Figure 8.12. EQUIPROBABLE SELECTING: Diagrammatic Illustrations

Figures 8.12a, 8.12b and 8.12c on the following pages 8.67, 8.69 and 8.71, together with the tabulated sample descriptions on pages 8.60, 8.62 and 8.64 facing each Figure, illustrate characteristics of equiprobable selecting {[EPS]; *i.e.*, a selecting process that makes equally probable [with probability $1/\binom{\mathbf{N}}{n}$] all samples of size n selected with*out* replacement from a respondent population of size $\mathbf{N}$}.

- When $\mathbf{N} = 500$ and $n = 25$ (as in this Figure 8.12), the number of possible samples is $\binom{500}{25} \simeq 10^{42}$.

The following matters are relevant to using this Figure 8.12.

- the numbers along the tops of the three upper diagrams on pages 8.67, 8.69 and 8.71 are respondent population *element numbers*, which are provided to facilitate identifying the particular population elements selected for each sample;
- because of space constraints on each page, the actual elements in the sample are shown *diagramatically* only for the *first* sample;
- the 16 horizontal lines (or 'bars') given under each population diagram represent the sample average (the central caret or 'hat') plus and minus one sample standard deviation (the 'bar') for each sample obtained from the population by EPS;
- the lowest diagram on each page shows 16 'blocks' representing the 16 sample averages; below the diagram is shown the average plus and minus one standard deviation for the 16 sample averages;
- on the three overleaf sides (pages 8.66, 8.68 and 8.70) which face Figures 8.12a, 8.12b and 8.12c, the 25 element numbers for each sample were taken from a table of equiprobable digits and then arranged in ascending order.

(*continued overleaf*)

[The diagrams which follow illustrate statistically useful properties (under repetition) of sampling carried out by *equiprobable* selecting:

- the sample average varies from sample to sample but its value is (usually) reasonably close to the population average;
- the sample standard deviation varies from sample to sample but its value is (usually) reasonably close to the population standard deviation;
- the distribution shape of the sample responses varies from sample to sample but is (usually) reasonably like that of the population responses;
- sample averages vary less than the population responses;
- the 'centre' of the distribution of sample averages is the population average (over the set of *all* possible samples);
- sample averages tend to approximately a normal distribution regardless of the (shape of the) distribution of population responses.

Figures 8.12e and 8.12f on pages 8.73 and 8.74 show histograms of sample averages and sample standard deviations for all possible samples of all possible sizes selected from a population of size 10 elements with integer response variate values of 1,...., 10.

- Under EPS, these histograms are the *distributions* of these two sample atttributes.

These histograms come from Figure 6.1 of these STAT 220 Course Materials, where there is background and discussion of them.]

Although this Figure 8.12 is concerned primarily with sample *selecting*, it can also be illustrative of (the useful) statistical properties of *repeated measuring* – see also Section 5 on page 6.5 in Figure 6.1 of these Course Materials.

Figure 8.12d. EQUIPROBABLE SELECTING: Diagrammatic Illustrations (page 8.72 continued) Practising Working with the Ideas

$\boxed{7}$ How does the *shape* of the distribution of the respondent population response variate affect the difficulty of estimating:

- its location
- its variation

from *sample* data. Where possible, illustrate your answer with values of appropriate numerical summaries.

$\boxed{8}$ Find the *median* response variate value for each respondent population.

- How does the value of each median compare with that of the corresponding *average*? Explain briefly.
- Which are more *variable*: the three averages or the three medians? Explain briefly.

$\boxed{9}$ Find the response variate *interquartile range* (IQR) for each respondent population.

- How does the value of each IQR compare with that of the corresponding standard deviations? Explain briefly.
- Which are more *variable*: the three IQR's or the three standard deviations? Explain briefly.

$\boxed{10}$ For each respondent *population*, find the proportion of its elements whose response variate value is within one, two and three standard deviations of the average. Tabulate your results in a way which displays them effectively, and comment briefly on the similarities and differences among them.

$\boxed{11}$ Give an example of a real-world population whose response variate value would have a distribution similar to that shown in each of the Diagrammatic Illustrations 1, 2 and 3.

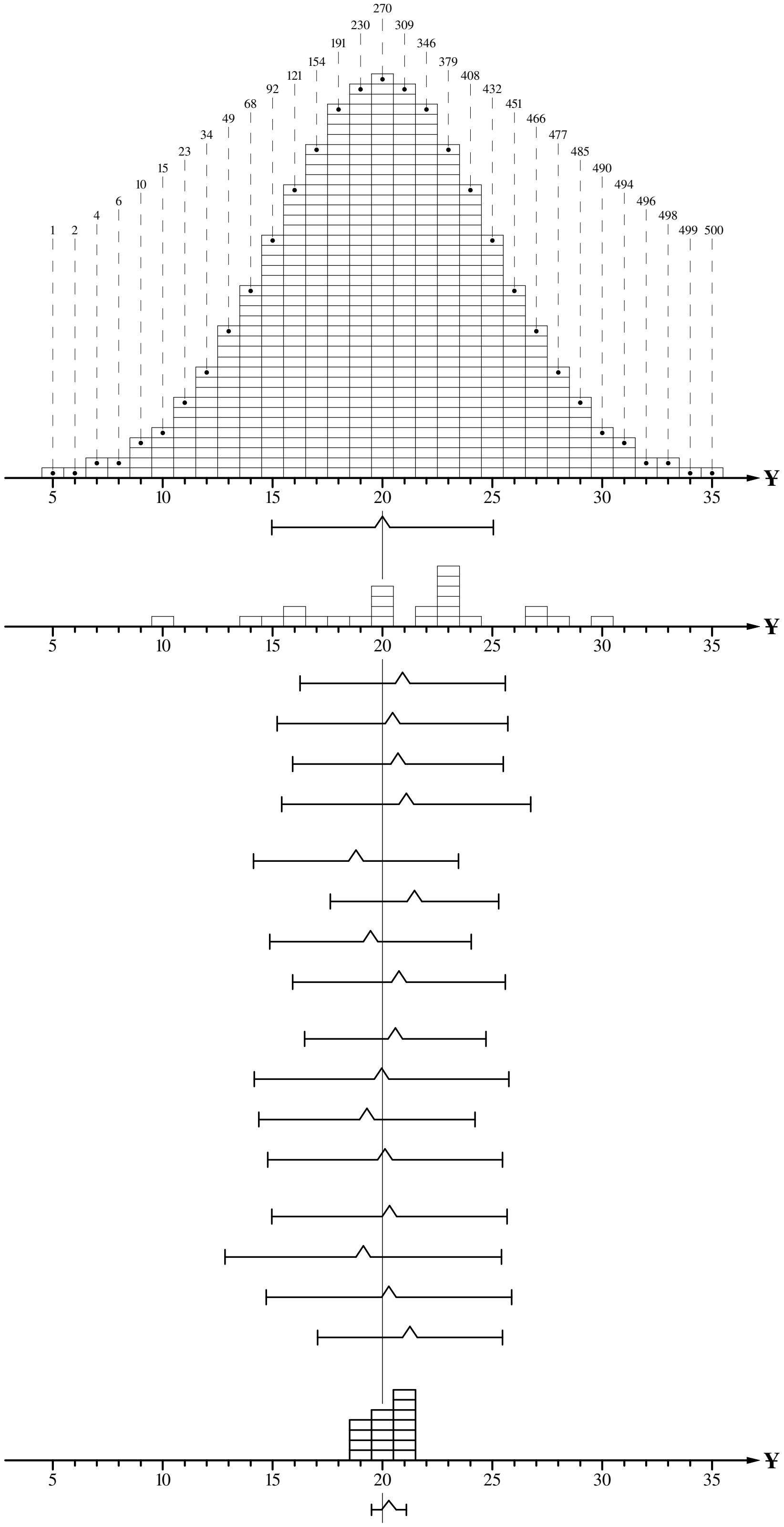
EQUIPROBABLE SELECTING from a ‘Normally’ Distributed Respondent Population
16 Samples of Size 25

