

財政部財政資訊中心綜合所得稅一般用途抽樣資料庫

抽樣方法及驗證規劃

一、緣起

因應新興科技及網路發展，多數先進國家已推動相關智慧政府發展計畫，共通性發展趨勢包含了「政府資料強化加值應用」、「資料分析優化政府決策」，財政部與時俱進，透過規劃建置能代表整體稅收環境之財稅抽樣資料庫，提供財稅政策分析及學術研究運用，藉以提升決策制定品質、活絡學術研究，激發稅務資料加值運用效益。

鑑於稅制改革統計及學界研究多以綜合所得稅為主，遂優先規劃提供綜合所得稅一般用途抽樣資料庫。

二、綜合所得稅一般用途抽樣資料庫建置規劃

(一)資料簡介

1. 一般統計用途資料庫包含我國綜合所得稅核定檔及相關核定分類檔(基本稅額、列舉明細、各類所得、基本稅額所得、核定稅籍)。我國課稅是以家戶為單位進行申報，所得課稅資訊來源於綜所稅申報核定檔，該檔案包含核定所得、各類扣除額、扣繳稅額與可扣抵稅額的總計資料，綜合所得稅一般用途抽樣資料庫建立於此檔案上。
2. 綜合所得稅申報核定檔為該年度核定家戶申報課稅資料，因此，倘若該課稅單位未達課稅申報門檻，即可不申報，故不會出現於該資料中。另外，該檔案並未包含以「個人」為單位的財產以及各類所得細項資料，如須進行更詳盡的財富分析，研究者可增加運用歷年所得資料檔或財產檔等資料，並以「個人」為主體，自行進行資料串連分析。

(二)抽樣設計

1. 分層簡單隨機抽樣：
 - (1) 抽樣母體綜合所得稅申報核定檔(IBET 100)為家戶單位申報資料，抽樣方式為分層抽樣，以「核定所得總額」為標準，將資料分為20個百分位數，即所得總額每5個百分位為1組。

- (2) 因資料變異程度易受極端值影響，因此先去除極端值（此定義為第 20 分位的前 1% 資料），以 95% 信心水準下依公式計算最高所得分位（即第 20 分位）所需樣本數，並驗證其抽樣結果於母體的代表性。為求樣本具代表母體的精確性並考量高所得資料最為離散，倘若驗證結果良好，則套用同樣樣本數至第 1 至 19 分位。
- (3) 驗證整體資料的抽樣結果，包含抽樣資料累積機率分布與母體資料的差異程度，並輔以 Kolmogorov-Smirnov 檢定（顯著水準 0.05）驗證抽樣資料的母體代表性。

2. 抽取樣本數：

- (1) 計算樣本數大小 (sample size)，首先依據信賴區間公式找出可接受之誤差範圍。此外，為求抽樣後，樣本特性符合母體特徵，定義誤差範圍為該分位所得總額平均數之 1%，以求精確。

$$\bar{X} \pm Z_{\alpha/2} * \frac{s}{\sqrt{n}} \quad (i)$$

$\bar{X}$ 為樣本平均數， s 為樣本標準差， α 為信心水準， $Z_{\alpha/2}$ 為 95% 信心水準下常態標準值， $Z_{\alpha/2} * \frac{s}{\sqrt{n}}$ 此項即誤差範圍（或稱誤差界限），後稱做 ε 。

- (2) 由於母體已知，並且若要從母體推論所需樣本數，將 $\bar{X}$ 以及 s 看做母體平均數與母體標準差，並計算誤差範圍 (所得總額平均數之 1%)。

(3)

經 (i) 式中移項後，各分位樣本數大小計算公式如下：

$$n \geq \frac{Z_{\alpha/2}^2 * s^2}{\varepsilon^2} \quad (ii)$$

- (4) 第 20 分位去除前 1% 極端值後，依公式計算所得樣本數為 11,851，取整數以 12,000 為樣本數。
- (5) 整體抽樣檔樣本數為 $12,000 \times 20 = 240,000$ 。
- (6) 整體抽樣檔依所得總額分位隨機分為 10 個小抽樣檔提供，即每個小抽樣檔為 24,000 筆家戶資料。

(三)驗證規劃

1. 驗證欄位

欄位名稱	中文名稱	備註
APL_INCM_T	核定所得總額	單位:千元
GENE_DED_AMT	一般或列舉扣除額合計	單位:千元
特別扣除額合計	特別扣除額合計	此欄位加總以下 IBET100 欄位: SLRY_DED (薪資所得特別扣除額) PROP_LOSS_DED (財產交易損失扣除額) SAVING_DED (儲蓄投資特別扣除額) DISAB_PERS_CNT (身心障礙人數) * 128,000 TUITION_DED (教育學費特別扣除額) CHILD_EDU_DED (幼兒學前特別扣除額)
APL_INCM_N	核定所得淨額	單位:千元
TPS_INCM_TOT	納稅義務人薪資所得總額	單位:千元
SS_INCM_TOT	配偶薪資所得總額	單位:千元
FLG_TPS_INCM_T	納稅義務人各類所得總額	單位:千元
FLG_SPOUS_INCM_T	配偶各類所得總額	單位:千元

2. 驗證方法：

- (1)觀察敘述統計值。
- (2)母體及抽樣資料累積機率分布比較。
- (3)Kolmogorov-Smirnov 檢定驗證抽樣資料的母體代表性。

3. 驗證檔案：

- (1)所得總額第 20 分位隨機抽樣 12,000 筆家戶資料(本階段實施 3 次抽樣)之母體代表性驗證。
- (2)整體抽樣檔即 240,000 筆家戶資料(本階段實施 3 次抽樣)之母體代表性驗證。
- (3)隨機提供 1 組小抽樣檔即 24,000 筆家戶資料之母體代表性驗證。

第 20 分位驗證資料

1. 敘述統計

核定所得總額 (apl_incm_t)				
組 別	平均值	標準差	中位數	樣本數
母體	4,670	20,562	3,298	313,974
抽樣 1	4,693	16,796	3,300	12,000
抽樣 2	4,625	12,612	3,283	12,000
抽樣 3	4,543	7,925	3,290	12,000
一般或列舉扣除額合計 (gene_ded_amt)				
組 別	平均值	標準差	中位數	樣本數
母體	346	3,783	238	313,974
抽樣 1	326	323	239	12,000
抽樣 2	328	417	234	12,000
抽樣 3	328	419	238	12,000
特別扣除額合計 (spe_cnt)				
組 別	平均值	標準差	中位數	樣本數
母體	351	151	334	313,974
抽樣 1	351	151	333	12,000
抽樣 2	348	151	329	12,000
抽樣 3	349	153	330	12,000
核定所得淨額 (apl_incm_n)				
組 別	平均值	標準差	中位數	樣本數
母體	3,560	17,798	2,303	313,974
抽樣 1	3,604	16,709	2,312	12,000
抽樣 2	3,539	12,431	2,289	12,000
抽樣 3	3,453	7,806	2,298	12,000
納稅義務人薪資所得總額 (tps_incm_tot)				
組 別	平均值	標準差	中位數	樣本數
母體	285	1,131	0	313,974
抽樣 1	282	1,047	0	12,000
抽樣 2	273	1,001	0	12,000
抽樣 3	285	1,020	0	12,000
配偶薪資所得總額 (ss_incm_tot)				
組 別	平均值	標準差	中位數	樣本數
母體	968	2,625	786	313,974
抽樣 1	966	2,465	778	12,000
抽樣 2	980	2,510	800	12,000
抽樣 3	955	1,589	796	12,000
納稅義務人各類所得總額 (flg_tps_incm_t)				

組 別	平均值	標準差	中位數	樣本數
母體	3,033	18,550	2,197	313,974
抽樣 1	2,902	5,755	2,222	12,000
抽樣 2	2,955	9,803	2,176	12,000
抽樣 3	2,938	6,593	2,188	12,000
配偶各類所得總額 (flg_spous_incm_t)				
組 別	平均值	標準差	中位數	樣本數
母體	1,529	5,787	1,107	313,974
抽樣 1	1,682	13,879	1,100	12,000
抽樣 2	1,576	5,386	1,118	12,000
抽樣 3	1,507	3,910	1,101	12,000

2. 資料分佈及 Kolmogorov-Smirnov 檢定 - 第 20 分位

(一) 所得總額

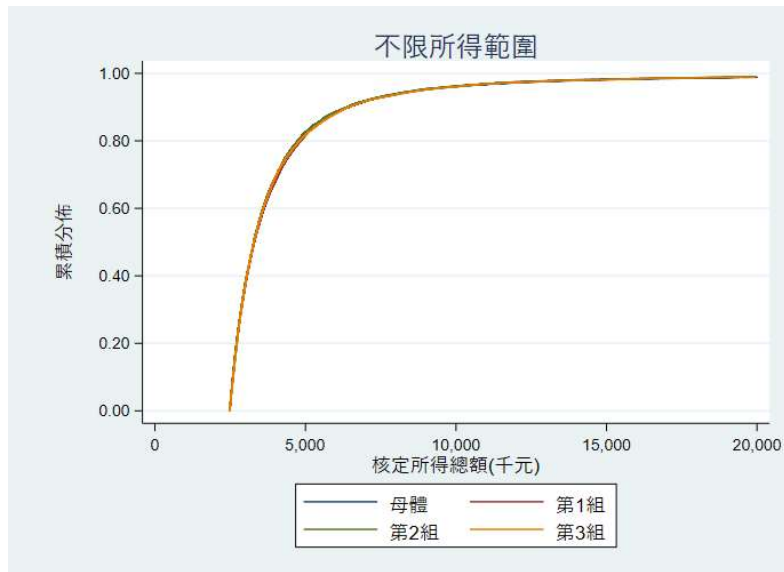


圖 1. ECDF : 所得總額 (0~2000 萬)

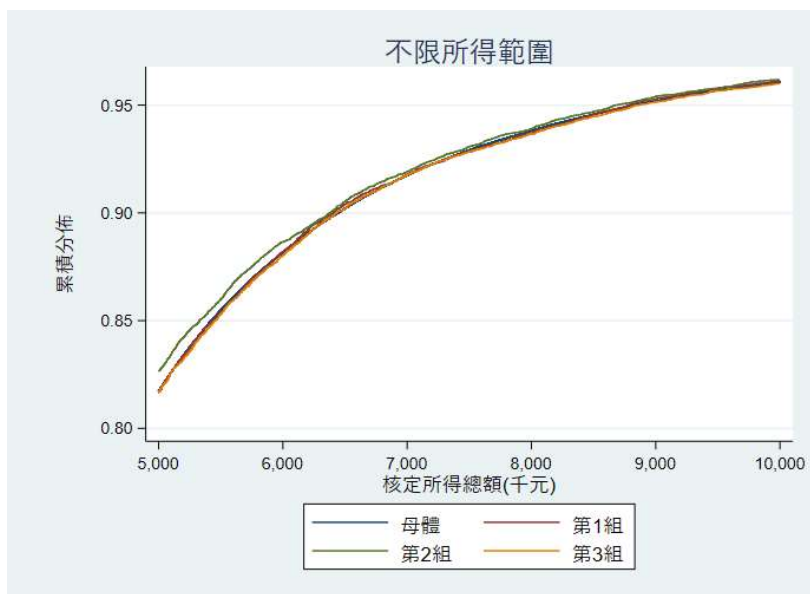


圖 2. ECDF : 所得總額 (500 萬~1000 萬)

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0070	0.326
第二組	0.0099	0.102
第三組	0.0084	0.197

(二) 標準扣除額

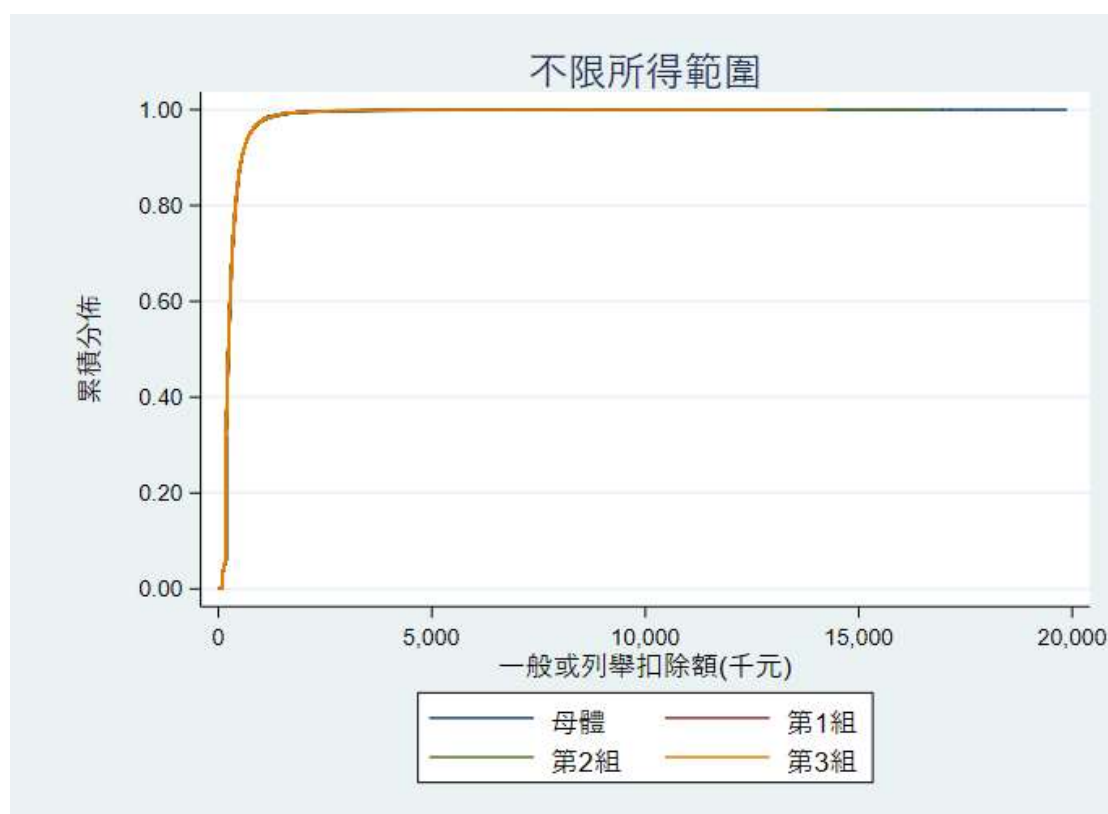


圖 3. ECDF：標準扣除額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0098	0.111
第二組	0.0118	0.040
第三組	0.0063	0.397

(三) 特別扣除額

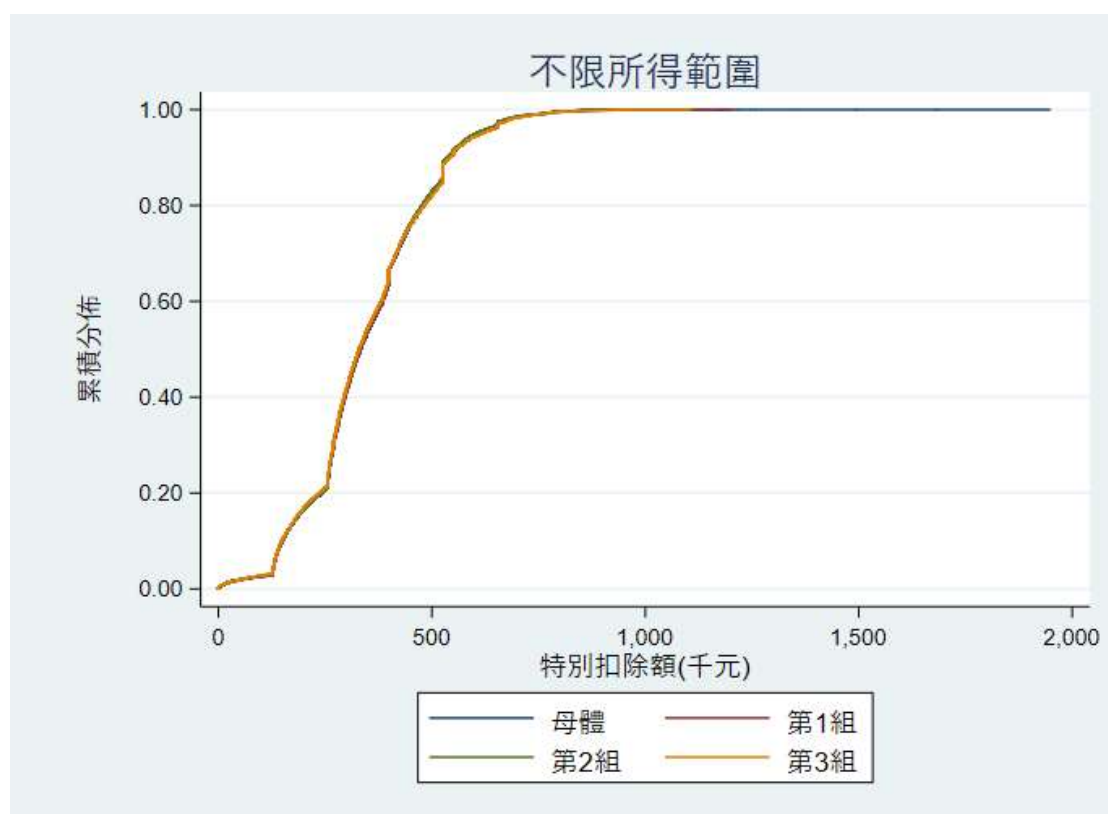


圖 4. ECDF：特別扣除額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0057	0.478
第二組	0.0119	0.038
第三組	0.0124	0.028

(四) 所得淨額

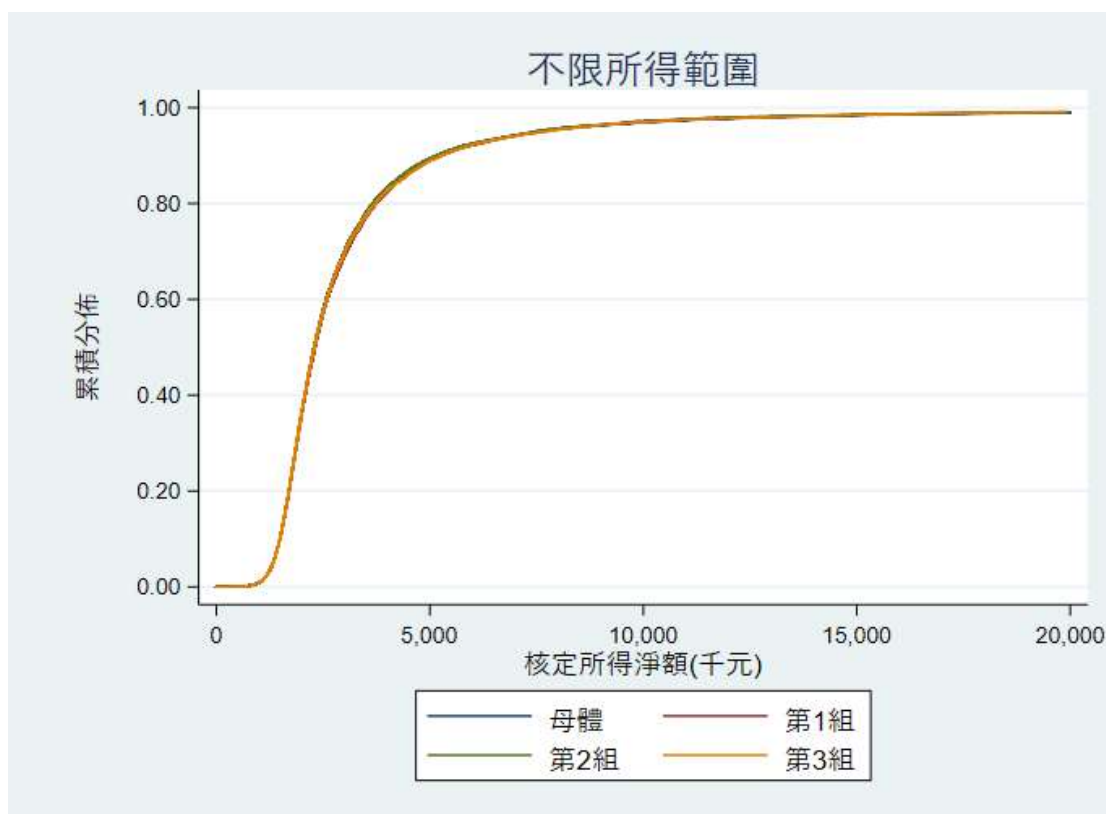


圖 5. ECDF : 所得淨額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0054	0.512
第二組	0.0096	0.118
第三組	0.0061	0.418

(五) 納稅義務人薪資所得額

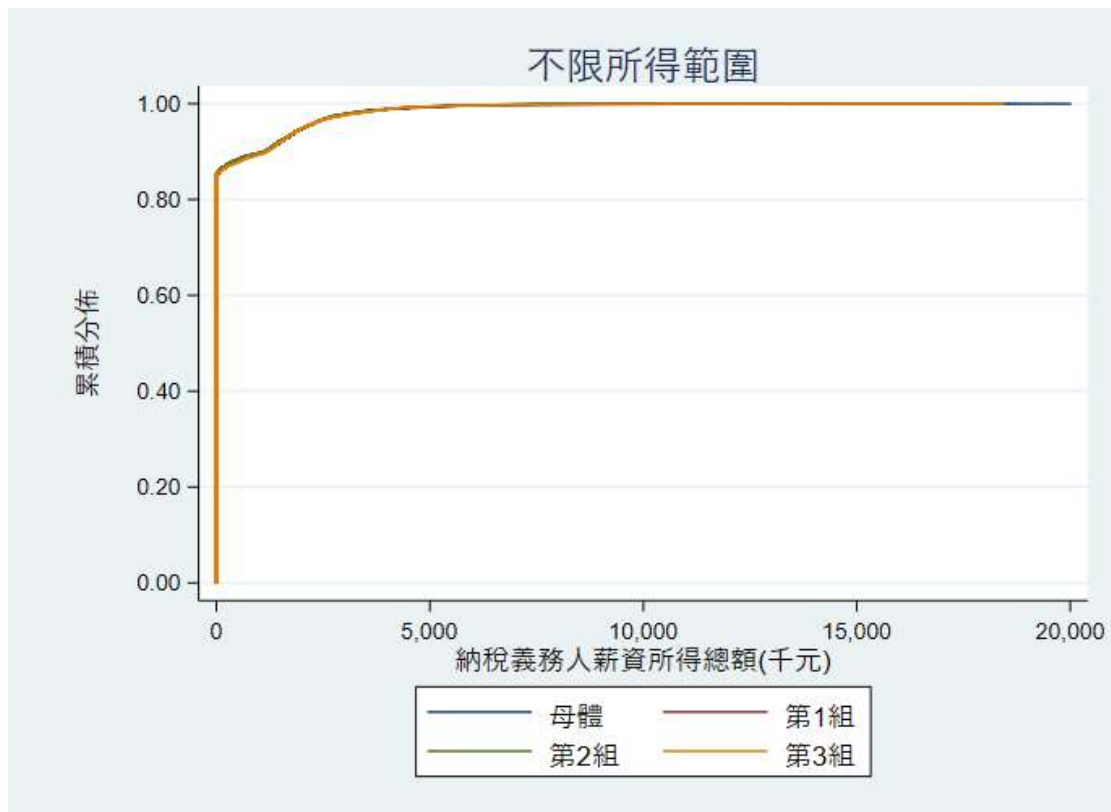


圖 6. ECDF：納稅義務人薪資所得額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0023	0.888
第二組	0.0059	0.451
第三組	0.0033	0.775

