

改進我國農業普查資料整理方法之研究

林元興

(作者爲本校地政系專任教授)

摘要

農業普查爲制定農業政策之基本大計，惟以往在普查後整理資料時，因爲檢誤工作耽擱時日，以致未能及時發表，況且在檢誤時，如遇有資料錯誤多以人爲主觀觀念修正之，頗不合統計原則。

本研究旨在利用電腦程式檢誤，俾能增進其時效性，並且藉迴歸模式以進行錯誤資料的修正，期能使我國農業普查工作更臻完善。

壹、緒言

近二十餘年來，臺灣的經濟開發已爲開發中國家楷模，而農業則爲臺灣經濟發展之基礎。政府推行各項土地改革政策雖對臺灣農業有著重大貢獻，但近年來由於工業快速成長，農業受生產條件限制，農業比重顯著下降。此種農工經濟結構之偏差，如不予適時改善，勢必造成農業部門之萎縮。

爲解決農業問題，必先了解目前農業經營現況，蒐集全國農業基本統計資料、資源分佈、經營狀況、農業生產設備及生產結構之變動等資料。針對問題癥結；謀求解決方案。這些基本資料主要來源之一即爲農業普查。

資料爲分析之基礎，錯誤過時之資料將導致政策的偏差及有效需要。因此在擬定政策目標之前，須將原始資料加以系統化整理，使之合乎時效性，經濟性之原則。農業普查範圍廣大，項目繁雜，如何將這數量龐大的普查資料，迅速整理，及時提出成果報告供各界所需，乃普查資料整理工作之目標，也是本文研究之目的。

爲達上述目的，必須建立一套合理客觀之資料檢誤方法，以更正偏誤資料。當然事前預防偏誤資料產生之措施亦可加速資料處理，故此種措施是否得當，也在本文探討之列。

由於六十九年臺閩地區農業普查之訪問調查工作自民國七十年一月五日起至二月十五日止結束，各縣市之農戶調查表已送交行政院普查委員會點收完畢，委由電腦公司整理，本研究受時間限制，乃選擇普查前所舉辦的第二次試查（以下簡稱本試查

）資料作爲本文研究測試對象。

本試查資料係於六十九年五月五日至十七日調查而得，凡屬靜態資料以民國六十八年十二月三十一日之情況爲準，動態資料以民國六十八年一月至十二月之數據爲準。試查對象爲臺灣本島在六十八年十二月三十一日從事農業生產活動之農戶主，採用「分層三段隨機抽樣法」，由二十個縣市，抽出八十四個鄉鎮，一百個村（里），每一村（里）抽出八戶，共計八百個樣本農戶。其資料委由中華電腦公司錄成磁帶。

貳、我國歷次農業普查資料整理方法之探討

甲、農業普查工作之一般程序

爲了解資料整理工作佔整個農業普查過程之地位，首先概略說明普查工作之作業程序。

普查工作可分爲下列五個階段進行：

一、普查之設計階段

蒐集國內外有關普查資料，比較研究，先釐訂重要原則性計劃，如調查範圍、對象、抽樣方法、調查項目、各業調查表、母體調查名冊之編製與判定等，然後舉辦第一次試查，獲取實地經驗，以補設計之不週。（註一）

二、調查前之準備階段

包括地方普查機構之組成、調查人員之遴選及訓練、宣傳、編製調查名冊、蒐集參考資料、釐訂普查進度表等，均須事先妥為籌備，以利調查。

三、普查之執行階段

調查人員根據調查名冊，實地進行訪問調查。為控制資料之正確度應同時分區嚴格督導，以便如期獲得資料。

四、普查資料之整理階段

普查表由各地區彙集於中央，點收之後即予系統管理，然後為人工之審核或為機器之處理，俟資料更正無誤之後，即可編製結果表。

五、普查報告之編輯階段

普查資料經分析整理後，應即提出普查初步報告，普查總報告及專題研究報告，以應各界研究之需，及供政府決策之參考。

乙、農業普查資料處理方法

以上已對普查過程作概略說明，現專就其中第四階段「普查資料之整理」進行探討。

農業普查資料整理與一般統計資料之整理方法類似，惟其資料數量內容繁雜，故整理工作益顯艱鉅。

整理之目的，乃是消除原始資料之雜亂不全，矛盾偏誤等，並加以分類、統計分析，迅速產生精確可靠之結果報告。資料整理應兼顧「時效性」及「精確性」二大原則。其程序包括資料之點收、審核、過錄整理、分類、計算製表等項目，凡此皆應事先規劃，明訂工作程序進度、人員、經費之調配、各種表格之設計，以期依計劃順利完成。

資料整理之方法，依資料數量多寡可分為：

- (一) 人工整理法：調查資料數量少，人力許可範圍內，可採人工劃記法或卡片整理法。
- (二) 機器整理法：調查資料甚多，人工整理耗時易錯，故科技進步的現代，多採此法，以收迅速確實之效。

丙、我國歷次農業普查資料整理方法之演進

自民國四十五年以來，我國已舉辦過六次農業普查，最近一次已於七十年二月十五日始結束，整理工作正在進行，姑且不予以討論，現按主辦單位之不同，調查範圍之大小，可區分為兩個階段，第一階段為民國四十五年之選樣普查，四十九年之全面普查及五十四年之選樣調查，此階段係由臺灣省政府舉辦的地方性農業普查。第二階段為民國五十九年之全面普查與六十四年之選樣調查，此階段係由行政院主計處接辦的臺閩地區農業普查。

早期農業普查資料之整理方法，由於初辦經驗不足，範圍較小，加上受科技限制，主以人工整理法，進而以人工為主，機器為輔。最近則進步為以機器為主、人工為輔，由此可知我國農業普查資料整理兼採人工整理法與機器整理法。

丁、我國農業普查資料整理方法之檢討

依據資料整理之兩大原則「時效性」「精確性」的觀點，檢討宜改進之處。

(一) 資料整理方法

民國五十九年以前普查資料整理方法以人工為主，由於資料龐大繁雜，人工整理易錯耗時，使普查結果降低利用價值，故資料整理電腦化為現代國家普查之趨勢。我國乃於民國五十九年普查規劃利用電子計算機處理，採半自動作業方式，輔以人工調查表內容設計亦配合機器處理，例如以圈註符號代替人工填寫文字。除此更利用機器進行審查更正、計算、列表等作業。既正確又爭取時效。

民國六十四年資料登錄由打卡方式進步到磁帶登錄，即不由打卡、驗卡步驟，直接將調查表之資料登錄在磁帶上，不僅減少打卡驗卡錯誤，更使資料整理時間大為縮短，此為資料整理方法之一項革新。

(二) 資料登錄方法

由於資料登錄磁帶過程中難免發生人為錯誤，為消除此種錯誤，可仿效歐美國家所採行之方法（註二），使用縮影設備及光

學閱讀機。將調查表拍成縮影片，再透過光學閱讀機直接把資料輸入電腦內。此項科學技術頗值學習。

(三) 資料之檢誤

除採電腦化作業外，普查內容充實化亦為現代趨勢。隨著國家經濟繁榮，每次普查項目，難免會增列若干重要項目，以應各政府機關、研究機構之需求。例如民國六十四年普查除維持五十九年普查原有項目外，並增列農作物收獲量、現代農場設備……等項。對於增列項目可能產生之偏誤，如何防止與更正，皆需妥為規劃設計，否則徒費人力、財力，卻得到不正確結果，就違背增列原意。本文擬就六十四年增列「農作物收獲量」一項，探討對該項目可能發生之偏誤，是否採取事前預防，事後更正之措施？又該項措施是否合理客觀？容後詳論。

農業普查為基本國勢調查之一種，如何獲取「精確可靠」資料，作為有關當局決策之參考，為其主要目的。因此普查工作之每一階段皆非常重視如何保持原始調查資料之完整、正確、合理與一致。實施各種預防措施避免錯誤之產生，至於錯誤既已產生，如何合理客觀更正，更是整理工作之目的。簡言之，普查資料如何檢誤、更正以達「迅速確實」之目標，乃資料整理工作之重點，亦為本文討論之重心。

三、邏輯檢誤法

甲、設計目的

普查之調查資料，雖經由逐級審核制之嚴行，但仍無法保證資料完整無誤，或由於調查項目設計欠週，或由於人為之疏忽。當資料登錄磁帶後，並加以驗製，以期磁帶上資料與調查表完全相符，即可進行邏輯檢誤工作。