																											Av.	S.d.
Sample 1	*Element number*	*13*	*65*	*73*	*99*	*104*	*152*	*168*	*197*	*231*	*232*	*247*	*250*	*314*	*318*	*347*	*358*	*359*	*365*	*369*	*377*	*404*	*452*	*464*	*470*	*487*		
	Response value	10	14	15	16	16	17	18	19	20	20	20	20	22	22	23	23	23	23	23	23	24	27	27	28	30	20.92	4.68
Sample 2	*Element number*	*23*	*24*	*44*	*57*	*64*	*88*	*103*	*129*	*165*	*176*	*291*	*315*	*320*	*328*	*335*	*348*	*369*	*373*	*385*	*440*	*448*	*454*	*460*	*462*	*474*		
	Response value	11	12	13	14	14	15	16	17	18	18	21	22	22	22	22	23	23	23	24	26	26	27	27	27	28	20.44	5.25
Sample 3	*Element number*	*28*	*54*	*68*	*118*	*131*	*137*	*147*	*178*	*223*	*238*	*239*	*249*	*252*	*284*	*303*	*305*	*339*	*350*	*374*	*389*	*403*	*438*	*477*	*486*	*492*		
	Response value	12	14	14	16	17	17	17	18	19	20	20	20	20	21	21	21	22	23	23	24	24	26	28	30	31	20.72	4.78
Sample 4	*Element number*	*2*	*27*	*61*	*76*	*105*	*137*	*211*	*213*	*243*	*268*	*272*	*281*	*310*	*315*	*343*	*357*	*363*	*401*	*406*	*429*	*454*	*459*	*473*	*482*	*491*		
	Response value	6	12	14	15	16	17	19	19	20	20	21	21	22	22	22	23	23	24	24	25	27	27	28	29	31	21.08	5.67
Sample 5	*Element number*	*7*	*15*	*28*	*44*	*48*	*98*	*163*	*175*	*176*	*205*	*213*	*222*	*225*	*230*	*236*	*261*	*272*	*275*	*286*	*288*	*289*	*405*	*443*	*446*	*454*		
	Response value	9	10	12	13	13	16	18	18	18	19	19	19	19	19	20	20	21	21	21	21	21	24	26	27	27	18.80	4.66
Sample 6	*Element number*	*61*	*92*	*119*	*127*	*194*	*214*	*226*	*227*	*264*	*283*	*299*	*306*	*315*	*318*	*323*	*353*	*355*	*365*	*378*	*382*	*386*	*425*	*426*	*484*	*486*		
	Response value	14	15	16	17	19	19	19	19	20	21	21	21	22	22	22	23	23	23	23	24	24	25	25	29	30	21.44	3.83
Sample 7	*Element number*	*14*	*15*	*41*	*48*	*58*	*77*	*174*	*183*	*194*	*212*	*248*	*261*	*274*	*278*	*300*	*304*	*314*	*327*	*333*	*355*	*359*	*383*	*409*	*444*	*446*		
	Response value	10	10	13	13	14	15	18	18	19	19	20	20	21	21	21	21	22	22	22	23	23	24	25	26	26	19.44	4.60
Sample 8	*Element number*	*28*	*45*	*61*	*69*	*72*	*82*	*172*	*202*	*209*	*211*	*250*	*255*	*282*	*307*	*342*	*380*	*386*	*398*	*423*	*424*	*428*	*439*	*463*	*467*	*472*		
	Response value	12	13	14	15	15	15	18	19	19	19	20	20	21	21	22	24	24	24	25	25	25	26	27	28	28	20.76	4.82
Sample 9	*Element number*	*14*	*26*	*130*	*154*	*172*	*176*	*196*	*223*	*250*	*266*	*269*	*280*	*287*	*300*	*303*	*324*	*337*	*347*	*351*	*352*	*371*	*376*	*422*	*473*	*480*		
	Response value	10	12	17	17	18	18	19	19	20	20	20	21	21	21	21	22	22	23	23	23	23	23	25	28	29	20.60	4.13
Sample 10	*Element number*	*9*	*15*	*21*	*66*	*76*	*107*	*127*	*128*	*150*	*155*	*192*	*199*	*236*	*253*	*261*	*271*	*391*	*394*	*397*	*425*	*442*	*462*	*463*	*471*	*493*		
	Response value	9	10	11	14	15	16	17	17	17	18	19	19	20	20	20	21	24	24	24	25	26	27	27	28	31	19.96	5.79
Sample 11	*Element number*	*25*	*38*	*42*	*44*	*55*	*62*	*63*	*125*	*167*	*175*	*178*	*182*	*208*	*215*	*251*	*275*	*289*	*315*	*359*	*367*	*392*	*411*	*434*	*470*	*485*		
	Response value	12	13	13	13	14	14	14	17	18	18	18	18	19	19	20	21	21	22	23	23	24	25	26	28	29	19.28	4.91
Sample 12	*Element number*	*8*	*32*	*43*	*67*	*111*	*138*	*141*	*161*	*173*	*186*	*200*	*247*	*265*	*269*	*276*	*316*	*334*	*363*	*369*	*378*	*416*	*421*	*460*	*468*	*497*		
	Response value	9	12	13	14	16	17	17	18	18	18	19	20	20	20	21	22	22	23	23	23	25	25	27	28	33	20.12	5.34
Sample 13	*Element number*	*3*	*52*	*57*	*88*	*101*	*150*	*154*	*189*	*203*	*204*	*214*	*226*	*246*	*248*	*254*	*267*	*325*	*365*	*405*	*435*	*444*	*451*	*469*	*483*	*489*		
	Response value	7	14	14	15	16	17	17	18	19	19	19	19	20	20	20	20	22	23	24	26	26	26	28	29	30	20.32	5.35
Sample 14	*Element number*	*4*	*5*	*20*	*28*	*36*	*48*	*63*	*75*	*85*	*175*	*202*	*205*	*239*	*249*	*333*	*346*	*381*	*382*	*394*	*404*	*408*	*428*	*468*	*471*	*478*		
	Response value	7	8	11	12	13	13	14	15	15	18	19	19	20	20	22	22	24	24	24	24	24	25	28	28	29	19.12	6.29
Sample 15	*Random number*	*5*	*34*	*40*	*70*	*77*	*90*	*108*	*112*	*139*	*156*	*246*	*249*	*288*	*327*	*337*	*376*	*387*	*408*	*412*	*426*	*438*	*444*	*450*	*463*	*494*		
	Response value	8	12	13	15	15	15	16	16	17	18	20	20	21	22	22	23	24	24	25	25	26	26	26	27	31	20.28	5.60
Sample 16	*Element number*	*17*	*52*	*128*	*150*	*152*	*182*	*236*	*249*	*257*	*259*	*270*	*291*	*295*	*318*	*331*	*347*	*370*	*399*	*403*	*408*	*409*	*418*	*419*	*481*	*484*		
	Response value	11	14	17	17	17	18	20	20	20	20	20	21	21	22	22	23	23	24	24	24	25	25	25	29	29	21.24	4.20

