704	0.9698	0.9996	0.9996
1	0.5898	0.5884	0.9081	0.9073	0.9958	0.9958	1.0000	1.0000
2	0.7021	0.7020	0.9683	0.9676	0.9995	0.9995	1.0000	1.0000
3	0.8084	0.8078	0.9901	0.9896	1.0000	1.0000	1.0000	1.0000
4	0.8765	0.8755	0.9968	0.9967	1.0000	1.0000	1.0000	1.0000
5	0.9279	0.9275	0.9998	0.9997	1.0000	1.0000	1.0000	1.0000
6	0.9596	0.9595	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0.9775	0.9768	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0.9920	0.9920	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0.9976	0.9976	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0.9989	0.9988	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

圖 1 GCI 法與 TCI 法進行非劣性檢定的檢定力函數圖



伍、實例分析

一、研究對象

本研究的研究對象為修讀北部某私立綜合大學管理學院 109 學年度第一學期的「數理統計」課程之兩個班級學生共 125 人，一個班級學生為實驗組，另一個班級學生為控制組。實驗組學生人數 (n_E) 為 71 人，約占 56.80%；控制組學生人數 (n_C) 為 54 人，約占 43.20%。

二、研究時點

本研究的研究時點為 109 學年度第一學期第四週開始至第十七週止 (2020/10/05~2021/01/08)，雖在防疫期間內，然因彼時國內本土疫情尚在可控制中，該校為執行超前防疫佈署演練，鼓勵教師先行實施遠距教學。研究者利用課程前二至三週分別向實驗組與控制組修課學生做課程教學方式的說明及準備，而在研究時點期間內，實驗組學生均採非同步遠距教學方式進行教學，而控制組學生則採現場面授教學方式進行教學。

三、教學策略

研究中的實驗組班級以 EverCam 軟體進行錄製教學影片，透過教學影片教學法作為學習機制導入，進行非同步遠距教學的教學策略。而控制組班級則採實體現場面授教學，透過傳統面授教學法作為學習機制的導入，進行現場實體教學的教學策略。

四、教學設計

本研究「數理統計」課程的教學設計分成課前、課程進行中、課後，以及測驗評量等部分，詳細說明請參見表 3。

五、教學內容

本研究以數理統計中的機率模型單元 (即第一學期「數理統計」課程內容) 作為實驗內容，教學內容如表 4 所示。

表 3 本研究「數理統計」課程的教學設計

課程進行階段	教學設計	非同步遠距教學	現場教學
課前	閱讀教材	指定教科書及放置在 TronClass 中的講義	指定教科書及講義
課程進行中	課程內容講解	放置在 TronClass 中的教學影片	電子化講桌、投影機、黑板
	師生互動討論	TronClass 的討論與互動功能	實體師生互動
	出缺席情形	TronClass 觀看教學影片的紀錄功能	實體課堂點名
課後	課程內容複習	教學影片、教科書及筆記	教科書及筆記
	助教實習課	放置在 TronClass 中的助教實習課影片	電子化講桌、投影機、黑板
	練習題演練	放置在 TronClass 中的練習題	教科書習題演練
	學習叮嚀	TronClass 公告功能	TronClass 公告功能
測驗評量	考試測驗	實體考試	實體考試

註：TronClass 為臺灣智園有限公司所建置的行動學習與翻轉課堂平臺。

表 4 109 學年度第一學期「數理統計」課程內容

週次	主題	
1	Introduction	
2	Probability(1)	
3	Probability(2)	
4	Discrete random variables and their probability distributions (1)	
5	Discrete random variables and their probability distributions (2)	
6	Discrete random variables and their probability distributions (3)	
7	Continuous random variables and their probability distributions (1)	
8	Continuous random variables and their probability distributions (2)	
9	Mid-term Exam	
10	Continuous random variables and their probability distributions (3)	
11	Multivariate probability distribution (1)	機 率 模 型 單 元
12	Multivariate probability distribution (2)	
13	Multivariate probability distribution (3)	
14	Functions of random variables (1)	
15	Functions of random variables (2)	
16	Functions of random variables (3)	
17	Functions of random variables (4)	
18	Final Exam	