(六) 配偶薪資所得額

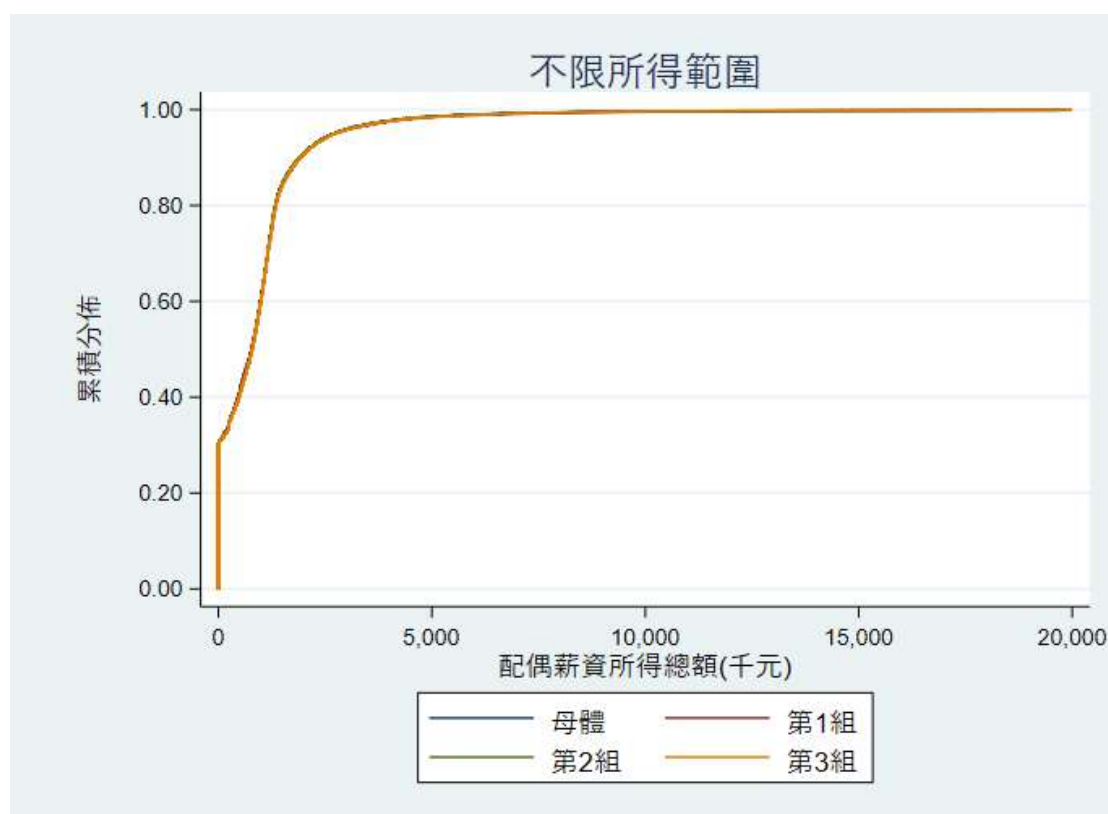


圖 7. ECDF：配偶薪資所得額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0089	0.163
第二組	0.0066	0.367
第三組	0.0090	0.154

(七) 納稅義務人各類所得額

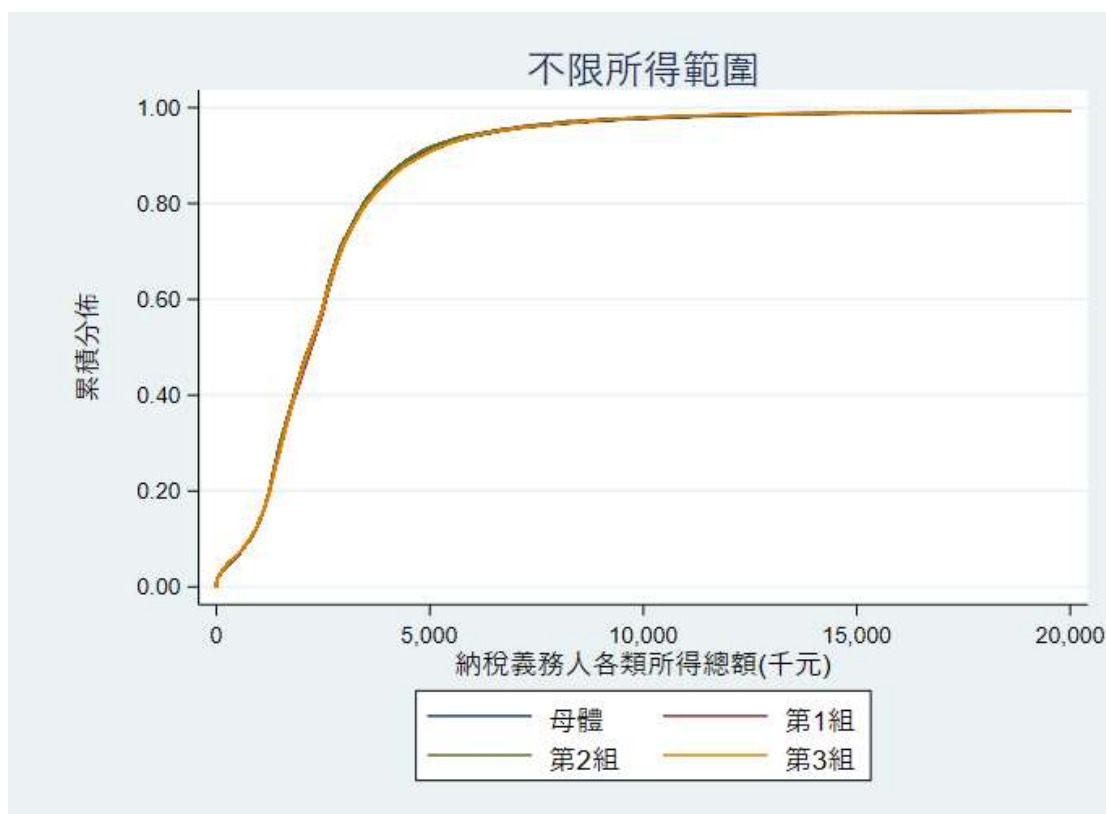


圖 8. ECDF：納稅義務人各類所得額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0086	0.180
第二組	0.0103	0.088
第三組	0.0072	0.304

(八) 配偶各類所得額

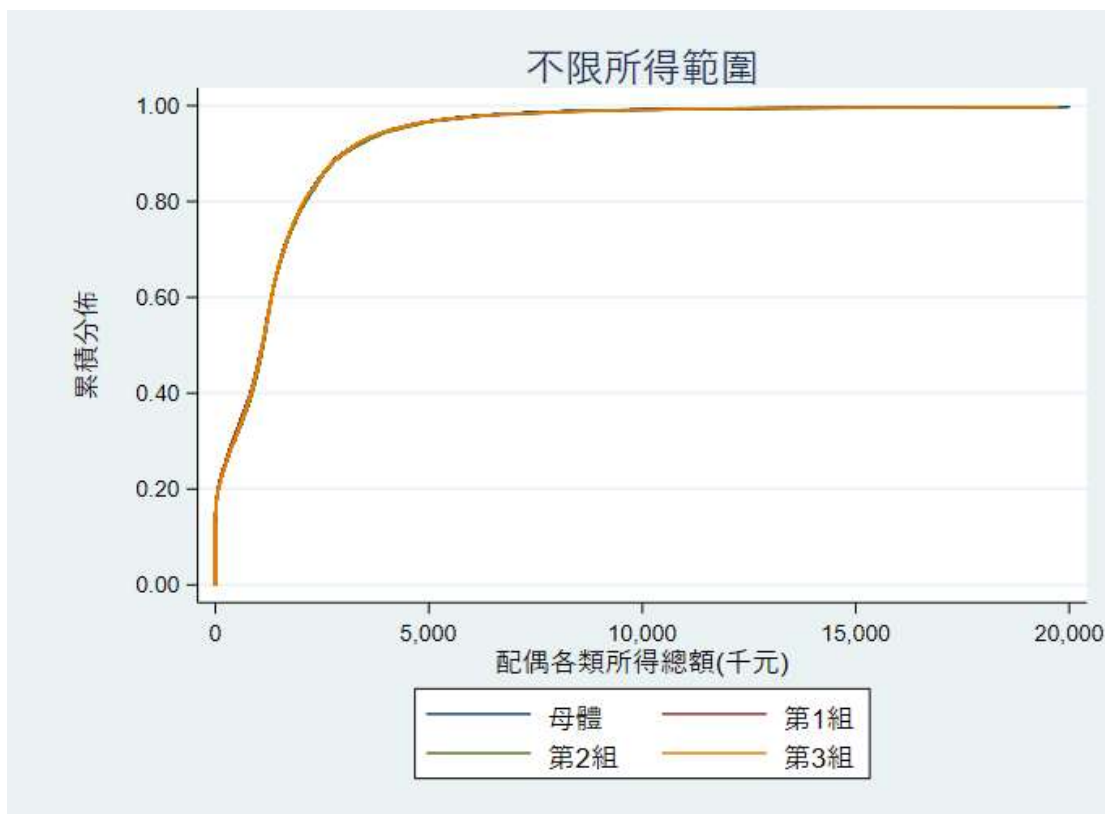


圖 9. ECDF：配偶各類所得額

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0088	0.168
第二組	0.0092	0.142
第三組	0.0043	0.651

第 20 分位驗證結果說明

欄 位	說 明
核定所得總額	所得 500 萬開始各組結果與母體有些許差異，但驗證結果仍算良好。
一般或列舉扣除額	驗證結果良好。
特別扣除額	驗證結果良好。
核定所得淨額	驗證結果良好。
納稅義務人薪資所得總額	驗證結果良好。
配偶薪資所得總額	驗證結果良好。
納稅義務人各類所得總額	驗證結果良好。
配偶各類所得總額	驗證結果良好。

➔ 第 20 分位所得驗證結果良好，各分位套用相同樣本數進行抽樣並驗證其結果。

整體抽樣檔驗證資料

(一) 所得總額

1. 敘述統計

X 不限所得範圍				
組別	平均值	標準差	中位數	樣本數
母體	923	4,705	611	6,279,484
抽樣 1	924	3,887	611	240,000
抽樣 2	921	2,989	611	240,000
抽樣 3	917	2,022	611	240,000
所得總額前 1%				
母體	10,440	45,507	6,444	62,794
抽樣 1	10,439	36,704	6,368	2,436
抽樣 2	10,550	28,059	6,482	2,294
抽樣 3	9,858	16,710	6,444	2,386
所得總額大於 1000 萬				
母體	27,141	100,847	14,629	12,357
抽樣 1	27,913	81,947	13,922	463
抽樣 2	27,326	59,926	14,497	458
抽樣 3	23,706	33,808	14,449	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0008	0.764
第二組	0.0006	0.863
第三組	0.0006	0.848

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0496	0.112
第二組	0.0302	0.446
第三組	0.0456	0.146

2. 資料分佈

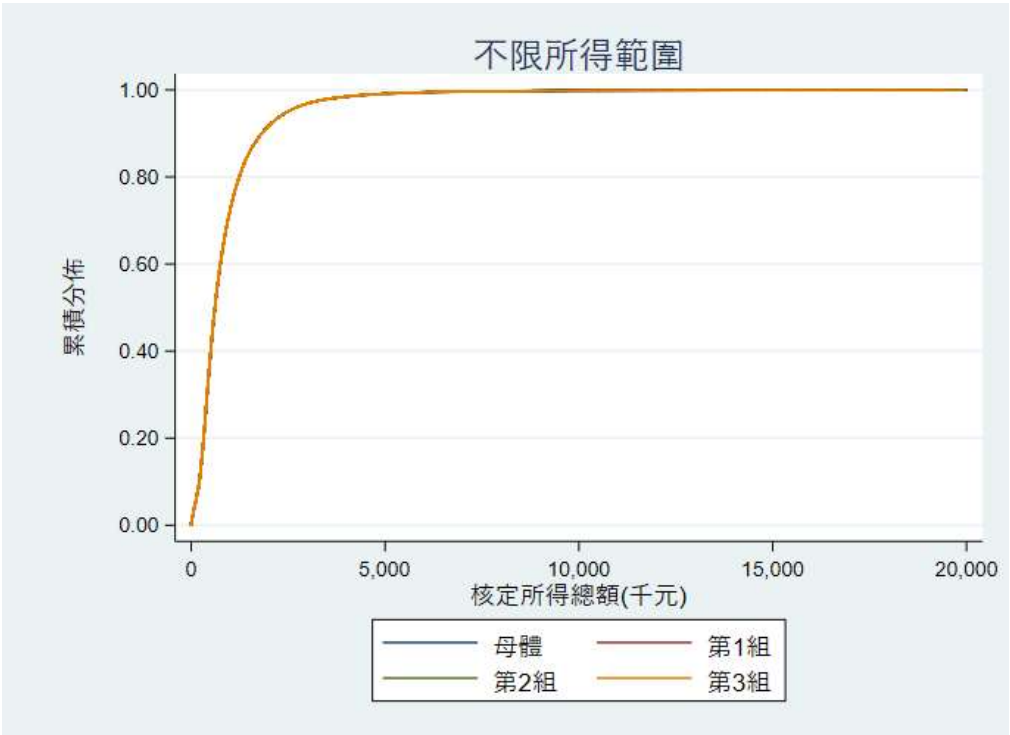


圖 10. ECDF : 所得總額 - 所有資料

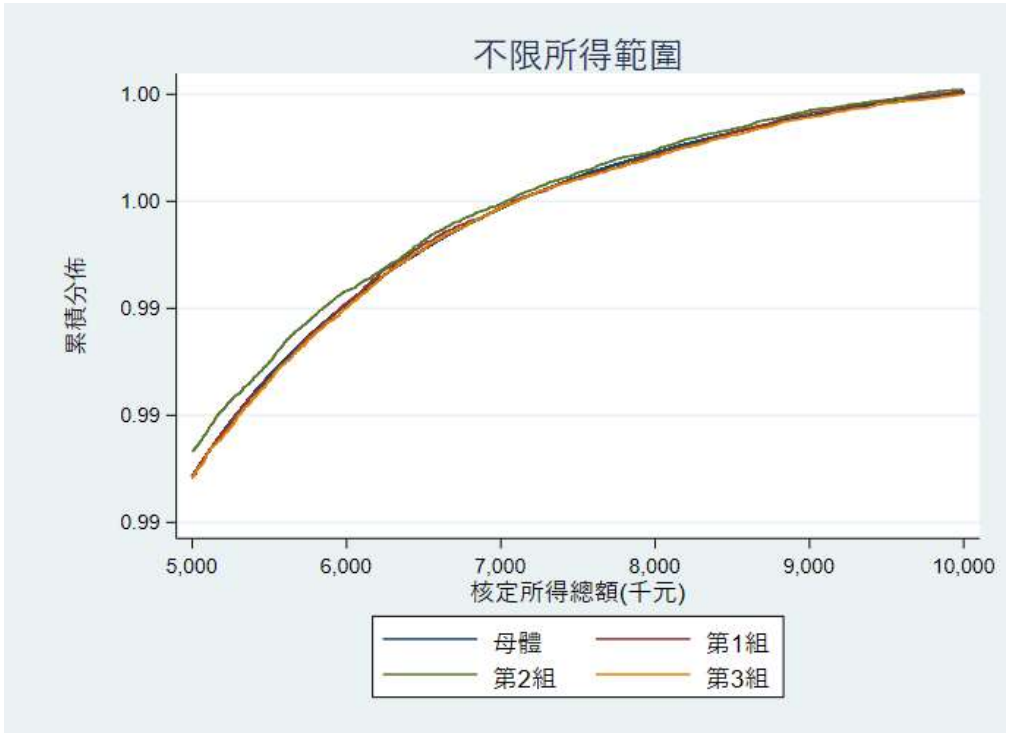


圖 11. ECDF : 所得總額 - 所有資料 (500 萬~1000 萬)

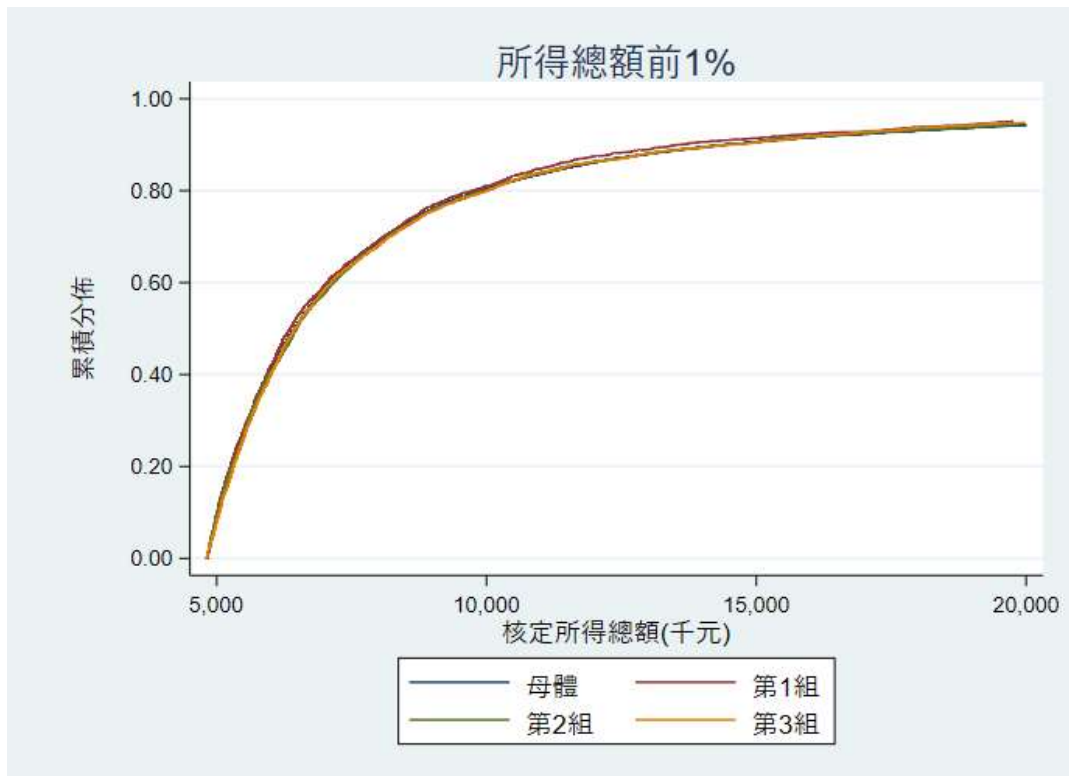


圖 12. ECDF：所得總額 - 所得總額前 1%

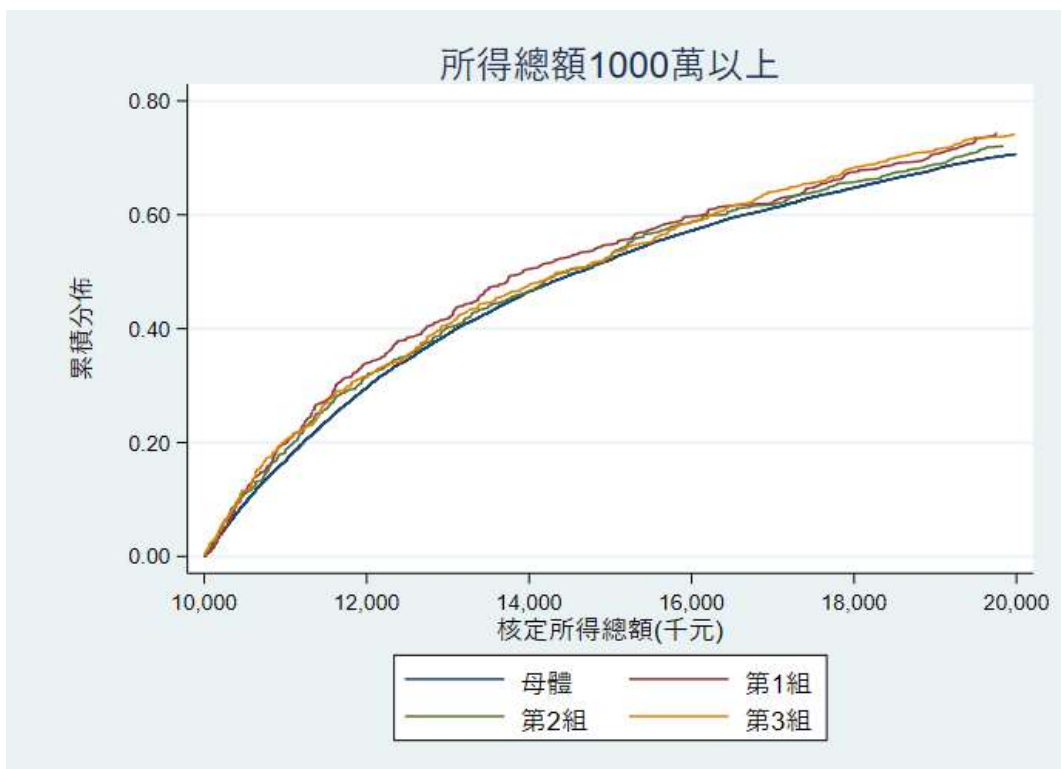


圖 13. ECDF：所得總額 - 所得總額 1000 萬以上

(二) 一般或列舉扣除額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	150	850	122	6,279,484
抽樣 1	150	124	125	240,000
抽樣 2	149	121	121	240,000
抽樣 3	150	122	121	240,000
所得總額前 1%				
母體	634	8,446	369	62,794
抽樣 1	528	602	374	2,436
抽樣 2	564	856	376	2,294
抽樣 3	548	845	372	2,386
所得總額大於 1000 萬				
母體	1,466	19,004	607	12,357
抽樣 1	939	1,122	612	463
抽樣 2	1,061	1,697	623	458
抽樣 3	1,009	1,682	596	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0016	0.296
第二組	0.0008	0.738
第三組	0.0007	0.777

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0388	0.262
第二組	0.0385	0.270
第三組	0.0415	0.204