(*continued*)

Figure 8.12a. EQUIPROBABLE SELECTING: Diagrammatic Illustration 1 of 3

Selecting from a ‘Normally’ Distributed Respondent Population



Respondent population attributes:

$\not{N} = 500$

$\not{\bar{Y}} = 20$

$\not{S} = 5.045$

Sample attributes:
(n = 25 in each case)

$\bar{y}_1 = 20.92$;	$s_1 = 4.68$
$\bar{y}_2 = 20.44$;	$s_2 = 5.25$
$\bar{y}_3 = 20.72$;	$s_3 = 4.78$
$\bar{y}_4 = 21.08$;	$s_4 = 5.64$
$\bar{y}_5 = 18.80$;	$s_5 = 4.66$
$\bar{y}_6 = 21.44$;	$s_6 = 3.83$
$\bar{y}_7 = 19.44$;	$s_7 = 4.60$
$\bar{y}_8 = 20.76$;	$s_8 = 4.82$
$\bar{y}_9 = 20.60$;	$s_9 = 4.13$
$\bar{y}_{10} = 19.96$;	$s_{10} = 5.79$
$\bar{y}_{11} = 19.28$;	$s_{11} = 4.91$
$\bar{y}_{12} = 20.12$;	$s_{12} = 5.34$
$\bar{y}_{13} = 20.32$;	$s_{13} = 5.35$
$\bar{y}_{14} = 19.12$;	$s_{14} = 6.29$
$\bar{y}_{15} = 20.28$;	$s_{15} = 5.60$
$\bar{y}_{16} = 21.24$;	$s_{16} = 4.20$

$\bar{\bar{y}} = 20.28$; $s_{\bar{y}} = 0.785$

NOTE: The equiprobable numbers and the corresponding response variate values for the 25 elements in each sample are tabulated on the facing page 8.66.

(*continued overleaf*)

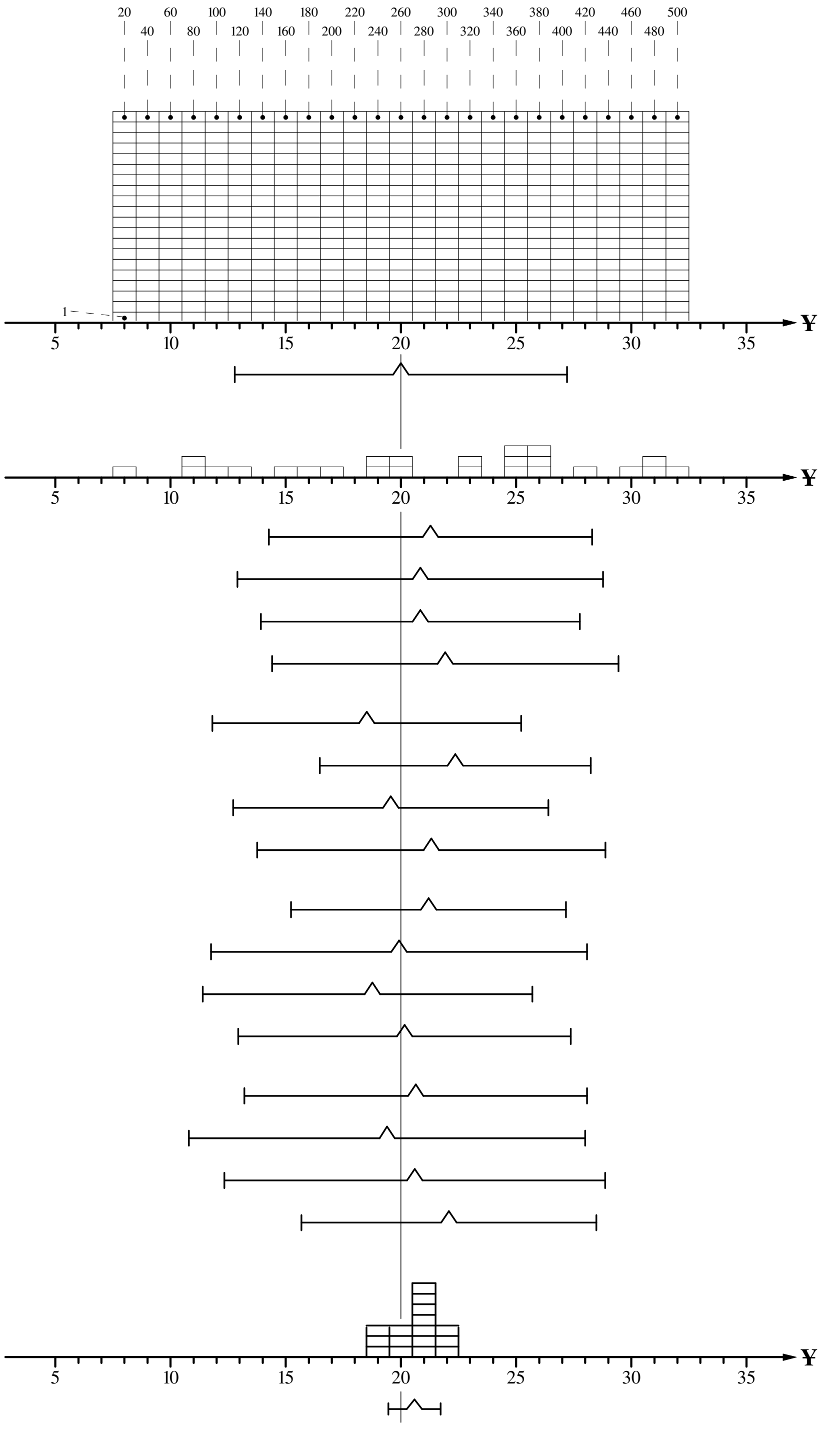
EQUIPROBABLE SELECTING from a ‘Uniformly’ Distributed Respondent Population
16 Samples of Size 25

