Stratified Systematic Random Sampling

Jason Ellwood 2016

This example uses the ‘Kiwi’ population data set.

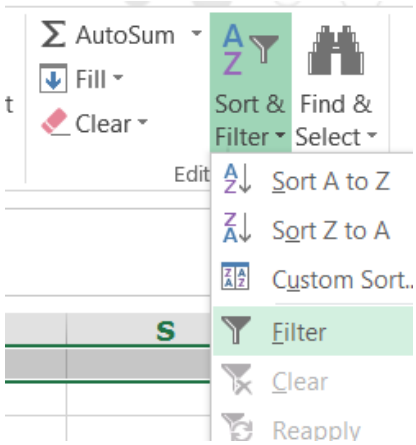
It is based on investigating the question: “I wonder if the median weight of male kiwis is heavier than the median weight of female kiwis”

1. Add a column to the data set and name it ‘sample’

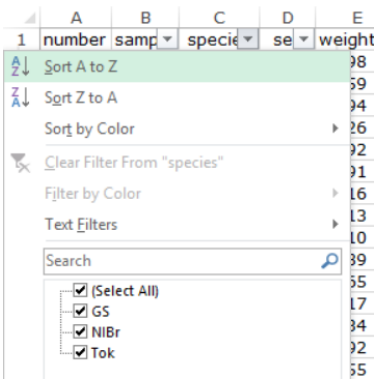
	A	B	C	D	E	F	G	H
1	number	sample	species	sex	weight	height	location	
2	1		Tok	F	2.398	40.3	SF	
3	2		NIBr	M	2.159	38.3	N	
4	3		Tok	M	2.294	37.0	StI	
5	4		Tok	M	1.826	36.6	NF	
6	5		Tok	F	2.292	39.8	StI	

2. Select header cells **excluding the numbering variable** (B1:F1) and turn filters on

	A	B	C	D	E	F	G	H
1	number	sample	species	sex	weight	height	location	
2	1		Tok	F	2.398	40.3	SF	
3	2		NIBr	M	2.159	38.3	N	
4	3		Tok	M	2.294	37.0	StI	
5	4		Tok	M	1.826	36.6	NF	
6	5		Tok	F	2.292	39.8	StI	
7	6		NIBr	F	2.691	37.7	N	
8	7		GS	M	2.216	44.3	CW	
9	8		GS	F	3.213	48.5	EC	
10	9		NIBr	F	2.810	42.1	N	
11	10		Tok	M	1.689	36.4	StI	
12	11		GS	M	2.165	44.2	EC	
13	12		Tok	F	2.317	39.8	SF	
14	13		Tok	F	3.234	37.6	StI	
15	14		GS	F	2.992	46.4	NWN	

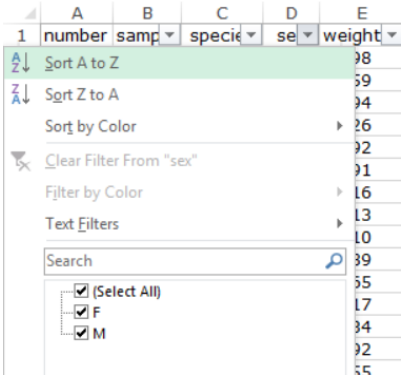


3. Using the filter drop downs, sort the data by each of the variables you believe will influence the variable of interest (in this case, species may influence the weight of a Kiwi)



	A	B	C	D	E	F	G
1	number	samp	specie	se	weight	heigh	locati
2	1		GS	M	2.216	44.3	CW
3	2		GS	F	3.213	48.5	EC
4	3		GS	M	2.165	44.2	EC
5	4		GS	F	2.992	46.4	NWN
6	5		GS	M	2.246	45.0	CW
7	6		GS	M	2.834	43.9	EC
8	7		GS	F	3.311	46.9	EC
9	8		GS	F	3.186	47.2	EC
10	9		GS	F	2.878	48.8	EC
11	10		GS	M	2.050	44.1	CW
12	11		GS	M	2.022	44.3	EC
13	12		GS	M	2.497	44.9	CW
14	13		GS	F	3.580	47.0	NWN
15	14		GS	M	1.984	43.7	NWN