所謂邏輯檢誤亦即針對調查表上各問項是否遺漏，細數項之和是否等於總數欄，各項數字合理與否，相關問項有無矛盾之處……等加以檢誤，並合理更正。作業時為達「時效性」之要求，應以機器處理為原則。此外為達到「精確性」之要求，應注意下列兩點：

- (一) 資料之完備性：若有遺漏未列之處，應儘量根據以往發展趨勢或其他相關項目加以推敲，間接求得合理答案補充之。
- (二) 資料之正確性：應考慮調查表上所填數字是否合理？各細數項之和是否與總數項符合？相關問項之間是否矛盾？

乙、檢誤範圍

邏輯檢誤法測試的對象為六十九年第二次試查資料，本試查調查表共分十二節，由於時間限制，茲選其中前四節問項試行邏輯檢誤工作。即第一節農業人口及其從業員工，第二節土地及灌溉，第三節農業主要設備與農機具及其使用情形，第四節作物種植面積與產銷量，由於調查表上第四節 4.1 問項欄上方有水田、旱田總面積，該欄數字雖與 2.1 問項耕地面積有關，應納入邏輯檢誤條件內，但因總面積數字未登錄在磁帶上，無法進行邏輯檢誤，故第四節僅就全年自用量、銷售量進行檢誤。至於作物收穫量，則留待第肆節進行廻歸檢誤。

丙、邏輯檢誤條件之試擬

茲根據六十九年農業普查試查手冊第陸節農業調查表審核方法及普查委員會審查六十九年農漁業普查有關電腦檢誤之項目、內容及各項應如何處理之會議紀錄，並參照六十四年普查報告第一卷總說明第陸節所列機器審核條件之處理原則，試擬邏輯檢誤條件如下：

子、農業人口及其從業人員工作情形

本節共有 1.1 問項至 1.7 問項，由於試查只限於一般農戶，故農場填用的 1.5 至 1.7 問項未列入檢誤條件之內。為說明方便，將第一節有關表格節錄部分於後：

1.1 請問貴戶在69年底人口共有幾個人？

(包括經常在家及共同生活的人口，不含離家就法人口。)

	男	女	合計
	男	女	未滿15歲
	男	女	15歲以上

(包括專辦家事、服役、在學、老病殘障及無工作者。)

代號	①稱謂	②性別		③實足年齡	④農業工作指揮者	⑤參加自家農業工作主要工作者	⑥全年參加自家農業工作日數	⑦從事其他工作者		⑧其他人口
		男	女					受雇用	自營	
01		1	2			9	1	2		9
02		1	2			9	1	2		9
03		1	2			9	1	2		9
04		1	2			9	1	2		9
05		1	2			9	1	2		9
06		1	2			9	1	2		9
07		1	2			9	1	2		9
08		1	2			9	1	2		9
09		1	2			9	1	2		9
10		1	2			9	1	2		9

農業工作指揮者教育程度

教育程度	農業學校		其他學校			未受教育
	大專	職校	大專	中學	小學	
代號	1	2	3	4	5	6

(限圈一個)

1.4 請問貴戶維持家庭生活主要是依靠什麼收入？

專業戶 兼業戶 (有人從事其他工作者應屬兼業戶)	全部依靠農業收入			1 2 3 (限圈一個)	
	主要依靠農業收入				
	主要依靠其他收入				

民國五十七年實施九年國民教育制度以來，農村子弟接受中等教育機會已增加，但因為農業經營辛苦，所得偏低，致農村青年紛紛轉向都市謀生，真正留在農村指揮農事者，仍以小學教育程度者居多。

〔1.2〕問題係滿十五歲以上人口工作形，故第③欄「實足年齡」應大於或等於十五。若有小於十五歲者，應予以剔除，以免核計總人數時發生錯誤。空白者，由機器挑出，人工依照①「稱謂」判定更正。

〔1.1〕問題係年底全戶人口總數，其中⑤「十五歲以上男」數字應與1.2問題②「性別」中「男」（資料代號為1）的總計數相者。倘不相等時，以1.2問題的總數為準，更改1.1問題⑤的數字。因為1.2問題乃係就個人資料逐一填寫，其人數總計數值的可信度較高。

〔1.1〕問題⑥「十五歲以上女」的處理原則同上，其中「男」改為「女」。

〔1.1〕問題①「合計男」應與該問題③「未滿十五歲男」及⑤「十五歲以上男」之和相等。倘不相等時，更改①「合計男」之數值。

〔1.1〕問題②「合計女」的處理原則同上，其中「男」改為「女」。

〔1.2〕問題④「農業工作指揮者」之資料型態應為「9或空白」，且全戶只能圈選一個，亦即該欄資料全戶只能出現「9」。倘有其他數值出現，自動更正為「空白」；倘全戶該欄資料全為空白，則以該戶年齡在九十九歲以下最大年齡者，更正為「9」；倘該戶圈選超過一人時，以第一個出現「9」者為準，其餘皆更改為「空白」。

〔7〕問完④之資料後，須繼續轉問「農業工作指揮者教育程度」，此欄資料型態應為「1~6」，倘有其他數字出現或空白時，自動更正為「5」，即小學教育程度。因為根據六十四年農業普查報告顯示，農民教育水準常較其他產業偏低，雖⑥的資料，當⑤為「1」時，補填⑥為「200」，當⑤為「2」時，補填⑥為「100」，又⑥的天數按常理應小於「365」，若⑥大於365，則按上述原則更改之。使全年參加自家農業的工作天數與工作身份相配合。

〔8.1.2〕問題⑤「參加自家農業工作」，資料型態應為「1」或「2」或「空白」。倘有其他數字出現或重複圈選時，則以

⑥「全年參加自家農業工作天數」為準。當工作天數大於二百天，則更正⑤為「1」，即「主要工作者，當工作天數小

於一百天，則更正⑤為「2」，即「農業工作幫助者」。倘⑤為空白，而⑥有資料，則依上述標準補填⑤的資料。

(九) 1.2 問項⑤「參加自家農業工作」有資料，則必有⑥「標準工作天數」。倘⑤有資料，⑥無資料時，則以⑤為準，加填
(十) 1.2 問項若⑥「參加自家農業工作標準天數」有資料，且⑦「從事其他工作者」亦有資料時，當⑥大於200天，則⑦的
資料只可能「2」——「臨時受雇者」，與「3」——「自營者」的型態，倘⑦出現「1」——「固定受雇者」之情形
，則更改⑦為「2」。

(十一) 1.2 問項⑧「其他人口」，係指專辦家事、服役、在學、老病殘障及無工作者，此項資料型態為「9」或「空白」，必
須前面④、⑤、⑥、⑦皆無資料，⑧才可能為「9」，倘④、⑤、⑥、⑦任何一欄有資料，而⑧亦有資料，則更正⑧為
空白。

又倘④、⑤、⑥、⑦、⑧皆無資料，而②「性別」，③「年齡」有資料時，依前面②、③資料推敲更正，倘③小於十九歲
或大於六十五歲時，更正⑧為「9」，將其列入其他人口。倘③在二十歲至六十四歲之間，則進一步依性別判斷，②為「1」
男性者，更正⑤為「2」，⑥為「100」，亦即將其列入農業幫助者。若②為「2」女性者，更正⑧為「9」，即將其視為專
辦家事，列入其他人口。

(十二) 1.4 問項關於專兼戶的判斷，其資料型態為「1」或「2」或「3」，此問項與1.2問項⑦「從事其他工作者」相關。
倘1.4問項出現其他數字或空白時，則視1.2問項⑦欄有無資料而更正。若⑦無任何資料，表示無從事農業外行業者，則
更改1.4問項為「1」，即全部依靠農業收入的專業戶，反之，若⑦有資料，則更改1.4問項為「2」，即主要依靠農業
收入的兼業戶。又若⑦有資料，而1.4問項出現「1」，顯係專業戶判斷錯誤，因為依照試查表之說明「有人從事其他
工作者應屬兼業戶」，故比照上述方法，更改1.4問項為「2」，主要依靠農業收入的兼業戶，而非專業戶。

丑、土地及灌溉

2.5 請問貴戶（場）有灌溉之耕地面積若干？.....→

水 源 別	合 計			自家地下抽水			由水利會輸水			其 他 水 源		
	代號	公 噌	公 敝	代號	公 噌	公 敝	代號	公 噌	公 敝	代號	公 噌	公 敝
灌 濑 面 積	145			150			151			152		