																											Av.	S.d.
Sample 1	*Element number*	13	65	73	99	104	152	168	197	231	232	247	250	314	318	347	358	359	365	369	377	404	452	464	470	487		
	Response value	8	11	11	12	13	15	16	17	19	19	20	20	23	23	25	25	25	26	26	26	28	30	31	31	32	21.28	7.01
Sample 2	*Element number*	23	24	44	57	64	88	103	129	165	176	291	315	320	328	335	348	369	373	385	440	448	454	460	462	474		
	Response value	9	9	10	10	11	12	13	14	16	16	22	23	23	24	24	25	26	26	27	29	30	30	30	31	31	20.84	7.93
Sample 3	*Element number*	28	54	68	118	131	137	147	178	223	238	239	249	252	284	303	305	339	350	374	389	403	438	477	486	492		
	Response value	9	10	11	13	14	14	15	16	19	19	19	20	20	22	23	23	24	25	26	27	28	29	31	32	32	20.84	6.93
Sample 4	*Element number*	2	27	61	76	105	137	211	213	243	268	272	281	310	315	343	357	363	401	406	429	454	459	473	482	491		
	Response value	8	9	11	11	13	14	18	18	20	21	21	22	23	23	25	25	26	28	28	29	30	30	31	32	32	21.92	7.52
Sample 5	*Element number*	7	15	28	44	48	98	163	175	176	205	213	222	225	230	236	261	272	275	286	288	289	405	443	446	454		
	Response value	8	8	9	9	10	12	16	16	16	18	18	19	19	19	19	21	21	21	22	22	22	28	30	30	30	18.52	6.70
Sample 6	*Element number*	61	92	119	127	194	214	226	227	264	283	299	306	315	318	323	353	355	365	378	382	386	425	426	484	486		
	Response value	11	12	13	14	17	18	19	19	21	22	22	23	23	23	24	25	25	26	26	27	27	29	29	32	32	22.36	5.87
Sample 7	*Element number*	14	15	41	48	58	77	174	183	194	212	248	261	274	278	300	304	314	327	333	355	359	383	409	444	446		
	Response value	8	8	10	10	10	11	16	17	17	18	20	21	21	21	22	23	23	24	24	25	25	27	28	30	30	19.56	6.84
Sample 8	*Element number*	28	45	61	69	72	82	172	202	209	211	250	255	282	307	342	380	386	398	423	424	428	439	463	467	472		
	Response value	9	10	11	11	11	12	16	18	18	18	20	20	22	23	25	26	27	27	29	29	29	29	31	31	31	21.32	7.56
Sample 9	*Element number*	14	26	130	154	172	176	196	223	250	266	269	280	287	300	303	324	337	347	351	352	371	376	422	473	480		
	Response value	8	9	14	15	16	16	17	19	20	21	21	21	22	22	23	24	24	25	25	25	26	26	29	31	31	21.20	5.97
Sample 10	*Element number*	9	15	21	66	76	107	127	128	150	155	192	199	236	253	261	271	391	394	397	425	442	462	463	471	493		
	Response value	8	8	9	11	11	13	14	14	15	15	17	17	19	20	21	21	27	27	27	29	30	31	31	31	32	19.92	8.16
Sample 11	*Element number*	25	38	42	44	55	62	63	125	167	175	178	182	208	215	251	275	289	315	359	367	392	411	434	470	485		
	Response value	10	9	10	10	10	11	11	14	16	16	16	17	18	18	20	21	22	23	25	26	27	28	29	31	32	18.76	7.36
Sample 12	*Element number*	8	32	43	67	111	138	141	161	173	186	200	247	265	269	276	316	334	363	369	378	416	421	460	468	497		
	Response value	8	9	10	11	13	14	15	16	16	17	17	20	21	21	21	23	24	26	26	26	28	29	30	31	32	20.16	7.22
Sample 13	*Element number*	3	52	57	88	101	150	154	189	203	204	214	226	246	248	254	267	325	365	405	435	444	451	469	483	489		
	Response value	8	10	10	12	13	15	15	17	18	18	18	19	20	20	20	21	24	26	28	29	30	30	31	32	32	20.64	7.43
Sample 14	*Element number*	4	5	20	28	36	48	63	75	85	175	202	205	239	249	333	346	381	382	394	404	408	428	468	471	478		
	Response value	8	8	8	9	9	10	11	11	12	16	18	18	19	20	24	25	27	27	27	28	28	29	31	31	31	19.40	8.60
Sample 15	*Elememt number*	5	34	40	70	77	90	108	112	139	156	246	249	288	327	337	376	387	408	412	426	438	444	450	463	494		
	Response value	8	9	9	11	11	12	13	13	14	15	20	20	22	24	24	26	27	28	28	29	29	30	30	31	32	20.60	8.26
Sample 16	*Element number*	17	52	128	150	152	182	236	249	257	259	270	291	295	318	331	347	370	399	403	408	409	418	419	481	484		
	Response value	8	10	14	15	15	17	19	20	20	20	21	22	22	23	24	25	26	27	28	28	28	28	28	32	32	22.08	6.40

(*continued*)

Figure 8.12b. EQUIPROBABLE SELECTING: Diagrammatic Illustration 2 of 3

Selecting from a ‘Uniformly’ Distributed Respondent Population



Respondent population attributes:

$N = 500$

$\overline{Y} = 20$

$S = 7.218$

Sample attributes:
(n = 25 in each case)

$\overline{y}_1 = 21.28;$	$s_1 = 7.01$
$\overline{y}_2 = 20.84;$	$s_2 = 7.93$
$\overline{y}_3 = 20.84;$	$s_3 = 6.93$
$\overline{y}_4 = 21.92;$	$s_4 = 7.52$
$\overline{y}_5 = 18.52;$	$s_5 = 6.70$
$\overline{y}_6 = 22.36;$	$s_6 = 5.87$
$\overline{y}_7 = 19.56;$	$s_7 = 6.84$
$\overline{y}_8 = 21.32;$	$s_8 = 7.56$
$\overline{y}_9 = 21.20;$	$s_9 = 5.97$
$\overline{y}_{10} = 19.92;$	$s_{10} = 8.16$
$\overline{y}_{11} = 18.76;$	$s_{11} = 7.36$
$\overline{y}_{12} = 20.16;$	$s_{12} = 7.22$
$\overline{y}_{13} = 20.64;$	$s_{13} = 7.43$
$\overline{y}_{14} = 19.40;$	$s_{14} = 8.60$
$\overline{y}_{15} = 20.60;$	$s_{15} = 8.26$
$\overline{y}_{16} = 22.08;$	$s_{16} = 6.40$

$\overline{\overline{y}} = 20.59;$ $s_{\overline{y}} = 1.137$

NOTE: The equiprobable numbers and the corresponding response variate values for the 25 elements in each sample are tabulated on the facing page 8.68.

