



11/13統計學實習課

114-1統計學實習課

第三次作業講解—抽樣方法

114-1統計學實習課

第三次作業第六題

6. The sampling methods often have a big impact on the behavior of the sample average. Based on the following data (the height of a singer, in inches), use simple random sampling, stratified random sampling, cluster sampling, and systematic sampling to draw 20 observations. Continue the process 10 times and discuss the results you have.

女高音36位、女低音35位、男高音20位、男低音39位，女性共71位，男性共59位，總共130筆身高的資料。
➤ 分別用簡單隨機抽樣、分層隨機抽樣、群集抽樣、系統抽樣，從130筆資料中抽出20筆，計算其平均值及變異數，並重複此動作10次，分析不同抽樣方法的結果差異。

Soprano	Alto	Tenor	Bass
64	65	69	72
62	62	72	70
66	68	71	72
65	67	66	69
60	67	76	73
61	63	74	71
65	67	71	72
66	66	66	68
65	63	68	68
63	72	67	71
67	62	70	66
65	61	65	68
62	66	72	71
65	64	70	73
68	60	68	73
65	61	73	70
63	66	66	68
65	66	68	70
62	66	67	75
65	62	64	68
66	70		71
62	65		70
65	64		74
63	63		70
65	65		75
66	69		75
65	61		69
62	66		72
65	65		71
66	61		70
65	63		71
61	64		68
65	67		70
66	66		75
65	68		72
62			66
			72
			70
			69