六、研究變項

本研究的研究目的是透過 GCI 法進行非劣性檢定，以評估教學者採用非同步遠距教學的教學模式時，學習者的學習成效並不會較面授教學模式下的學習者學習表現差。

本研究的自變項為教學方法，實驗組的教學方法為非同步遠距教學（即教學影片教學法），控制組的教學方法為現場面授教學（即傳統面授教學法），並在第三週舉行統計學基礎測驗作為本研究的前測，前測結果呈現實驗組與控制組學生在基礎統計學測驗的平均成績無統計顯著差異（ p 值約為 0.84），因此兩組學生在修讀「數理統計」課程前所具備的統計學基礎能力具有一致性。本研究的依變項為修習機率模型單元（109 學年度第一學期「數理統計」課程內容）後的學生修課平均成績，分別利用廣義區間估計法（GCI 法）及 t 區間估計法（TCI 法）進行評估執行非劣性統計檢定程序。研究的控制變項整理如表 5 所示。

表 5 研究的控制變項

控制變項	實驗組 (E)	控制組 (C)
授課教師	由同一位教師擔任「數理統計」課程全部的教學	
實驗學生	北部某私立綜合大學管理學院 109 學年度第一學期修讀「數理統計」課程學生，人數為 71 人	北部某私立綜合大學管理學院 109 學年度第一學期修讀「數理統計」課程學生，人數為 54 人
教學時間	每週三上午 9:00~12:00	每週二上午 9:00~12:00
教學內容	數理統計中的機率模型單元 (即第一學期「數理統計」課程內容)	
測驗評量	兩組另在共同時間（星期三下午 1:40~3:30）舉行考試測驗	

七、試驗效度

（一）內在效度

本研究控制同一位授課教師、相同教材（教科書及講義）、相同教學內容、相同教學時間（一週正課均為三節課以及一節助教實習課）、同一考試

測驗時間，以及同一份考題測驗題目等因素影響，在研究中有比較的控制組，以避免統計迴歸現象。

（二）外在效度

本研究以學生修課班級作為分組試驗，不再另行重新分組。在進行實例的試驗過程中，受試者較不容易自覺正在接受試驗研究的感受，較易控制試驗安排所可能產生的反作用效果。

此外，考量不同參數組合下執行統計模擬研究呈現，GCI 法與 TCI 法的型 I 誤差比率均落在 0.0457 至 0.0543 之容許誤差範圍內，利用 GCI 法與 TCI 法進行非劣性檢定時均能嚴謹且適切控制型 I 誤差。再者，在不同樣本數組合下，GCI 法與 TCI 法的檢定力表現雖差異不大，但特別在小樣本下，GCI 法的檢定力表現則是略優於 TCI 法。

綜上所述，本研究的內在及外在效度應可獲較佳控制。

八、敘述統計及傳統獨立樣本 t 檢定

本研究採非同步遠距教學（實驗組）與現場教學（控制組）學生在機率模型單元的學習表現，其摘要統計列於表 6。採兩種不同教學方式的學生在機率模型單元之平均成績分別為 $\bar{x}_E=47.11$ 與 $\bar{x}_C=42.24$ ，而標準差則分別為 $s_E=24.48$ 與 $s_C=23.78$ 。此外，兩種不同教學方式的學生在機率模型單元之平均成績的 Shapiro-Wilk 常態性檢定 p 值分別為 0.24 與 0.37，均大於 $\alpha=0.05$ ，表示兩種不同教學方式下的機率模型單元之學生成績均符合常態分配假設。

表 6 非同步遠距教學與現場教學學生在機率模型單元成績摘要統計及 Shapiro-Wilk 常態性檢定 p 值

學習成效指標	教學方式	學生人數	平均數	標準差	Shapiro-Wilk 常態性檢定 p 值
機率模型單元 平均成績	非同步遠距教學 (實驗組)	71	47.11	24.48	0.37
	現場教學 (控制組)	54	42.24	23.78	0.24

