3.3: Measures of Central Tendency and Dispersion from Grouped Data

Sometimes we only have access to a frequency distribution of the data, and not to the raw data. (This is often called grouped data.) For grouped data, we can approximate the mean, but cannot calculate the mean exactly. Smaller class widths (intervals) generally result in a better approximation.

Approximating the mean of a frequency distribution:

The Mean: Grouped Data:
A data set of *n* measurements is grouped into *k* classes in a frequency table. If x_i is the midpoint of the *i*th class interval and f_i is the *i*th class frequency, then the *mean* of the grouped data is

$$[\text{mean}] = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{n} \quad \text{where}$$

$\overline{x} = [\text{mean}]$ if data set is a sample
$\mu = [\text{mean}]$ if data set is the population

Example 1: Find the mean for the grouped data.

midpoint	Height (in inches)	Frequency
64	63–65	3
67	66–68	6
70	69–71	7
73	72–74	4
76	75–77	3
		sum: 23

To get the midpoint, average the 2 boundary points:

$$\frac{63+65}{2} = \frac{128}{2} = 64$$

We are assuming the following data pts:

64 64 64
67 67 67 67 67 67
70 70 70 70 70 70 70
73 73 73 73
76 76 76

} 23 data points

To add these, we multiply the frequencies by the midpoints:

$$\bar{x} = \frac{3(64)+6(67)+7(70)+4(73)+3(76)}{23} \approx \boxed{69.74}$$

(not done during class)

Example 2: Find the mean for the grouped data.

$\frac{69+60}{2} = 64.5$

midpoints	Score	Frequency
64.5	60–69	12
74.5	70–79	7
84.5	80–89	14
94.5	90–99	10
	Sum	43

$\bar{x} = \frac{64.5(12) + 74.5(7) + 84.5(14) + 94.5(10)}{43}$

$= \frac{3423.5}{43} \approx \boxed{79.616}$

we are skipping this. (I will put a bonus question on the test)

Approximating the variance and standard deviation of a frequency distribution:

Variance and standard deviation for grouped data:

Suppose a data set of n sample measurements is grouped into k classes in a frequency table, where x_i is the midpoint and f_i is the frequency of the ith class interval.

Then the *sample variance* s^2 for the grouped data (with mean $\bar{x}$) is given by

$$s^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n-1} \text{ or } s^2 = \frac{\sum_{i=1}^{k} f_i x_i - n\bar{x}^2}{n-1} \text{ (both equivalent)}$$

where $n = \sum_{i=1}^{k} f_i$ = total number of measurements .

Then the *sample standard deviation* s for the grouped data (with mean $\bar{x}$) is given by

$$s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n-1}} \text{ or } s = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i - n\bar{x}^2}{n-1}} \text{ (both equivalent)}$$

where $n = \sum_{i=1}^{k} f_i$ = total number of measurements .

Recall:

Standard Deviation $= \sqrt{\text{Variance}}$

Example 1: Suppose that the price-earning ratios of some randomly selected stocks are given in the following table. Find the mean, variance and standard deviation for the sample.

Midpoint	Interval	Frequency
2	-0.5-4.5	7
7	4.5-9.5	53
12	9.5-14.5	22
17	14.5-19.5	14
22	19.5-24.5	0
27	24.5-29.5	4
32	29.5-34.5	3

Sum = 103

$$\frac{-0.5+4.5}{2} = \frac{4}{2} = 2$$

$$\frac{4.5+9.5}{2} = \frac{14}{2} = 7$$

$$\bar{x} = \frac{2(7)+7(53)+12(22)+17(14)+22(0)+27(4)+32(3)}{103}$$

$$= \frac{1091}{103} \approx 10.5922$$ (Stored in calculator to avoid rounding error)

$$\boxed{\text{mean: } \bar{x} = 10.5922}$$

use stored value for $\bar{x}$ (instead of typing 10.59 into calculator)

