APPENDIX B

Grouped Frequency Distributions and Central Tendency

Summary

When a data set is large, compiling the unorganized scores into a *grouped frequency distribution*, rather than a *simple frequency distribution*, makes understanding the scores much easier. One rule-of-thumb for deciding if a data set should be presented as a *grouped frequency distribution* is if the range of scores is greater than 20. A *grouped frequency distribution* consists of *class intervals* and *frequencies* and can be presented as a table, a graph, or both. Each *class interval* covers the same number of *X* scores, which is *not* the same as the number of *frequencies*. The number of scores in an interval is symbolized *i.*

There are four conventions for creating *class intervals*:

1. The number of *class intervals* should be between 6 and 12.
2. The size of *i* should be odd, which makes the midpoint of the class interval a whole number. Popular interval sizes include 3, 5, 10, and multiples of 10.
3. The *lower limit* of each *class interval* should be a multiple of *i,* except when *i* = 5 (in that case, reserve a multiple of 5 for the midpoint of each class interval, which reads nicely on a graph).
4. Place the higher scores at the top of the table.

Measures of *central tendency* can also be derived from *grouped frequency distributions.* To find the *mean* of a grouped frequency distribution, multiply each class-interval midpoint by the frequency in the interval. Sum these products, and divide by *N* (which is the sum of the *f* column).

To find the *median*, first use the formula $\frac{N+1}{2}$ to find the location of the middle score. Add frequencies from the bottom of the distribution until you locate the class interval that contains the number produced by the median location formula. The midpoint of the class interval that contains the median location is the *median*.

The *mode* is the midpoint of the *class interval* that occurs with the greatest *frequency*.

Problems

1. A school district administrator gathered all ACT scores for the most recent graduates from her (very small) district. Arrange the data into an appropriate frequency distribution. From the distribution, find the mean, median and mode. Write a sentence of interpretation for the data.

21	14	20	23	20	30	21	31	20	19
16	19	25	20	22	14	34	13	20	15
23	27	21	21	20	16	20	26	20	13

2. A class of college students was asked to list all the states in the USA. The numbers of states they accurately produced are shown below. Arrange them into an appropriate frequency distribution and from the distribution, find the mean, median and mode. Write a sentence of interpretation for the data.

18	34	25	32	18	40	12	38	26	40	23	29	38
23	26	42	21	9	35	25	17	15	25	37	24	12
45	28	13	22	16	28	32	25	41	27	19	24	18
16	46	28	39	11	25	45	20	29	33	21		

3. Arrange the set of scores into a grouped frequency distribution. Find the mean, median, and mode of the distribution.

50	40	55	57	59	42	44	44
39	36	42	31	52	50	40	43
39	35	61	57	49	57	60	67
51	50	40	37	42	52	62	50

4. Arrange the set of scores into a grouped frequency distribution. Find the mean, median, and mode of the distribution.

29	39	49	27	37	47	52	39	17	18
21	60	42	51	12	19	29	42	61	4
8	27	68	38	17	0	17	14	0	3
47	48	29	21	17					

5. Arrange the set of scores into two different grouped frequency distributions: one with $i = 3$ and one with $i = 5$. Find the mean, median, and mode based on each grouped frequency distribution.

31	22	27	41	37	13	33	35	31	36	36	32	15
31	40	35	34	24	36	30	18	25	37	27	30	34
35	38	37	28	34	19	37	32	29	33	21	26	31
38												

ANSWERS

Problems

1.

ACT Scores (class interval)	Midpoint (X)	f
33-35	34	1
30-32	31	2
27-29	28	1
24-26	25	2
21-23	22	7
18-20	19	10
15-17	16	3
12-14	13	4
		$N = 30$

Mean: 20.6
Median: 19
Mode: 19

On average, the most recent graduates in the school district scored 20.6 on the ACT. There is a slight positive skew to the data, with the median (19) being less than the mode. The most frequent class interval was 18-20, producing a mode of 19.

Because the range for these data was 21 (34 – 13), and 21 is divisible by 3 (producing 7 class intervals), i = 3 is a good interval width to use.

ACT Scores (class interval)	Midpoint (X)	f	fX
33-35	34	1	34
30-32	31	2	62
27-29	28	1	28
24-26	25	2	50
21-23	22	7	154
18-20	19	10	190
15-17	16	3	48
12-14	13	4	52
		$N = 30$	$\Sigma fX = 618$

$$\mu = \frac{\sum fX}{N} = \frac{618}{30} = 20.6$$

$$\text{Median location} = \frac{N+1}{2} = \frac{30+1}{2} = 15.5$$

Here, the population mean was calculated, because the scores are from all of the graduates in the district. The score in position 15 and the score in position 16 are both in the 18-20 interval, making 19 the reported median. The class interval with the most scores is also 18-20, making 19 the mode.

2.

Number of States Recalled (class interval)	Midpoint (X)	f
43-47	45	3
38-42	40	7
33-37	35	4
28-32	30	7
23-27	25	12
18-22	20	8
13-17	15	5
8-12	10	4
		$N = 50$

Mean: 26.8
Median: 25
Mode: 25

On average, the students in this class accurately recalled 26.8 states. There is a slight positive skew to the data, with the median (25) being less than the mean. The most frequent class interval was 23-27, producing a mode of 25.

The range for these data was 37 (46 – 9). Using $i = 5$ produces 8 intervals, which is within the recommended range. Recall that, for $i = 5$, it is preferable to use multiples of 5 for the midpoint, instead of the lower limit of the first interval.