本研究利用傳統獨立樣本 t 檢定比較兩種不同教學方式的學生於機率模型單元之學習平均成績是否存在顯著差異，而由表 7 中的變異數相等 Levene F 檢定結果可知，兩種不同教學方式下的學生成績變異數並無統計上顯著差異 ($p = 0.93$)，表示兩種不同教學方式下的學生成績變異數為相等之情形。接續進行傳統獨立樣本 t 檢定，以比較兩種不同教學方式的學生在機率模型單元之學習平均成績是否有顯著差異，在變異數相等的情形下，由表 7 中的平均數相等 t 檢定結果可知， t 值未達顯著水準 ($t_{0.025}(123) = 1.12$, $p = 0.27$)，僅能表示採非同步遠距教學與現場教學兩種不同教學方式的學生在機率模型單元之學習平均成績於統計上無顯著差異。有可能採用兩種教學方式下的學生學習表現均一樣好，抑或兩種教學方式的學生學習表現都不理想。基於此，本研究引用臨床試驗中探討學名藥對於專利藥療效與安全性的非劣性檢定概念，將其應用至教學場域中，進行遠距教學模式與傳統面授教學等兩種教學方法的學生學習表現之非劣性檢定，藉以評估遠距教學的學生學習表現雖不一定能顯著優於面授教學下的學生學習表現，但透過非劣性檢定則能夠說明遠距教學的學生學習表現不會較傳統面授教學的學生學習表現差。

表 7 非同步遠距教學與現場教學的學生在機率模型單元成績獨立樣本 t 檢定

		變異數相等的 Levene 檢定		平均數相等的 t 檢定						
		F 檢定	顯著性	t	自由度	顯著性 (雙尾)	平均 差異	標準 誤差異	差異的 95% 信賴區間	
									下界	上界
機率模 型單元 成績	符合變異 數相等的 假設	0.01	0.93	1.12	123	0.27	4.87	4.37	-3.77	13.51

九、非同步遠距教學與現場教學方式的學生學習表現非劣性檢定

研究者根據非同步遠距教學與現場教學兩種不同教學方式的學生在機率模型單元之學習成績，利用 GCI 法進行 (2) 式

$$H_0: \mu_E - \mu_C \leq -M \quad v.s. \quad H_1: \mu_E - \mu_C > -M \quad (M > 0)$$

的非劣性檢定， $-M$ 為非劣性檢定臨界值（margin）。

Chow 與 Shao（2006）定義 M 為：

$$M = \left[(z_{1-\alpha} + z_{1-\beta}) \sqrt{\frac{\sigma_C^2}{n_C} + \delta^2} \right] - \left[z_{1-\varepsilon} \sqrt{\frac{\sigma_E^2}{n_E} + \delta^2} \right] \quad (7)$$

其中， $\delta = \text{Min}\left(\frac{\sigma_E}{\sqrt{n_E}}, \frac{\sigma_C}{\sqrt{n_C}}\right)$ ， α 是顯著水準， β 為進行非劣性檢定犯型 II 誤差（type II error）的機率， ε 表示標準常態分配的累積機率；而 $z_{1-\alpha}$ 、 $z_{1-\beta}$ 與 $z_{1-\varepsilon}$ 分別表示累積機率為 $1-\alpha$ 、 $1-\beta$ 與 $1-\varepsilon$ 的標準常態分配臨界值。透過 Chow 與 Shao（2006）的定義，計算該實例進行非劣性檢定的臨界值，並利用 GCI 法與 TCI 法的 95% 信賴區間下界進行（2）式之非劣性檢定。

研究者考慮 $\alpha=0.05$ 、 $\beta=0.2$ 與 $\varepsilon=0.019$ ，經（7）式 M 的計算後，可得（2）式的非劣性檢定臨界值為 -2.2896 。而利用 GCI 法及 TCI 法構建兩種不同教學方式的學生成績之 95% 信賴區間下界分別為 -2.2604 及 -2.3641 ，其中 GCI 法的 95% 信賴區間下界大於 -2.2896 ，透過非劣性檢定，其檢定結果請見表 8。