2. 資料分佈

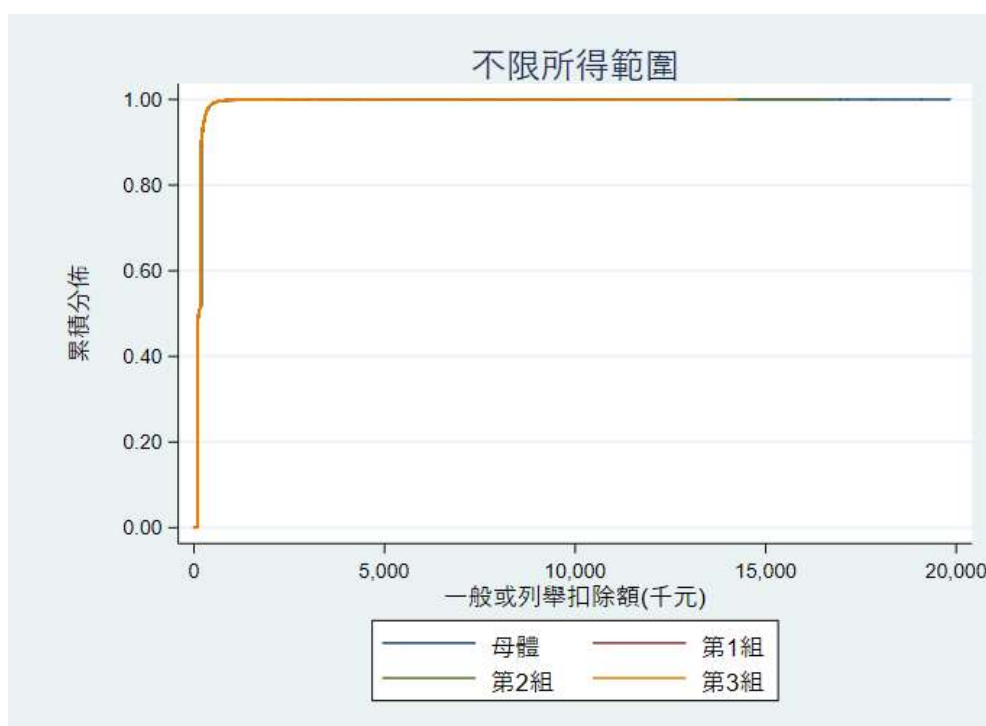


圖 14. ECDF : 標準扣除額 - 所有資料

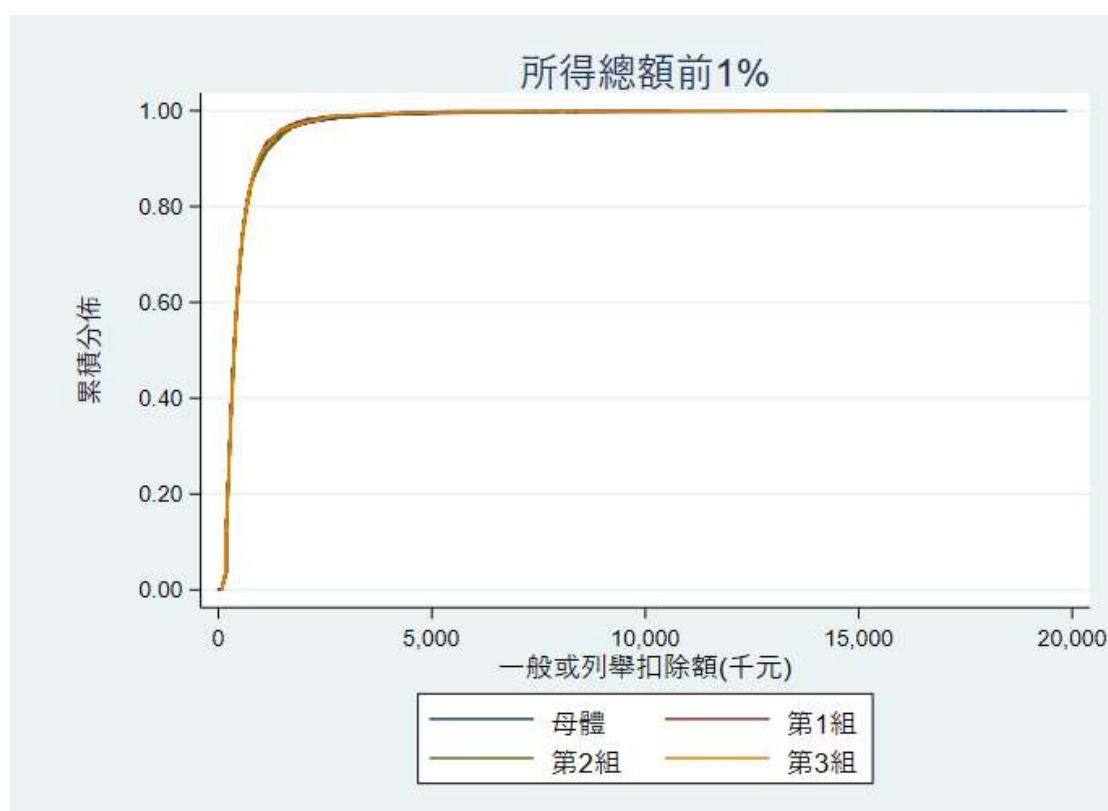


圖 15. ECDF : 標準扣除額 - 所得總額前 1%

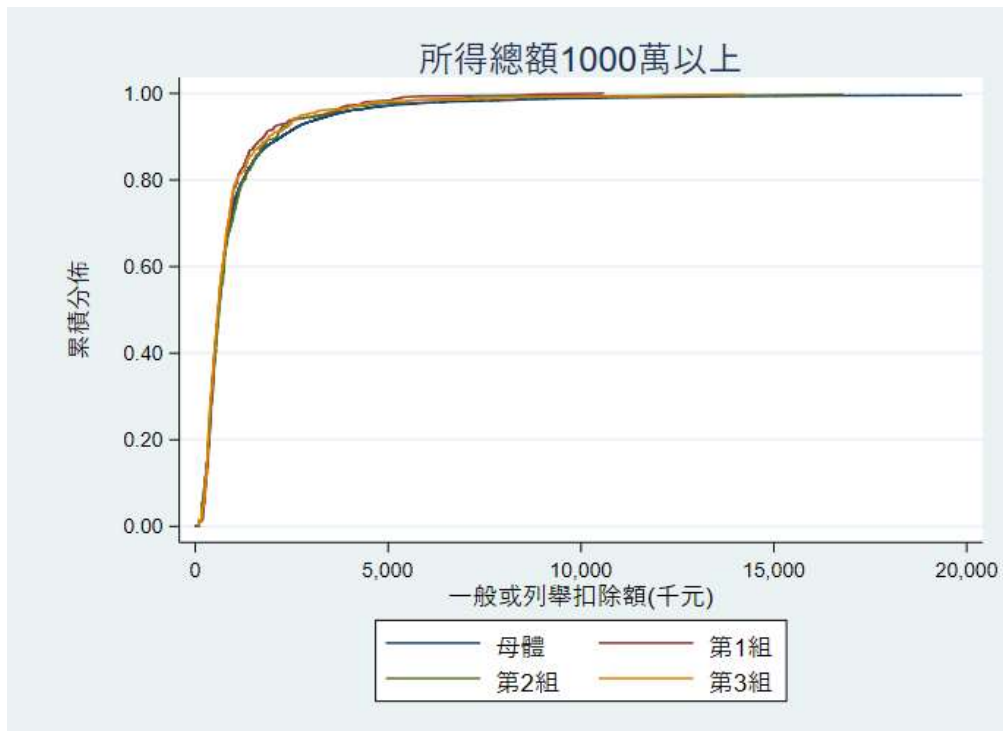


圖 16. ECDF : 標準扣除額 - 所得總額 1000 萬以上

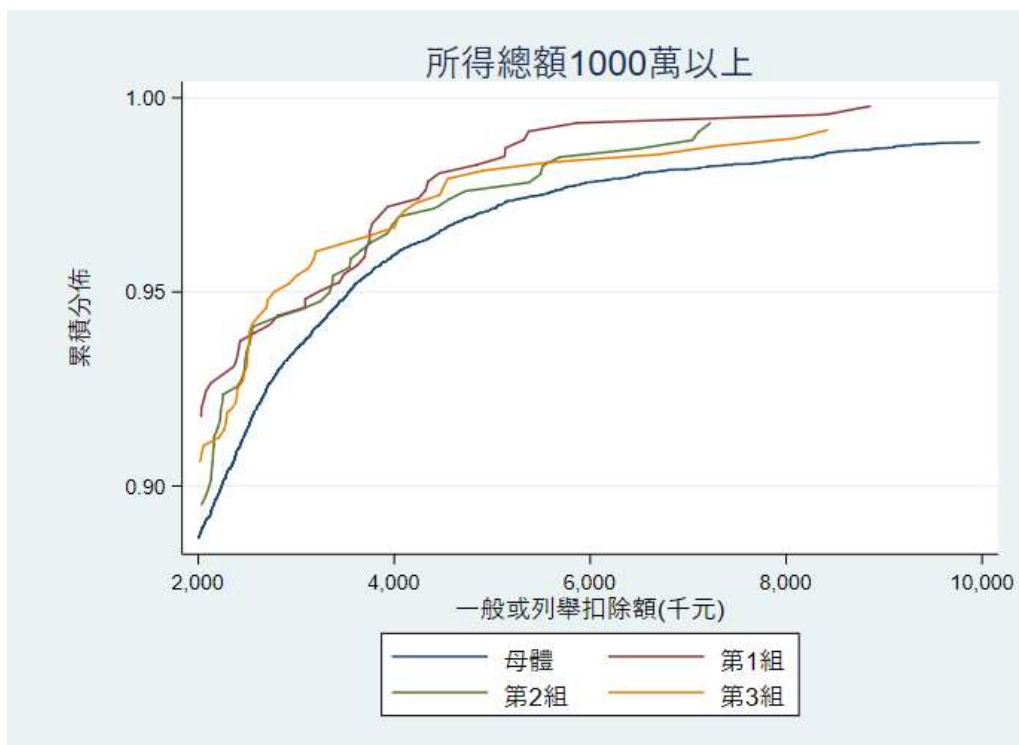


圖 17. ECDF : 標準扣除額 - 所得總額 1000 萬以上

(三) 特別扣除額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	198	122	147	6,279,484
抽樣 1	198	122	147	240,000
抽樣 2	198	121	147	240,000
抽樣 3	198	122	147	240,000
所得總額前 1%				
母體	356	158	344	62,794
抽樣 1	356	156	342	2,436
抽樣 2	354	167	336	2,294
抽樣 3	355	159	338	2,386
所得總額大於 1000 萬				
母體	356	156	352	12,357
抽樣 1	347	152	344	463
抽樣 2	353	155	343	458
抽樣 3	362	157	375	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0012	0.497
第二組	0.0010	0.602
第三組	0.0013	0.478

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0294	0.463
第二組	0.0287	0.484
第三組	0.0533	0.072

2. 資料分佈

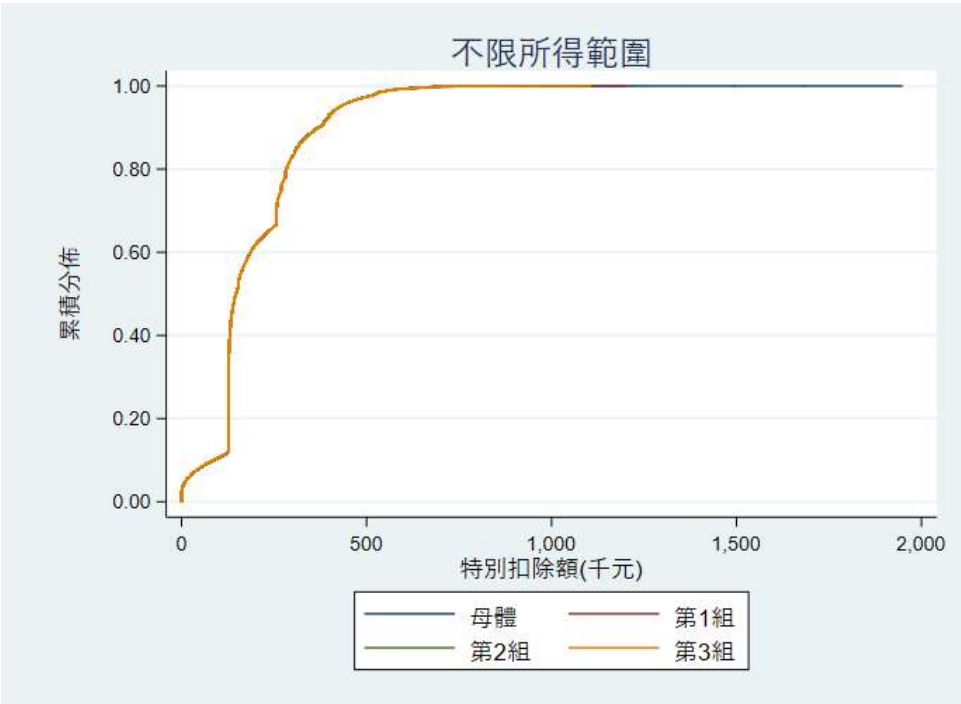


圖 18. ECDF：特別扣除額 - 所有資料

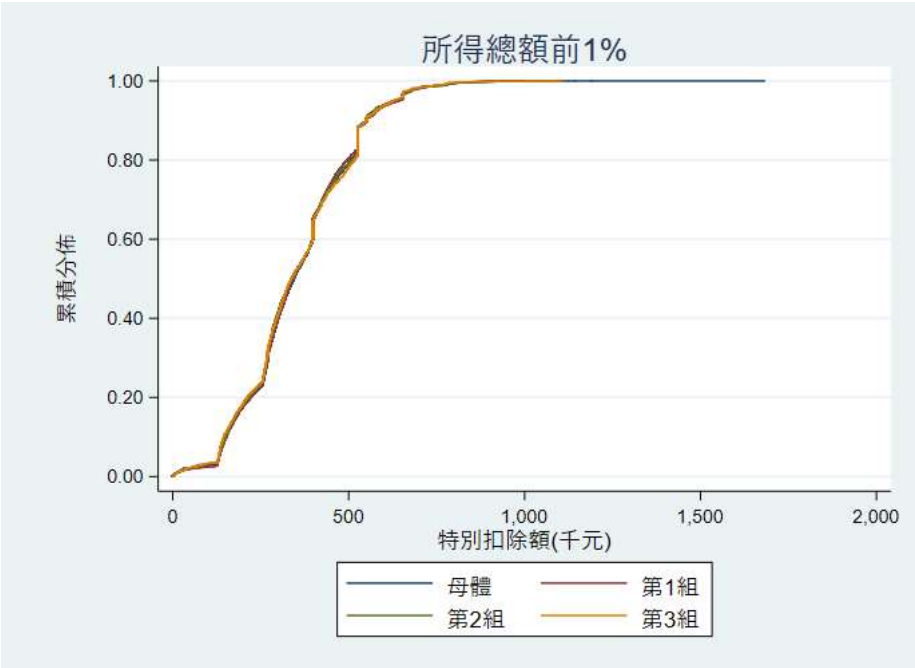


圖 19. ECDF：特別扣除額 - 所得總額前 1%

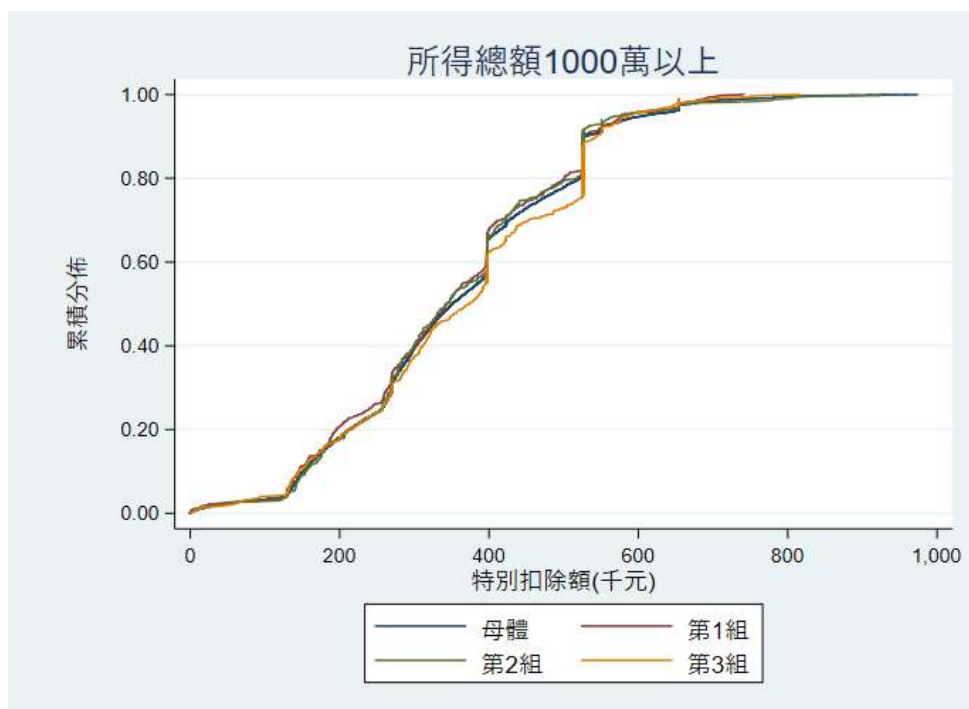


圖 20. ECDF：特別扣除額 - 所得總額 1000 萬以上

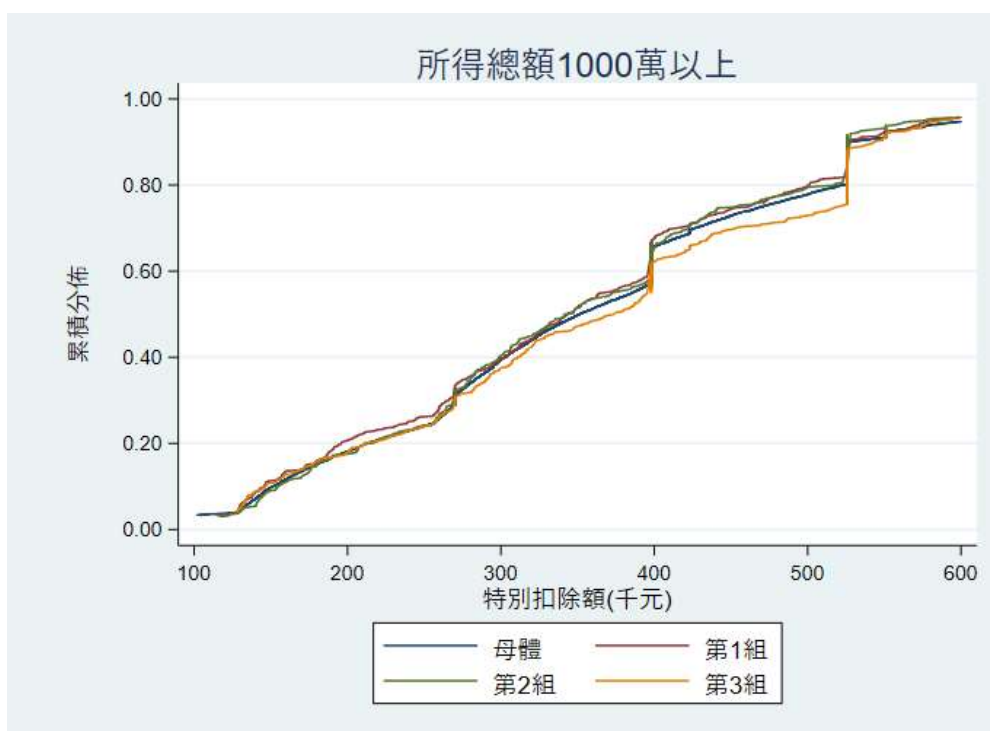


圖 21. ECDF：特別扣除額 - 所得總額 1000 萬以上

(四) 核定所得淨額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	400	4,059	97	6,279,484
抽樣 1	402	3,822	97	240,000
抽樣 2	399	2,891	97	240,000
抽樣 3	395	1,910	96	240,000
所得總額前 1%				
母體	9,045	39,300	5,273	62,794
抽樣 1	9,153	36,542	5,240	2,436
抽樣 2	9,235	27,689	5,271	2,294
抽樣 3	8,561	16,498	5,298	2,386
所得總額大於 1000 萬				
母體	24,950	86,757	13,087	12,357
抽樣 1	26,284	81,651	12,475	463
抽樣 2	25,589	59,200	12,839	458
抽樣 3	21,996	33,490	12,949	480

Kolmogorov-Smirnov 檢定

抽樣組別	最大差異程度 D	P-Value
第一組	0.0016	0.286
第二組	0.0010	0.610
第三組	0.0012	0.499

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)