(*continued overleaf*)

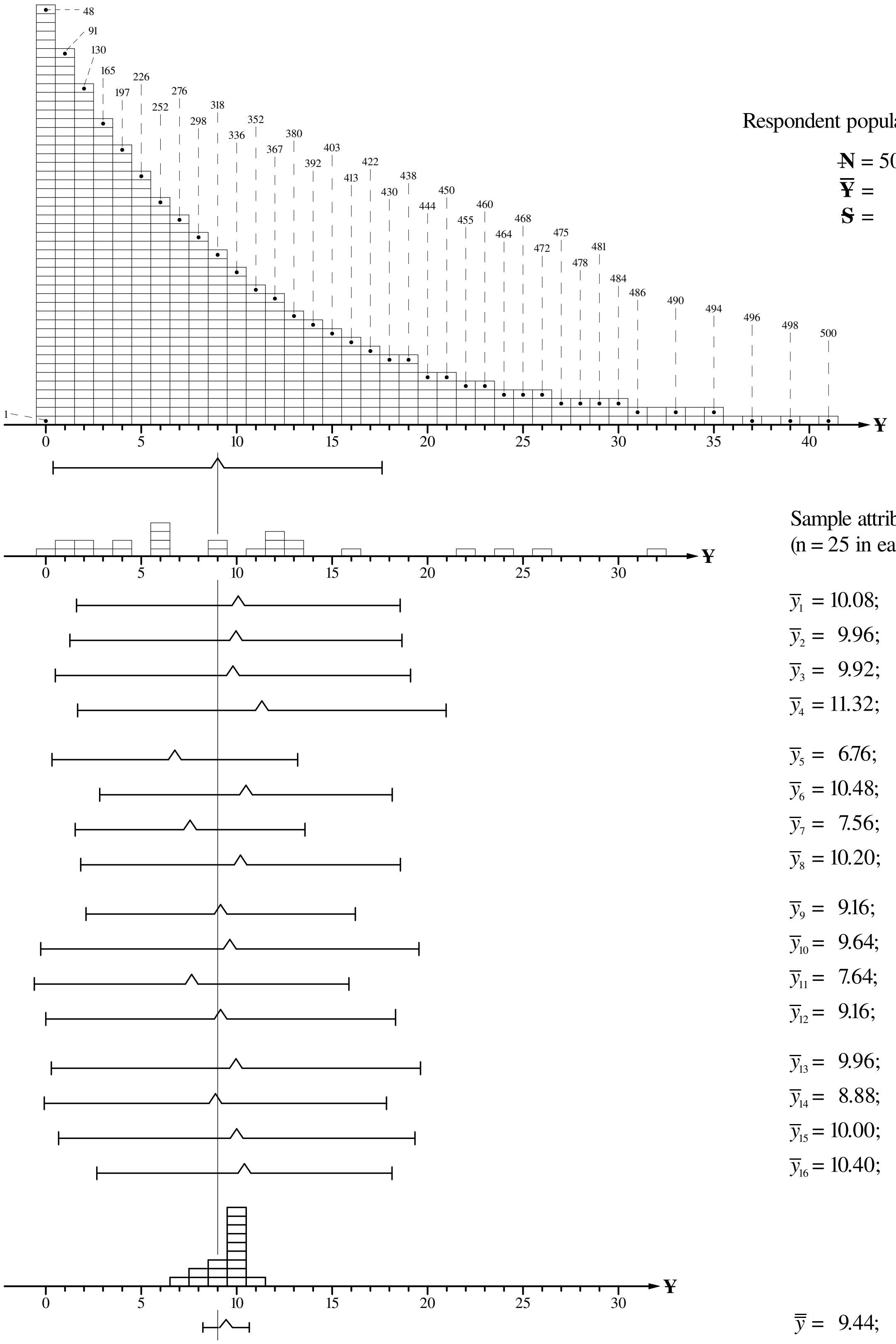
EQUIPROBABLE SELECTING from an ‘Exponentially’ Distributed Respondent Population
16 Samples of Size 25