Number of States Recalled (class interval)	Midpoint (X)	f	fX
43-47	45	3	135
38-42	40	7	280
33-37	35	4	140
28-32	30	7	210
23-27	25	12	300
18-22	20	8	160
13-17	15	5	75
8-12	10	4	40
		$N = 50$	$\Sigma fX = 1{,}340$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}340}{50} = 26.8$$

$$\text{Median location} = \frac{N+1}{2} = \frac{50+1}{2} = 25.5$$

A second grouped frequency distribution for these data is also shown below. This version uses $i = 3$, which produces 13 intervals (one more than is commonly recommended). It is sometimes useful to test out different sizes of i before deciding which one to report.

Number of States Recalled (class interval)	Midpoint (X)	f	fX
45-47	46	3	138
42-44	43	1	43
39-41	40	4	160
36-38	37	3	111

33-35	34	3	102
30-32	31	2	62
27-29	28	6	168
24-26	25	9	225
21-23	22	5	110
18-20	19	5	95
15-17	16	4	64
12-14	13	3	39
9-11	10	2	20
		$N = 50$	$\Sigma fX = 1{,}337$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}337}{50} = 26.74$$

You'll notice that the interval size slightly altered the mean in these two versions. However, in each version of the grouped frequency distribution, the score in position 25 and the score in position 26 are both in the interval that has 25 as the midpoint, making 25 the reported median for both distributions. The class interval with the most scores is also the interval that has 25 as the midpoint in both distributions, making 25 the mode.

3.

Class Interval	Midpoint (X)	f
63-67	65	1
58-62	60	4
53-57	55	4
48-52	50	8
43-47	45	3
38-42	40	8
33-37	35	3
28-32	30	1
		$N = 32$

Mean: 47.34
Median: 50
Mode: 40 and 50

The range for these data was 36 (67 – 31). Using $i = 5$ produces 8 intervals, which is within the recommended range. Recall that, for $i = 5$, it is preferable to use multiples of 5 for the midpoint, instead of the lower limit of the first interval.

Class Interval	Midpoint (X)	f	fX
63-67	65	1	65
58-62	60	4	240
53-57	55	4	220
48-52	50	8	400
43-47	45	3	135
38-42	40	8	320
33-37	35	3	105
28-32	30	1	30
		$N = 32$	$\Sigma fX = 1{,}515$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}515}{32} = 47.34$$

$$\text{Median location} = \frac{N+1}{2} = \frac{32+1}{2} = 16.5$$

A second grouped frequency distribution for these data is also shown below. This version uses $i = 3$, which produces 13 intervals (one more than is commonly recommended). It is sometimes useful to test out different sizes of i before deciding which one to report.

Class Interval	Midpoint (X)	f	fX
66-68	67	1	67
63-65	64	0	0
60-62	61	3	183
57-59	58	4	232
54-56	55	1	55
51-53	52	3	156
48-50	49	5	245
45-47	46	0	0
42-44	43	6	258
39-41	40	5	200
36-38	37	1	37
33-35	34	1	34
30-32	31	1	31
		$N = 32$	$\Sigma fX = 1{,}498$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}498}{32} = 46.81$$

$i = 3$
Mean: 46.81
Median: 49
Mode: 43

For these data, the two distributions result in slightly different central tendency estimates. When $i = 5$, the mean is 0.53 higher and two modes are identified, instead of one (40 and 50 versus 43). Because $i = 3$ results in 13 intervals (one more than is recommended) and $i = 5$ results in midpoints in multiples of 5 (easy to understand), the $i = 5$ approach would be the better option.

4.

Class Interval	Midpoint (X)	f
60-69	65	3
50-59	55	2
40-49	45	6
30-39	35	4
20-29	25	7
10-19	15	8
0-9	5	4
		$N = 35$

Mean: 29.43
Median: 25
Mode: 15

The range for these data was 68 (68 – 0). Using $i = 5$ would produce too many intervals (15). Using $i = 10$ produces 7 intervals, which is within the recommended range, and allows the midpoint to be a multiple of

5, which is easy to understand.

Class Interval	Midpoint (X)	f	fX
60-69	65	3	195
50-59	55	2	110
40-49	45	6	270
30-39	35	4	140
20-29	25	7	175
10-19	15	8	120
0-9	5	4	20
		$N = 35$	$\Sigma fX = 1{,}030$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}030}{35} = 29.43$$

$$\text{Median location} = \frac{N+1}{2} = \frac{35+1}{2} = 18$$

5.

Class Interval	Midpoint (X)	f
39-41	40	2
36-38	37	9
33-35	34	8
30-32	31	8
27-29	28	4
24-26	25	3
21-23	22	2
18-20	19	2
15-17	16	1
12-14	13	1
		$N = 40$

AND

Class Interval	Midpoint (X)	f
38-42	40	4
33-37	35	15
28-32	30	10
23-27	25	5
18-22	20	4
13-17	15	2
		$N = 40$

$i = 3$
Mean: 30.78
Median: 31
Mode: 37

$i = 5$
Mean: 30.5
Median: 30
Mode: 35

For $i = 3$

Class Interval	Midpoint (X)	f	fX
39-41	40	2	80
36-38	37	9	333
33-35	34	8	272
30-32	31	8	248
27-29	28	4	112
24-26	25	3	75
21-23	22	2	44
18-20	19	2	38
15-17	16	1	16

12-14	13	1	13
		$N = 40$	$\Sigma fX = 1{,}231$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}231}{40} = 30.78$$

$$\text{Median location} = \frac{N+1}{2} = \frac{40+1}{2} = 20.5$$

For $i = 5$

Class Interval	Midpoint (X)	f	fX
38-42	40	4	160
33-37	35	15	525
28-32	30	10	300
23-27	25	5	125
18-22	20	4	80
13-17	15	2	30
		$N = 40$	$\Sigma fX = 1{,}220$

$$\bar{X} \text{ or } \mu = \frac{\sum fX}{N} = \frac{1{,}220}{40} = 30.5$$