抽樣組別	最大差異程度 D	P-Value
第一組	0.0606	0.038
第二組	0.0274	0.516
第三組	0.0399	0.229

2. 資料分佈

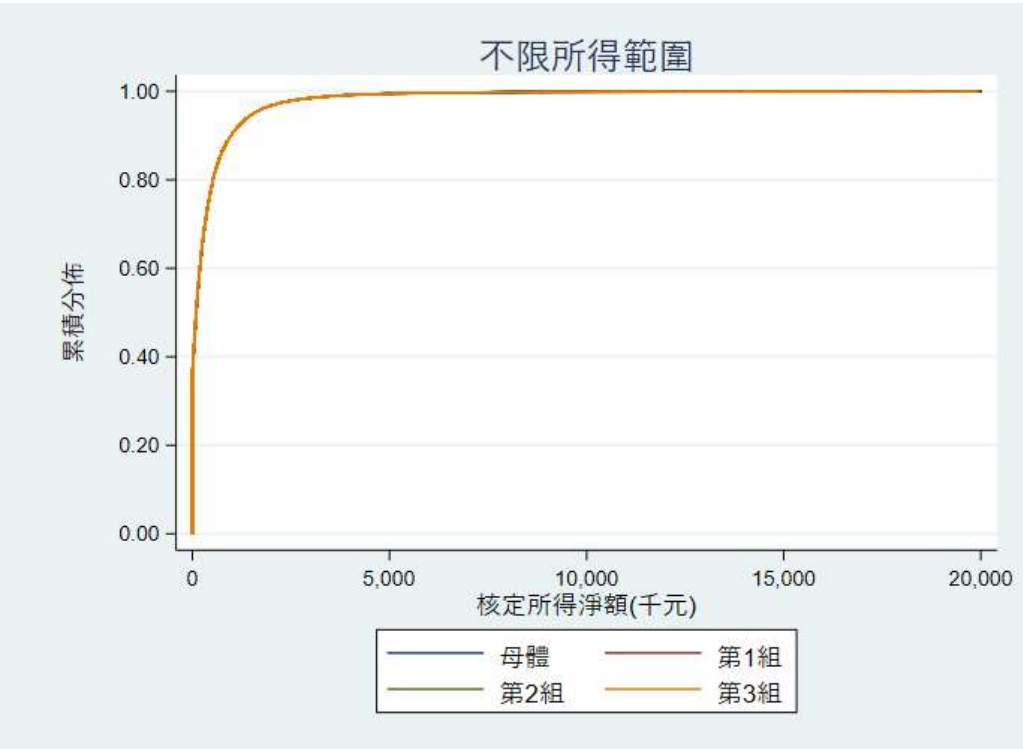


圖 22. ECDF : 核定所得淨額 - 所有資料

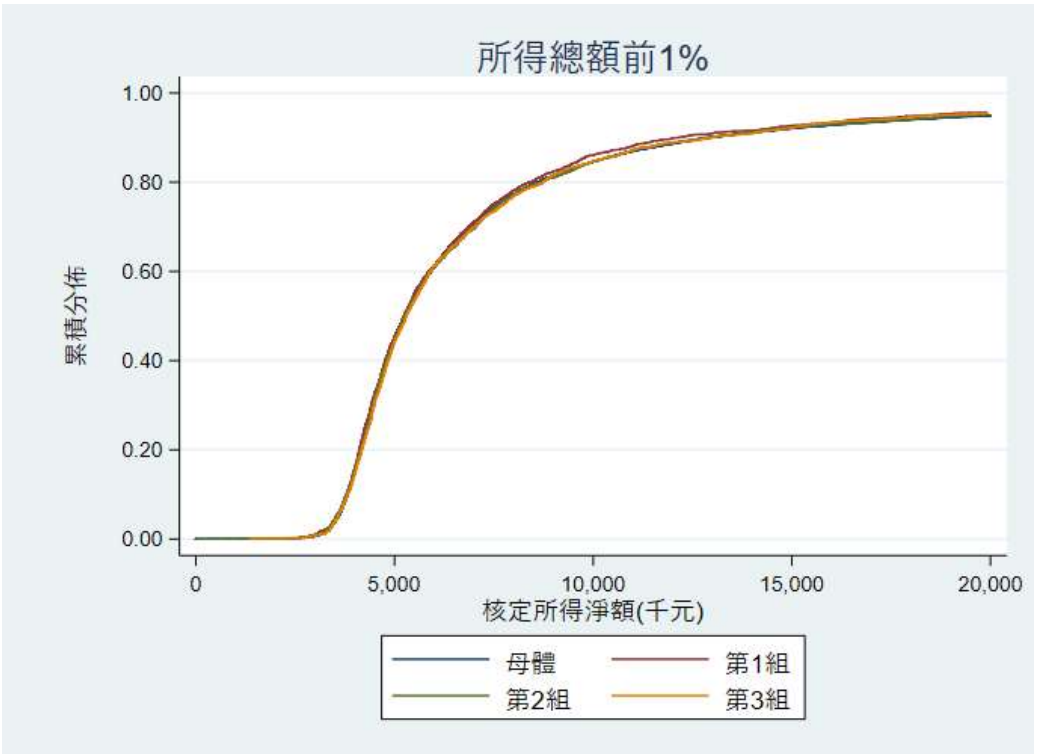


圖 23. ECDF : 核定所得淨額 - 所得總額前 1%

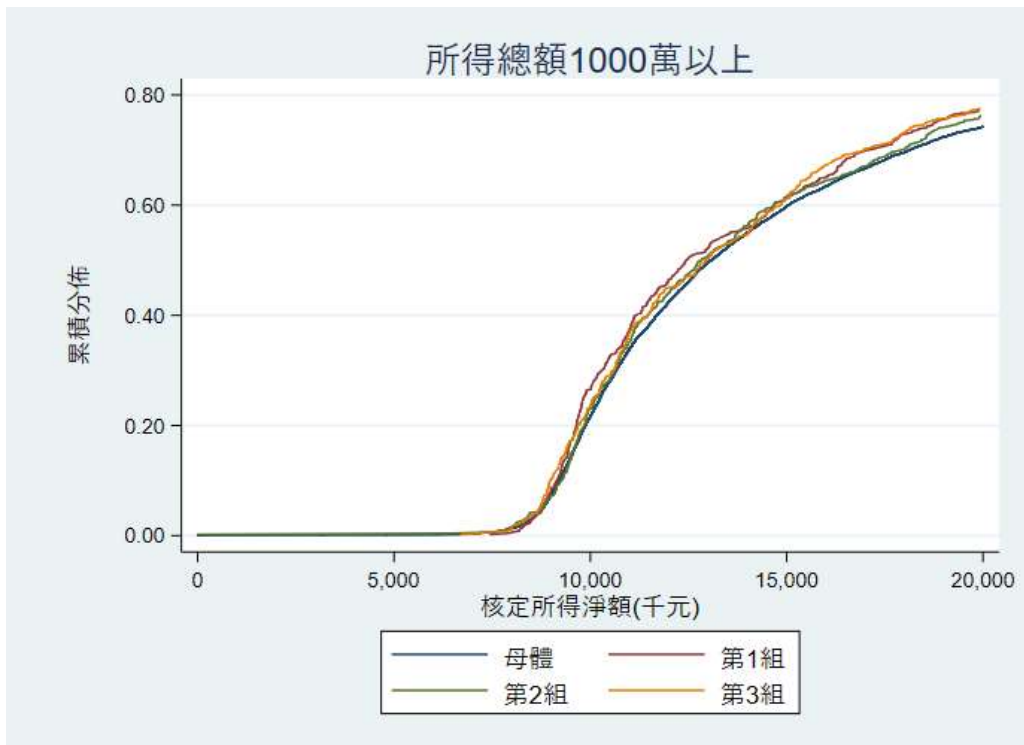


圖 24. ECDF：核定所得淨額 - 所得總額 1000 萬以上

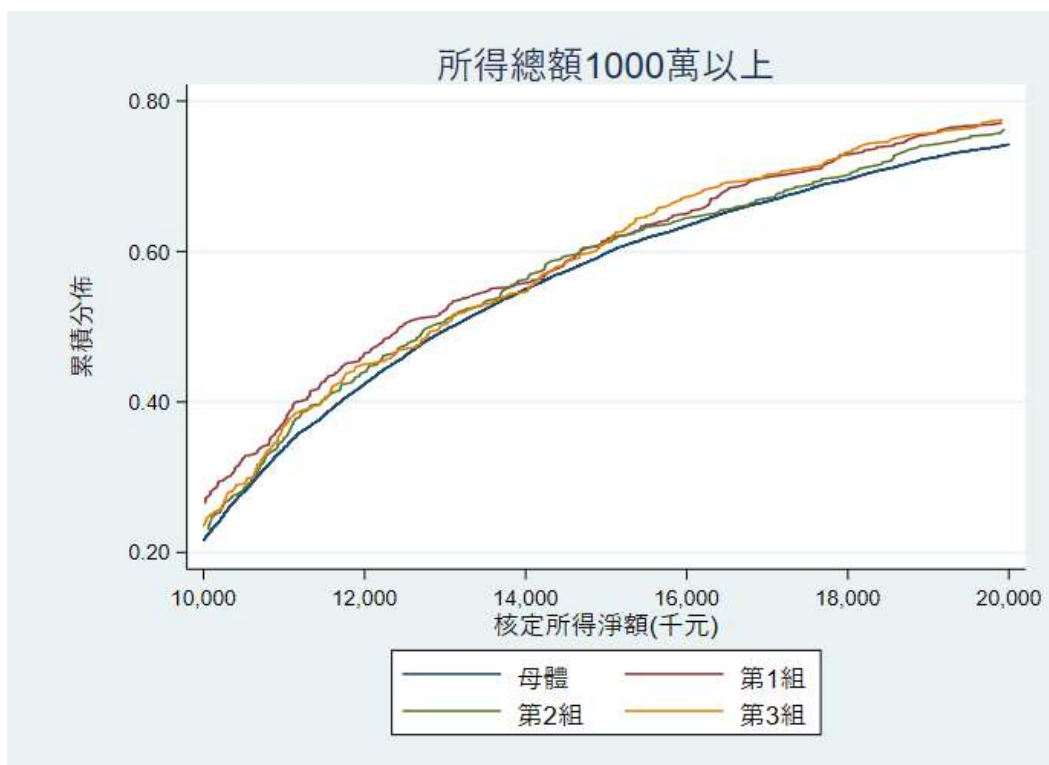


圖 25. ECDF：核定所得淨額 - 所得總額 1000 萬以上

(五) 納稅義務人薪資所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	25	282	0	6,279,484
抽樣 1	25	265	0	240,000
抽樣 2	25	256	0	240,000
抽樣 3	25	260	0	240,000
所得總額前 1%				
母體	576	2,173	0	62,794
抽樣 1	562	1,940	0	2,436
抽樣 2	568	1,878	0	2,294
抽樣 3	572	1,877	0	2,386
所得總額大於 1000 萬				
母體	883	3,992	0	12,357
抽樣 1	906	3,412	0	463
抽樣 2	714	3,053	0	458
抽樣 3	845	3,145	0	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0003	0.960
第二組	0.0004	0.936
第三組	0.0002	0.973

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0119	0.881
第二組	0.0252	0.571
第三組	0.0138	0.838

2. 資料分佈

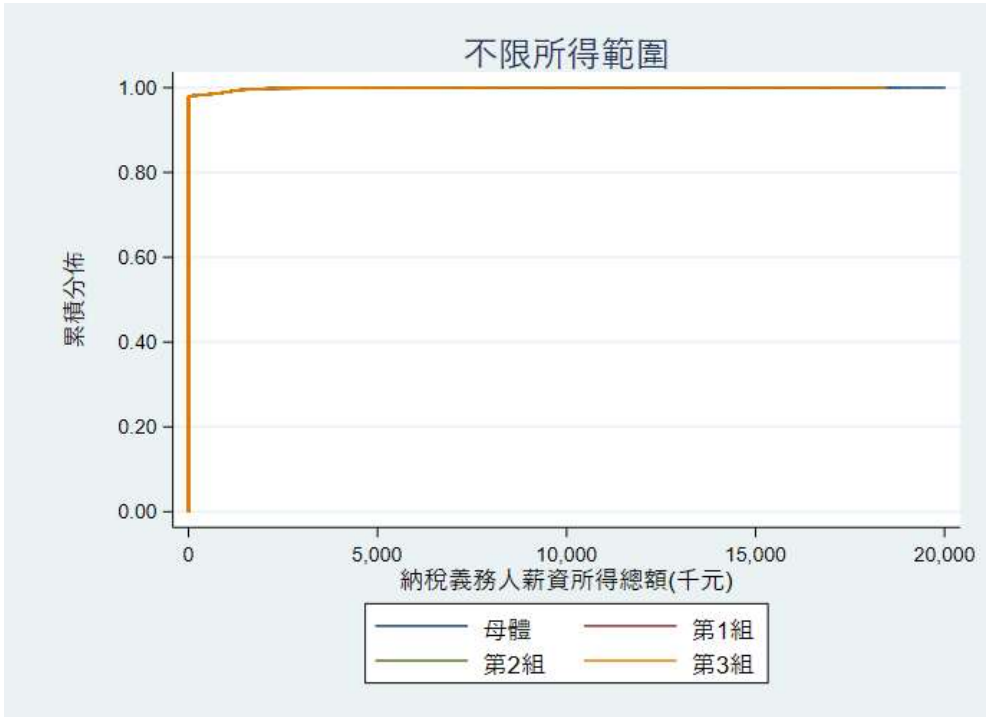


圖 26. ECDF：納稅義務人薪資所得總額 - 所有資料

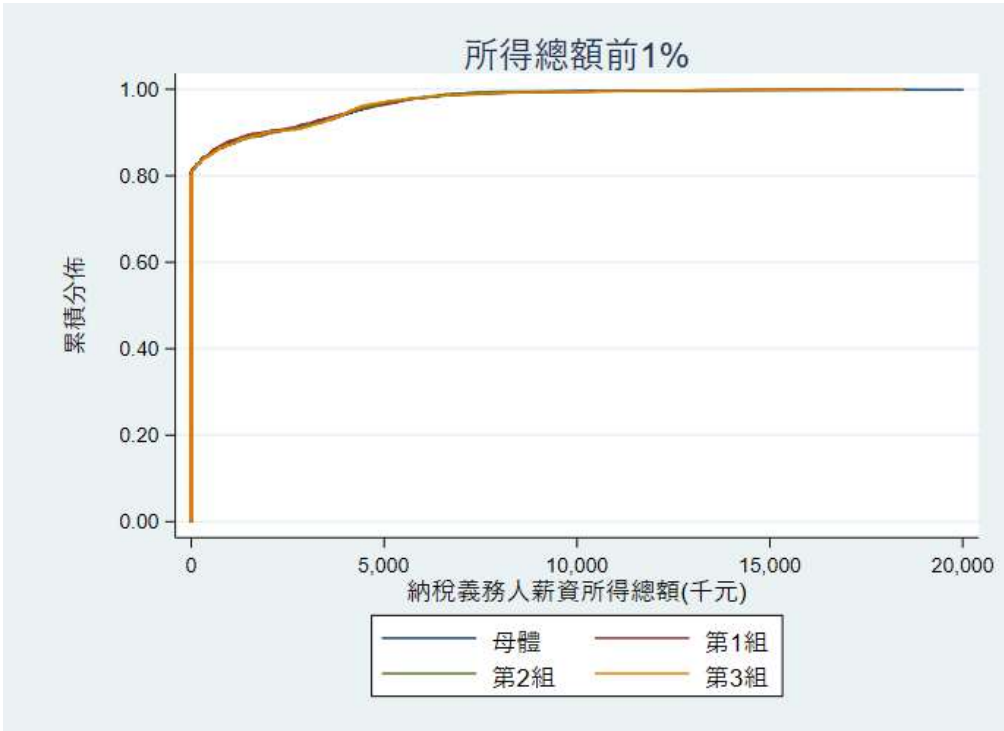


圖 27. ECDF：納稅義務人薪資所得總額 - 所得總額前 1%

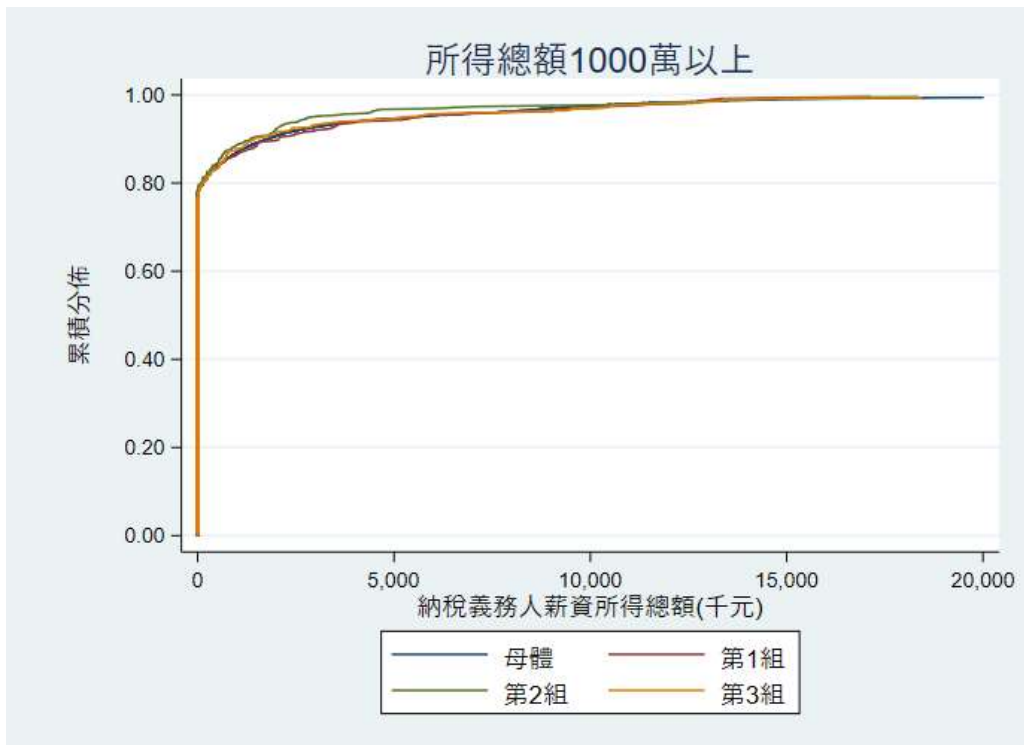


圖 28. ECDF：納稅義務人薪資所得總額 - 所得總額 1000 萬以上

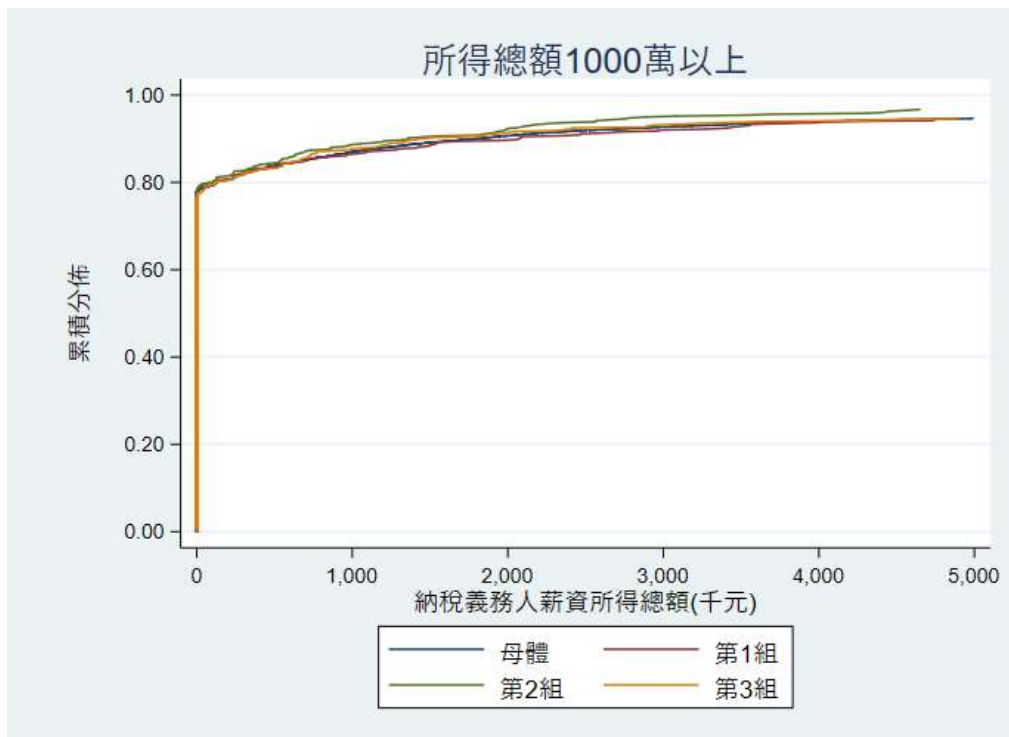


圖 29. ECDF：納稅義務人薪資所得總額 - 所得總額 1000 萬以上

(六) 配偶薪資所得總額

1. 敘述統計

X 不限所得範圍				
組別	平均值	標準差	中位數	樣本數
母體	95	643	0	6,279,484
抽樣 1	95	611	0	240,000
抽樣 2	96	621	0	240,000
抽樣 3	94	441	0	240,000
所得總額前 1%				
母體	1,920	5,623	969	62,794
抽樣 1	1,902	5,209	914	2,436
抽樣 2	2,022	5,457	1,002	2,294
抽樣 3	1,826	3,159	994	2,386
所得總額大於 1000 萬				
母體	4,289	12,013	1,348	12,357
抽樣 1	4,279	11,205	1,464	463
抽樣 2	4,669	11,465	1,298	458
抽樣 3	3,663	6,002	1,362	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0007	0.773
第二組	0.0010	0.652
第三組	0.0006	0.840

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0254	0.562
第二組	0.0156	0.806
第三組	0.0250	0.562

2. 資料分佈

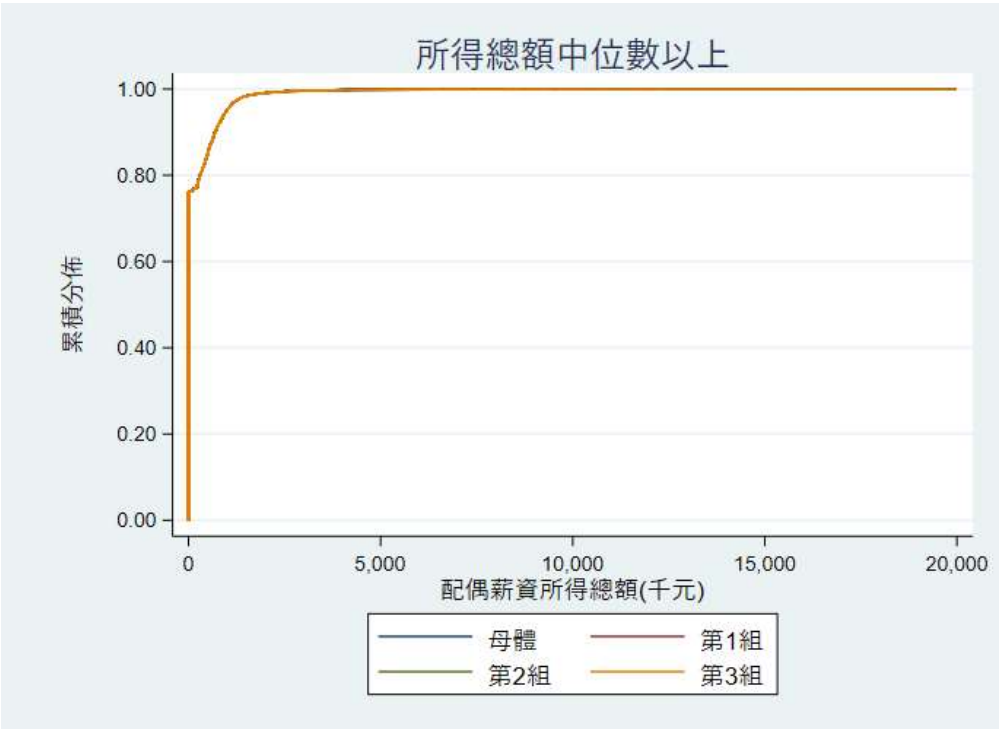


圖 30. ECDF：配偶薪資所得總額 - 所有資料

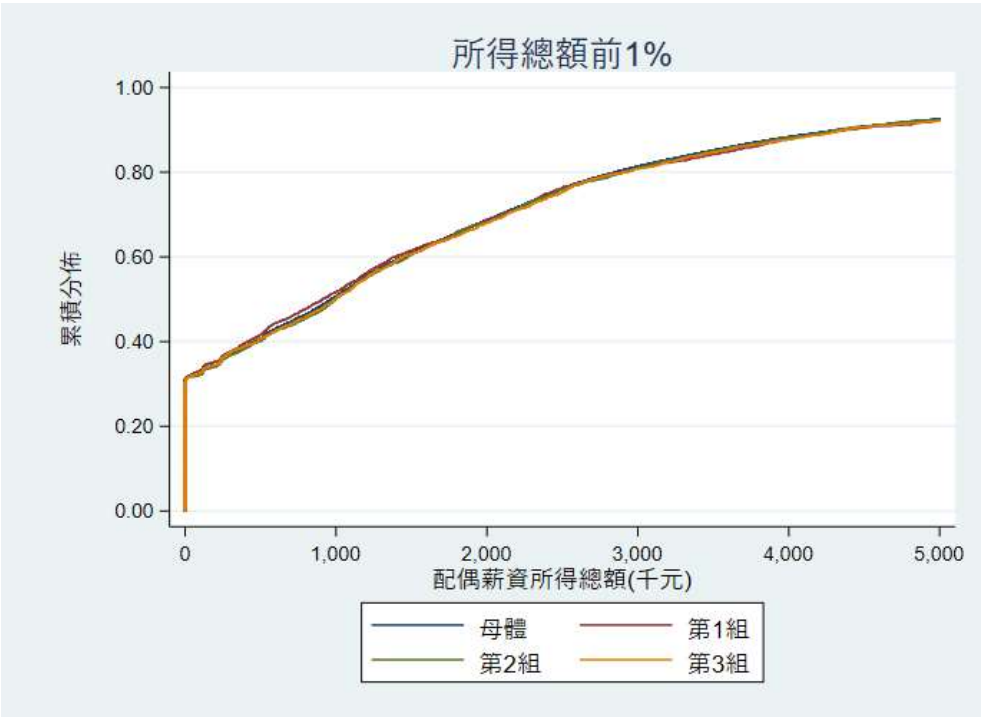


圖 31. ECDF：配偶薪資所得總額 - 所得總額前 1%

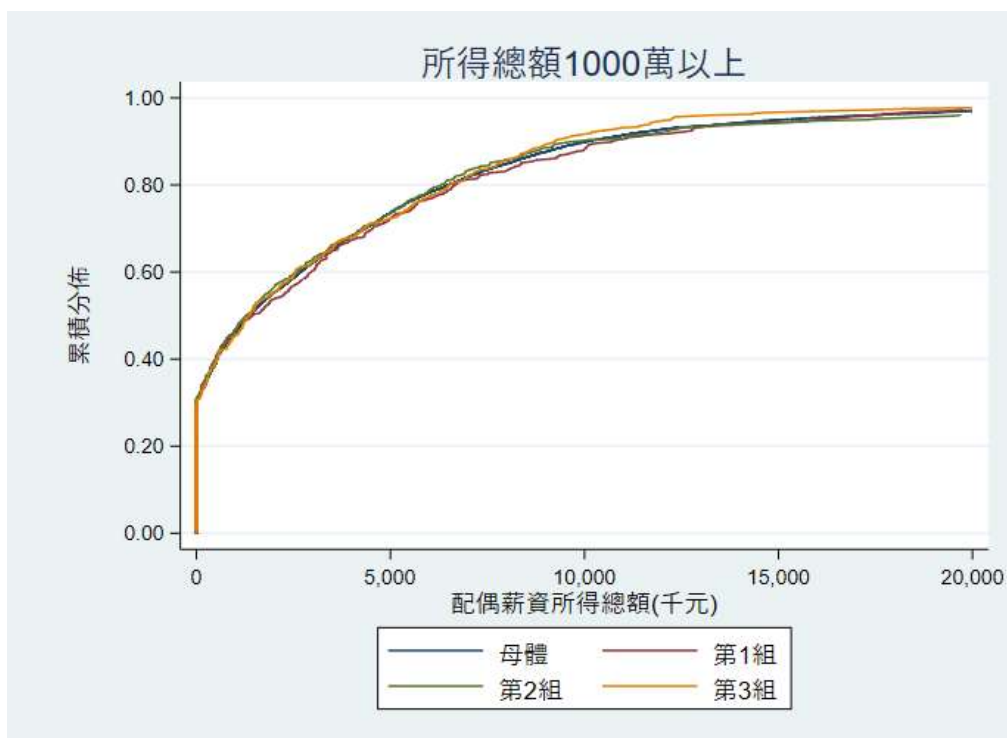


圖 32. ECDF：配偶薪資所得總額 - 所得總額 1000 萬以上

(七) 納稅義務人各類所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	675	4,201	485	6,279,484
抽樣 1	668	1,439	485	240,000
抽樣 2	671	2,287	486	240,000
抽樣 3	670	1,611	485	240,000
所得總額前 1%				
母體	7,025	41,189	4,793	62,794
抽樣 1	6,274	12,039	4,786	2,436
抽樣 2	6,847	21,902	4,661	2,294
抽樣 3	6,592	14,066	4,852	2,386
所得總額大於 1000 萬				
母體	18,778	91,811	10,929	12,357
抽樣 1	15,488	25,242	10,587	463
抽樣 2	18,010	47,225	10,408	458
抽樣 3	16,072	29,196	10,377	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0018	0.214
第二組	0.0012	0.536
第三組	0.0012	0.507

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0443	0.174
第二組	0.0483	0.127
第三組	0.0614	0.031