雖然傳統獨立樣本 t 檢定比較兩種不同教學方式的學生於機率模型單元之學習平均成績並無顯著差異，但使用 GCI 法進行非劣性檢定則可進一步說明在 COVID-19 的疫情影響而不得不採用非同步遠距教學下，評估非同步遠距教學的學生成績表現，非同步遠距教學的學生學習表現至少不會比現場教學下的學生學習表現差。

表 8 GCI 法與 TCI 法構建的 95% 信賴區間下界及非劣性檢定結果

信賴區間估計法	95% 信賴區間下界	非劣性檢定結果
GCI	- 2.2604	通過非劣性檢定
TCI	- 2.3641	沒有通過非劣性檢定

陸、結論

一、非劣性檢定的研究結果

在停課不停學的政策措施下，教師的教學方式不得不改採遠距教學模式取代傳統現場面授教學方式，來因應 COVID-19 的疫情對教師教學與學生學

習所造成的影響。除了教學法外，本研究在其他影響非同步遠距教學和傳統現場面授教學方式的學習成效因素控制下，針對數理統計中的機率模型單元，評估非同步遠距教學方式與實體面授教學的學生學習表現。而當研究者欲進行非同步遠距教學的學生平均成績與實體面授教學的學生平均成績是否有差異之假設檢定程序，若使用獨立樣本 t 檢定的檢定，結果呈現無法推翻兩種教學方式下的學生學習表現不具統計上顯著差異，此時統計上無法說明非同步遠距教學的學生學習表現是否優或劣於實體面授教學的學生學習表現。遂此，本研究應用非劣性檢定來評估採非同步遠距教學下的學生學習表現不劣於實體面授教學下的學生學習表現。研究中提出利用 GCI 的 95% 區間下界來進行非劣性檢定程序，並與 TCI 進行比較，以說明非同步遠距教學的學生學習表現並不劣於現場教學方式下的學生學習表現。再者，透過統計模擬研究結果發現，在不同參數組合下，利用 GCI 法及 TCI 法進行非劣性檢定的型 I 誤差比率均落在 0.0457 至 0.0543 之容許誤差範圍內，均能嚴謹且適切控制型 I 誤差。又在不同樣本數組合下，GCI 法與 TCI 法的檢定力表現雖差異不大，但特別在小樣本下，GCI 法進行非劣性檢定的檢定力表現則是略優於 TCI 法，具穩定檢定結果。在教學場域中樣本數不大的情境下，GCI 法亦適合推薦進行非劣性檢定，以評估兩種不同教學法的學生學習成效。

數理統計內容中涵蓋許多數理公式的推導，透過非同步遠距教學方式，學生一方面可重複觀看、聆聽遠距數位教學影片，以了解相關理論意義外，對數理統計中所講述的公式推導步驟可更詳實理解。在 COVID-19 的疫情肆虐下，教師的教學方式不得不改採遠距教學模式進行教學，利用非劣性檢定進行評估非同步遠距教學方式下的學生學習表現應有其應用價值。

二、討論

與其他管理學院的課程相比，「數理統計」課程的目標強調統計理論之學習，此外更強調統計理論的講述及推導。修讀學生在學習上常因課程內容較為理論難懂，致使學習過程常發生困難。雖然有配置教學助教講授的實習課，但實習課並不是針對學生個人進行教學輔導，倘若學生因個人因素而致使自身產生學習困難，即使有實習課輔助，但適性教學的成效有限。再者，「數理統計」課程內容較為廣泛，受限學期時數，融入的單元實則有限。

然而，透過 EverCam 軟體進行教學影片錄製，以教學影片教學法作為學習機制導入進行非同步遠距教學，從學習效益的觀點來看，修課學生能有彈

性且充足的時間不斷重複觀看課程講授內容，相較實體面授課程的教學時間限制，以教學影片教學法作為非同步遠距教學，能讓學生有更充足時間去理解統計理論的相關公式及定理推導，提升學習統計理論的學習效益。