2.1 請問貴○（場）經營的耕地有多少？ 2.2 請問貴○（場）經營的非耕地有多少？

(凡可耕種的土地，不論已否種有作物或暫時休閒及休耕，均應包括在耕地內。但不包括出租或委託他人經營之耕地，係以自己經營者為限。)

(指耕地以外的各種土地，包括原為耕地已植樹成林或改為建築基地。但不包括出租或委託他人經營之非耕地，係以自己經營者為限。)

(為防止耕地面積答復偏低或遺漏，應運用耕地參考資料」加以研判。)

自 耕 農	1	半 自 耕 農	2	半 佃 農	3	佃 農	4	非 耕 種 農	5
-------	---	---------	---	-------	---	-----	---	---------	---

耕 地 別	代 號	合 计		自 有 土 地 (包括山地保留地)		租 入 土 地 包括河川地、未登錄地	
		公 噌	公 敝	公 噌	公 敝	公 噌	公 敝
耕 地	兩 期 作 水 田	121					
	單 期 作 水 田	122					
	輪 作 田	123					
	平 地 旱 田	124					
	山 坡 地 旱 田	125					
	合 计	120					
非 耕 地	建 築 基 地 及 晒 場	131					
	魚 池 (塭)	132					
	山 林 地	133					
	草 生 地	134					
	其 他	135					
	合 计	130					

[T.1] 問項六十九年計底經營的耕地面積，先核算其橫列之合計數是否與細數項之和相等。耕地共分爲五種類別，首先核計每一類別土地的②「自有土地」面積與③「租入土地」面積之和應與該類土地的①「合計」相等。倘不相等時，挑出由人工更正。

然後再核算縱行之合計數是否與各細項之和相等，亦即分別計算①、②、③中各類土地面積和是否與合計欄相等，倘不相等時，處理原則同上。

[T.2.2] 問項六十九年底經營的非耕地面積，分別計算橫列、縱行之合計項是否與各細數項之和相等。其處理原則與上述相同。

[T.2.1] 問項上方關於農戶種類，其資料型態爲「1~5」，限圈選一個，倘有其他數字出現或空白時，依相關問項2.1中②「自有土地合計」與③「租入「土地合計」的數值大小關係判定。倘②合計有資料，③合計無資料，表示全爲自有土地，則更正農戶種類爲「1」，即自耕農；倘②、③皆有資料，而②大於③，表示大部份爲自有耕地，小部份租入土地，則更正爲「2」，即半自耕農，反之，若②小於③，則表示大部份向他人承租土地，更正農戶種類爲「3」，即半佃農，倘②無資料，③有資料，表示全部租他人土地，則更正爲「4」，即佃農，倘②、③皆無資料，則更正爲「5」，即非耕種農。

[T.2.5] 問項灌溉的耕地面積，核算橫列合計項是否與各細項之和相等。倘不相等，則更改合計項。

[T.2.5] 問項灌溉耕地面積合計數應小於或等於2.1問項耕地面積之合計。倘灌溉面積大於農戶所經營之耕地面積，則將其間之誤差，先由「其他水源」減去，倘仍不足，再由「水利會輸水」減去差額。

寅、農業主要設備與農機具及其使用情形

[T.3.2] 問項六十九年內使用動力農機具耕作面積，此問項與3.1問項主要農機具之數量相關。3.2問項若犁田整地、中耕除草的①「使由自家農機具」有資料，則3.1問項的「耕耘機」、「曳引機」應有資料，若後者無資料，則將3.2問項

3.1 請問貴○（場）在下列主要農機具中各有多少？

農機別	耕耘機			聯合收穫機	曳引機	
	未滿10馬力	10—15馬力	15馬力以上		未滿70馬力	70馬力以上
代號	153	154	155	156	157	158
(台)						

烘乾機 (含乾燥機)	動脫穀機	動插秧機	動噴霧(粉)機	農用動力車	農抽水機
159	160	161	162	163	164

3.3 請問貴○（場）在下列主要設備中各有多少？

3.2 請問貴○（場）使用農機具情形？

(係以累計面積計算)

中①的機耕面積加入②「包給他人人機耕」之機耕面積，更改①爲空白。

13.2 問項播種插秧的①有資料，則3.1 問項的「動力插秧機」應有資料，若後者無資料，處理原則同上。

卯、作物種植面積、收穫面積與產量

4.1 請問農戶（婦）在下列耕種作物中，種植玉米及其產銷量各有多少？

4.3 請問費用（總）在下列各項作物中種植面積、收穫面積與產量的最多有誰？

米		稻		黑		白		金黄白米		糙米		生米		熟米	
类别	作物名稱	代	公	代	公	代	公	代	公	代	公	代	公	代	公
		1. 精米	2. 粗米	3. 糙米	4. 黑米	5. 白米	6. 金黄白米	7. 糙米	8. 精米	9. 糙米	10. 黑米	11. 白米	12. 金黄白米	13. 糙米	14. 精米
谷	稻 4011							4012							
稻	4021							4022							
稻	4031							4032							
稻	4041							4042							
稻	4051							4052							
								4053							
								4054							

〔1〕若有作物收穫量，則必有「全年自用量」或「全年出售量」的資料，倘兩者皆空白時，則將「全年出售量」更改為該作物各期收穫量之和，亦即將當年收穫量視為全部出售。

〔2〕合計項之數值應與各細數項之和相等的原則於此欄亦適用之。

一、測試結果

根據上面所擬之檢誤條件，利用 FORTRAN 電腦語言設計程式，與本試查資料一併輸入電腦中，進行邏輯檢誤條件的測試，茲將測試結果節錄部分說明於下：

第一節 農業人口及其從業人員工作情形

(一)

1.2.3	AGE	LESS	THAN	15,	NO.=770121604	0
1.2.3	AGE	LESS	THAN	15,	NO.= 770121604	0
1.2.3	AGE	LESS	THAN	15,	NO.= 770120408	0
1.2.3	AGE	LESS	THAN	15,	NO.= 770120408	0
1.2.3	AGE	LESS	THAN	15,	NO.= 770120408	0
1.2.3	AGE	LESS	THAN	15,	NO.= 770121303	0
1.2.3	AGE	LESS	THAN	15,	NO.= 770121303	0
1.2.3	AGE	LESS	THAN	15,	NO.=3260020901	14
1.2.3	AGE	LESS	THAN	15,	NO.=3260020901	12

表中 1.2.3 檢測 1.2 訊項③「年齡」，年齡資料空白者，由機器列出檢誤表及農戶編號，由人工判定更正。若 1.2.3 資料小於「15」，則將之剔除，故第八、九筆資料應剔除，不加入計算滿十五歲之人口。

(二)

1.1.A	NO.=1110170803	4	1.2.A	NO.=1110170803	3-MAL
1.1.A	NO.=1110170803	3	1.2.B	NO.=1110170803	4FEM
1.1.A	NO.=1320121003	4	1.2.A	NO.=1320121003	5-MAL
1.1.A	NO.=1310110424	4	1.2.B	NO.=1310110424	3FEM
1.1.A	NO.=1460040319	3	1.2.A	NO.=1460040319	2-MAL
1.1.A	NO.=1460040319	1	1.2.B	NO.=1460040319	2FEM

表中 1.1A 表示 1.1 問項中^⑤「十五歲以上男」或^⑥「十五歲以上女」，1.2.A 表示 1.2 問項^②「性別」中「男」的加總，1.2.B 表示 1.2 問項^②「性別」中「女」的加總。檢證條件係 1.1.A 與 1.2.A (MAL) 或與 1.2.B (FEM) 相等，故第一列資料 (NO.=111017803) 顯示 1.2 問項十五歲以上男有三人，而 1.1 問項十五歲以上男有四人，與檢誤條件不符，故更正時依前述檢誤條件更正第一筆資料 1.1.A 的「4」為「3」，同理，第二筆資料 1.1.A 的「3」應更正為「4」。

(三)

1.1.C	NO.=1310110424	ITEMS=	3	1	3	MALE	
1.1.D	NO.=1310110424	ITEMS=	5	0	4	FEMALE	
1.1.D	NO.=1210061202	ITEMS=	3	0	2	FEMALE	
1.1.C	NO.=22	1808	ITEMS=	2	1	2	MALE
1.1.D	NO.=22	21808	ITEMS=	3	0	2	FEMALE

表中 1.1C 表示 1.1 問項^①「合計男」，後者 ITEMS 表示「合計男」，「未滿十五歲男」、「十五歲以上男」三欄數字，原則上「十五歲以上男」已依 1.2 問項校正過，故「合計男」數字與後兩欄之和不符時，更改前者。又表中 1.1.D 表示 1.1 問項^②「合計女」，後者 ITEMS 表示「合計女」、「未滿十五歲女」、「十五歲以上女」，處理原則同上。故第