2. 資料分佈

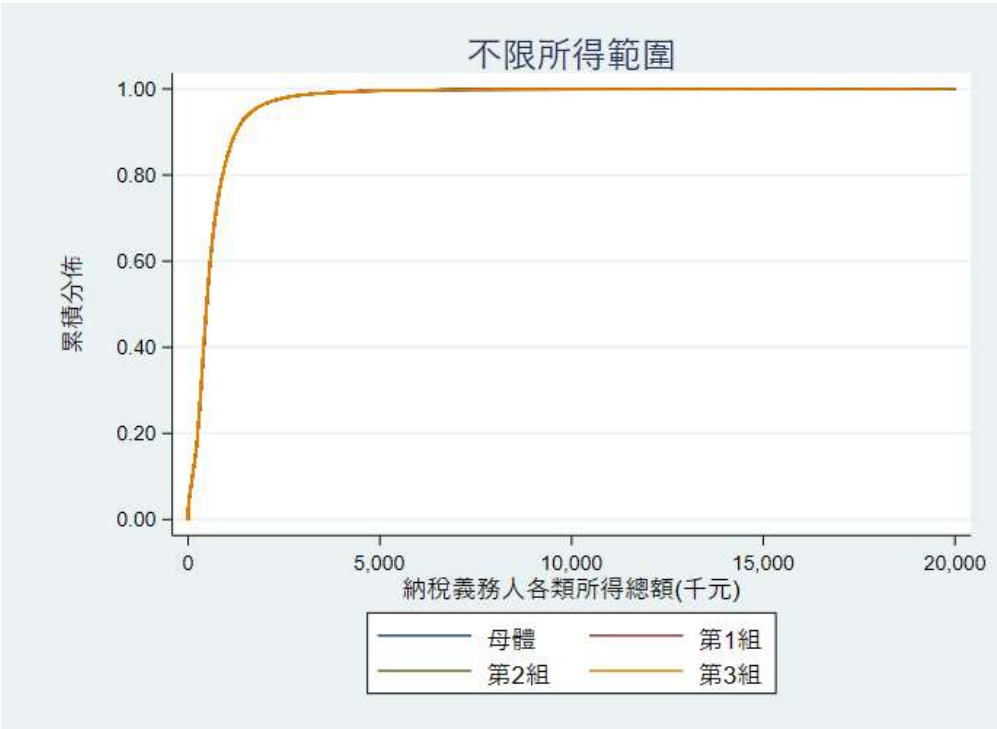


圖 33. ECDF：納稅義務人各類所得總額 - 所有資料

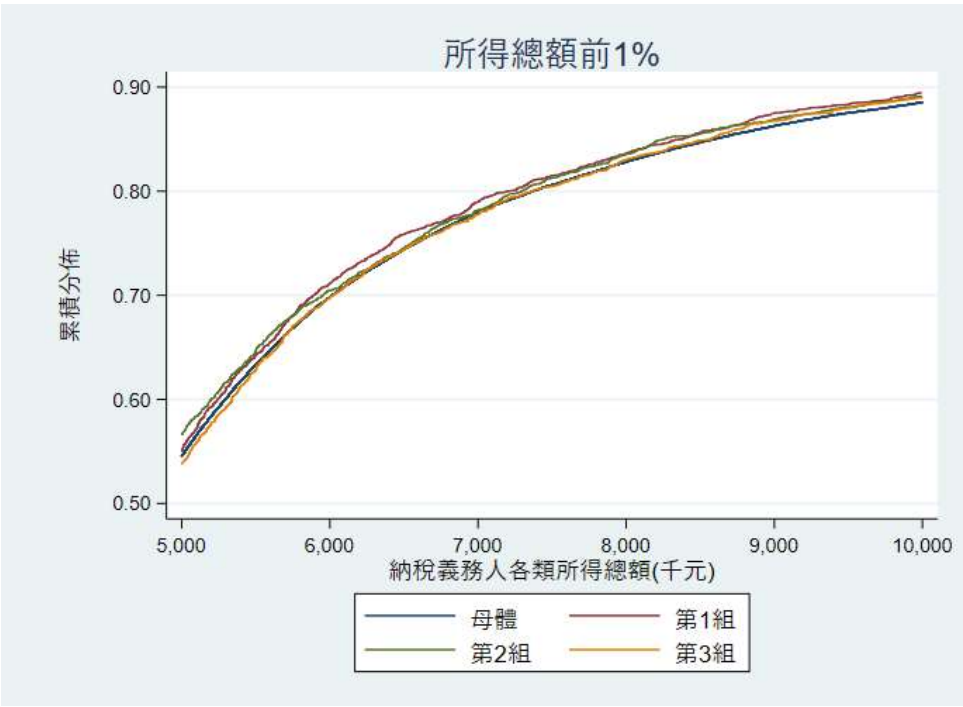


圖 34. ECDF：納稅義務人各類所得總額 - 所得總額前 1%

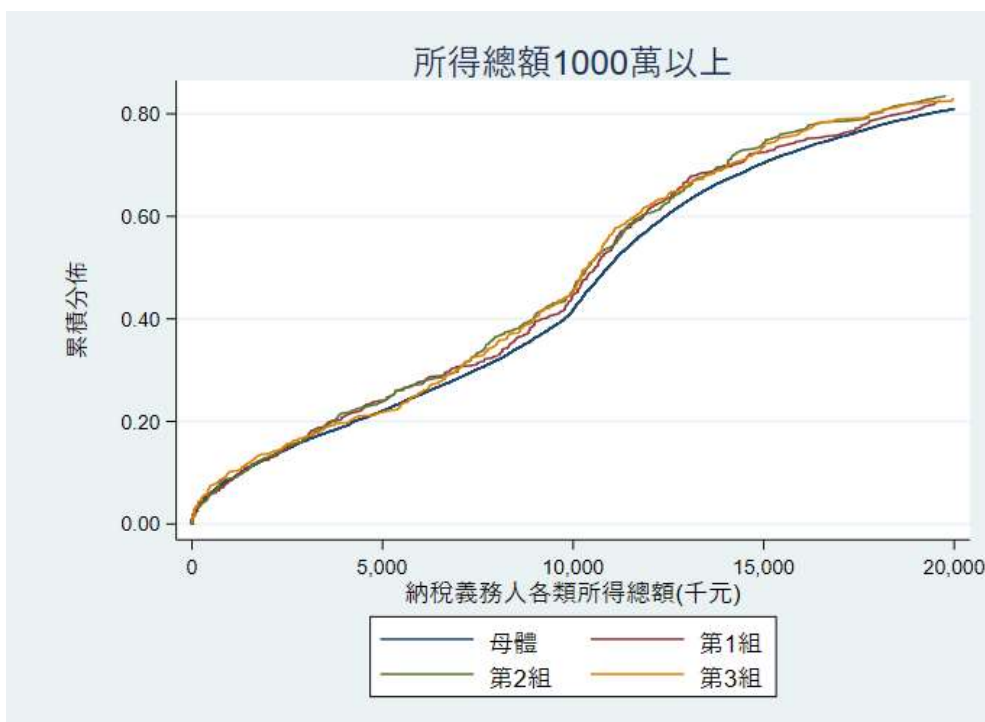


圖 35. ECDF：納稅義務人各類所得總額 - 所得總額 1000 萬以上

(八) 配偶各類所得總額

1. 敘述統計

X 不限所得範圍				
組別	平均值	標準差	中位數	樣本數
母體	220	1,360	0	6,279,484
抽樣 1	228	3,135	0	240,000
抽樣 2	222	1,278	0	240,000
抽樣 3	218	968	0	240,000
所得總額前 1%				
母體	3,217	12,667	1,534	62,794
抽樣 1	3,954	30,647	1,445	2,436
抽樣 2	3,553	11,970	1,716	2,294
抽樣 3	3,117	8,381	1,519	2,386
所得總額大於 1000 萬				
母體	7,796	27,771	2,355	12,357
抽樣 1	11,804	69,672	1,920	463
抽樣 2	8,941	25,778	3,321	458
抽樣 3	7,328	17,617	2,729	480

Kolmogorov-Smirnov 檢定		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0017	0.258
第二組	0.0013	0.437
第三組	0.0014	0.430

Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
抽樣組別	最大差異程度 D	P-Value
第一組	0.0460	0.151
第二組	0.0537	0.078
第三組	0.0377	0.268

2. 資料分佈

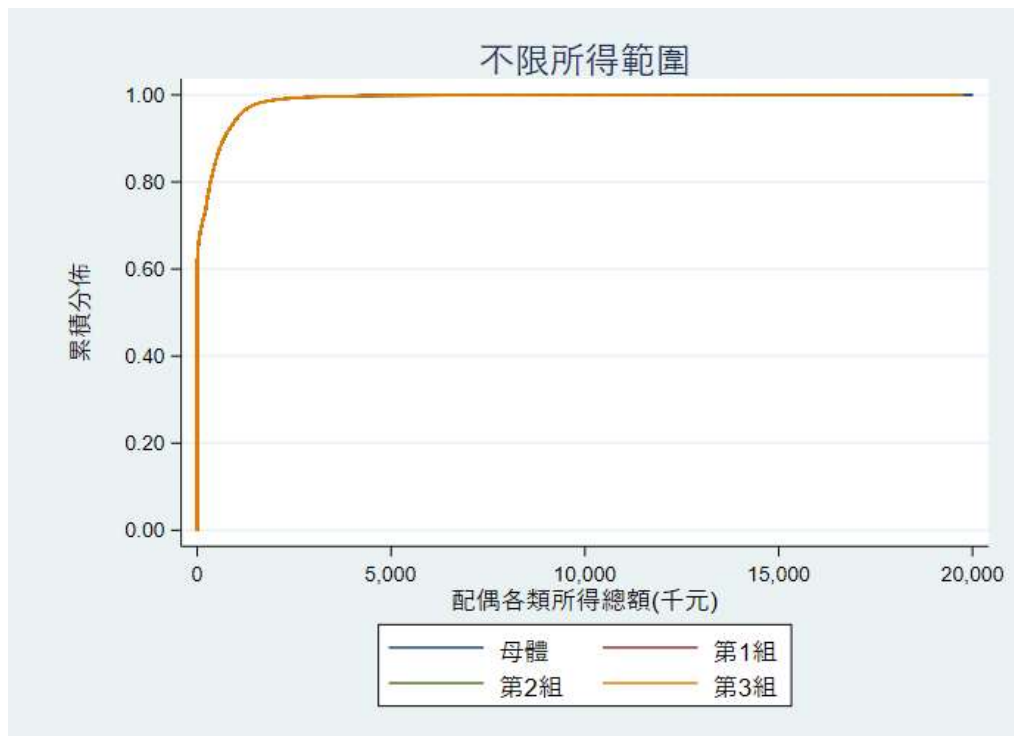


圖 36. ECDF：配偶各類所得總額 - 所有資料

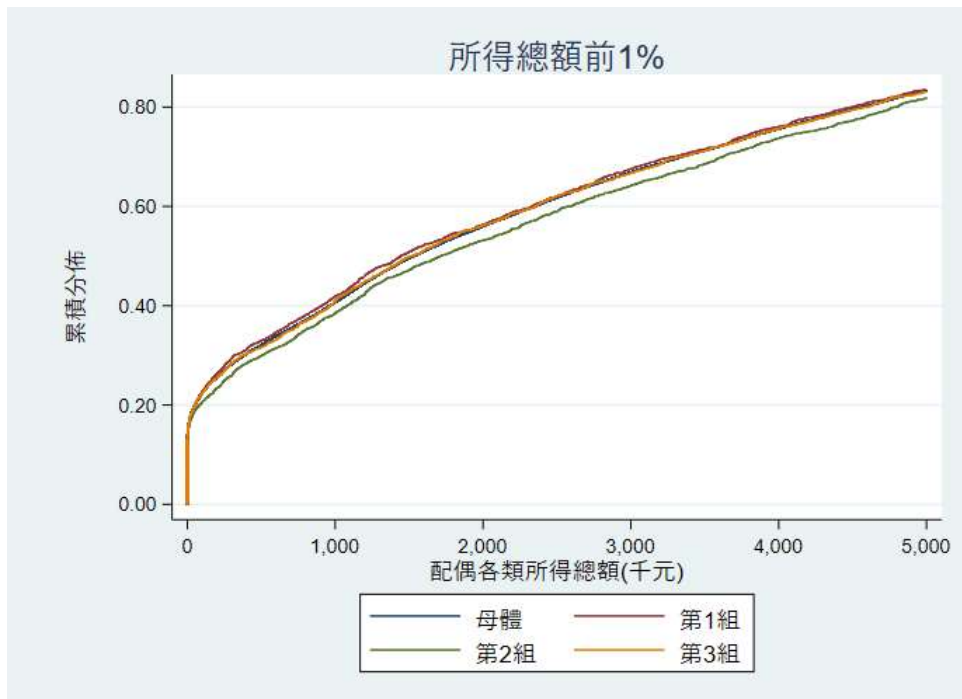


圖 37. ECDF：配偶各類所得總額 - 所得總額前 1%

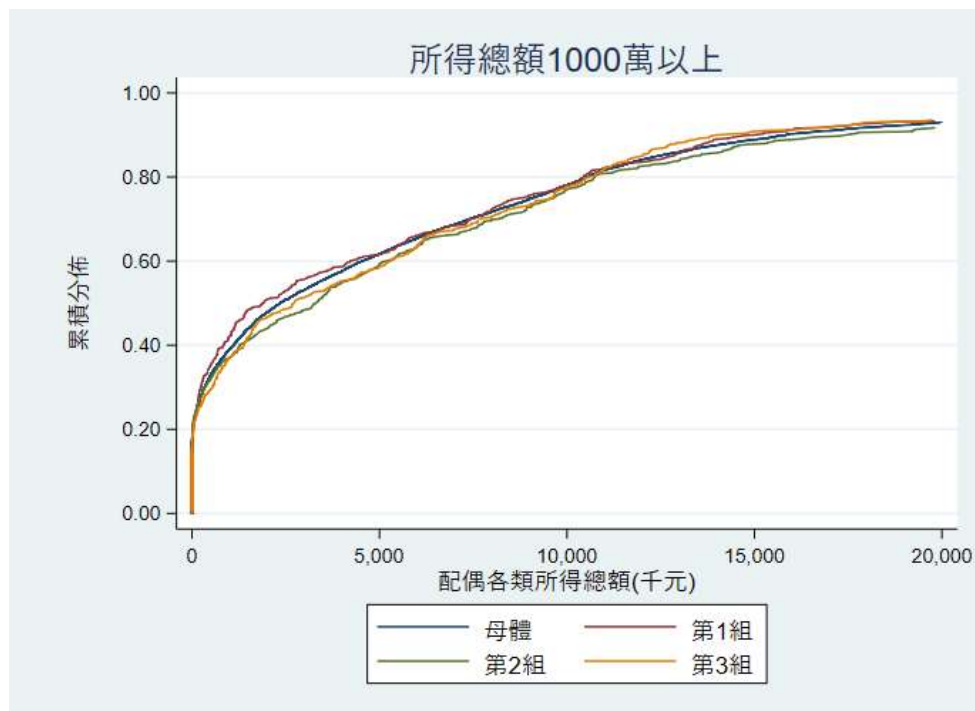


圖 38. ECDF : 配偶各類所得總額 -所得總額大於 1000 萬

整體抽樣檔驗證結果概述

欄 位	說 明
核定所得總額	當所得大於 1000 萬時，各組抽樣所得到的平均數與母體相比有落差，但因平均數易受極端值影響，因此觀察其資料分布，仍與母體有落差，因此所得大於 1000 萬抽樣結果須注意；前 1%所得資料分布也與母體有些許差異，但差異不大，所得小於 1000 萬時驗證結果良好。
標準扣除額	當所得大於 1000 萬時，各組抽樣所得到的平均數與母體相比有落差，但因平均數易受極端值影響，因此觀察其資料分布，與母體落差較大 (約 0.04%)，因此所得大於 1000 萬抽樣結果須注意；所得小於 1000 萬時驗證結果良好。
特別扣除額	所得 1000 萬以上時，特別扣除額在所得資料分布上與母體有差異但不大，而所得 1000 萬以下並無差異，故驗證結果良好。
核定所得淨額	當所得大於 1000 萬時，各組抽樣所得到的平均數與母體相比有落差，但因平均數易受極端值影響，因此觀察其資料分布，仍與母體有落差，因此所得大於 1000 萬抽樣結果須注意；前 1%所得資料分部也與母體有些許差異，但差異不大，故所得小於 1000 萬時驗證結果良好。
納稅義務人薪資所得總額	前 1%所得資料以及所得 1000 萬以上資料分布與母體相比較不一致，但差異小，整體驗證結果視為良好。
配偶薪資所得總額	前 1%所得資料以及所得 1000 萬以上資料分布與母體相比較不一致，但差異小，整體驗證結果視為良好。
納稅義務人各類所得總額	前 1%所得資料以及所得 1000 萬以上資料分布與母體相比較不一致，但差異小，整體驗證結果視為良好。
配偶各類所得總額	前 1%所得資料以及所得 1000 萬以上資料分布與母體相比較不一致，但差異小，整體驗證結果視為良好。

小抽樣檔驗證資料

(一) 所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	923	4,705	611	6,279,484
24 萬	924	3,887	611	240,000
小樣本	922	1,978	611	24,000
所得總額前 1 %				
母體	10,440	45,507	6,444	62,794
24 萬	10,439	36,704	6,368	2,436
小樣本	9,872	15,349	6,568	259
所得總額大於 1000 萬				
母體	27,141	100,847	14,629	12,357
24 萬	27,913	81,947	13,922	463
小樣本	27,566	33,014	18,090	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0008	0.764
小樣本	0.0021	0.816
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0496	0.112
小樣本	0.1681	0.094

3. 圖形分佈

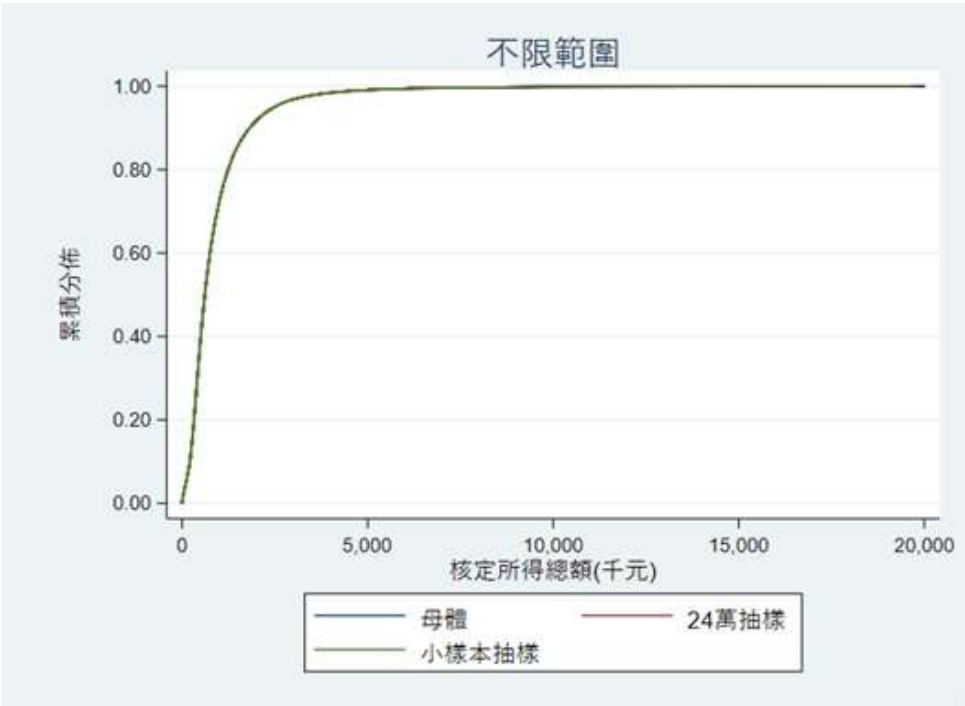


圖 1. ECDF : 所得總額 - 所有資料

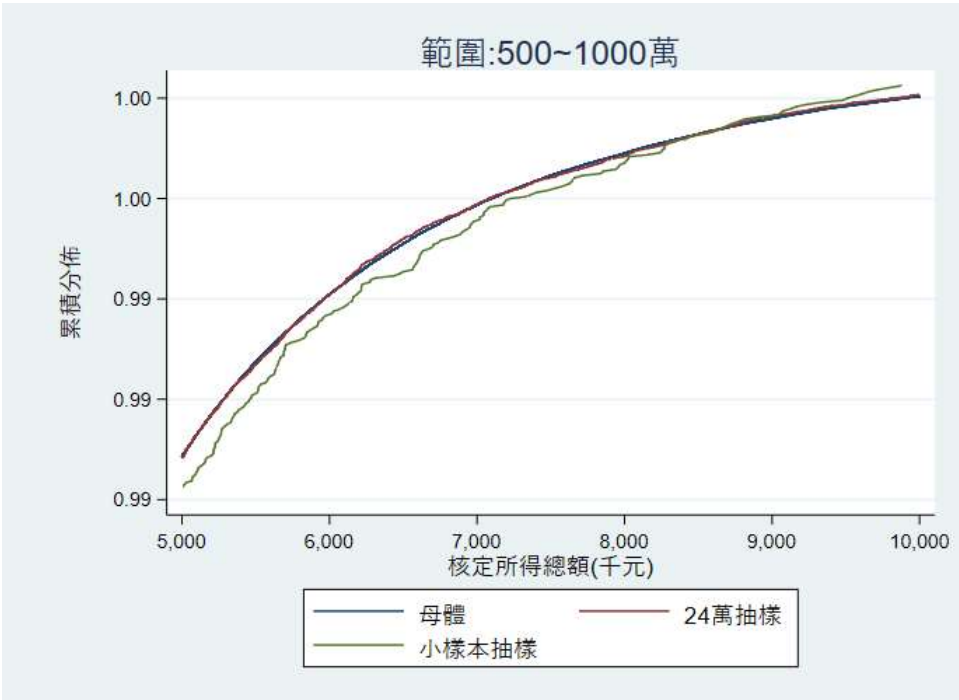


圖 2. ECDF : 所得總額 - 所有資料 (500 萬~1000 萬)