4. Next, sort the data by the grouping variable being investigated (in this case gender).



	A	B	C	D	E	F	G
1	number	samp	specie	se	weight	heigh	locati
2	1		GS	F	3.213	48.5	EC
3	2		GS	F	2.992	46.4	NWN
4	3		GS	F	3.311	46.9	EC
5	4		GS	F	3.186	47.2	EC
6	5		GS	F	2.878	48.8	EC
7	6		GS	F	3.580	47.0	NWN
8	7		GS	F	3.291	48.0	EC
9	8		GS	F	3.595	49.9	EC
10	9		GS	F	3.436	46.8	CW
11	10		GS	F	3.374	48.2	CW
12	11		GS	F	3.163	47.3	NWN
13	12		GS	F	3.497	46.4	CW
14	13		GS	F	3.503	50.4	NWN
15	14		GS	F	3.749	50.6	NWN

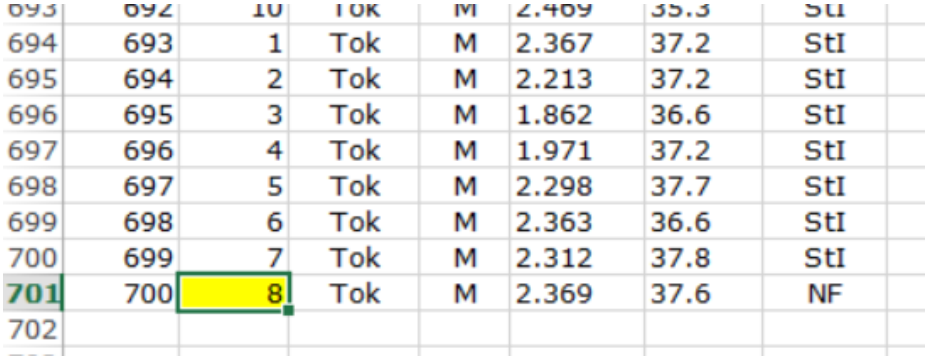
5. Now, identify the number of items in your population (n = 700 kiwis) – ask about how to do this in Excel.
6. Generate a random start point for your sample – type =randbetween(1,n) into any cell.
7. Figure out the interval size for your systematic sample. I like to work on around 30 pieces of data per sample, so oversample 70 items from the population. Provided that the number of items in each level of the grouping variable is roughly equal and because we sorted by this variable last, this should work out fine. The interval size becomes the population size divided by 70 (10 in our case). This potentially produces an unequal number in each sample (an unequal number of male and female kiwis), which again challenges the students' understanding of the inferential process.
8. Start at the random start point (73 in the example). Enter the number 1 in the cell directly next to this item number. Number down from her until you reach your interval size (10 in this case).

73	72		GS	F	3.773	48.6	NWN
74	73	1	GS	F	3.007	45.9	EC
75	74	2	GS	F	3.381	48.7	CW
76	75	3	GS	F	3.448	48.5	EC
77	76	4	GS	F	3.386	49.2	CW
78	77	5	GS	F	3.313	48.0	EC
79	78	6	GS	F	3.573	47.4	CW
80	79	7	GS	F	3.290	45.9	EC
81	80	8	GS	F	3.940	49.6	EC
82	81	9	GS	F	2.994	48.6	NWN
83	82	10	NIBr	F	2.691	37.7	N
84	83		NIBr	F	2.810	42.1	N