1 筆資料 (NO.=1310110424) 應將第一個數字「۲」改為後兩欄之和「۴」。

(四)

1.2.4	NO.=18 260714	DIRECTORMISSING, CHOOSEMAXAGE	49
1.2.4	NO.=1790260606	DIRECTORMISSING, CHOOSEMAXAGE	40
1.2.4	NO.=1840260616	DIRECTORMISSING, CHOOSEMAXAGE	41
1.2.4	NO.=1430110609	DIRECTORMISSING, CHOOSEMAXAGE	44
1.2.4	NO.=1430110209	DIRECTORMISSING, CHOOSEMAXAGE	47
1.2.4	NO.=1570240952	DIRECTORMISSING, CHOOSEMAXAGE	62

表中 1.2.4 表示 1.2 問項④「農業工作指揮者」，其資料全以只能圈選一個，且至少須有一個，故全以該欄資料皆空白也 (DIRECTOR MISSING)，選擇該項年齡在九十歲以下最高者，自動更正為「۹」，故第一筆資料 (NO.=1800260714) 選擇該項農業人口年齡四十九歲者為農業工作指揮者。

(五)

1.2.4.1	EDUCATION	NO.=2870101803	0
1.2.4.1	EDUCATION	NO.=2970011201	0
1.2.4.1	EDUCATION	NO.=3290310301	0

表中 1.2.4.1 表示 1.2 問項④「農業工作指揮者」的附帶問題「指揮者的教育程度」。當此欄資料空白時，自動更正為「۰」小學程度，理由已於上述。故表中所列資料其 1.2.4.1 問項皆應加填「۰」。

(六)

1.2.5.6	NO.=240360508	2 DAYS400
1.2.5.6	NO.=240360601	2 DAYS902

1.2.5.6 NO.=240360603 1 DAYS802

表中 1.2.5.6 表示 1.2 問項⑤「參加自家農業工作」與⑥「標準工作天數」兩項資料。依檢誤條件⑥不應大於 365 天，而表中資料的天數皆超過 365 天，顯不合理，須依⑤的資料，加以更改。故前兩筆資料 (NO.240360508 與 NO.240360601) 的②為「2」的工作幫助者，依此資料自動更正⑥為「100」天，第三筆資料 (NO.240360603) 的⑤為「1」主要工作者，乃自動將⑥更正為「200」天。

(k)

1.4.	NO.=1520890507	0	1.2.7	NO.=1520890507	0
1.4.	NO.=1680140611	3	1.2.7	NO.=1680140611	0
1.4.	NO.=1680140703	3	1.2.7	NO.=1680140703	0
1.4.	NO.=1730020607	1	1.2.7	NO.=1730020607	1
1.4.	NO.=1770180111	2	1.2.7	NO.=1770180111	0

表中 1.4 表示 1.4 問項維持家庭主要收入來自農業或農外。1.2.7 表示該且 1.2 問項⑦「從事其他工作者」，此兩者互相關。1.4 問項資料型態應為「1~3」，倘出現空白時，依 1.2.7 有無資料判定，表中第一筆資料 1.2 問項⑦無資料，表示無從事其他工作者，故更正 1.4 為「1」表專業農戶。同理，1.2.7 無資料時，1.4 只可能為「1」的情形，故第二、三筆資料，其 1.4 資料「3」應更正為「1」，又當 1.2.7 有資料時，1.4 不可能出現「1」，故第四筆資料 1.4 問項的「1」，應更正為「2」。

(l)

1.2.7 NO.=29 160502 K6=293 K7=1

表中 1.2.7 表示 1.2 問項⑦「從事其他工作者」，K6 表示⑥「標準工作天數」，K7 即為 1.2.7 的資料。依檢誤條件⑥、⑦皆有資料時，⑥工作天數超過 200 天，而⑦又為「1」從事其他工作的固定受雇者，顯然互相矛盾，故將⑦改為「2」

」臨時工作者。報表中K7資料，應更正為「2」。

(4)

1.2.8	NO.= 590120703	K1,K2,K3,K4,K5,K6,K7,K8=4	2	16	0	0	0	0
1.2.8	NO.=2710140404	K1,K2,K3,K4,K5,K6,K7,K8=1	2	51	0	0	0	0
1.2.8	NO.= 5	60405	K5,K6,K8=2	120	9			
1.2.8	NO.= 770120605		K5,K6,K8=2	20	9			
1.2.8	NO.= 710030101	K5,K6,K8=2	15	9				

表廿 1.2.8 表示 1.2 問項⑧「其他人口」，4,5,6,7,8，歲，K1,K2,K3,K4,K5,K6,K7,K8 與 1.2 問項①「稱謂」，②「性別」，③「年齡」，④「農業指揮者」，⑤「參加自家農業工作」，⑥「工作天數」，⑦「從事其他工作」，⑧「其他人口」等各欄資料。依檢誤條件④、⑤、⑥、⑦，皆空白時，才圈⑧。反之，任一欄有資料，則⑧應為空白，故後三筆資料⑤、⑥皆有資料，則⑧自動更改「9」為「好丘」。又若④～⑧全空白，則依②、③資料來更正。倘③「年齡」在十九歲以下或六十五歲以上，則更正⑧為「9」，故表中第一筆資料的K8 應自動更正為「9」，反之，年齡在十九～六十五歲者，再依②「性別」判斷，若為「1」男性，則K5 改為「2」，K6 改為「100」，若為「2」女性，則改K8 為「9」，故第一筆資料中K8 應更正為「9」。

第一節 土地及灌溉

(1)

2.1.A	NO.=1240061305	124	40	0	0
2.2.A	NO.=1240061305	131	8	0	0
2.2.A	NO.=1740230710	133	5	5	5

表中 2.1.A 表示 2.1 問項耕地每一類土地橫列之資料，2.2.A 表示非耕地每一類土地橫列之資料。第一筆資料表示「平

地旱田」（試查表編號 124）的合計欄為「40公畝」，而「自有土地」與「租入土地」皆無資料。第一筆資料表示「建築基地及晒場」（試查表編號 131）的合計欄為「∞公畝」，而「自有土地」與「租入土地」皆無資料。第二筆資料表示「山林地」（試查編號 133）的合計欄為「5公畝」，而「自有土地」與「租入土地」面積亦各為「15公畝」，依檢誤條件，第一欄數字應與後兩欄之和相等，故上述錯誤資料皆由機器挑出，再由人工更正。

(二)

2.1.B	NO.=1320120703	120=	50	50	0	OTHER=	10	10	0
2.1.B	NO.=1430110507	120=	70	70	0	OTHER=	0	0	0
2.1.B	NO.=1740230407	130=	12	4	8	OTHER=	4	4	0
2.1.B	NO.=1760040421	130=	30	30	0	OTHER=	3	3	0

表中 2.1.B 表示耕地面積縱行之合計數，「120=50,50,0」表示耕地合計欄（試查表編號 120.）的「合計」、「自有土地」、「租入土地」，各為「50」，「50」，「0」，表中「OTHE=10,10,0」表示三欄縱行各細數項之和，依檢誤條件，「120=」的三個數字與「OTHER=」的三個數字對應項須一致，否則由機器挑出，人工更正。

(三)

2.5.A	NO.=2480081008	149=	100	150+151+152=	0	
2.5.A	NO.=29	160502	149=	240	150+151+152=	50
2.5.A	NO.=2710140502	149=	20	150+151+152=	15	
2.5.A	NO.=2710141202	149=	20	150+151+152=	12	

表中 2.5.A 表示 2.5 問項的灌溉耕地面積，合計欄（試查表編號 149）的數字，應與「150+151+152」的細數項之和相等，倘不相等，由機器挑出，人工更正。

2.5.B	NO.=3570060203	149=	100	120=	96
-------	----------------	------	-----	------	----

2.5.B	NO.=3570060208	149=	100	120=	82
2.5.B	NO.=3570060908	149=	73	120=	62
2.5.B	NO.=3490010400	149=	70	120=	42

表中 2.5.B 的「149」表示灌溉耕地面積的合計數，「120」表示 2.1 問項中耕地面積的總計，前者應小於或等於後者，第一筆資料 (NO.=3570060203) 的誤差為「 $100 - 96 = 4$ 」公頃，則先由「其他水源」減去「4」，若仍不夠，再由「水利會輪水」減不夠之數額。原則上資料應先進行上一檢誤條件「 $149 = 150 + 151 + 152$ 」無誤之後，再進行此一檢誤條件。

