LESSON 8: ESTIMATING MEANS and WEIGHTED MEANS
What if Some Values Count More than Others?

PART A: ESTIMATING THE MEAN FROM A FREQUENCY TABLE

Example 1 (Estimating the Mean: Presidents' Ages)

Here is the **frequency table** for presidents' ages from Lesson 4, Example 1.

Age class (years)	Frequency
35-39	0
40-44	2
45-49	7
50-54	13
55-59	12
60-64	7
65-69	3
70-74	1
	Sum = N = 45

Let's say we lose the individual ages (the raw data). Based on this frequency table, estimate the mean age on becoming president.

§ Solution

We will take the limited, "grainy" information about the data set from the table and "boil down" the classes into their **class marks**.

• The age class "60-64 years" has as its class limits 60 years and 64 years, which are, respectively, the **lowest** and **highest** values permitted in the class.

• We will represent each age class with its class mark, which is the **average** of the class limits. It is the **midpoint** of the class. For the age class "60-64 years," the **class mark** would be:

$$\frac{60+64}{2} = 62 \text{ years}$$

• For simplicity, we will assume that all seven presidents in the age class "60-64 years" **were all 62 years old** when they became president.

• If that assumption seems too strange, we will get the same results if we assume that these seven presidents' ages **averaged 62 years**.

• This scheme will not automatically bias our estimated mean towards either the low end or the high end.

Here is the **frequency table** with the **class marks**:

Age class (years)	**Class marks** (x values)	**Frequency** (f values)
35-39	37	0
40-44	42	2
45-49	47	7
50-54	52	13
55-59	57	12
60-64	62	7
65-69	67	3
70-74	72	1
		$\sum f = N = 45$

We need to estimate the **sum** of the ages by using the **class marks**.

• A long way to estimate the sum would be to **add copies** of the class marks. If a class has frequency f and class mark x, then f copies of x are added together. For example, 2 copies of 42 are added together.

$$42 + 42 +$$

$$47 + 47 + 47 + 47 + 47 + 47 + 47 +$$

$$52 + 52 + 52 + 52 + 52 + 52 + 52 + 52 + 52 + 52 + 52 + 52 + 52 +$$

$$57 + 57 + 57 + 57 + 57 + 57 + 57 + 57 + 57 + 57 + 57 + 57 +$$

$$62 + 62 + 62 + 62 + 62 + 62 + 62 +$$

$$67 + 67 + 67 +$$

$$72$$

• The **estimated sum** would be 2480 years.

• Here's a much faster way to get the same **estimated sum**. For each class, **multiply** each class mark (x) by the class frequency $\left(f\right)$ to get the "**class sum**" $\left(f \cdot x\right)$, the sum of the ages in the class. Then, **add** these "class sums" to get the overall **estimated sum** $\sum f \cdot x$

$$
\begin{aligned}
&0(37) &&+\\
&2(42) &&+\\
&7(47) &&+\\
&13(52) &&+\\
&12(57) &&+\\
&7(62) &&+\\
&3(67) &&+\\
&1(72)
\end{aligned}
$$

• Imagine doing a **zigzag** between the x and f values through the table.

• Again, the **estimated sum** would be 2480 years.

We now **divide** the **estimated sum** by the **number of ages** to obtain our **estimated mean** of the ages.

$$
\begin{aligned}
\text{Estimated Mean} &= \frac{\text{Estimated Sum}}{N}, \quad \text{or} \quad \frac{\sum f \cdot x}{\sum f}\\
&= \frac{2480}{45} \quad \left(\text{We've processed the numerator before dividing.}\right)\\
&\approx 55.1 \text{ years}
\end{aligned}
$$

Our **estimated mean** for the presidents' ages is about 55.1 years.

• In fact, if we recover the individual ages, the mean is about 55.0 years.

(See Footnote 1 on rounding issues.)

§

PART B: WEIGHTED MEANS

In some data sets, some values are more important than others. We use weights to measure the relative importance of data values.

Think About It: If you get an "A" on a midterm worth 10% of a class and an "F" on a final worth 90% of the class, does that mean you will pass the class with a "C"?

Example 2 (Calculating a Weighted Mean: GPA)

If you receive the following grade report for a term, what is your grade point average (GPA) for the term?

Course	**Number of Units** (Weights, w)	**Grade**
Math	5	A-
Chemistry	4	D
Biology	3	C+

§ Solution

GPA Scheme

The grades are **ordinal** data values. They are nonnumeric, but there is a natural order among the possible grades: A+, A, A-, B+, etc.
We will **recode** the grades as **quantitative** numeric data as follows:

$$A = 4$$
$$B = 3$$
$$C = 2$$
$$D = 1$$
$$F = 0$$

The "+" modifier raises the value by 0.3.
The "-" modifier lowers the value by 0.3.