三、研究限制

本研究假設兩種教學方式下的學生成績服從常態分配，利用 GPQ 構建非同步遠距教學方式與現場面授教學方式的學生學習表現差異 GCI 之 95% 區間下界，應用非劣性檢定評估非同步遠距教學方式不劣於現場面授教學方式的學生學習表現。當學生的成績資料違反常態分配假設時，Lee 等（2016）的研究建議可採取 Box 與 Cox（1964）的轉換方式，對非常態分配資料進行轉換，使資料滿足常態分配假設後，再利用 GCI 的 95% 區間下界執行非劣性檢定程序。

此外，本研究是以「數理統計」課程做為研究，研究結果可能因不同課程、學生、校院系所等條件而有所不同，故對不同屬性學科不宜做過度推論。評估學生學習成效概括為知識、技能及態度三大面向，本研究僅以知識面向做為學生學習成效的評估，研究結果的推論係以數據分析為立論基礎，未進一步進行學生質性訪談，僅就量化資料呈現研究結果。此外，對於非本研究可控制之變數，如學生家境狀況、學生修課行為等可能影響學習表現的因素，均不納入分析結果的討論，屬本研究的限制，建議後續研究者可以針對技能及態度面向做進一步研究。

謝誌

研究者感謝主編及審查委員提供的建議與寶貴意見，讓本文更臻完善。

參考文獻

中文部分

- 田耐青、洪明洲（1998）。電腦中介溝通與高等教育革新。**臺北師院學報**，**11**，1-22。
- 何希慧、劉怡、吳佩真（2010）。美國加州大學洛杉磯校區（UCLA）高等教育研究機構（CIRP）執行美國大學生學習成效問卷調查之經驗分享。**評鑑雙月刊**，**23**，50-53。 <http://doi.org/10.6445/EB.201001.0012>
- 吳雅琪（2012）。不劣性試驗統計審查重點。**當代醫藥法規月刊**，**22**，1-6。
- 李忠謀、邱瓊芳（1999）。遠距教學環境：支援系統的發展與評估。**科學教育學刊**，**7**（1），49-69。 <http://doi.org/10.6173/CJSE.1999.0701.05>
- 李秋璿（2010）。應用網路教學於護理系學生暴力處置課程知識成效之探討（未出版之碩士論文）。中山醫學大學。
- 沈中偉（1998）。即時群播遠距教學之教學設計與教學策略探討。**遠距教育**，**7**，13-19。
- 范又方（2019）。針對成對二分資料利用強韌分數檢定衍生的信賴區間做等效性與非劣性檢定（未出版之碩士論文）。國立中央大學。
- 翁榮源、林宏樺（2013）。不同學習工具使用在網路化學教育之應用研究。**化學**，**71**（4），329-346。 <http://doi.org/10.6623/chem.2013032>
- 馬國華、張嘉恩、陳昆鴻、王雲充、吳耀欽（2021）。使用雲端影片教學對提升物理治療實習生學習成效探討。**物理治療**，**42**（2），129-130。 <http://doi.org/10.6215/FJPT.202106.P18>
- 陳思光、陳瑞松、藍萬烘、謝季全（1997）。牙科放射線學非同步遠距教學之短期成效。**醫學教育**，**1**（4），428-435。 [http://doi.org/10.6145/jme.199712_1\(4\).0007](http://doi.org/10.6145/jme.199712_1(4).0007)
- 陳昭珍、徐芝君、洪嘉馥、胡衍南（2021）。COVID-19 下臺師大的遠距教學經驗與省思。**當代教育研究季刊**，**29**（1），1-23。 [http://doi.org/10.6151/CERQ.202103_29\(1\).0001](http://doi.org/10.6151/CERQ.202103_29(1).0001)
- 陳瑜芬（2021）。後疫時代教師教學策略規劃與應用。**商管科技季刊**，**22**（4），539-550。
- 黃秀仍（2007）。學生對非同步遠距教學看法之探究。**遠東科大通識學報**，**1**，