(四)

3.2.	NO.= 560130909	167	740
3.2.	NO.= 540081306	165	104
3.2.	NO.= 630080101	168	103

表中 3.2 表示 3.2 問項使用自家農機具情形，應與 3.1 問項農戶所有的農機具相配合。第一筆資料「中耕除草使用自家農機具」的機耕面積有資料 (代號 167)，與 3.1 問項的「耕耘機」、「曳曳機」也應有資料，而第一筆資料的「耕耘機」、「曳曳機」並無資料，故應將「中耕除草使用自家農機具的 740 公頃」改為「空白」，而在「中耕除草包給他人機耕」的面積，加上 740 公頃。

第三節 作物產銷量

2.1	NO.=1720100611	2011=	5136	2012=	5778	2014=	0	2015=	0
2.1	NO.=1720100201	2011=	3600	2012=	1000	2014=	0	2015=	0

表中「2011」表示 1 期作水稻收穫量，「2012」表示 1 期作水稻收穫量「2014」表示全年自用量，「2015」表示全年出售量，當「2011」或「2012」有收穫量時，不可能自用量與出售量皆無資料，依檢誤條件，則將「2015」加填「 $2011 + 2012$ 」的合計數。故第一筆資料的「2015」應更正為「10914」，第一筆資料則更正為「4600」。

以上所列爲試查資料經過邏輯檢誤後部份錯誤資料，除須人工判定更正之項目外，皆爲機器自動更正項目。本文測試僅以文字說明更正後之正確資料型態，若將來實際檢誤時，可依上述原則直接更正磁帶上之錯誤資料，印出正確結果的資料。

戊、測試結果之檢討

邏輯檢誤工作，應儘量利用機器自動更正處理，但仍不免有些項目須以人工加以判定更正，例如耕地面積之橫計、縱計與細數項不符時，或合計欄有資料而細數項無資料、究竟合計項填錯或各細數項漏填，無法就該單一項目判斷，仍須人工判斷。此時應儘可能將錯誤原因以代號表示，並列出相關問項資料，以利人工直接利用檢誤表判斷。人工判斷更正工作，應分組爲之，每組專就某一項目處理，以期熟能生巧，加速更正時效。

肆、迴歸檢誤法

甲、設計目的

迴歸檢誤是指利用統計方法上迴歸方程式之建立，將普查表上某項資料所發生之偏差，進行合理客觀之更正以求其精確性。

民國五十四年普查項目中之農作物收穫量一項，由於農民顧慮稅負加重，調查結果數字偏低甚多，連帶影響其他基本項目之正確性。因此五十九年普查時將此項目刪除。至民國六十四年普查時，又引起一番爭論。反對者認爲收穫量不易調查確實，往往偏低甚多，影響其他資料確實性，徒增普查工作之負擔。而贊成者則認爲依世界糧農組織FAO之建議，各國皆已列爲主要調查項目，爲配合國際間比較研究，應增列收穫量項目，至於結果偏低現象，可採預防方法補救之。經過多次討論結果，決定增列此項目。至於其補救措施有下列七項：（註三）

（一）六十四年農業普查係採抽樣調查，樣本戶數少，可精選優秀調查員辦理。

〔二〕加強調查人員訓練酌予延長一天。

〔三〕事先蒐集各鄉鎮區「各種作物之最高、最低及平均產量表」，作為調查時之參考。

〔四〕運用生產面、支出面及訪問技巧等調查技術，加以反覆推敲印證。

〔五〕重新規定調查員、指導員、審核員及督導員等四級審核方法。

〔六〕結果仍有偏低可用「迴歸調整法」加以調整修正。

〔七〕實施必要之控制複查，以考查調查偏誤程度。

上述設計可謂非常週全。至於是否應增列「收穫量」則不在本文討論之列，本文所欲討論之重點在於增查收穫量後，如果發生離奇偏誤，應如何檢誤更正。雖然六十四年普查已提出「迴歸調整法」但實際作業時，僅以最高、最低產量為正確資料上下限，而以平均產量更正超出此標準者，並未採用「迴歸調整法」修正。本文即試圖建立迴歸檢誤法以修正錯誤的收穫量資料。

乙、各種檢誤條件及更正標準之比較

統計學上，根據已知變數預測未知變數，常以迴歸程式研究之。所謂迴歸檢誤法主要透過尋找下列問題之答案而建立的：

〔一〕何謂錯誤資料：錯誤資料如何找出？

〔二〕針對錯誤資料如何更正之？

一、檢誤條件的建立（第一個問題）

由於迴歸檢誤法係為更正偏誤的收穫量而設計，所以首先判定普查所得資料中那些是錯誤的資料。

民國六十四年普查，係計算整個普查地區內各鄉鎮區各種作物平均最高產量，平均最低產量及平均產量之平均數，作為檢誤收穫量之條件，亦即全部普查地區採同一檢誤標準。凡普查所得收穫量超過平均最高產量或低於平均最低產量，即為錯誤資料，此為目前所行之方法。本文擬利用統計為上常態分配之觀念，建立另一種檢誤條件。

由於宇宙間各種現象常呈常態分配或近似常態。因此土地條件相近者其收穫量均有向平均數集中之趨勢。產量過低或过高僅為少數。根據常態分配之性質（註四）樣本值若在平均數上下一個標準差之間者，佔總樣本數的六八·二六%，若在平均數上下三個標準差之間者，佔總樣本數的九九·七二%，因此可知若有單位產量超過平均產量三個標準差以外者，可謂離奇偏誤，因其機率只有〇·二八%。

所以本文所採之檢誤條件為以各種作物之平均單位產量加減三個標準產的範圍為判定標準。超過此標準，可謂收穫量偏高或偏低，可利用電腦計算比較，列出收穫量偏高或偏低資料，然此資料仍非為更正對象。因為該戶該年度若遭受水、旱、蟲等災害可能導致偏低，或某區為芒果之特產區則該區芒果收穫量自會偏高。所以由機器列出偏高、偏低資料後、必須加以人工判定、視其調查表之附記欄有否附帶說明理由，若無附帶說明原因則此偏高、偏低現象為不合理，應列入錯誤資料更正。

二、更正標準之建立

六十四年普查以各種作物之平均單位產量為標準，乘上該戶種植該種作物面積，即為更正後的收穫量。六十九年普查，據行政院主計處普查會資料整理組之構想，超過最高收穫量之平均數者，按平均最高收穫量更正，低於最低收穫量之平均數者，按平均最低收穫量更正，有種植面積而漏填收穫量者，按各鄉鎮正常收穫量平均數更正。

除上述更正標準外，本文另提數種可行更正標準：

(一) 比例配賦法

將正確資料按單位產量大小分組，計算每一組資料個數所佔之比例。例如：

組別	一	二	三	四
單位 (公斤／公畝)	二〇—三〇	三〇—四〇	四〇—五〇	五〇—六〇
佔 正 確 資 料 所 比 例	一〇%	四〇%	三〇%	二〇%

最後則自動將錯誤資料由小到大按百分之十，四〇、三〇、二〇的比例配賦到每一組內，按該組的平均數為標準更正之。

〔二〕相似情形代替法

發現錯誤資料，可利用電腦程式之設計在正確資料中尋找與錯誤農戶生產條件相近之資料，即以該正確資料代替錯誤資料更正。

〔三〕迴歸檢誤法

迴歸方程式乃表示變數之間相互關係之數式，故可依據已知變數來推測未知變數。根據此理，可令作物收穫量為Y，與收穫量相關之因素很多，現以肥料施用量為自變數X，根據正確資料中收穫量與肥料用量之數值，求出 $Y = a + bX$ 式中的a與b後，即可依此方程式來更正錯誤資料，將錯誤資料之肥料用量代入方程式，即可求出合理之收穫量Y而更正之。

上述三種方法，迴歸檢誤最為複雜，但卻最合理客觀，可以考慮各別農戶情形。今日電腦科技進步，複雜計算公式可利用電腦迅速求出，因此本文擬選擇合理客觀之迴歸檢誤法加以探討，以為更正標準。

丙、迴歸檢誤法建立之步驟

關於以迴歸法來更正各種農作物收穫量的錯誤資料，季聲國博士曾在「民國六十九年同國服務綜合報告」中「六十九年臺灣地區農業普查資料檢誤與調整更正方法大綱」一文提到，基本估計之迴歸方程式可設計出二種分別適用短表與長表資料之直線迴歸方程式，本文參考其原理，加以修正，進行直線迴歸方程式之試驗，並加以擴充其他可能迴歸方程式之測試。