• The "A-" grade in Math, for example, corresponds to: $4 - 0.3 = 3.7$

Solution to Example 2

Course	**Number of Units** (Weights, w)	**Grade**	**Grade Recoded as Grade Points** (x values)
Math	5	A-	3.7
Chemistry	4	D	1.0
Biology	3	C+	2.3
	Sum, $\sum w = \mathbf{12}$		

- Because the Math class is worth **five** units, imagine writing 3.7 on **five** slips of paper.

- Similarly, imagine writing 1.0 (corresponding to the "D") on **four** slips.

- Imagine writing 2.3 (corresponding to the "C+") on **three** slips.

Our weighted mean (or weighted average) will then simply be the arithmetic mean of the numbers on all 12 slips.

The **"grand sum"** here is obtained by doing a **zigzag** between the w and x values through the table. This is similar to the **zigzag** we did for the **estimated sum** in Example 1.

Row Sums: $w \cdot x$, where:

$w =$ number of units, and

$x =$ grade value

Slips	Row Sums
3.7 3.7 3.7 3.7 3.7	$(5)(3.7)$
1.0 1.0 1.0 1.0	$(4)(1.0)$
2.3 2.3 2.3	$(3)(2.3)$

$$\begin{aligned}\text{Grand sum} &= \sum w \cdot x \\ &= (5)(3.7) + (4)(1.0) + (3)(2.3) \\ &= 29.4 \text{ grade points}\end{aligned}$$

$$
\begin{aligned}
\text{GPA} &= \frac{\text{Grand sum}}{\text{Sum of all weights (that is, the total number of units)}} \\
&= \frac{\sum w \cdot x}{\sum w} \\
&= \frac{(5)(3.7)+(4)(1.0)+(3)(2.3)}{5+4+3} \\
&= \frac{29.4}{12} \\
&= 2.45 \text{ grade points}
\end{aligned}
$$

Your GPA for the term would be 2.45 grade points.

• If a student stays at a 2.45 GPA throughout college, would this be good enough for medical school in the U.S.?

§

Example 3 (Course Grade)

(What is Needed to Obtain [At Least] a Particular Weighted Mean: "What Do I Need on the Final to …")

You receive the following scores in a class:

Exam	**% of Course Grade**	**Score (out of 100 points)**
Quiz 1	10%	75
Quiz 2	10%	80
Quiz 3	10%	100
Midterm	25%	95
Final	45%	a

You have not yet taken the Final, so we have labeled the Final score with a hopeful "a."

What must you get on the Final to get at least 90% in the course overall (that is, at least 90 points as a weighted mean, or weighted average)?

§ Solution

Again, we are asking about a weighted mean. Observe that the Final is worth much more than any other exam.

We will write the exam weights as decimals: 0.10 for 10%, etc.

$$\begin{aligned}\text{Course average} &= \frac{\sum w \cdot x}{\sum w} \\ &= \frac{(0.10)(75)+(0.10)(80)+(0.10)(100)+(0.25)(95)+(0.45)(a)}{1} \\ &= 49.25+0.45a\end{aligned}$$

WARNING 1: Do **not** combine the unlike terms 49.25 and 0.45*a* as 49.70a; that is wrong!

We have expressed your course average as a function of *a*, the score on the Final.

• If you **don't take the Final** … or if you are caught cheating and you get a "0" on it, then your course average will be 49.25 points (or 49.25% of the maximum possible performance).

Note: If we had used 10, 10, 10, 25, and 45 as our weights, then $\sum w = 100$, and we would have obtained the same function:

$$\begin{aligned}\text{Course average} &= \frac{\sum w \cdot x}{\sum w} \\ &= \frac{(10)(75)+(10)(80)+(10)(100)+(25)(95)+(45)(a)}{100} \\ &= \frac{4925+45a}{100} \\ &= 49.25+0.45a\end{aligned}$$

Let's rephrase the question. What scores (values for *a*) on the Final will yield a weighted score average of 90 points for the course? (Use 90, not 0.90.) We solve the inequality:

$$49.25 + 0.45a \geq 90$$

Multiply both sides by 100.

$$4925 + 45a \geq 9000$$

$$45a \geq 4075$$

$$a \geq 90.6 \text{ points}$$

You need to score at least 90.6 points on the Final to get at least 90% (of the maximum possible performance) in the course overall.

- If the grader only grades in integers, then, as a practical matter, we could say that you need at least 91 points.

§

FOOTNOTES (OPTIONAL)

#1) Class marks and rounding issues. One could argue that the class marks in Example 1 should really be 37.5 years, 42.5 years, 47.5 years, etc. For example, the class mark for the age class "60-64" years could be said to be 62.5 years, since that age class would include presidents who just turned 60 years old and also those who are just about to turn 65 years old. When we report ages, we usually round them down; for example, a 64.9-year-old president would be called 64 years old. The new class marks are all 0.5 years higher than the old class marks used in Example 1. If the new class marks are used, then our estimate for the mean would rise 0.5 years as well.