																											Av.	S.d.
Sample 1	*Element number*	13	65	73	99	104	152	168	197	231	232	247	250	314	318	347	358	359	365	369	377	404	452	464	470	487		
	Response value	1	1	1	2	2	3	4	4	6	6	6	6	9	9	11	12	12	12	13	13	16	22	24	26	32	10.08	8.48
Sample 2	*Element number*	23	24	44	57	64	88	103	129	165	176	291	315	320	328	335	348	369	373	385	440	448	454	460	462	474		
	Response value	0	0	0	1	1	1	2	2	3	4	8	9	10	10	10	11	13	13	14	20	21	22	23	24	27	9.96	8.70
Sample 3	*Element number*	28	54	68	118	131	137	147	178	223	238	239	249	252	284	303	305	339	350	374	389	403	438	477	486	492		
	Response value	0	1	1	2	3	3	3	4	5	6	6	6	6	8	9	9	11	11	13	14	15	19	28	31	34	9.92	9.30
Sample 4	*Element number*	2	27	61	76	105	137	211	213	243	268	272	281	310	315	343	357	363	401	406	429	454	459	473	482	491		
	Response value	0	0	1	1	2	3	5	5	6	7	7	8	9	9	11	12	12	15	16	18	22	23	27	30	34	11.32	9.65
Sample 5	*Element number*	7	15	28	44	48	98	163	175	176	205	213	222	225	230	236	261	272	275	286	288	289	405	443	446	454		
	Response value	0	0	0	0	0	2	3	4	4	5	5	5	5	6	6	7	7	7	8	8	8	16	20	21	22	6.76	6.44
Sample 6	*Element number*	61	92	119	127	194	214	226	227	264	283	299	306	315	318	323	353	355	365	378	382	386	425	426	484	486		
	Response value	1	2	2	2	4	5	5	6	7	8	9	9	9	9	10	12	12	12	13	14	14	18	18	30	31	10.48	7.66
Sample 7	*Elememt number*	14	15	41	48	58	77	174	183	194	212	248	261	274	278	300	304	314	327	333	355	359	383	409	444	446		
	Response value	0	0	0	0	1	1	4	4	4	5	6	7	7	8	9	9	9	10	10	12	12	14	16	20	21	7.56	6.01
Sample 8	*Element number*	28	45	61	69	72	82	172	202	209	211	250	255	282	307	342	380	386	398	423	424	428	439	463	467	472		
	Response value	0	0	1	1	1	1	4	5	5	5	6	7	8	9	11	13	14	15	18	18	18	20	24	25	26	10.20	8.38
Sample 9	*Element number*	14	26	130	154	172	176	196	223	250	266	269	280	287	300	303	324	337	347	351	352	371	376	422	473	480		
	Response value	0	0	2	3	4	4	4	5	6	7	7	8	8	9	9	10	11	11	11	11	13	13	17	27	29	9.16	7.05
Sample 10	*Element number*	9	15	21	66	76	107	127	128	150	155	192	199	236	253	261	271	391	394	397	425	442	462	463	471	493		
	Response value	0	0	0	1	1	2	2	2	3	3	4	5	6	7	7	7	14	15	15	18	20	24	24	26	35	9.64	9.91
Sample 11	*Element number*	25	38	42	44	55	62	63	125	167	175	178	182	208	215	251	275	289	315	359	367	392	411	434	470	485		
	Response value	0	0	0	0	1	1	1	2	4	4	4	4	5	5	6	7	8	9	12	12	14	16	19	26	31	7.64	8.24
Sample 12	*Element number*	8	32	43	67	111	138	141	161	173	186	200	247	265	269	276	316	334	363	369	378	416	421	460	468	497		
	Response value	0	0	0	1	2	3	3	3	4	4	5	6	7	7	7	9	10	12	13	13	17	17	23	25	38	9.16	9.16
Sample 13	*Element number*	3	52	57	88	101	150	154	189	203	204	214	226	246	248	254	267	325	365	405	435	444	451	469	483	489		
	Response value	0	1	1	1	2	3	3	4	5	5	5	5	6	6	7	7	10	12	16	19	20	22	26	30	33	9.96	9.66
Sample 14	*Element number*	4	5	20	28	36	48	63	75	85	175	202	205	239	249	333	346	381	382	394	404	408	428	468	471	478		
	Response value	0	0	0	0	0	0	1	1	1	4	5	5	6	6	10	11	14	14	15	16	16	18	25	26	28	8.88	8.96
Sample 15	*Element number*	5	34	40	70	77	90	108	112	139	156	246	249	288	327	337	376	387	408	412	426	438	444	450	463	494		
	Response value	0	0	0	1	1	1	2	2	3	3	6	6	8	10	11	13	14	16	16	18	19	20	21	24	35	10.00	9.33
Sample 16	*Element number*	17	52	128	150	152	182	236	249	257	259	270	291	295	318	331	347	370	399	403	408	409	418	419	481	484		
	Response value	0	1	2	3	3	4	6	6	7	7	7	8	8	9	10	11	13	15	15	16	16	17	17	29	30	10.40	7.73

(*continued*)

Figure 8.12c. EQUIPROBABLE SELECTING: Diagrammatic Illustration 3 of 3

Selecting from an ‘Exponentially’ Distributed Respondent Population



Respondent population attributes:

~~N~~ = 500
$\overline{Y}$ = 9
~~S~~ = 8.629

Sample attributes:
(n = 25 in each case)

$\overline{y}_1$ = 10.08;	s_1 = 8.48
$\overline{y}_2$ = 9.96;	s_2 = 8.70
$\overline{y}_3$ = 9.92;	s_3 = 9.30
$\overline{y}_4$ = 11.32;	s_4 = 9.65
$\overline{y}_5$ = 6.76;	s_5 = 6.44
$\overline{y}_6$ = 10.48;	s_6 = 7.66
$\overline{y}_7$ = 7.56;	s_7 = 6.01
$\overline{y}_8$ = 10.20;	s_8 = 8.38
$\overline{y}_9$ = 9.16;	s_9 = 7.05
$\overline{y}_{10}$ = 9.64;	s_{10} = 9.91
$\overline{y}_{11}$ = 7.64;	s_{11} = 8.24
$\overline{y}_{12}$ = 9.16;	s_{12} = 9.16
$\overline{y}_{13}$ = 9.96;	s_{13} = 9.66
$\overline{y}_{14}$ = 8.88;	s_{14} = 8.96
$\overline{y}_{15}$ = 10.00;	s_{15} = 9.33
$\overline{y}_{16}$ = 10.40;	s_{16} = 7.73

$\overline{\overline{y}}$ = 9.44; $s_{\overline{y}}$ = 1.213

NOTE: The equiprobable numbers and the corresponding response variate values for the 25 elements in each sample are tabulated on the facing page 8.70.

(*continued overleaf*)

Figure 8.12d. EQUIPROBABLE SELECTING: Diagrammati c Illustrations
Practising Working with the Ideas

The following questions refer to Figures 8.12a, 8.12b and 8.12c and the matters they illustrate.

⊙ The set of all possible samples that can be selected from a population of size $\mathbf{N} = 10$ are illustrated on the following pages 8.73 and 8.74 and they are discussed in more detail in Figure 6.1 of the STAT 220 Course Materials.

$\boxed{1}$ For each shape of distribution of the respondent population response variate, the sample averages *vary*.

- What are the *reasons* for this variation?
  - Are the reasons the *same* in the three cases? Explain briefly.
- What is the 'centre' or 'average' for this variation? How is it the *same* and how is it *different* in the three cases?
- What factor(s) determine the magnitude of the variation?
  - How does this magnitude differ among the three shapes of distribution?
- Why is it legitimate in the three cases to model the sample average as a *random variable*?
  - What is the *source* of the 'randomness'?
  - What factor(s) determine whether the sample average obtained in an *actual* survey can properly be modelled as a random variable?
    - ○ What advantage(s) arise if a sample survey average *can* properly be modelled in this way?

$\boxed{2}$ For each distribution of the respondent population response variate, the sample standard deviations *vary*.

- What are the *reasons* for this variation?
  - Are the reasons the *same* in the three cases? Explain briefly.
- What factor(s) determine the magnitude of this variation?
  - How does this magnitude differ among the three shapes of distribution?
- Why is it legitimate in the three cases to model the sample standard deviation as a *random variable*?
  - What is the *source* of the 'randomness'?
  - What factor(s) determine whether the sample standard deviation obtained in an *actual* survey can properly be modelled as a random variable?
    - ○ What advantage(s) arise if a sample survey standard deviation *can* properly be modelled in this way?

$\boxed{3}$ Rank the three respondent populations in order of *increasing* variation of their response variates.