由圖1可知男性、女性身高差異很大，抽樣時應考慮性別差異調整樣本的男女性人數。

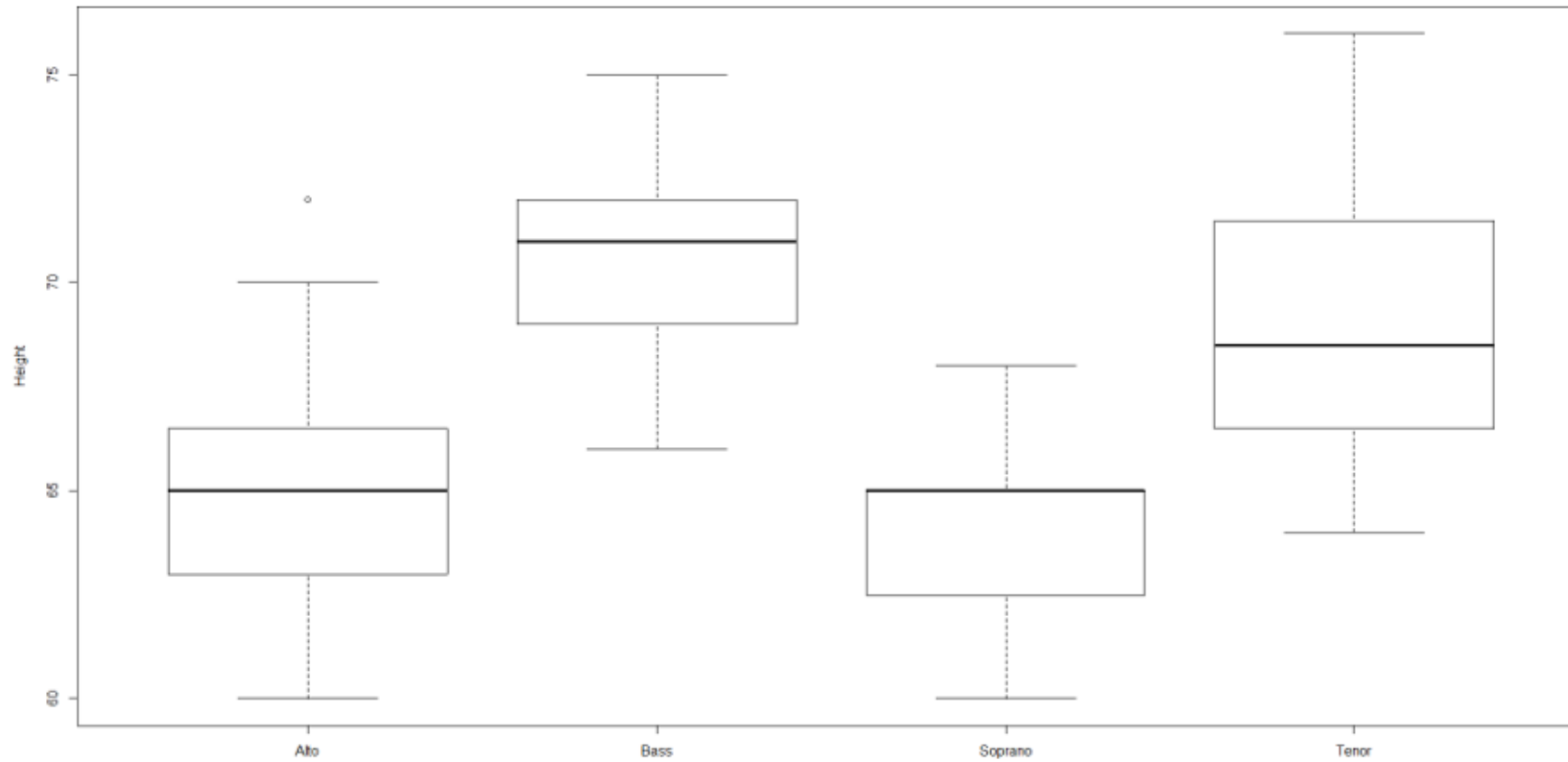


圖 1、四個聲部成員身高箱型圖←

1. 簡單隨機抽樣

→ 簡單隨機抽樣需先將所有成員排成一行（清單），再從中隨機抽出（不放回）

20 位，數值代表被抽出成員的相對位置，例如：

1 10 11 36 47 58 65 70 81 82 83 87 91 92 101 108 111 116 123 126

抽出的 20 個觀察值平均身高為 67.35 英吋，與母體平均值略有差異。

清單	64	62	66	65	60	61	65	66	65	...
----	----	----	----	----	----	----	----	----	----	-----



抽20位

2. 分層隨機抽樣

→分層隨機抽樣需要先決定「層」，原則上要求「層內同質、層間異質」，由於性別差異將母體拆成「男」、「女」兩組(層)，兩組在母體比例分別為 45.4%、54.6%。預計抽出 20 位成員中男女性分別為 9 位($\approx 20 * 45.4\%$)、11 位($\approx 20 * 54.6\%$)，男女性樣本抽取則採簡單隨機抽樣。模擬結果得出平均身高 67.3 英吋。

女性	64	62	66	65	60	61	65	66	65	...	→ 抽11位
男性	69	72	71	66	76	74	71	66	68	...	→ 抽9位

3. 群集抽樣

→ 群集抽樣與分層隨機抽樣類似，但要求群與群的差異很小，每次抽出其中幾群，配合樣本數 20 個的設定，因此將母體分成 13 組（群）、每群 10 個，抽樣時只需選擇其中兩群。每群 10 個的分組原則上要求男女比例和母體(45.4%、54.6%)接近，無法取得整數，因此某些群的男女個數為 4 及 6（共有 6 群）、某些為 5 及 5（共有 7 群）。實際分群方式請同學自行調整，模擬結果的平均身高為 66.85 英吋。

第1群	64	62	66	65	60	61	69	72	71	66	→ 女6男4
第2群	65	66	65	63	67	65	76	74	71	66	
第3群	62	65	68	65	63	65	68	67	70	65	
					.						
					.						
					.						
第13群	61	63	64	67	66	72	66	72	70	69	→ 女5男5

抽2群

4. 系統抽樣

→系統抽樣為固定每隔幾個抽出一個觀察值，較直接的作法仿造簡單隨機抽樣列出一張清單，預計從母體 130 位成員抽出 20 個觀察值，亦即每隔 $130/20 \approx 7$ 個抽出一個。操作方法為先從 1 到 130 隨機選擇一個數字，接著間隔 6 抽出觀察值，我抽出了第 20 位，因此抽出的樣本順序為：

20→27→34→41→...→118→125→132→139→146→153

2→ 9→ 16→ 23

超過 130 之後再減去 130。這組樣本的平均身高為 67.25 英吋。

清單	64	62	66	65	60	61	65	66	65	...
----	----	----	----	----	----	----	----	----	----	-----

→ 抽第1位

之後每次間隔6抽出下一個觀察值

Excel 抽樣(參考9/9實習課)

Ex：以簡單隨機抽樣抽出一人

- 使用函數： INDEX()、RANDBETWEEN() 與 ROWS()。
- 一般公式：=INDEX(名單範圍,RANDBETWEEN(1,ROWS(名單範圍)))
- 公式邏輯：這個公式是先以 ROWS 函數取得名單的數量，接著靠著 RANDBETWEEN 函數產生一個介於 1 到名單數量之間的隨機整數，最後依照這個隨機的整數作為索引值，以 INDEX 函數取出對應位置的資料。

Excel抽樣(參考9/9實習課)

Ex：以簡單隨機抽樣抽出五人

- 使用函數：使用函數： INDEX() 與 RANK()。
- 一般公式：=INDEX(名單範圍,RANK(亂數值,亂數範圍))
- 公式邏輯：這個公式會使用 RANK 函數從亂數資料中，產生隨機的整數，而由於每次指定的亂數值都是不同的，所以產生的隨機索引值也就不會重複，最後將索引值交給 INDEX 函數抽出不重複的名單，將公式打好後，再下拉表格至自己希望抽樣的數量即可。