一、變數之選定

依變數當然為作物收穫量，至於自變數一方面考慮與收穫量有關之各種因素，一方面又須考慮該因素之資料來源。有些因素收集不易、或不易量化皆無法當作自變數，故最好從普查表中選定之。

六十九年普查全面普查採用短表，調查項目僅為基本及重要項目，抽樣調查採用長表，調查項目較為詳盡，故可採用之變數亦較多，原來擬定之自變數有四項，即作物種植面積，所在鄉鎮某種作物生產力排列名次，某種作物每公頃使用肥料量及

某作物每公頃平均使用農藥金額，其中除生產力名次由某種作物正常產量決定以外，其餘三項皆可由調查表獲得。但目前關於土地生產力之資料甚為缺乏，雖然農發會已於六十七年開始訂定，但目前尚未完成，無法採用。臺灣地區原有地目等則，應可為生產力之指標但因年代久遠，環境變遷，多已失實，故無法做為判斷標準（註五）。此外土地單位面積收穫量可為生產力最簡單而常用之衡量標準（註六），故本文擬利用正確資料單位面積收穫高低，排定名次，作為土地生產力名次。

至於短表所選用自變數，只有作物種植面積及土地生產力名次兩項。

二、迴歸模式之試擬

依變數收穫量與自變數之間的關係，究為直線關係或為曲線關係，尚不得而知，擬藉其他迴歸樣式，進行測試，以選擇一最佳模式。

在母體情況未知時，一般可由普查資料得知母體或由抽查資料推知母體。由於本試查僅抽取八百戶的樣本進行調查，故母體迴歸方程式（P R F）須藉樣本迴歸方程式（S R F）來推知。

$$\text{I} \quad \text{直線迴歸方程式} \quad \hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$$

$$\text{II} \quad \text{曲線迴歸方程式} \quad \hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3^2$$

$$\text{III} \quad \text{半對數迴歸方程式} \quad \hat{Y} = b_0 + b_1 \log X_1 + b_2 \log X_2 + b_3 \log X_3 + b_4 \log X_4$$

三、迴歸方程式之確立與評估

其中 X_1 = 作物種植面積（公畝）， X_2 = 土地生產力名次（以其代替土地等則）， X_3 = 肥料施用量（公斤）， X_4 = 農藥使用金額（元），因為 X_1 若為 0，則 b_0 必為 0，故可將其省略。原則上，各種迴歸方程式係依土地種類（水田、旱田）、作物種類、耕作時期（一期作、二期作、裏作）分別求出，式中 X_1 、 X_3 、 X_4 資料均可由調查表上獲得，但須為正確無誤資料。而 X_2 （土地生產名次）須利用調查表上各種作物之單位產量間接求得。至於 b_0 、 b_1 、 b_2 、 b_3 、 b_4 各係數則須由正確資料依最小平方法估計求得，亦即使收穫量之觀測值 Y 與收穫量之估計值 \hat{Y} 之間誤差之平方和為最小；求得最適合線。當同一種作物的直線迴歸、曲線迴歸、半數對迴歸方程式皆求得後，即以 F 考驗測知複相關係數（R）是否超過所要求顯著性水準之臨界值，

然後在 R 已達顯著水準模式中，選擇一判定係數 (R^2) 最大，或估計標準誤（註七）最小之模式為最佳之模式。

四、錯誤資料之更正

確定一最佳模式，即可將偏誤資料農戶之種植面積、肥料施用量、農藥使用金額、土地生產力名次代入模式中，即可計出合理收獲量並更正之。

以上為原則性說明、實際測試結果於下面詳論。

丁、測試結果

一、測試範圍

原則上每種作物依其耕作時期皆應建立一迴歸方程式，現因作物種類多及受資料限制，故^{4.1}問項短期作物水田選擇稻米、甘蔗，旱田選擇甘蔗。其中稻米分為六十九年第二期（秋作）（註八），六九年第一期（春作），甘蔗分為六十八／六九年裏作，六九年／七十年第二期作，六十九／七十年第一期作，六十九／七十年裏作及七十／七一年第一期作計十二種類型。^{4.3}問項長期作物則水田、旱田皆選擇香蕉與鳳梨，計四種類型。

二、資料轉錄

資料轉錄結果，發現甘蔗有許多筆有種植面積，無收獲量之資料，因甘蔗為一年至一年半生作物，其生長期常跨越數期，而磁帶上並無存錄特殊符號，致無法判斷究為漏填或跨越數期之情形。鑑於上述原因，乃選擇甘蔗作物資料較齊全之六十八／六九年期裏作，分水田旱田進行測試。長期作物水田無資料，故只有旱田的香蕉、鳳梨進行測試。由上所述，資料轉錄後，真正進行測試只有水田稻米二期作、一期作、甘蔗六十八／六九年裏作及旱田甘蔗六十八／六九年裏作。園藝作物為旱田的香蕉、鳳梨，總計六種類型。

三、分組檢誤

由於各地生長條件不同，故各地作物收獲量亦有高低之分，如僅建立一個全國單位面積收獲量之檢誤標準，其檢誤標準最

高與最低間差距必大。所以本文將調查區分為數組分別按作物種類建立檢誤標準，分組方法係將試查的八百戶，先按縣市分組，原則上該縣市調查戶數若少於十六戶者併入相鄰縣市，惟臺東縣因地理環境較獨立，故單獨編為一組。若該縣市戶數多於五十戶者，將地理區位相近且土地類別相似之鄉鎮併為一組，總計分為二十四組。

茲以二期作水稻為例說明，分組之後，即以組別為單位，先除去產量為○的資料，然後計算單位面積收獲量之平均數與標準差，將平均數加減三個標準差之範圍即為檢誤標準，凡超過此標準為錯誤資料；錯誤資料經查對原調查表之附記欄皆無特殊原因說明，故此錯誤資料即為須更正之錯誤資料。

四、迴歸方程式之建立與評估

本試查係用抽樣調查表，故應採用長表適用之迴歸方程式，由於迴歸方程式不僅按作物別，且按耕作時期分類故同為水稻則有二期作與一期作之分，而試查表上肥料與農藥用量，僅列「稻作用」一種，並無分別列出一期作與二期作各施用多少，故無按耕作時期建立迴歸方程式。甘蔗與鳳梨亦遭遇同樣問題。如果將一期作與二期作資料合併建立「水稻用迴歸方程式」其所預估為水稻總收獲量，事實上一期作與二期作收獲量各應更正多少則無法判定。由於上述原因，只能選取作物種植面積及土地生產力名次兩項變數，建立短表用迴歸方程式。

迴歸模式之建立，係利用 BMD02R 逐步迴歸方程式，根據此法，自變數進入方程式的次序，是以誤差平方和減少最多者為原則，亦即與依變數淨相關最大者最先進入方程式，除此之外，變數亦可以強制方法控制其進入之順序。本文係迴歸方程式之不同，事先控制進入方程式之自變數個數，再依其中與依變數相關最大者最先進入方程式，餘此類推，其最後結果即為本文所探討的複迴歸方程式，以下所列者為逐步迴歸法最後所建立的複迴歸方程式。

$$\hat{Y} - \text{作物收獲量之估計值 (公斤)} = \text{因變項}$$

$$X_1 - \text{作物種植面積 (公頃)}$$

$$X_2 - \text{土地生產力名次 (1 2 3 24)}$$

$$R - \text{複相關係數}$$

R²—判定係數

SEE—估計標準誤

F—變異數分析表中的 F 值

N—樣本總數

茲將輸出報表之結果節錄於下

(+)二期作水稻迴歸方程式 (N=594)