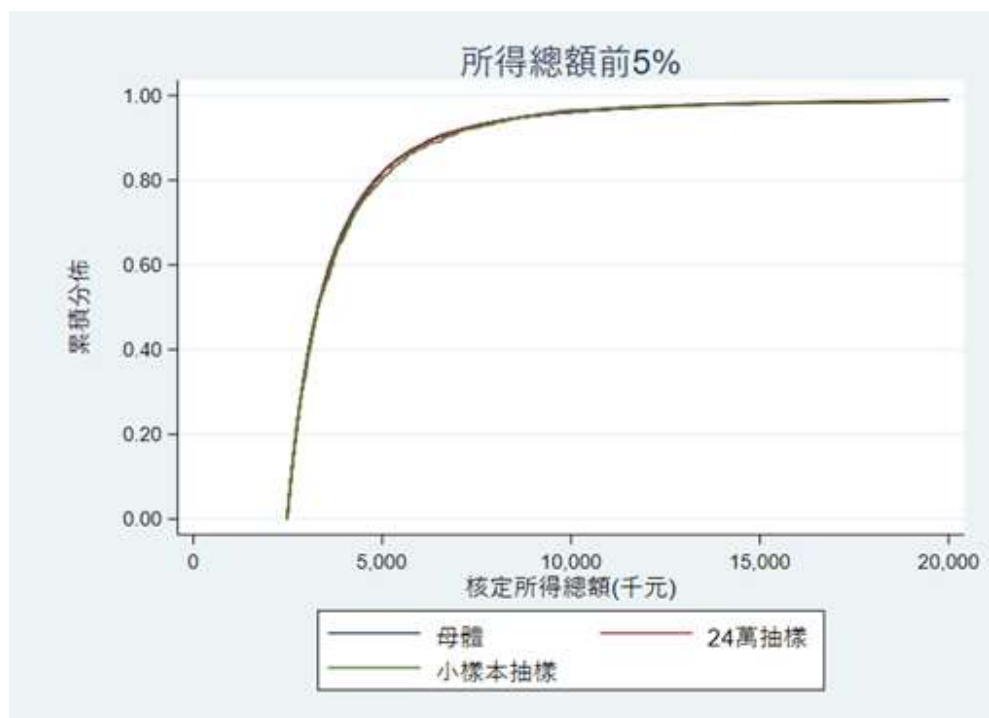


圖 3. ECDF : 所得總額 - 所得總額前 5%

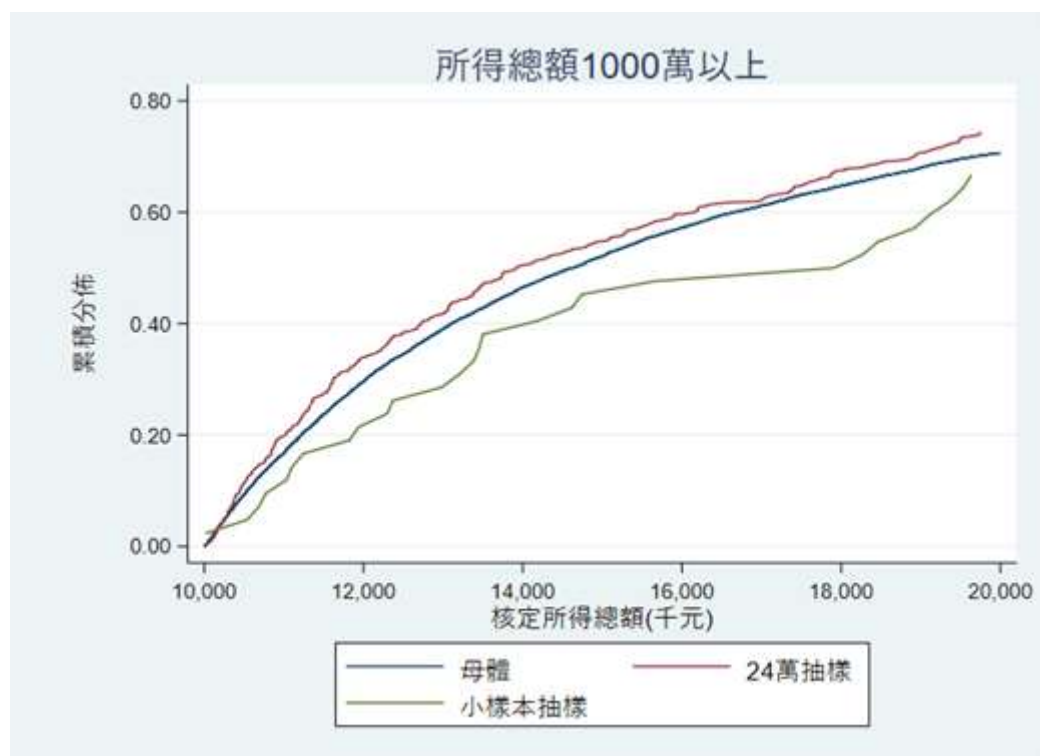


圖 4. ECDF : 所得總額 - 所得總額 1000 萬以上

(二) 一般或列舉扣除額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	150	850	122	6,279,484
24 萬	150	124	125	240,000
小樣本	149	112	116	24,000
所得總額前 1%				
母體	634	8,446	369	62,794
24 萬	528	602	374	2,436
小樣本	549	721	372	259
所得總額大於 1000 萬				
母體	1,466	19,004	607	12,357
24 萬	939	1,122	612	463
小樣本	1,219	1,519	714	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0016	0.296
小樣本	0.0052	0.27
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0388	0.262
小樣本	0.1535	0.139

3. 圖形分佈

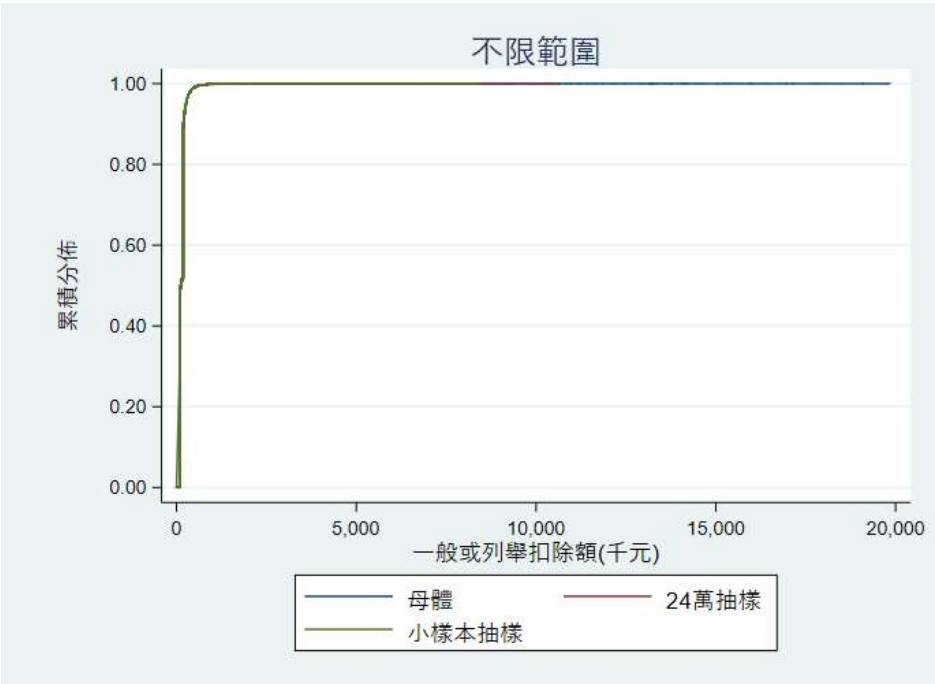


圖 5. ECDF：標準扣除額 - 所有資料

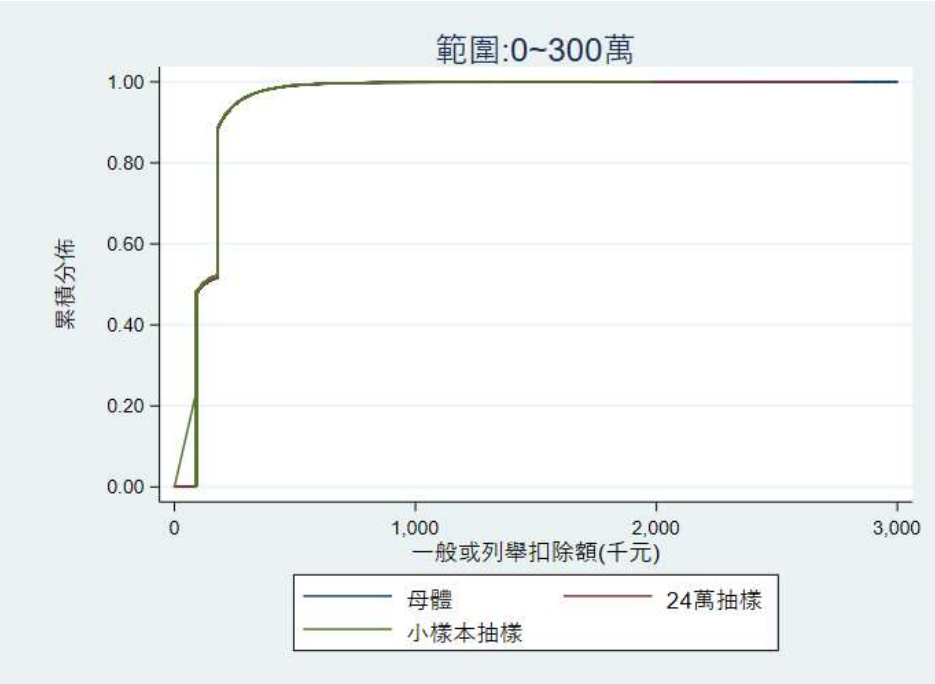


圖 6. ECDF：標準扣除額 - 所有資料 (0 萬~300 萬)

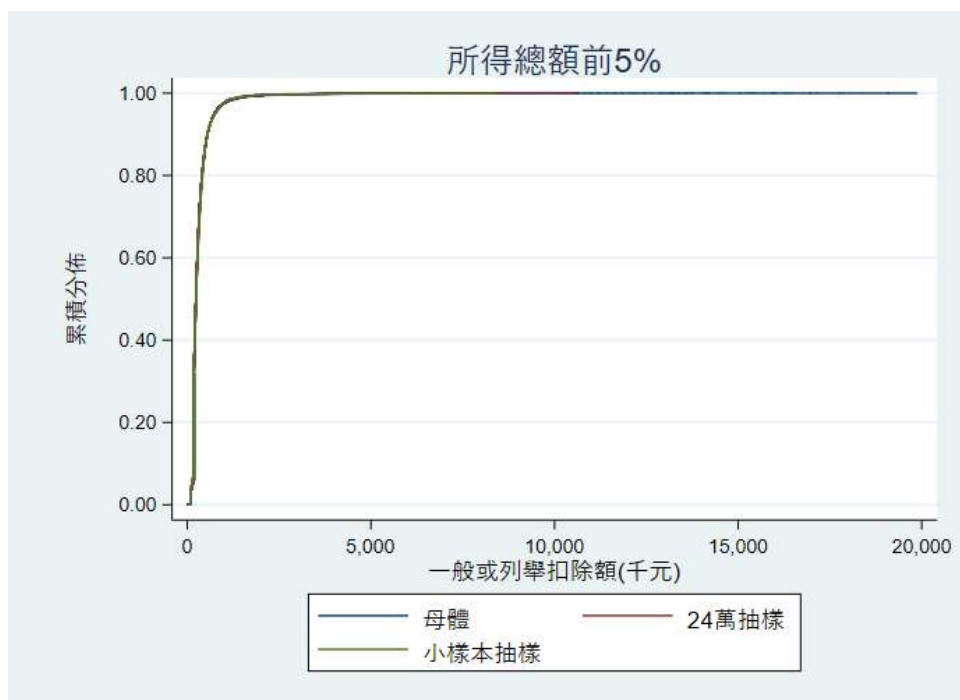


圖 7. ECDF：標準扣除額 - 所得總額前 5%

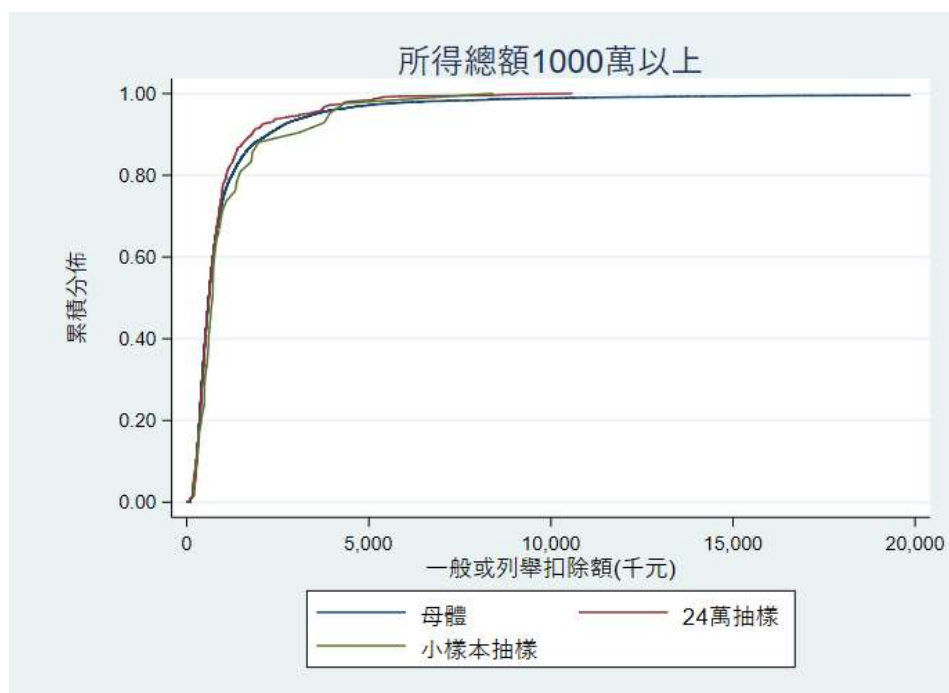


圖 8. ECDF：標準扣除額 - 所得總額 1000 萬以上

(三) 特別扣除額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	198	122	147	6,279,484
24 萬	198	122	147	240,000
小樣本	198	122	146	24,000
所得總額前 1%				
母體	356	158	344	62,794
24 萬	356	156	342	2,436
小樣本	348	138	331	259
所得總額大於 1000 萬				
母體	356	156	352	12,357
24 萬	347	152	344	463
小樣本	358	129	356	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0012	0.497
小樣本	0.0042	0.436
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0294	0.463
小樣本	0.0945	0.473

3. 圖形分佈

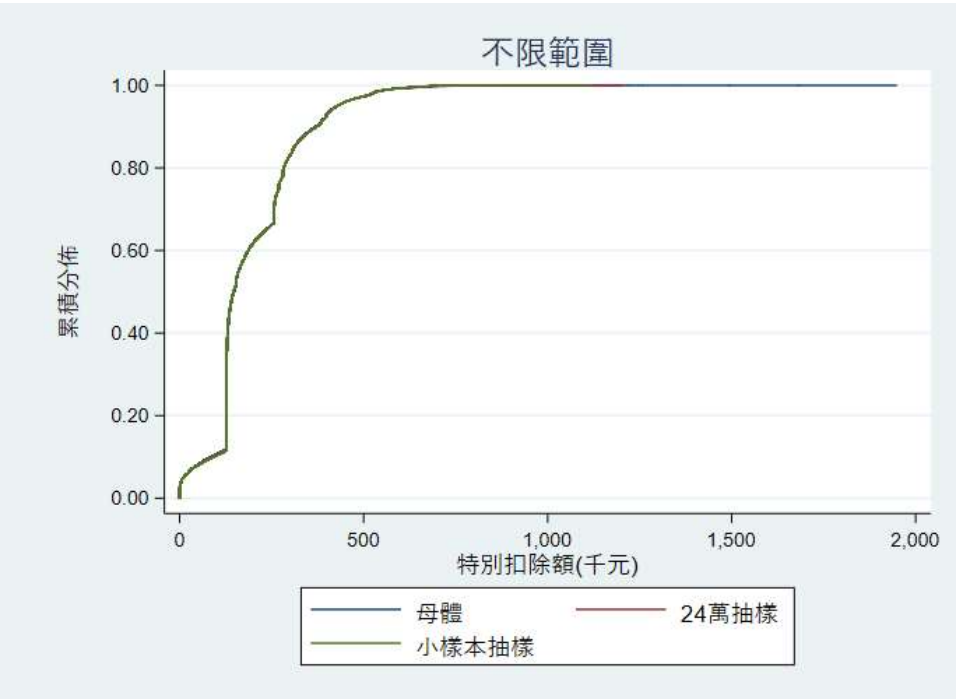


圖 9. ECDF：特別扣除額 - 所有資料

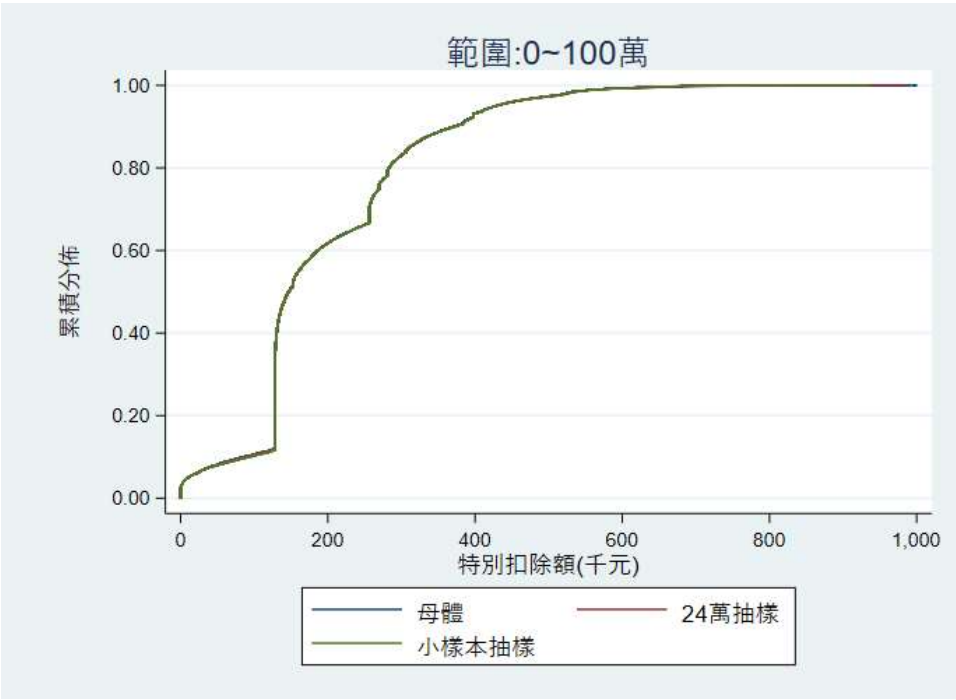


圖 10. ECDF：特別扣除額 - 所有資料 (0 萬~100 萬)

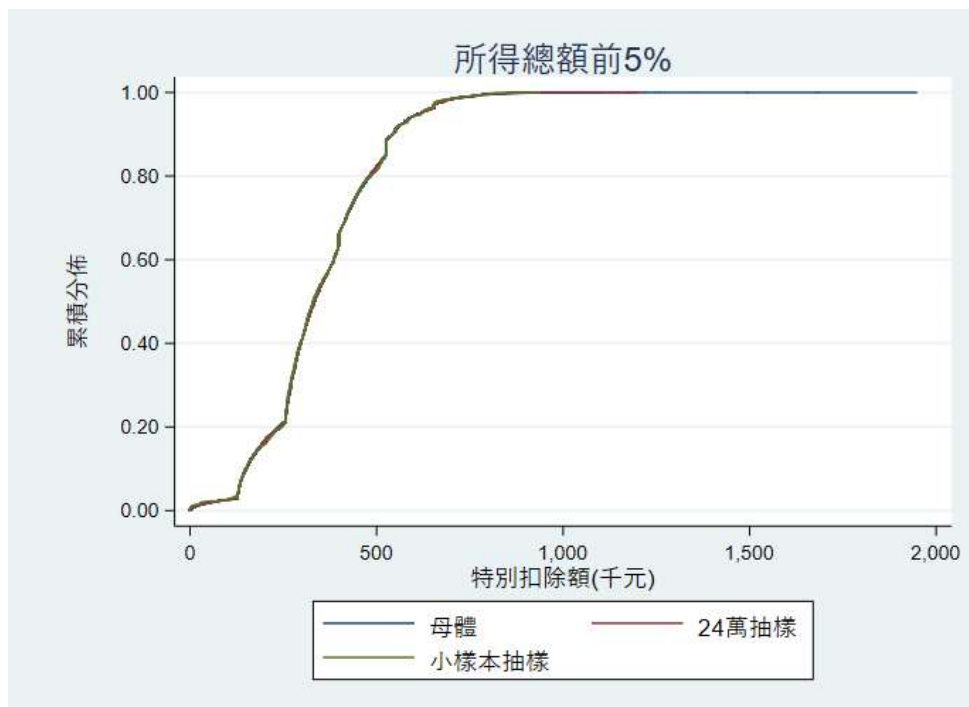


圖 11. ECDF：特別扣除額 - 所得總額前 5%

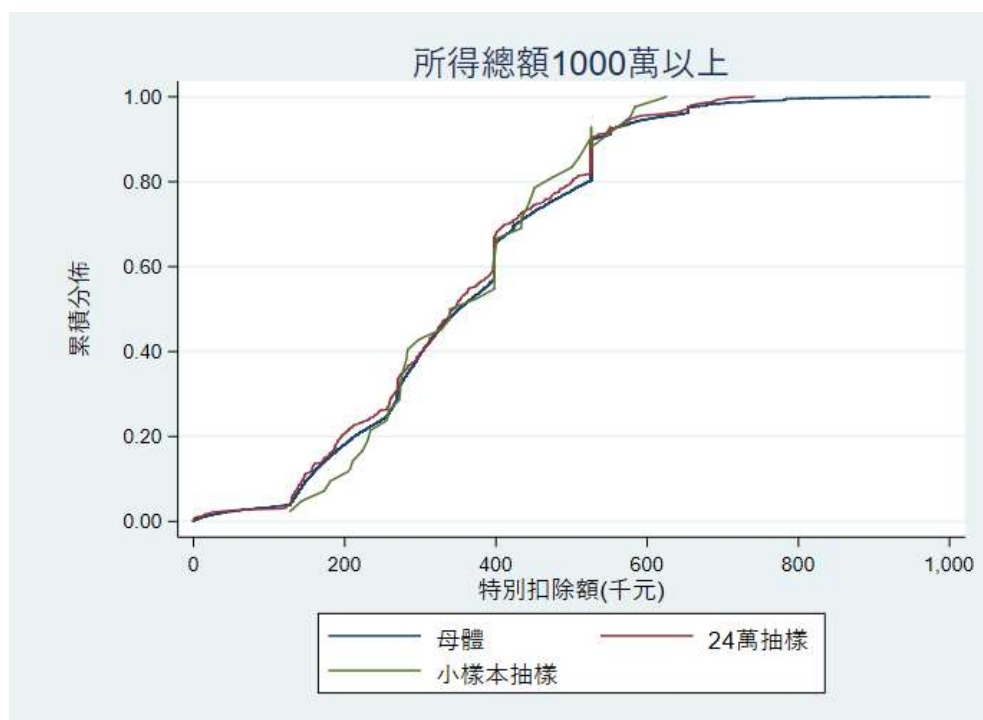


圖 12. ECDF：特別扣除額 - 所得總額 1000 萬以上

(四) 核定所得淨額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	400	4,059	97	6,279,484
24 萬	402	3,822	97	240,000
小樣本	402	1,859	98	24,000
所得總額前 1%				
母體	9,045	39,300	5,273	62,794
24 萬	9,153	36,542	5,240	2,436
小樣本	8,576	15,100	5,322	259
所得總額大於 1000 萬				
母體	24,950	86,757	13,087	12,357
24 萬	26,284	81,651	12,475	463
小樣本	25,670	32,673	16,033	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0016	0.286
小樣本	0.0037	0.515
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0606	0.038
小樣本	0.1493	0.155