Variance:

$$s^2 = \frac{7(2-10.59)^2 + 53(7-10.59)^2 + 22(12-10.59)^2 + 14(17-10.59)^2 + 0 + 4(27-10.59)^2 + 3(32-10.59)^2}{103-1}$$

$$\approx \frac{4270.873786}{102} \approx 41.8713163$$

Variance: $s^2 \approx 41.87$

Std. Dev. $s \approx 6.47$

standard deviation:

$$s = \sqrt{s^2} \approx \sqrt{41.8713163} \approx 6.470804558$$

The weighted mean:

When calculating the *mean* of a set of scores, each score carries the same weight. To calculate the mean, we add all the scores, and divide by the number of scores. Therefore, no score carries more weight/importance than any other score.

In a weighted mean, some scores are weighted more heavily than others. We multiply the scores by the corresponding weights, then divide by the total of the weights.

The Weighted Mean:
Suppose a data set has *n* measurements, and that each measurement x_i is associated with a weight w_i. Then the *weighted mean* of the set of measurements is

$$[\text{mean}] = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_n w_n}{w_1 + w_2 + \ldots + w_n} \qquad \text{where}$$

$\bar{x} = [\text{mean}]$ if data set is a sample

$\mu = [\text{mean}]$ if data set is the population

Example 2: Suppose the components of a college class are weighted as follows: Homework (15%), Quizzes (15%), Computer Lab (10%), Midterm Exam (30%), and Final Exam (30%). Suppose a student's grades for homework, quizzes, lab, midterm, and final were 92, 80, 100, 75, and 76, respectively. Calculate the student's course average.

Sum of weights: $\sum w_i = 0.15 + 0.15 + 0.10 + 0.30 + 0.30 = 1$

Course average $= 0.15(92) + 0.15(80) + 0.10(100) + 0.30(75) + 0.30(76)$

$= 81.1$

Divide by $\Rightarrow$ 81.1% course average

Example 3: Using the weights in the above example, and using the same homework, quizzes, lab, and midterm grades (92, 80, 100, 75), what grade must the student receive on the final exam to earn an 80% course average and thus a B in the course? x = final exam grade

$0.15(92) + 0.15(80) + 0.10(100) + 0.30(75) + 0.30x = 80$

$56.8 + 0.30x = 80$

$0.30x = 23.2$

$\frac{0.3x}{0.3} = \frac{23.2}{0.3}$

$x = 77.3$

$\boxed{\text{Needs a } 77.3}$

Example 4: Suppose a student takes ten courses during the academic year, earning the following grades. Calculate the student's grade point average (GPA).

Calculus I (4 credit hours): A
Physics I (5 credit hours): C
History (3 credit hours): B
English Literature (3 credit hours): B
Racquetball (1 credit hour): B
Calculus II (4 credit hours): B
Physics I (5 credit hours): C
Government (3 credit hours): A
Speech (3 credit hours): D
Psychology (3 credit hours): A

A ⇒ 4 grade pts
B ⇒ 3 grade pts
C ⇒ 2 grade pts
D ⇒ 1 grade pts

Total credit hours: $4+5+3+3+1+4+5+3+3+3 = 34$

credit hrs ↓ Grade ↓

$$4(4)+5(2)+3(3)+3(3)+1(3)+4(3)+5(2)+3(4)+3(1)+3(4)=96$$

$$\text{GPA} = \frac{96}{34} = 2.82$$

Example 5: Suppose 30% of a car manufacturer's vehicles get 22 miles per gallon (MPG), 27% get 35 MPG, 8% get 49 MPG, and 35% get 29 MPG. What is the average MPG for all the company's cars?

(We did not do this one in class)

Total weights: $0.30+0.27+0.08+0.35 = 1$ ✓

If the weights are percentages, they should always add up to 1.

$$\text{Average MPG} = 0.30(22)+0.27(35)+0.08(49)+0.35(29)$$

$$= \boxed{30.12 \text{ MPG}}$$

or, $\frac{30.12}{1} = 30.12$