- Explain in terms of the *shapes* of the distributions why they fall in the order you have given.
- How does the *sample* variation relate to the *population* variation?
- How does the *variation of the sample average* compare with, and relate to, the population variation?
  - How does the variation of the sample *average* compare with, and relate to, the *sample* variation?
    - ○ Explain this behaviour of sample averages.
      - + Discuss briefly why this behaviour of sample averages is of practical importance.

$\boxed{4}$ For the symbolic statement given at the right: $\overline{Y} \dot{\div} N[\overline{\mathbf{Y}}, \mathbf{S}\sqrt{\frac{1}{\mathrm{n}} - \frac{1}{\mathbf{N}}}]$;

- explain its meaning in *words*; separate clearly the different components of the meaning;
- indicate briefly how the three Diagrammatic Illustrations illustrate each component of the meaning.

$\boxed{5}$ Calculate the average $\overline{\mathrm{s}}$ and the standard deviation (s_s) of the 16 sample standard deviations given in each Diagrammatic Illustration; tabulate your results in a way that displays them effectively.

- How close is each value of $\overline{\mathrm{s}}$ to the corresponding value of $\mathbf{S}$? Compare the three cases.
- Which shows greater variation: $\overline{\mathrm{y}}$ or s? Compare the three cases.
  - Instead of comparing directly the *standard deviations* of $\overline{\mathrm{y}}$ and s, calculate their *coefficient of variation* (in %) as defined at the right; *i.e.*, calculate $\mathrm{CV}_{\overline{\mathrm{y}}}$ and CV_s in the three cases and tabulate the six values in an effective way. $\mathrm{CV} = \frac{\text{st.dev.}}{\text{average}} \times 100$
    - ○ Discuss briefly the relative variation of $\overline{\mathrm{y}}$ and s on *this* basis.
      - + Which comparison – of standard deviations or coefficients of variation – is more meaningful? Explain briefly.

$\boxed{6}$ Each respondent population response has attributes of *shape*, *location* and *variation*.

- Arrange the three attributes in order of *increasing* difficulty of estimating from *sample* data; where possible, illustrate your answer with values of appropriate numerical summaries.

Additional questions 7 to 11 are given on the lower half of page 8.65. (*continued*)

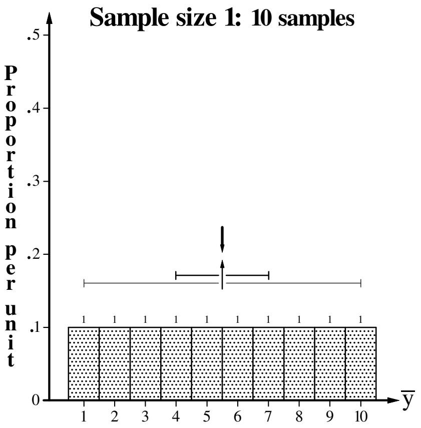
Figure 8.12e. EQUIPROBABLE SELECTING: Histograms of Sample Averages

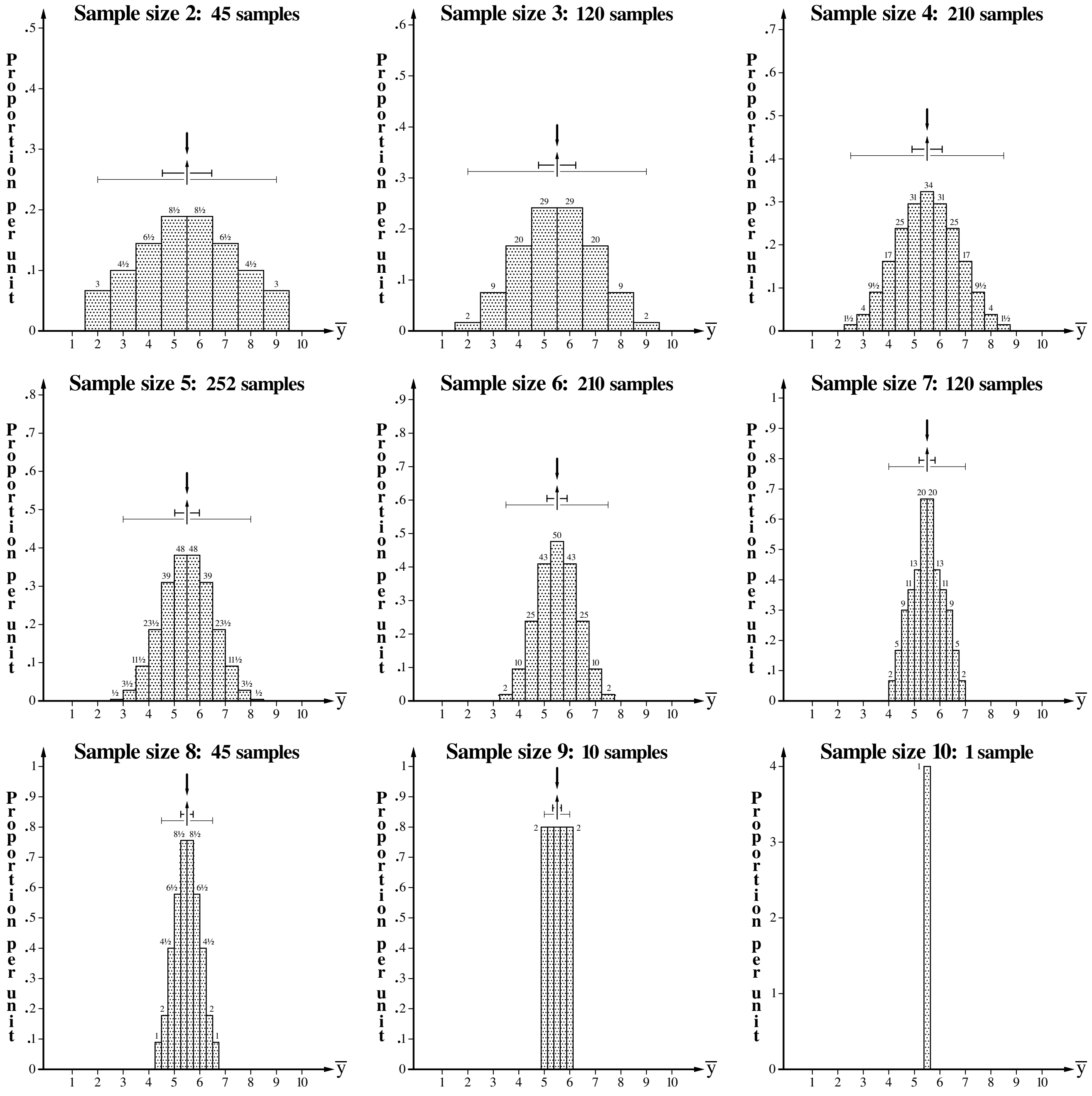
These ten histograms show the distribution of the *averages* of all possible samples of a given size that can be selected without replacement from a population of $\mathbf{N} = 10$ units, whose response variate values are $\mathbf{Y} = 1, 2, ..., 10$. [A histogram of these ten population *responses* would be like that given at the right for the averages of the ten possible samples of size 1.]