3. 圖形分佈

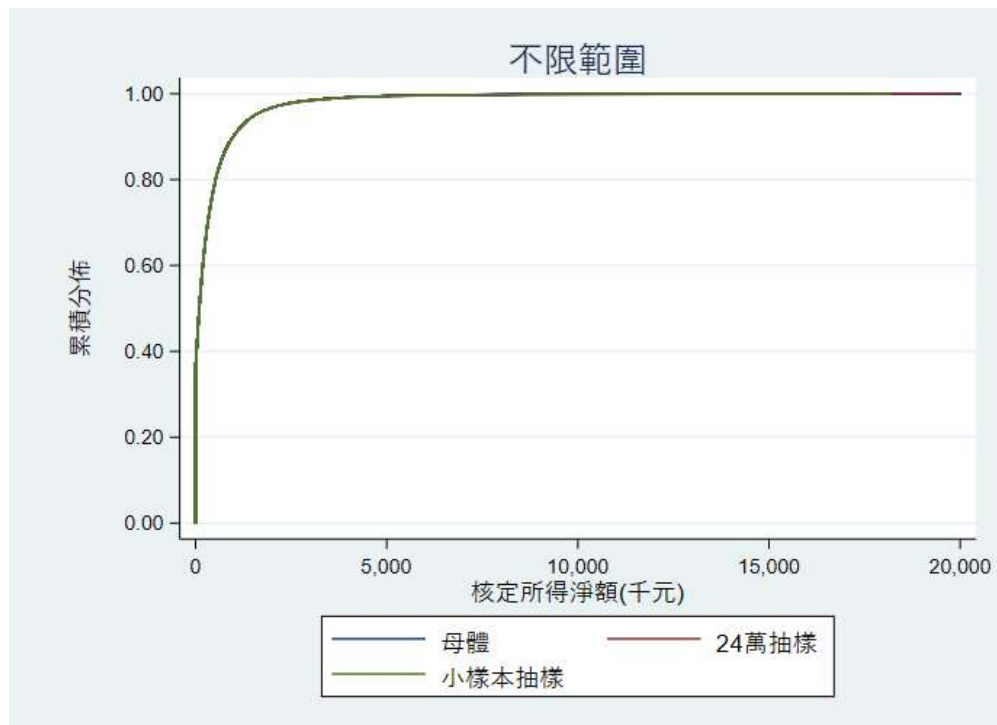


圖 13. ECDF：核定所得淨額 - 所有資料

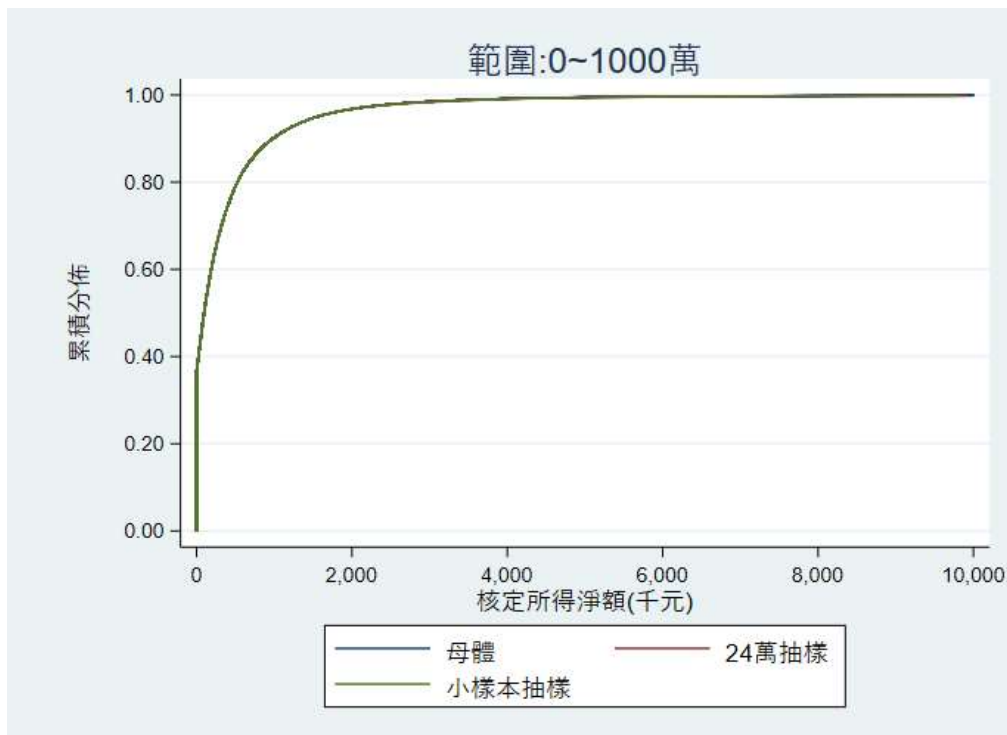


圖 14. ECDF：核定所得淨額 - 所有資料 (0 萬~1000 萬)

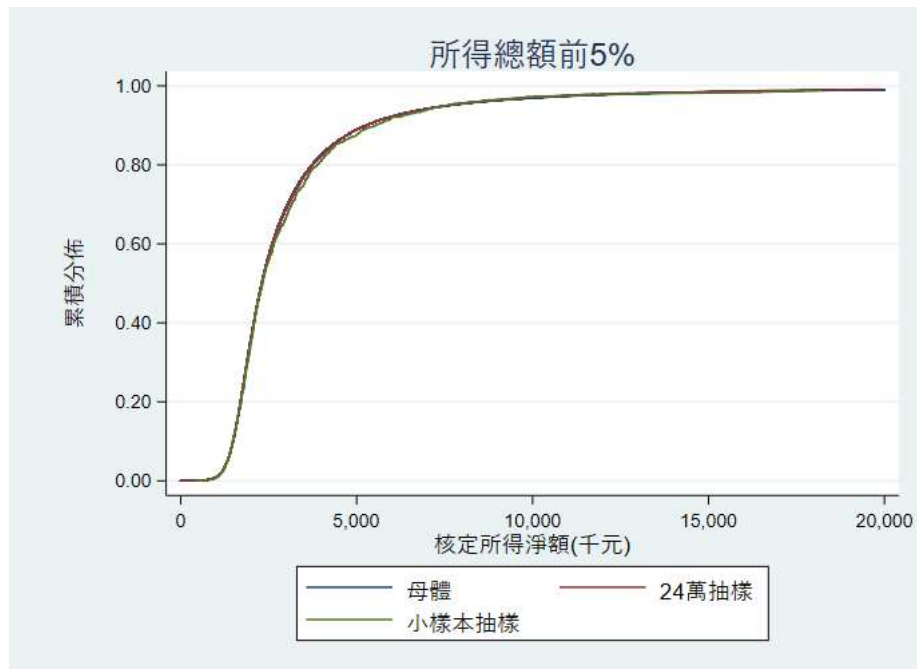


圖 15. ECDF：核定所得淨額 - 所得總額前 5%

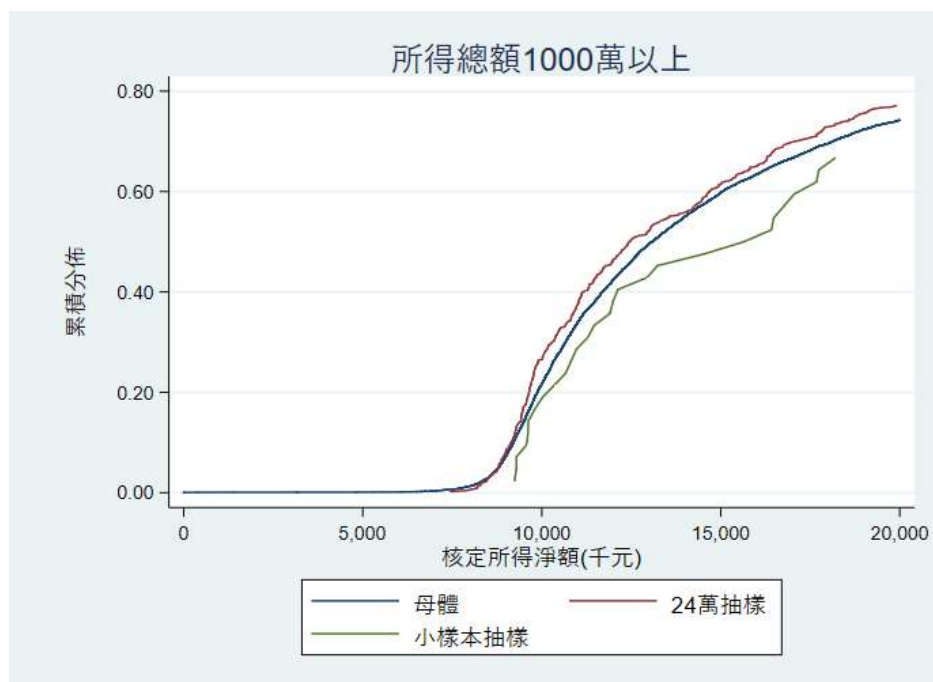


圖 16. ECDF：核定所得淨額 - 所得總額 1000 萬以上

(五) 納稅義務人薪資所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	25	282	0	6,279,484
24 萬	25	265	0	240,000
小樣本	28	359	0	24,000
所得總額前 1%				
母體	576	2,173	0	62,794
24 萬	562	1,940	0	2,436
小樣本	693	2,970	0	259
所得總額大於 1000 萬				
母體	883	3,992	0	12,357
24 萬	906	3,412	0	463
小樣本	1,984	6,651	0	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0003	0.96
小樣本	0.001	0.956
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0119	0.881
小樣本	0.1192	0.304

3. 圖形分佈

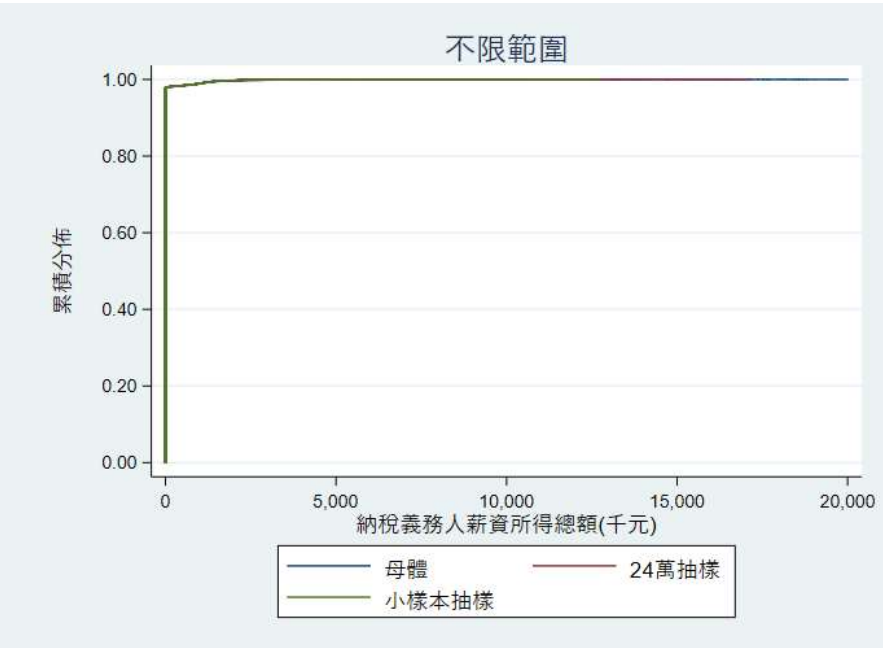


圖 17. ECDF：納稅義務人薪資所得總額 - 所有資料

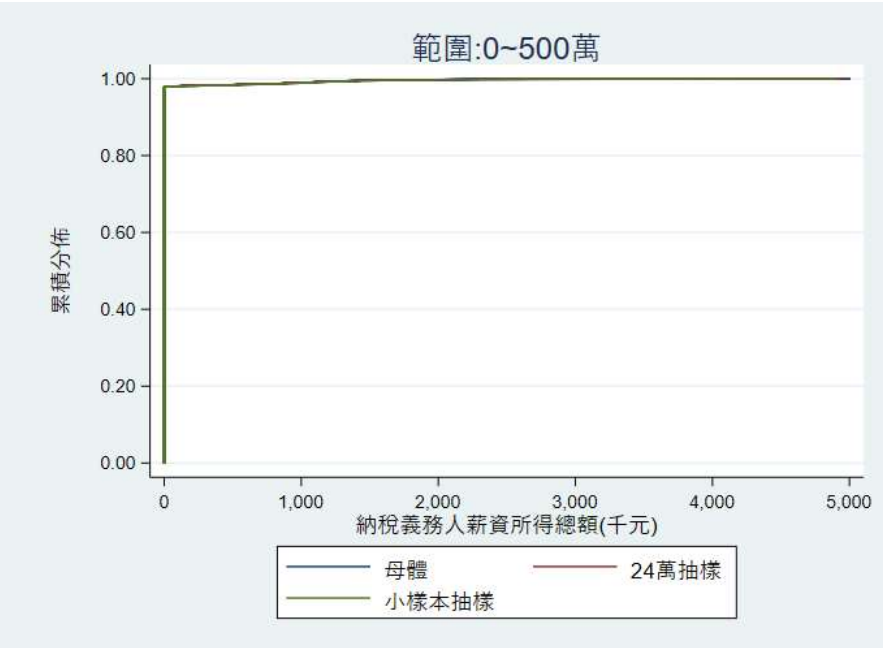


圖 18. ECDF：納稅義務人薪資所得總額 - 所有資料 (0 萬~500 萬)

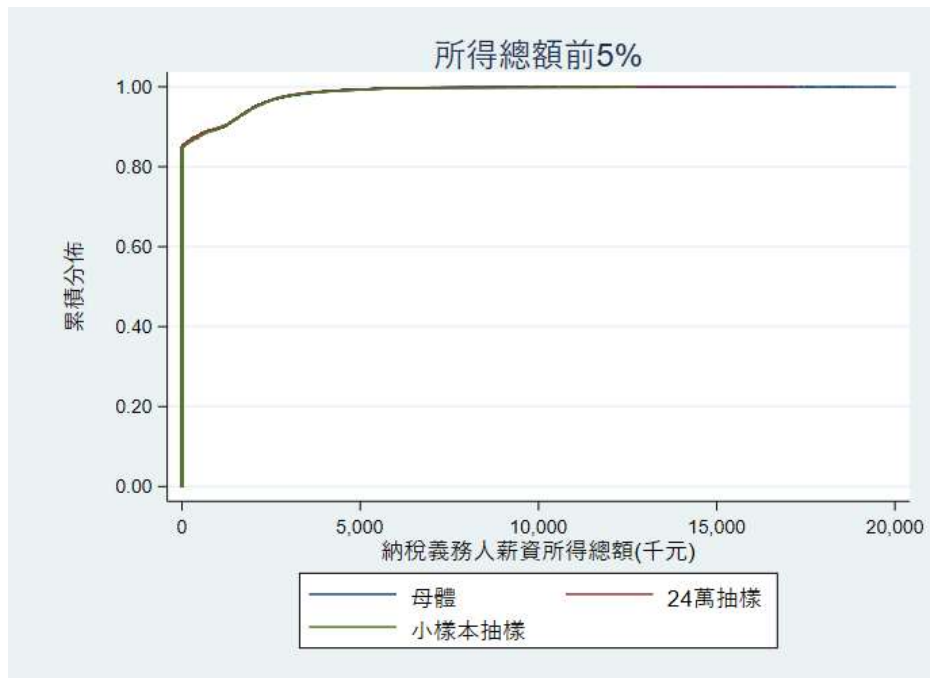


圖 19. ECDF：納稅義務人薪資所得總額 - 所得總額前 5%

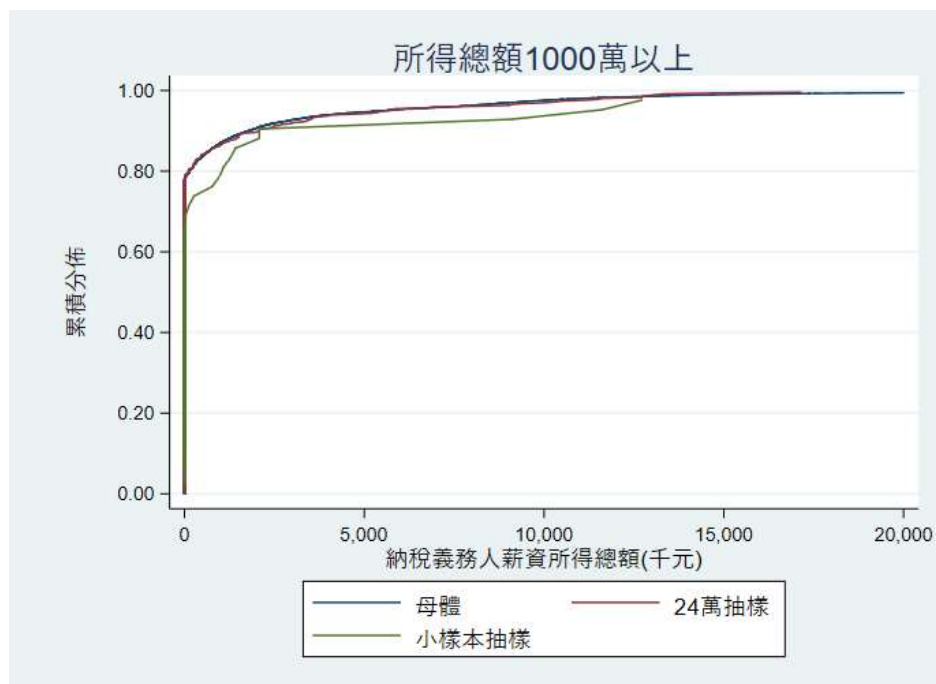


圖 20. ECDF：納稅義務人薪資所得總額 - 所得總額 1000 萬以上

(六) 配偶薪資所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	95	643	0	6,279,484
24 萬	95	611	0	240,000
小樣本	97	520	0	24,000
所得總額前 1%				
母體	1,920	5,623	969	62,794
24 萬	1,902	5,209	914	2,436
小樣本	2,177	3,868	1,074	259
所得總額大於 1000 萬				
母體	4,289	12,013	1,348	12,357
24 萬	4,279	11,205	1,464	463
小樣本	5,649	7,990	2,720	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0007	0.773
小樣本	0.003	0.647
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0254	0.562
小樣本	0.1627	0.109

3. 圖形分佈

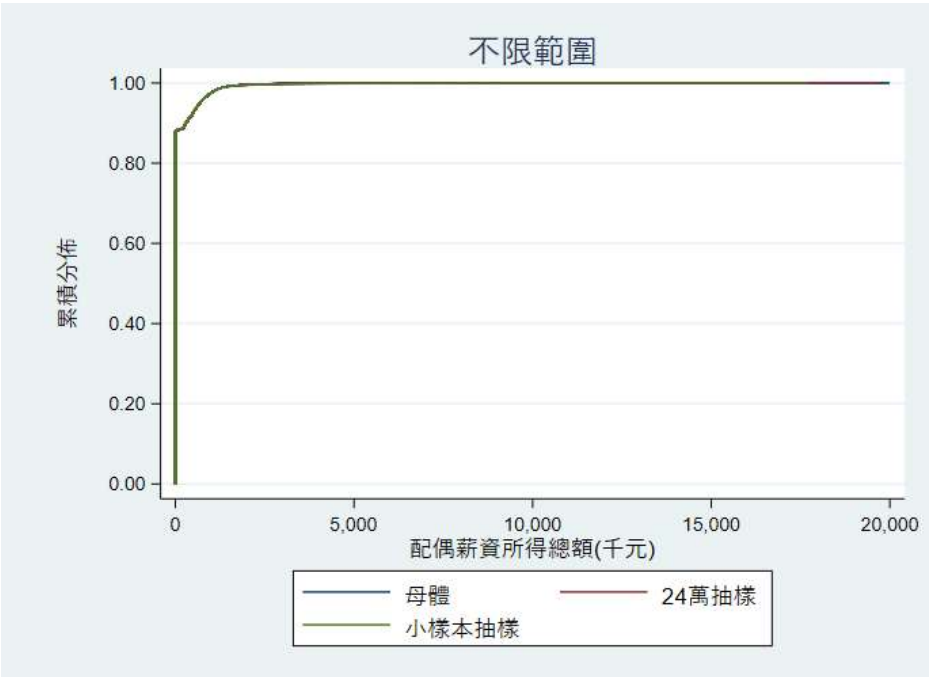


圖 21. ECDF：配偶薪資所得總額 - 所有資料

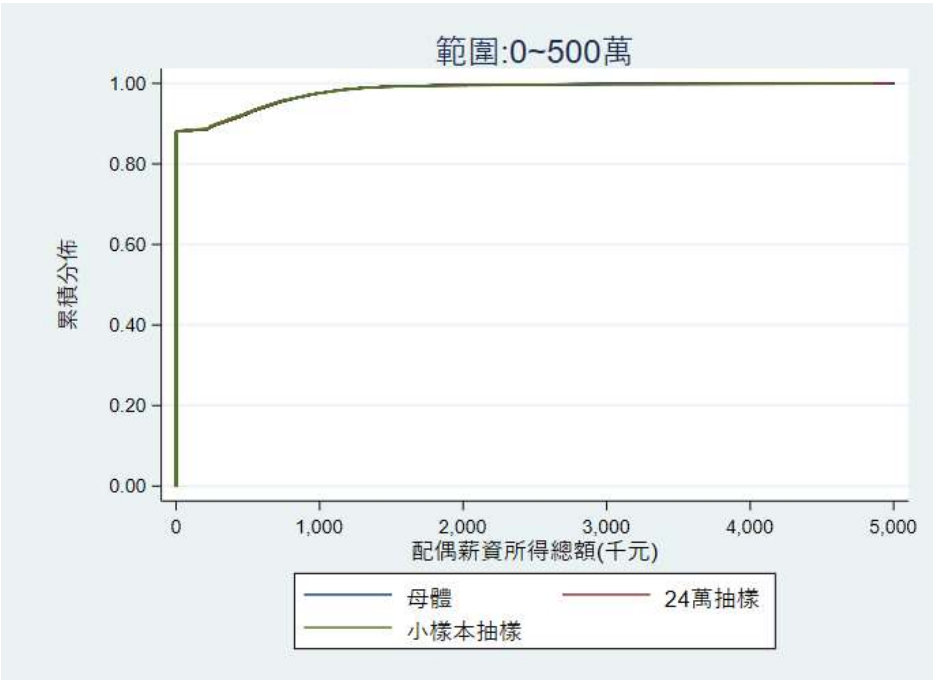


圖 22. ECDF：配偶薪資所得總額 - 所有資料 (0 萬~500 萬)

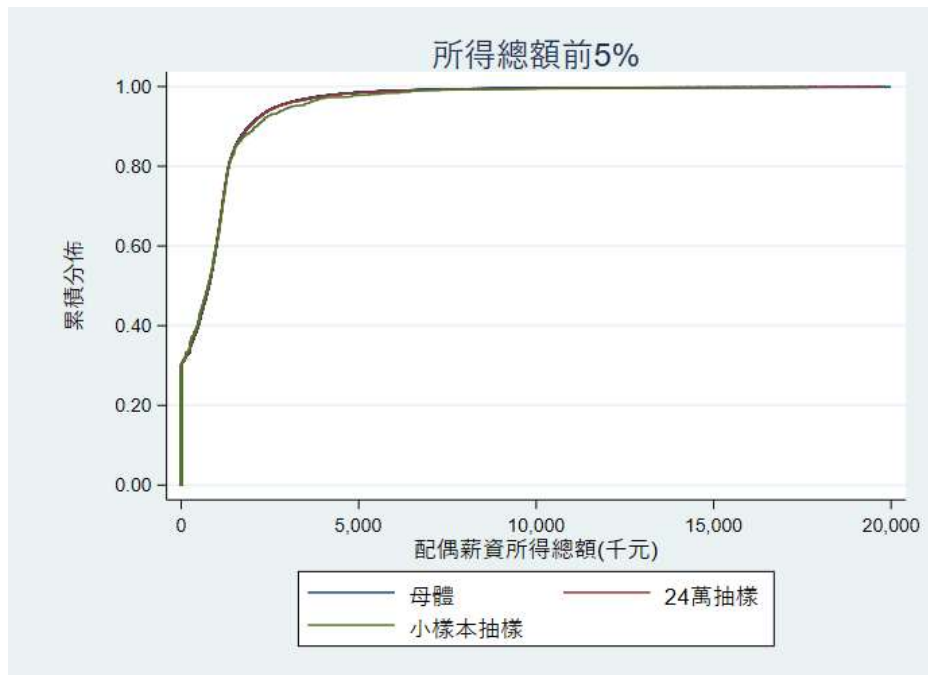


圖 23. ECDF：配偶薪資所得總額 - 所得總額前 5%

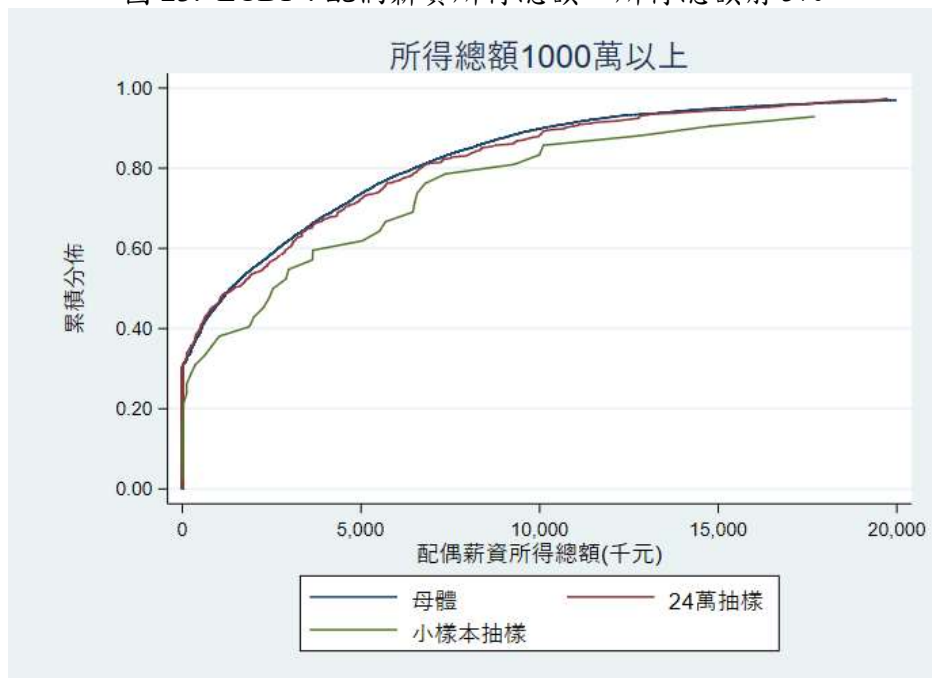


圖 24. ECDF：配偶薪資所得總額 - 所得總額 1000 萬以上

(七) 納稅義務人各類所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	675	4,201	485	6,279,484
24 萬	668	1,439	485	240,000
小樣本	674	1,393	483	24,000
所得總額前 1%				
母體	7,025	41,189	4,793	62,794
24 萬	6,274	12,039	4,786	2,436
小樣本	6,369	11,033	4,639	259
所得總額大於 1000 萬				
母體	18,778	91,811	10,929	12,357
24 萬	15,488	25,242	10,587	463
小樣本	18,812	23,372	12,036	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0018	0.214
小樣本	0.0032	0.605
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.0443	0.174
小樣本	0.1231	0.281