9. In the next cell down, reference the cell where you entered the number 1 (in the example, type =B74)

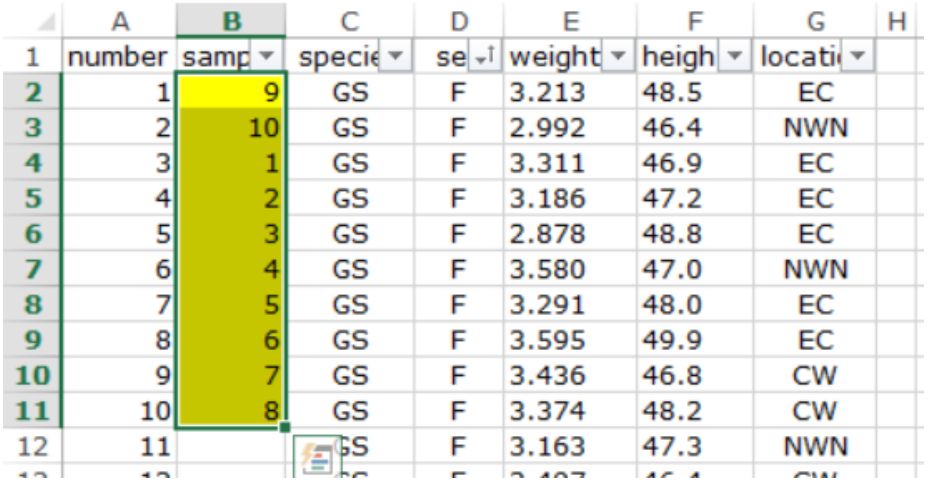
73	72		GS	F	3.773	48.6	NWN
74	73	1	GS	F	3.007	45.9	EC
75	74	2	GS	F	3.381	48.7	CW
76	75	3	GS	F	3.448	48.5	EC
77	76	4	GS	F	3.386	49.2	CW
78	77	5	GS	F	3.313	48.0	EC
79	78	6	GS	F	3.573	47.4	CW
80	79	7	GS	F	3.290	45.9	EC
81	80	8	GS	F	3.940	49.6	EC
82	81	9	GS	F	2.994	48.6	NWN
83	82	10	NIBr	F	2.691	37.7	N
84	83	=B74	NIBr	F	2.810	42.1	N
85	84		NIBr	F	2.541	38.5	F

10. Fill this formula to the bottom of the population. Identify which number you are up to at the final population item (try Ctrl+↓)



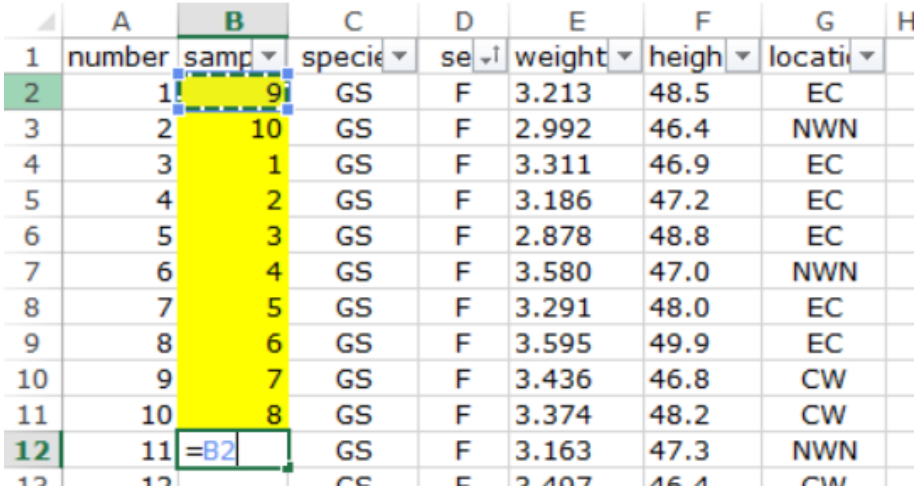
693	692	10	Tok	M	2.469	35.3	StI
694	693	1	Tok	M	2.367	37.2	StI
695	694	2	Tok	M	2.213	37.2	StI
696	695	3	Tok	M	1.862	36.6	StI
697	696	4	Tok	M	1.971	37.2	StI
698	697	5	Tok	M	2.298	37.7	StI
699	698	6	Tok	M	2.363	36.6	StI
700	699	7	Tok	M	2.312	37.8	StI
701	700	8	Tok	M	2.369	37.6	NF
702							