- 152-159。 <http://doi.org/10.6795/JFEUGE.200707.0152>
- 楊奕農、柴蕙質（2002）。非同步網路學習成效及影響因素之計量分析：經濟學課程個案研究。*科學教育學刊*，**10**（2），193-210。 <http://doi.org/10.6173/CJSE.2002.1002.05>
- 楊建民（2010）。*探究式教學法與講述式教學法在國小 Scratch 程式教學學習成效之研究*（未出版之碩士論文）。國立屏東教育大學。
- 溫嘉榮、張建原（2014）。通識教育採用遠距教學成效及滿意度之研究：以計算機概論為例。*樹德科技大學學報*，**16**（1），133-145。
- 董力華（2021）。COVID-19 疫情影響下的通識課：網路合作學習經驗與實施成效探究。*人文社會科學研究*，**15**（4），81-110。 [http://doi.org/10.6284/NPUSTHSSR.202112_15\(4\).4](http://doi.org/10.6284/NPUSTHSSR.202112_15(4).4)
- 鄒景平（2005）。*數位學習新視界：上網教與學*。鼎茂。
- 劉通國（2002）。*學習風格與電腦自我效能對非同步遠距學習成效及滿意度之研究*（未出版之碩士論文）。國立高雄師範大學。
- 鄭韻如、高毓秀（2012）。探討癌症照護課程不同教學方式之成效。*護理雜誌*，**59**（5），38-48。 <http://doi.org/10.6224/JN.59.5.38>
- 盧宏益（2020）。數理統計創新教學：資訊科技融入與數位學習網站實作之教學實踐歷程。*大學教學實務與研究學刊*，**4**（2），97-121。 [http://doi.org/10.6870/JTPRHE.202012_4\(2\).0004](http://doi.org/10.6870/JTPRHE.202012_4(2).0004)
- 羅芳吟、陳政煥（2021）。COVID-19 疫情下同步與非同步資訊科技輔助的大學遠距英語文教學。*當代教育研究季刊*，**29**（1），69-114。 [http://doi.org/10.6151/CERQ.202103_29\(1\).0003](http://doi.org/10.6151/CERQ.202103_29(1).0003)

英文部分

- Allgood, S., & Walstad, W. B. (1999). The longitudinal effect of economic education on teachers and their students. *Journal of Economic Education*, **30**, 99-111. <https://doi.org/10.1080/00220489909595946>
- Boghikian-Whitby, S., & Mortagy, Y. (2008). The effect of student background in e-learning longitudinal study. *Issues in Informing Science and Information Technology*, **5**, 107-126. <https://doi.org/10.28945/3203>
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society: Series B*, **26**, 211-246. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Chang, F. C., Yeh, S. Y., & Hsieh, H. N. (2013). Generalized confidence interval estimation

- for the difference in paired areas under the ROC curves in the absence of a gold standard. *Communications in Statistics-Simulation and Computation*, 42, 2056-2072. <https://doi.org/10.1080/03610918.2012.690483>
- Chizmar, J. F., & Walbert, M. S (1999). Web-based learning environments guided by principle of good teaching practice. *Journal of Economic Education*, 30, 248-259. <https://doi.org/10.2307/1183061>
- Chou, H. W. (2001). Influences of cognitive style and training method on training effectiveness. *Computer & Education*, 37, 11-25. [https://doi.org/10.1016/S0360-1315\(01\)00028-8](https://doi.org/10.1016/S0360-1315(01)00028-8)
- Chow, S. C., & Shao, J. (2006). On non-inferiority margin and statistical tests in active control trials. *Statistics in Medicine*, 25, 1101-1113. <https://doi.org/10.1002/sim.2208>
- Davies, R., & Berrow, T. (1998). An evaluation of the use of computer supported peer review for developing higher level skills. *Computer & Education*, 30, 111-115. [https://doi.org/10.1016/S0360-1315\(97\)00086-9](https://doi.org/10.1016/S0360-1315(97)00086-9)
- Gregor, S. D., & Cuskelly, E. F. (1994). Computer mediated communication in distance education. *Journal of Computer Assisted Learning*, 10, 168-181. <https://doi.org/10.1111/j.1365-2729.1994.tb00293.x>
- Hiltz, S. R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM*, 40(9), 44-49. <https://doi.org/10.1145/260750.260764>
- Hsieh, H. N., & Lu, H. Y. (2020). The generalized inference on the ratio of mean differences for fraction retention noninferiority hypothesis. *PLoS ONE*, 15(6), e0234432. <https://doi.org/10.1371/journal.pone.0234432>
- Hsieh, H. N., Su, H. Y., & Lu, X. (2009). A generalized inference on assessment of similarity between dissolution profiles. *Advances and Applications in Statistics*, 12, 163-190. http://www.pphmj.com/article.php?act=art_download&art_id=4133
- Lee, H. I., Chen, H. Y., Kishino, H., & Liao, C. T. (2016). A reference population-based conformance proportion. *Journal of Agricultural, Biological, and Environmental Statistics*, 21, 684-697. <https://doi.org/10.1007/s13253-016-0268-z>
- Li, C. R., Liao, C. T., & Liu, J. P. (2008a). A non-inferiority test for diagnostic accuracy based on the paired partial areas under ROC curves. *Statistics in Medicine*, 27, 1762-1776. <https://doi.org/10.1002/sim.3121>
- Li, C. R., Liao, C. T., & Liu, J. P. (2008b). On the exact interval estimation for the difference in paired areas under the ROC curves. *Statistics in Medicine*, 27, 224-242. <https://doi.org/10.1002/sim.2760>

