3.2: Math Refresher

Statistics is not math. It does, however, use math as a tool. This section will review some of the tools that you'll use in statistical equations, with an ending section on calculating percentages. You learned much of this in junior high or early high school, but that was a long time ago for most of us!

Order of Operations

Example 3.2.1

What does “Please Excuse My Dear Aunt Sally” or “PEMDAS” mean?

Solution

It's the order in which you should do each mathematical operation.

“Please Excuse My Dear Aunt Sally” or PEMDAS tells you what kind of calculations must be completed before other calculations. However, this can still be overwhelming. Although you must follow the PEMDAS order, you can start at the top left, and go right and down, just like how you read. In fact, PEMDAS is sorta like a sentence structure but with equations.

Note

- P- You must complete the calculations within parentheses before you can use those numbers in the rest of the equation.
- E- You must complete the exponents (which will always just be squaring a number in statistics) before you can use that number in the rest of the equation.
- M- You must complete the multiplication before you can add or subtract it with other numbers in the equation. I actually think that you're supposed to divide before you multiply?
- D- You must complete the division before you can add or subtract it with other numbers in the equation. I actually think that you're supposed to divide before you multiply?
- A- Once all of the other mathematical calculations are completed, you can now add or subtract. I actually think that you're supposed to subtract before you add?
- S- Once all of the other mathematical calculations are completed, you can now add or subtract. I actually think that you're supposed to subtract before you add?

What might throw you off in actual equations that we'll use is square rooting. Square rooting is actually the last operation because you have to get down to one number.

Notation & Symbols

Table 3.2.1 will cover some of the major mathematical symbols used in statistical equations. The column on the right is where you can write explanations or examples in your own words.

Table 3.2.1- Common Mathematical Symbols

Symbol	Read As	Example or Description
>	"greater than"	
<	"less than"	
* or XY	"multiply" or "times"	
$()^2$	"square" or "times itself"	
√	"square root" or "find the number that, multiplied by itself, equals the number in the square root sign"	
|#|	“absolute value” or "make it a positive number" (the lines, not the #-sign)	

Summing (Sigma or Σ)

Many statistical formulas involve summing numbers. Fortunately there is a convenient notation for expressing summation. This section covers the basics of this summation notation. Let's say we have a variable X that represents the weights (in grams) of four grapes show in Table 3.2.2:

Table 3.2.2- Grape Weights in Grams

Grape	X
A	4.6
B	5.1
C	4.9
D	4.4

The following formula means to sum up the weights of the four grapes:

$$\sum X$$

The Greek letter Σ indicates summation. The X is a placeholder for all of the scores. Therefore,

$$\sum X = X_A + X_B + X_C + X_D = 4.6 + 5.1 + 4.9 + 4.4 = 19$$

Sums of Squares

Many formulas involve squaring numbers before they are summed. This is indicated as

$$\begin{aligned}\left(\sum X\right)^2 &= 4.6^2 + 5.1^2 + 4.9^2 + 4.4^2 \\ &= 21.16 + 26.01 + 24.01 + 19.36 = 90.54\end{aligned}$$

Notice that:

$$\left(\sum \mathrm{X}\right)^2 \neq \sum \left(\mathrm{X}^2\right)$$

because the expression on the left means to sum up all the values of X and then square the sum ($19^2 = 361$), whereas the expression on the right means to square the numbers and then sum the squares (90.54, as shown).

Multiplication (Products)

Some formulas involve the sum of cross products. Below are the data for variables X and Y. The cross products (XY, or $X \times Y$) are shown in the third column. The sum of the cross products is 3 + 4 + 21 = 28.

Table 3.2.2- Table of Scores and Products

Participant	X	Y	XY
A	1	3	3
B	2	2	4
C	3	7	21
$\sum$	$\sum = 1+2+3 = 6$	$\sum = 3+2+7 = 12$	$\sum = 3+4+21 = 28$

Percentages

Providing the percentages is often more useful than merely displaying frequency information, especially when comparing distributions that have different sample sizes (different total amount of units/numbers).

Example 3.2.2

1. Having 1 red-head in a group of 10= _____%
2. Having 1 red-head in a group of 100= _____%

Solution

1. Having 1 red-head in a group of $10 = 10\%$
2. Having 1 red-head in a group of $100 = 1\%$

In Example $3.2.1$, showing the frequency of one doesn't tell us as much as the percentage does because although 10% isn't a large proportion, it's way bigger than 1%. The only time when providing frequencies (counts) is better than percentages is when you have a very small population. For example, having 1 out of 3 people like your cologne is 33%, but it that doesn't really tell you anything meaningful about how you smell.

Once you get to a decimal point, you then multiply by 100 to get the percentage.

- $0.17 = 17\%$
- $0.94 = 94\%$

You probably all do this in your head, but try to write it out for now so that you don't skip a step.

Example 3.2.3

Imagine that there are 35 students in your class and 5 of them had earned full credit on all assignments. What's the percentage of students in your class earning an A+?

Solution

1. Division: $\frac{5}{35} = 0.1429$ (We would normally round decimal points to 2 numbers after the decimal, but for percentages we should include 4 numbers because we'll need to multiply them.
2. Multiplication: $0.1429 \times 100 = 14.29\%$
3. Conclusion: The percentage of students in your class earning an A+ was 14.29%

As we move on to the actual calculations used to describe distributions of data, you can check out this Crash Course video on data visualization which has a bit on percentages.