1. 直線迴歸方程式

$$\hat{Y} = b_1x_1 + b_2x_2$$

$$\hat{Y} = 35,44067x_1 + 29,59787x_2$$

$$R = 0.9772$$

$$R^2 = 0.9548$$

$$SEE = 725.1777$$

$$F = 6257.340$$

2. 半對數迴歸方程式

$$\hat{Y} = b_1 \log x_1 + b_2 \log x_2$$

$$\hat{Y} = 887.35889 \log x_1 - 275.67311 \log x_2$$

$$R = 0.8603$$

$$R^2 = 0.740$$

$$SEE = 1739.7451$$

$$F = 842.625$$

3.曲線迴歸方程式

$$\hat{Y} = b_1x_1 + b_2x_2 + b_3x_1^2$$

$$\hat{Y} = 34.43652x_1 + 31.66333x_2 + 0.00457x_1^2$$

$$R^2 = 0.9772$$

$$R^2 = 0.9549$$

$$SEE = 725.0422$$

$$F = 4173.527$$

首先以變異數分析表的 F 值考驗複相關係數 R 是否達顯著性水準。在百分之五的顯著水準 $F_{0.05}(2, \infty) = 3.00$ $F_{0.05}(3, \infty) = 2.60$ ，顯然上述三個模式之 F 值均已超過其自由度下之臨界值，故皆已達顯著水準。其次比較 R^2 大小，以第三個模式之 $R^2 = 0.9549$ 為最大，由此可知可正確預測到 Y 的變數總變異的 95.49%，其餘 4.51% 是無法利用此迴歸方程式正確預測，所以評估模式時以 R^2 越大越好（註九）。經上述評估結果，以曲線迴歸方程式最佳，直線迴歸方程式次之，但兩者相差不大。但曲線迴歸較直線迴歸多一項 x_1^2 由變數，故直線迴歸方程式可謂曲線迴歸方程式之第二步的結果，從曲線迴歸方程式輸出報表最後的總表顯示，第二步的 R^2 為 0.9548（亦即直線迴歸方程式之 R^2 ）第三步的 R^2 為 0.9549（亦即曲線迴歸方程式之 R^2 ），茲以下列公式考驗 R^2 之增量是否達顯著水準：

$$F^* = \frac{(R_L^2 - R_s^2)/(m_L - m_s)}{(1 - R_L^2)/(N - m_L - 1)}$$

式中 R_s^2 = 第 i 步的判定係數 ($i=1, 2, \dots, N$)

R_L^2 = 第 $i+1$ 步的判定係數

m_s = 第 i 步的自變數個數

m_L = 第 $i+1$ 步的自變數個數

N = 樣本總數

$$F^* = \frac{(0.9549 - 0.9548)/(3-2)}{(1-0.9549)/(594-3-1)} = 1.308$$

在百分之五顯著水準 $F_{0.05}(3-2, 594-3-1) = 3.84$ ，由於 $F^* < F$ ，表示曲線迴歸之 R^2 雖較直線迴歸之 R^2 大，但所增加之 R^2 並未達顯著水準，亦即曲線迴歸方程式增加一項自變數 X_1^2 後，並未加強對收獲量之預測能力，若要以最少變項達到最大預測效果，則只需進行到第一步即可，且直線迴歸方程式之計算亦較簡便，故二期稻作收獲量以直線迴歸方程式為更正標準。

II 一期作水稻迴歸方程式 (N=526)

根據相同理論，採直線迴歸方程式為更正標準

$$\hat{Y} = b_1x_1 + b_2x_2$$

$$\hat{Y} = 39.07538x_1 + 48.39x_2$$

$$R = 0.9711$$

$$R^2 = 0.9430$$

$$SEE = 952.3936$$

$$F = 4337.602$$

〔三〕水田甘蔗六十八／六十九年期裏作迴歸方程式 (N=19)

亦採直線迴歸方程式為更正標準

$$\hat{Y} = b_1x_1 + b_2x_2$$

$$\hat{Y} = 992.65552x_1 + 273.63574x_2$$

$$R = 0.9863$$

$R^2 = 0.9728$

$SEE = 8600.9141$

$F = 304.198$

國旱田甘蔗六十八／六十九年期裏作迴歸方程式 ($N=19$)

採曲線迴歸方程式爲更正標準

$$\hat{Y} = b_1x_1 + b_2x_2 + b_3x_1^2$$

$$\hat{Y} = 1409.771x_1 - 529.88135x_2 - 2.92782x_1^2$$

$R = 0.9816$

$R^2 = 0.9636$

$SEE = 17020.8008$

$F = 141.066$

國旱田香蕉迴歸方程式 ($N=12$)

試查表上長期作物有「種植面積」與「收獲面積」兩種，種植面積係指該長期作物在六十九年內已種植之總面積，收獲面積係指六十九年內已收獲總面積，凡已具採收價值，而在六十九年內尚未採收者，仍應包括在內。而種植面積可能包括已種植而未成熟者，故預測收獲量時迴歸方程式中的 x_1 在長期作物係指「收獲面積」而言。

採直線迴歸方程式爲更正標準

$$\hat{Y} = b_1x_1 + b_2x_2$$

$$\hat{Y} = 250.00256x_1 - 55.92081x_2$$

$R = 0.9168$

$R^2 = 0.8405$

SEE = 3249.512

110

F = 26.354

五、錯誤資料之更正

在第三階段分組檢誤時，已確定須更正之錯誤資料，一期作水稻有十筆，二期作水稻有八筆。現即可將錯誤資料之有關自變數之數值代入最適迴歸模式，求得更正後收穫量之數值，茲將錯誤資料更正前後之情形列表於後：

(一) 一期作水稻收穫量錯誤資料之更正

評估最適迴歸模式為 $\hat{Y} = 35.44067x_1 + 29.59787x_2$

評估最適迴歸模式為 $\hat{Y} = 39.07538x_1 + 48.39x_2$

一期作水稻迴歸檢誤更正表					
農戶編號	組別	種植面積(x_1)	土生產力(x_2)	調查表之偏誤收穫量(Y)	迴歸更正之收穫量(\hat{Y})
1520671106	9	168	14	80640	7242
2900160505	19	80	21	7600	4142
3570051107	23	20	3	7800	927
1720100612	12	121	17	10890	5551
1320120701	7	10	15	880	1117
1420011209	8	24	11	1800	1470
1240060704	5	30	9	1800	1608
1680191303	10	56	24	0	3350

二期作水稻迴歸檢誤更正表

農戶編號	組別	種植面積 (x_1)	土地生產力 (x_2)	調查表之偏誤收獲量 (Y)	迴歸更正之收獲量 (\hat{Y})
1800262008	13	20	19	1632	1271
3570051107	23	20	5	7600	857
1720100201	12	40	21	3600	2039
1520671106	9	168	7	70560	6161
1680191303	10	56	23	0	2665
1420011209	8	24	17	1900	1354
1320120701	7	10	15	750	798
3500011659	22	50	10	1300	2068
0770121604	2	47	4	2400	1784
2900160505	19	240	3	16200	8595

迴歸檢誤法經過實際測試結果，發現幾項問題，值得繼續研究。

戊、測試結果分析與檢討

一、檢誤條件之偏差

分組檢誤條件是以單位面積平均收獲量加減三個標準差為範圍，但某組內若有特別偏高或偏低資料，將使 \bar{X} 及 SD 增加，致 $\bar{X} \pm 3 \cdot SD$ 之範圍亦加入，錯誤資料也不易找出來，例如第九組二期稻作三八筆資料中有一筆高達四二〇公斤／公畝，另

有五筆在 10 公斤／公畝以下，若扣除此六筆資料不加入計算，則 $\bar{X}=39.3$ ， $SD=5.0$ ，若以全部資料計算 $\bar{X}=46.3$ ， $SD=62.11$ 。而且由於 X 之偏高或偏低將使根據 X 大小所排定之土地生產力名次亦不確實，所以應用迴歸檢誤法時，必須注意上述情形，某一組之 SD 與地理區位相近之組的 SD 相較之下，顯不合理時，可比較相鄰組之檢誤標準，挑出錯誤資料，再以正確資料，計算其土地生產力名次。