- Lin, T. Y., & Liao, C. T. (2006). A beta-expectation tolerance interval for general balanced mixed linear models. *Computational Statistics and Data Analysis*, 50(4), 911-925. <https://doi.org/10.1016/j.csda.2004.11.007>
- Lin, T. Y., & Liao, C. T. (2008). Prediction intervals for general balanced linear random models. *Journal of Statistical Planning and Inference*, 138(10), 3164-3175. <https://doi.org/10.1016/j.jspi.2008.01.001>
- Motiwalla, L., & Tello, S. (2000). Distance learning on the internet: An exploratory study. *The Internet and Higher Education*, 2(4), 253-264. [https://doi.org/10.1016/S1096-7516\(00\)00026-9](https://doi.org/10.1016/S1096-7516(00)00026-9)
- Noesgaard, S. S., & Ørngreen, R. (2015). The effectiveness of e-learning: An explorative and integrative review of the definitions, methodologies and factors that promote e-Learning effectiveness. *Electronic Journal of e-Learning*, 13(4), 278-290. <https://www.academic-publishing.org/index.php/ejel/article/view/1735/1698>
- Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environment: A research framework and a preliminary assessment of effectiveness in basic IT skill training. *MIS Quarterly*, 25(4), 401-426. <https://doi.org/10.2307/3250989>
- Shih, S. M., Wu, W. H., & Hsieh, H. N. (2016). A non-inferiority test for diagnostic accuracy in the absence of the golden standard test based on the paired partial areas under receiver operating characteristic curves. *Journal of Applied Statistics*, 43, 550-562. <https://doi.org/10.1080/02664763.2015.1070810>
- Tsai, J. R. (2010). Generalized confidence interval for the slope in linear measurement error model. *Journal of Statistical Computation and Simulation*, 80(8), 927-936. <https://doi.org/10.1080/00949650902838032>
- Webster, J., & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning. *Academy of Management Journal*, 40(6), 1282-1309. <https://doi.org/10.2307/257034>
- Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88, 899-905. <https://doi.org/10.2307/2290779>
- Wu, W. H., & Hsieh, H. N. (2014). Generalized confidence interval estimation for the mean of delta-lognormal distribution: An application to New Zealand trawl survey data. *Journal of Applied Statistics*, 41, 1471-1485. <https://doi.org/10.1080/02664763.2014.881780>
- Xiao, Y., Liang, Z. L., Li, Q., & Jia, R. J. (2017). Exploring the Feasibility of Video-Mediated Listening Test in a Nation-Wide Proficiency English Examination in China. *TESOL International Journal*, 12(2), 1-16. <http://files.eric.ed.gov/fulltext/EJ1247808.pdf>

Yu, C. Y., & Tsai, J. R. (2012). A comparison of approaches to interval estimation of mediated effects in linear regression models. *Communications in Statistics-Simulation and Computation*, 41, 1301-1316. <https://doi.org/10.1080/03610918.2011.601834>