The decreasing vertical axis scale unit down the page reduces the visual impact of increasing histogram height; bar *frequencies* are given by the numbers at the tops of the bars.

Above each histogram, except the one at the bottom right of the page:

- the upper bold arrow (↓) indicates the value of the *population* average $\overline{\mathbf{Y}} = 5.5$;
- the lower arrow (↑) indicates the *average* of the set of *sample* averages ($\overline{\overline{y}} = 5.5$);
- the upper bar (⊢—⊣) crossing the lower arrow has a length equal to the value of the *standard deviation* of the set of sample averages;
- the lower such bar shows the *range* of the set of sample averages.





(*continued overleaf*)

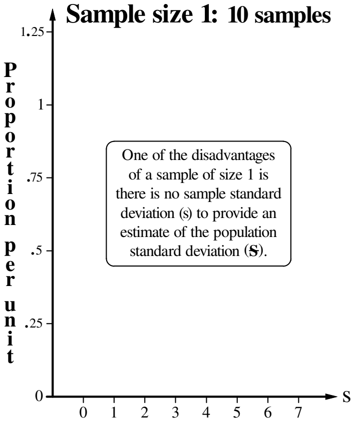
Figure 8.12f. SELECTING: Histograms of Sample Standard Deviations

These nine histograms show the distribution of the *standard deviations* of all possible samples of a given size that can be selected without replacement from a population of N = 10 units, whose response variate values are Y = 1, 2, ..., 10.
[A histogram of these ten population *responses* would be like that given at the top right of the facing page for the averages of the ten possible samples of size 1.]

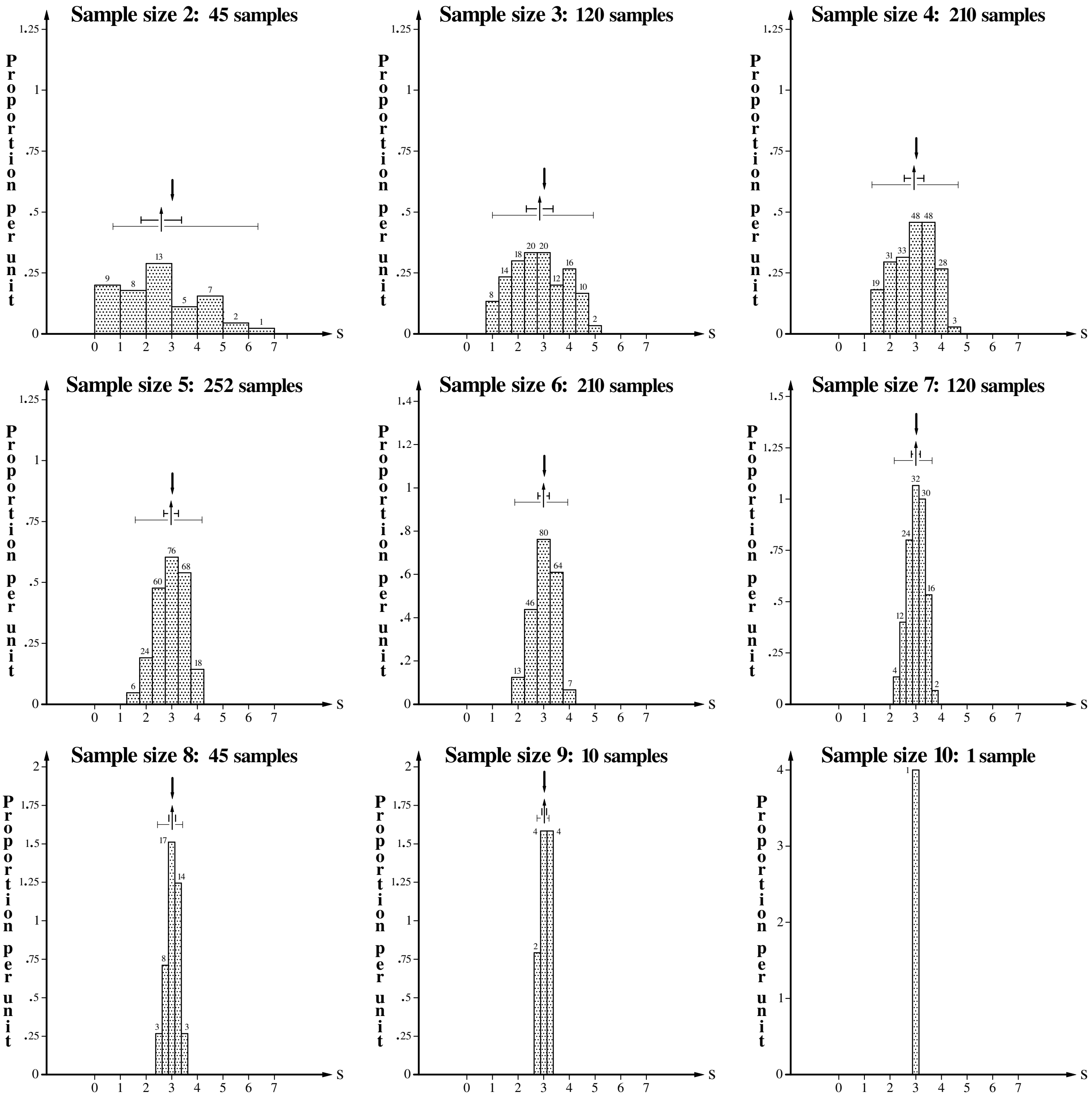
The decreasing vertical axis scale unit down the page reduces the visual impact of increasing histogram height; bar *frequencies* are given by the numbers at the tops of the bars.

Above each histogram, except the one at the bottom right of the page:

- the upper bold arrow (↓) indicates the value of the *population* standard deviation S = 3.027 650;
- the lower arrow (↑) indicates the *average* ($\overline{s}$) of the set of *sample* standard deviations;
- the upper bar (⊢—⊣) crossing the lower arrow has a length equal to the value of the *standard deviation* of the set of sample standard deviations;
- the lower such bar shows the *range* of the set of sample standard deviations.



One of the disadvantages of a sample of size 1 is there is no sample standard deviation (s) to provide an estimate of the population standard deviation (S).


Sample size 2: 45 samples
Sample size 3: 120 samples
Sample size 4: 210 samples
Sample size 5: 252 samples
Sample size 6: 210 samples
Sample size 7: 120 samples
Sample size 8: 45 samples
Sample size 9: 10 samples
Sample size 10: 1 sample
Proportion per unit
S