3. 圖形分佈

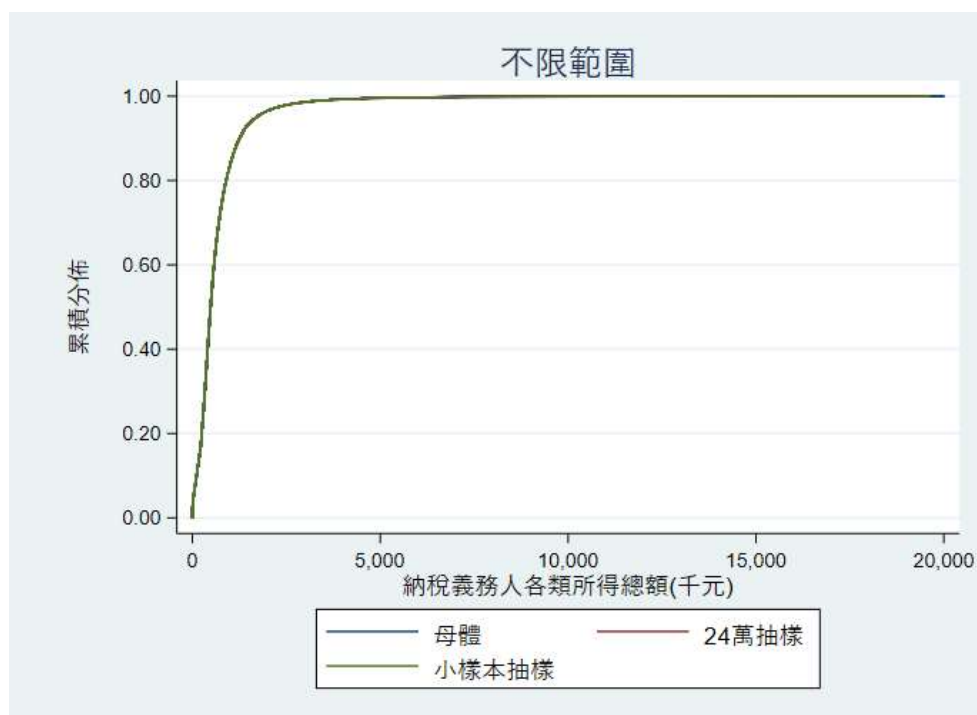


圖 25. ECDF：納稅義務人各類所得總額 - 所有資料

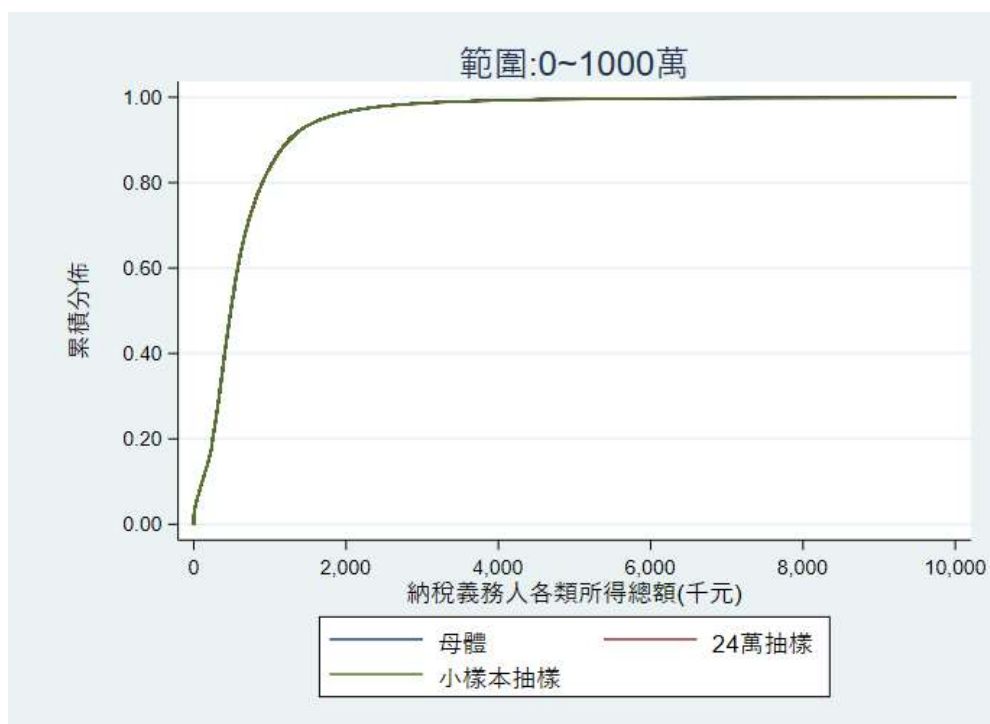


圖 26. ECDF：納稅義務人各類所得總額 - 所有資料 (0 萬~1000 萬)

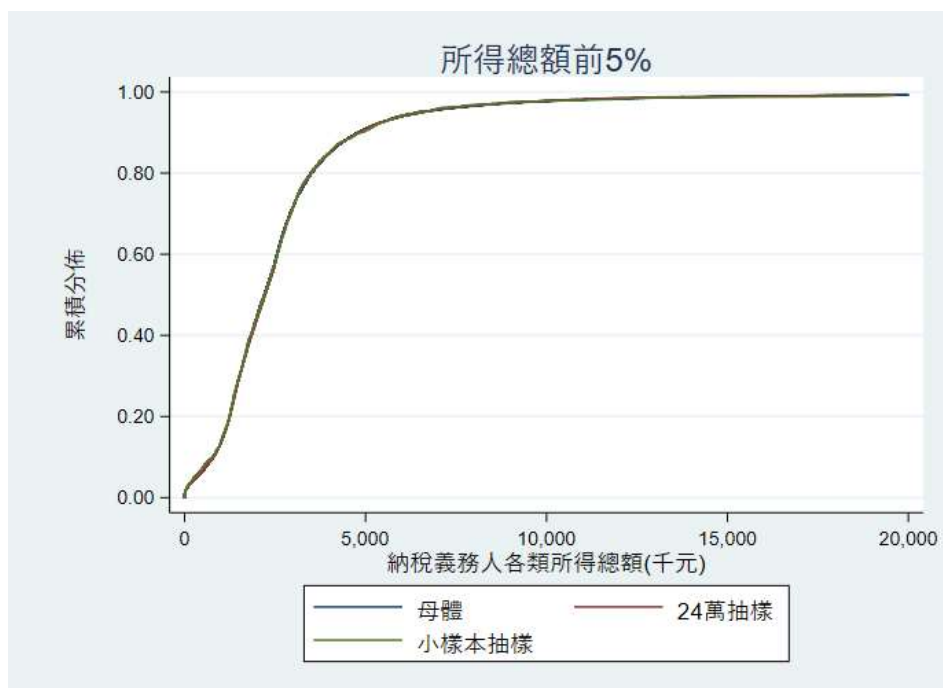


圖 27. ECDF：納稅義務人各類所得總額 - 所得總額前 5%

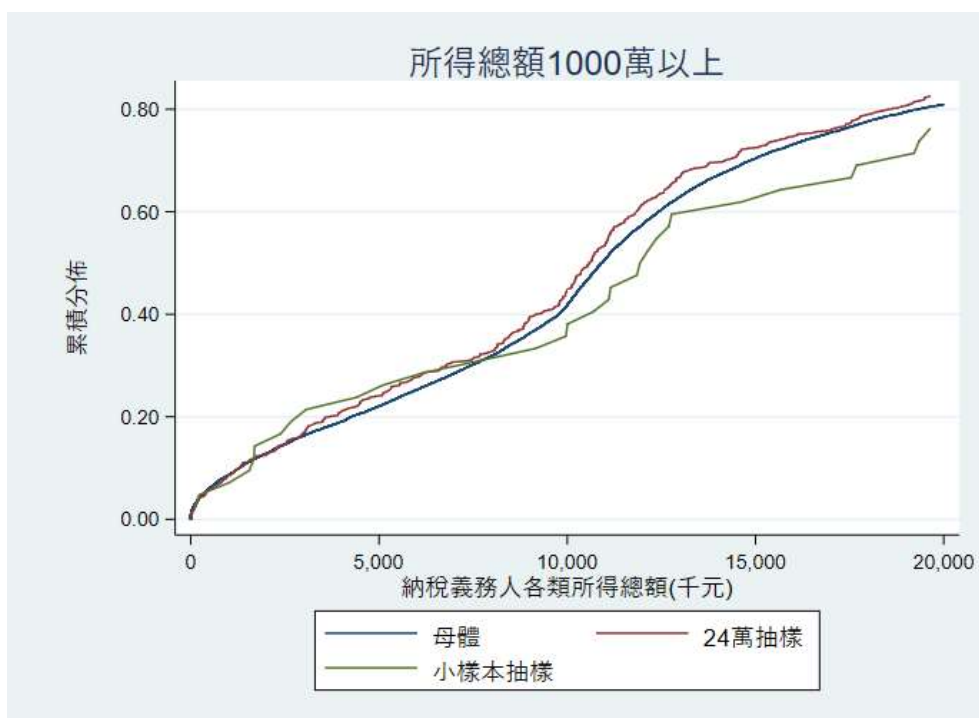


圖 28. ECDF：納稅義務人各類所得總額 - 所得總額 1000 萬以上

(八) 配偶各類所得總額

1. 敘述統計

X 不限所得範圍				
組 別	平均值	標準差	中位數	樣本數
母體	220	1,360	0	6,279,484
24 萬	228	3,135	0	240,000
小樣本	221	975	0	24,000
所得總額前 1%				
母體	3,217	12,667	1,534	62,794
24 萬	3,954	30,647	1,445	2,436
小樣本	3,367	8,015	1,916	259
所得總額大於 1000 萬				
母體	7,796	27,771	2,355	12,357
24 萬	11,804	69,672	1,920	463
小樣本	8,376	18,572	2,710	42

2. 檢定分佈

Kolmogorov-Smirnov 檢定		
母體 vs	最大差異	P-Value
24 萬樣本	0.0017	0.258
小樣本	0.0035	0.55
Kolmogorov-Smirnov 檢定 (所得總額 1000 萬以上)		
母體 vs	最大差異	P-Value
24 萬樣本	0.046	0.151
小樣本	0.0897	0.51

3. 圖形分佈

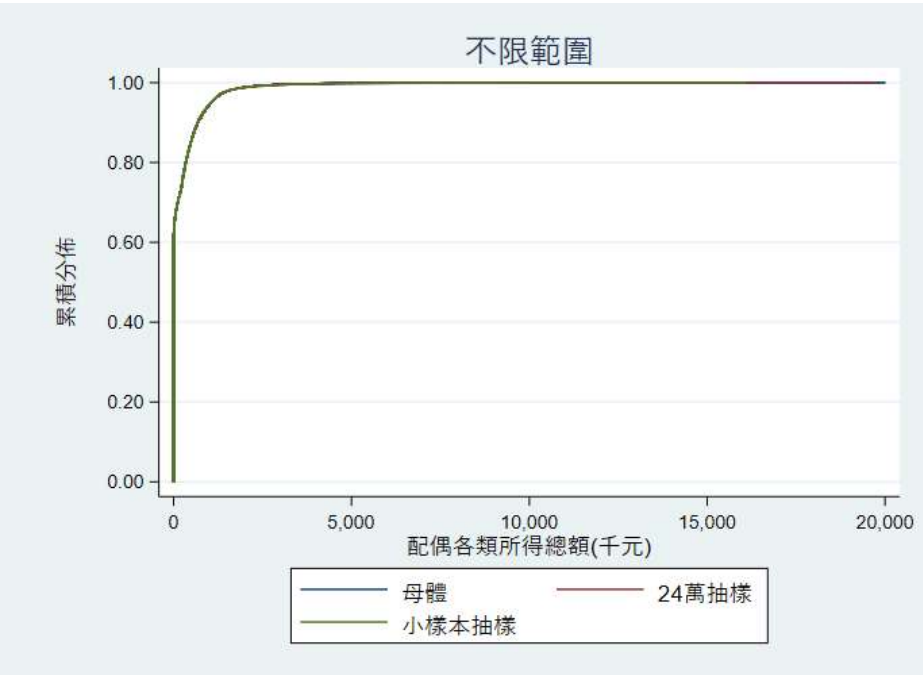


圖 29. ECDF：配偶各類所得總額 - 所有資料

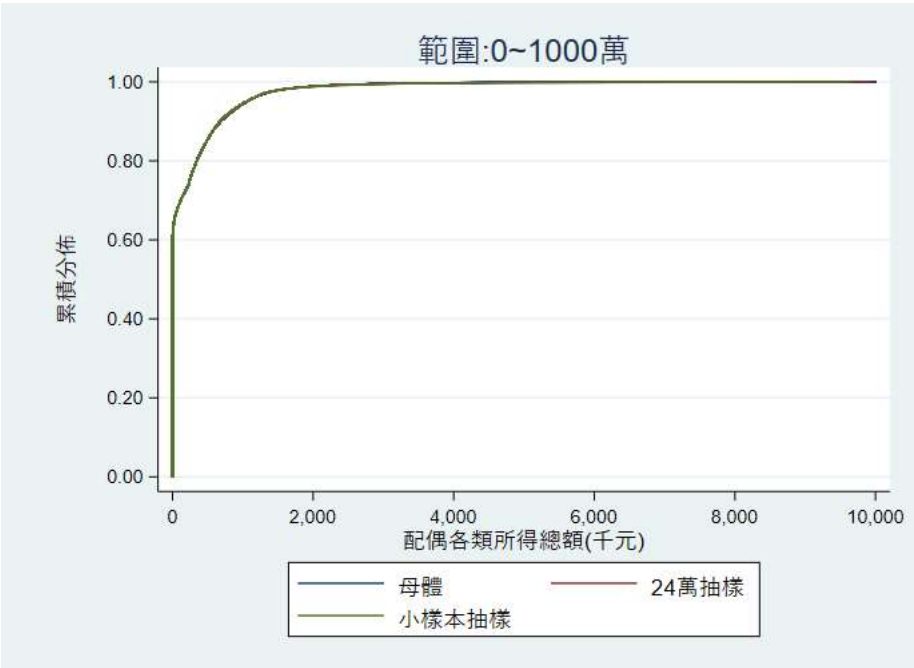


圖 30. ECDF：配偶各類所得總額 - 所有資料 (0 萬~1000 萬)

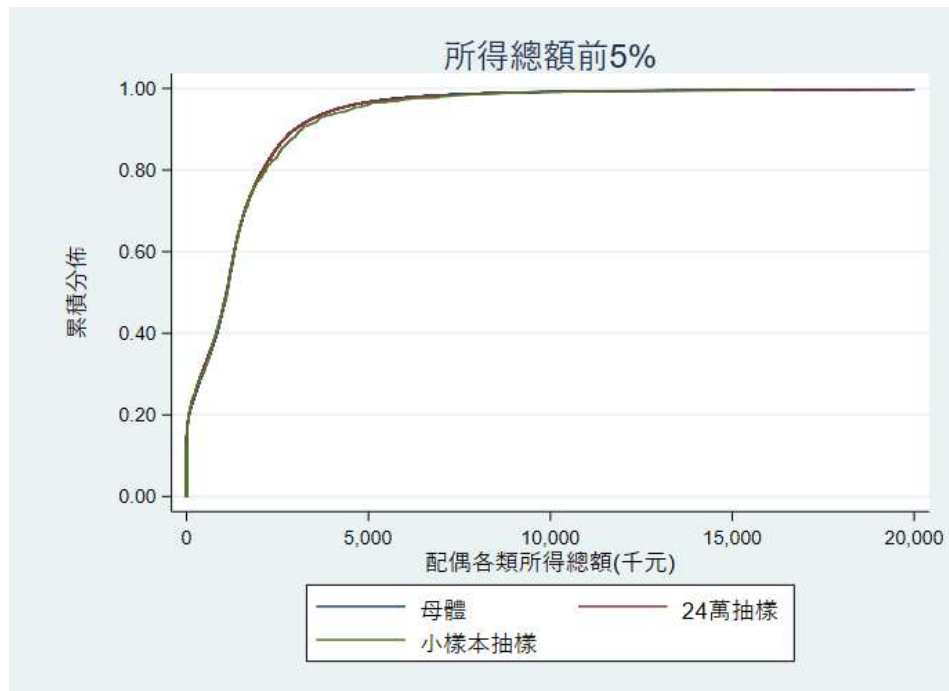


圖 31. ECDF：配偶各類所得總額 -所得總額前 5%

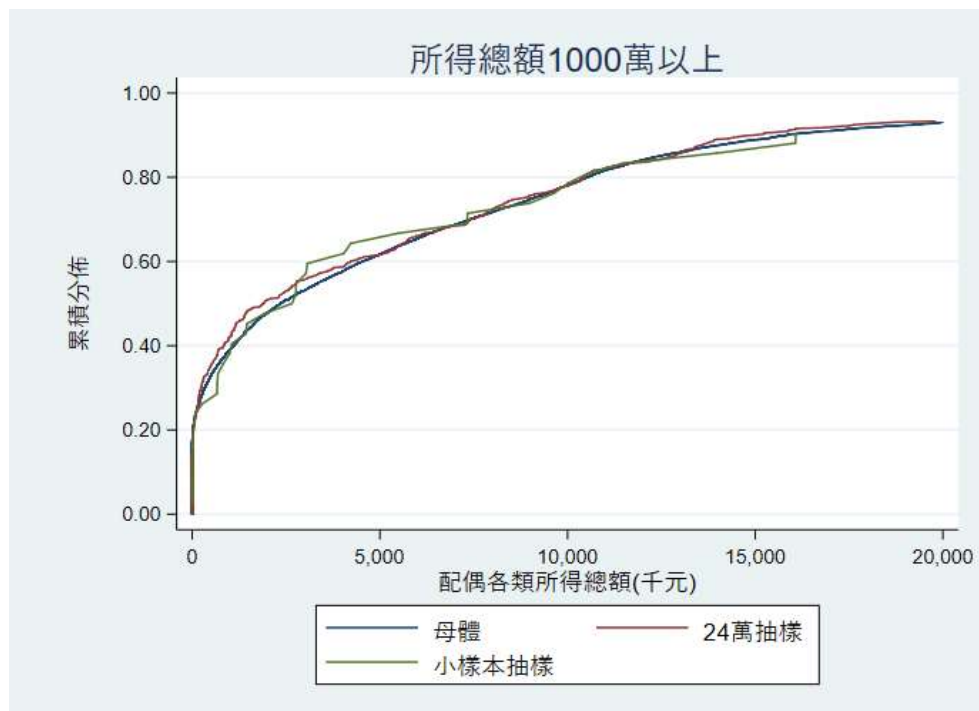


圖 32. ECDF：配偶各類所得總額 -所得總額大於 1000 萬

小樣本抽樣檔驗證結果概述

因平均數與標準差容易受到極端值影響，所得金額分佈又相當大，所以敘述統計的說明會以中位數的比較為基準。

欄位	說明
核定所得總額	1. 當所得前 1% 的資料，敘述統計顯示小樣本與母體差距程度約 20~30 萬差距；所得大於 1000 萬則差距來到 300 萬左右差距。 2. 觀察圖形所得分佈，所得大於 1000 萬時，24 萬樣本與母體有明顯差異，當將樣本分割成 10 組小樣本，即一組 2 萬 4 千筆資料，所得分佈與母體差異更大（見圖 4）。當所得小於 1000 萬，小樣本資料也與母體有差異，但每個所得點上累積機率分佈相差僅約 1%。（見圖 2） 3. 因此，小樣本的驗證結果與 24 萬樣本的驗證結果皆在所得前 1% 有疑慮，而所得大於 1000 萬則未通過，且小樣本驗證結果更差。
標準扣除額	1. 敘述統計顯示所得大於 1000 萬差距約 10 萬左右，其他範圍的標準扣除額差異皆不到 1 萬。 2. 觀察圖形所得分佈，所得大於 1000 萬時，小樣本所得資料分佈與母體差異明顯（見圖 8）。當所得總額小於 1000 萬的圖形分佈皆看不出差異。 3. 敘述統計顯示當所得大於 1000 萬時，小樣本與母體差異不大，但圖形分佈小樣本與母體具明顯差異；因此，驗證結果於所得 1000 萬以上時有疑慮，但整體差異不大。
特別扣除額	1. 敘述統計顯示不論所得大於或是小於 1000 萬特別扣除額差異皆不大。 2. 觀察圖形所得分佈，所得大於 1000 萬時，小樣本所得資料分佈對比 24 萬樣本與母體差異明顯（見圖 12）。當所得總額小於 1000 萬的圖形分佈皆看不出差異。 3. 雖然敘述統計顯示當所得大於 1000 萬時，小樣本與母體不具明顯差異，但圖形分佈可以看出不同百分位的差距（圖 12 顯示約 20 與 80 百分位部分差異明顯）；因此，驗證結果於所得 1000 萬以上時有疑慮，但整體差異不大。
核定所得淨額	1. 敘述統計顯示所得大於 1000 萬差異約 300 萬左右。 2. 觀察圖形所得分佈，所得大於 1000 萬時，小樣本與母體差異更甚 24 萬樣本與母體差異（見圖 16）。 3. 因此，小樣本的驗證結果在高所得族群（所得大於 1000 萬）未通過。
納稅義務人薪資所得總額	1. 觀察圖形所得分佈，資料皆集中在 80 百分位以上，且僅在所得大於 1000 萬時，小樣本與母體有明顯的差異（見圖 20）。 2. 因此，小樣本的驗證結果僅在高所得族群（所得大於 1000 萬）未通過。
配偶薪資所得總額	3. 敘述統計顯示所得大於 1000 萬差異約 100 萬左右，其他部分差異不明顯。 4. 觀察圖形所得分佈，在所得大於 1000 萬時，小樣本與母體有明顯的差異（見圖 24）。

	5. 因此，小樣本的驗證結果僅在高所得族群（所得大於 1000 萬）未通過。
納稅義務人各類所得總額	1. 敘述統計顯示所得大於 1000 萬差異約 100 萬左右，其他部分差異不明顯。 2. 觀察圖形所得分佈，在所得大於 1000 萬時，小樣本與母體有明顯的差異(見圖 28)。 3. 因此，小樣本的驗證結果僅在高所得族群（所得大於 1000 萬）未通過。
配偶各類所得總額	1. 觀察圖形所得分佈，僅在所得大於 1000 萬時，小樣本與母體有明顯的差異(見圖 32)。 2. 因此，小樣本的驗證結果僅在高所得族群（所得大於 1000 萬）未通過。