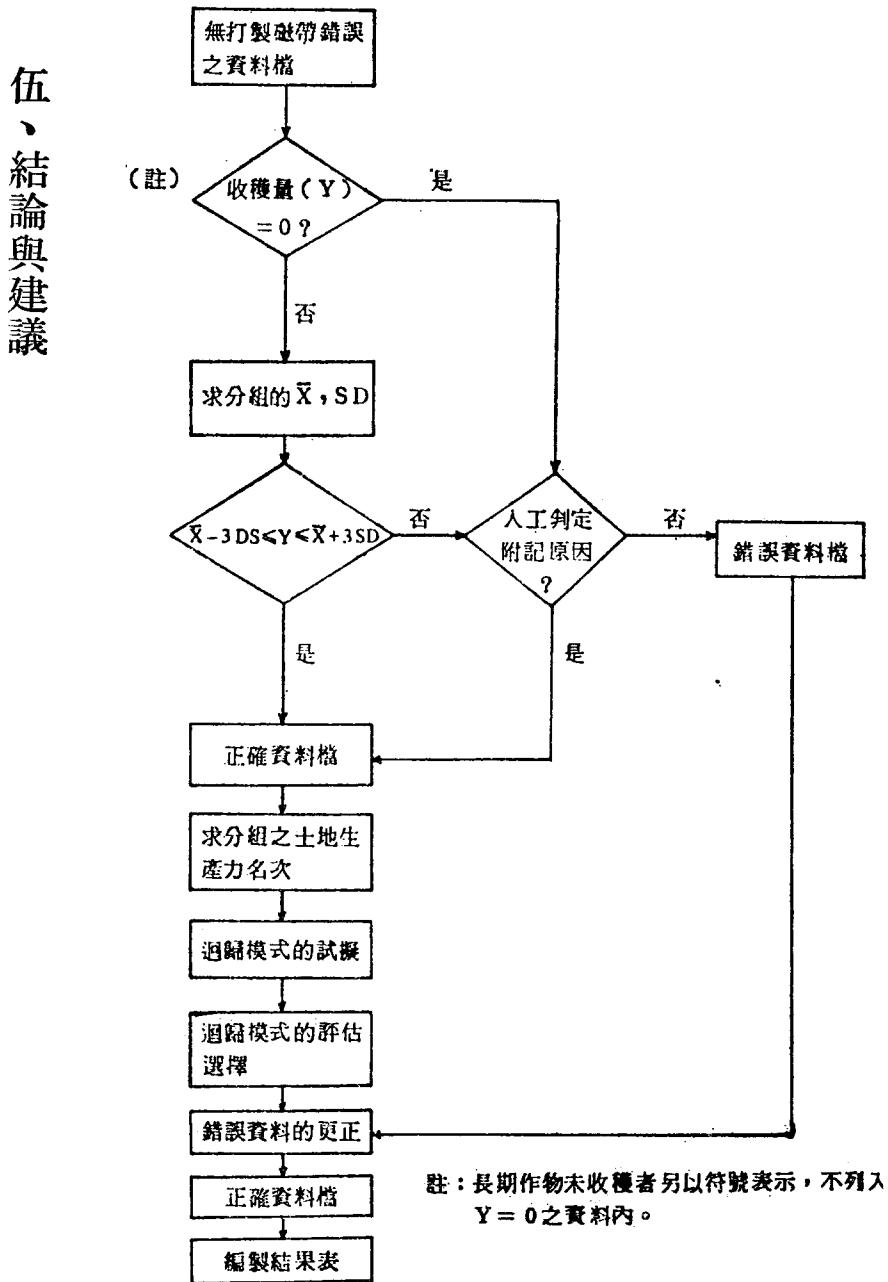
二、更正標準之信度

本文所討論之迴歸方程式，乃是按土地種類及耕作時期之不同，全省建立一更正標準，但此法仍未考慮到區域性條件之差異，因為南北地區只要農戶個別條件相似，則預估之收穫量亦相同，難免抹殺地區性作物生產之特性。故可將調查區分為北、中、東、南四區，每區各建立區域之迴歸方程式，甚至可細分為縣市或鄉鎮的迴歸方程式，則更可增加更正標準之可信度。

三、資料處理之時效性

進行測試時，利用電腦套裝程式，可即刻求出各類迴歸模式，惟因資料少，故探討評估模式的方法後，即以人工評選，並加以更正錯誤資料，將來全面普查，可將評估條件輸入電腦，由機器自動更改錯誤資料，除了由人工判定而無附記原因為外，均可利用機器大量處理。又將來若能以鄉鎮為單位建立各鄉鎮的迴歸模式，只要該縣市調查表點收齊全後，即可進行迴歸檢誤法，無須等全部資料收集後，才求全省各鄉鎮參考資料的平均數以作為檢誤條件及更正標準。故利用迴歸檢誤法，可將陸續送齊之調查資料，即刻交由機器處理，增進資料處理之時效性。

最後將迴歸檢誤法之整理流程圖表列於下：



伍、結論與建議

農業普查為基本國勢調查之一種，其主要目的為獲取精確可靠資料，以供農業決策之參考。故整個普查過程，非常重視資料的完整精確，不僅調查期間嚴行逐級審核制，並且在調查後要迅速整理以及更正錯誤。本文研究改進農業普查資料整理方法

，即欲採科學設備期能合理客觀進行錯誤資料之調整與估計。現就邏輯檢誤法與迴歸檢誤法之測試結果分析，提出下列幾點建議：

(一)普查資料存錄方式現採磁帶登錄方式，此法不僅耗費打驗製磁帶的時間，且會發生人爲打製錯誤。現歐美先進國家，多以光學閱讀機代替打卡法。即將調查表拍攝成縮影片，再經由特殊機器，直接將資料轉錄在磁帶上，此法不僅節省資料存放空間，且可免除人爲錯誤，縮減驗製磁帶時間，增進資料處理時效，故建議速購此項設備，以利將來資料庫之建立。

(二)普查資料邏輯檢誤方面儘可能加強電子計算機檢誤，並自動更正的作業，以減少人工更正困難，對於錯誤資料，除列出該欄資料之數值及錯誤原因外，並列出相關問項之資料，以便審核人員可直接就檢誤表判明更正。又邏輯檢誤條件必須在設計階段即予以測試修正，以便普查時大量資料之處理。

(三)迴歸檢誤法可利用套裝電腦程式，從試擬，評選最佳模式，到自動更正偏誤，全部機器處理。將來全面普查若能建立各鄉鎮之迴歸模式，不僅考慮個別條件，更顧及區域特性，而且可免除等待其他資料送達時間。

將來普查表上若調查時能將農戶使用之稻作用肥料、農藥，分別註明「二期作」及「一期作」之數量，甘蔗用肥料、農藥亦按耕作期分別註明，果樹用肥料、農藥則填上果樹種類，則肥料、農藥資料皆可選爲估計收穫量之自變數。甚至將來還可以將 1.2 項農戶工作人口情形，換算爲統一計算單位，亦可選爲一項自變數，凡此皆頗值得繼續研究探討。

(四)統計、電腦人才之培養：電腦化整理已爲現代普查趨勢，故普查單位平時即應注重人才之培養，使電腦設計、統計修養及農業常識合而爲一，則整個處理過程必能達到理想，否則統計人員與程式設計人員分立，爲互相協調配合或因財務問題或因時間因素，必須牽就讓步，使得原來設計理想，無法完全達成。且電腦人員未必了解普查資料性質，故設計程式前，必與普查統計人員多次商議。若普查統計人員本身即具有電腦方面訓練，則可順利設計適合機器處理之條件，減少不必要商討時間，無形中亦增加資料處理之時效性。

附 註

註 一・毛育剛、莊福典「農漁業普查之設計及檢討」，中國統計學報

19.2中國統計學社 70.4 P.2

註 二・吳清在「參加美國普查局統計訓練計畫報告」

主計月報 50.4 69.10 P. 29

汪 錦、黃子貞、張碧惠「出席國際統計專家會議暨考察瑞典

、丹麥、荷蘭、法國、西等等五國普查制度報告」主計月報

50.5 69.11 P.5。

註 三・「中華民國六十四年農業普查報告」，第一卷分析說明及綜合提

要表」P.6。

註 四・若非常態分配，至少可按“Chebysheff”定理求其分配。

$$S_{Y.XX_2} = \sqrt{S_{Y.1} - R^2_{Y.XX_2}}$$

註 五・林元興「臺灣土地地目等則問題之分析」

行政院研究發展考核委員會編印 67.5。

註 六・陳月娥「臺灣農業生產力之研究」

臺灣銀行季刊30.1 P.1。

註 七・估計標準誤 $S_{Y.X} = \sqrt{\sum(Y - \hat{Y})^2 / N}$ ，當母羣體 σ 未知，以樣本來估計時，須用下列不偏估計值 $\sqrt{\frac{N}{N-2}(Y - \hat{Y})^2}$

註 八・試查所調查實為六十八年作物收穫情形，但因本試查表乃模擬六十九年農業普查而設計的，故調查表上皆印為六九年，茲為說明方便，文中皆按調查表之格式。

註 九・若以估計標準誤來評估，誤差越小越好，由於估計標準誤 ($S_{Y.XX_2}$) 與複相關係數 (R^2) 之間關係為

$$S_{Y.XX_2} = \sqrt{S_{Y.1} - R^2_{Y.XX_2}}$$

式中 $S_{Y.XX_2}$ 與 R^2 成反比，故以 R^2 最大或以 $S_{Y.XX_2}$ 最小評估模式，其結果相同。