11. Now continue numbering at the top of the population. You will need to manually enter the first cycle.



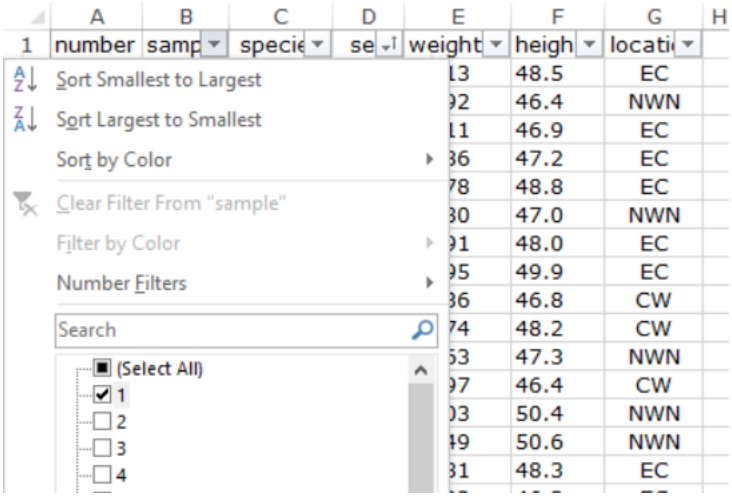
	A	B	C	D	E	F	G	H
1	number	samp	specie	se	weight	heigh	locati	
2	1	9	GS	F	3.213	48.5	EC	
3	2	10	GS	F	2.992	46.4	NWN	
4	3	1	GS	F	3.311	46.9	EC	
5	4	2	GS	F	3.186	47.2	EC	
6	5	3	GS	F	2.878	48.8	EC	
7	6	4	GS	F	3.580	47.0	NWN	
8	7	5	GS	F	3.291	48.0	EC	
9	8	6	GS	F	3.595	49.9	EC	
10	9	7	GS	F	3.436	46.8	CW	
11	10	8	GS	F	3.374	48.2	CW	
12	11		GS	F	3.163	47.3	NWN	

12. Fill in the blank cells as before.



	A	B	C	D	E	F	G	H
1	number	samp	specie	se	weight	heigh	locati	
2	1	9	GS	F	3.213	48.5	EC	
3	2	10	GS	F	2.992	46.4	NWN	
4	3	1	GS	F	3.311	46.9	EC	
5	4	2	GS	F	3.186	47.2	EC	
6	5	3	GS	F	2.878	48.8	EC	
7	6	4	GS	F	3.580	47.0	NWN	
8	7	5	GS	F	3.291	48.0	EC	
9	8	6	GS	F	3.595	49.9	EC	
10	9	7	GS	F	3.436	46.8	CW	
11	10	8	GS	F	3.374	48.2	CW	
12	11	=B2	GS	F	3.163	47.3	NWN	

13. Now, all the number 1's should be included in your sample. Filter them out and then copy paste them into a new workbook for import into iNZight.



	number	sample	species	sex	weight	height	location
4	3	1	GS	F	3.311	46.9	EC
14	13	1	GS	F	3.503	50.4	NWN
24	23	1	GS	F	3.249	49.0	CW
34	33	1	GS	F	4.008	49.5	NWN
44	43	1	GS	F	3.446	51.0	CW
54	53	1	GS	F	3.458	47.6	CW
64	63	1	GS	F	3.518	48.4	CW
74	73	1	GS	F	3.007	45.9	EC
84	83	1	NIBr	F	2.810	42.1	N
94	93	1	NIBr	F	2.871	39.9	E
104	103	1	NIBr	F	3.005	39.0	N
114	113	1	NIBr	F	2.874	38.5	W
124	123	1	NIBr	F	3.160	42.4	N
134	133	1	NIBr	F	3.176	41.3	W