Learning Objectives

- Understand and generate frequency tables of raw data
- Describe rationale for frequency tables, grouped distributions
- Balance between needed *simplicity* in data presentations vs. *accuracy* in data presentations
- Compare/Contrast different methods of visualizing data (including their implied meanings)

Raw Score Distribution (pp 48)

table 3.1 Scores from statistics exam ($N = 70$)

95	57	76	93	86	80	89
76	76	63	74	94	96	77
65	79	60	56	72	82	70
67	79	71	77	52	76	68
72	88	84	70	83	93	76
82	96	87	69	89	77	81
87	65	77	72	56	78	78
58	54	82	82	66	73	79
86	81	63	46	62	99	93
82	92	75	76	90	74	67

Frequency Distribution

table 3.2 Scores from Table 3.1 organized into a frequency distribution

Score	*f*	Score	*f*	Score	*f*	Score	*f*
99	1	85	0	71	1	57	1
98	0	84	1	70	2	56	2
97	0	83	1	69	1	55	0
96	2	82	5	68	1	54	1
95	1	81	2	67	2	53	0
94	1	80	1	66	1	52	1
93	3	79	3	65	2	51	0
92	1	78	2	64	0	50	0
91	0	77	4	63	2	49	0
90	1	76	6	62	1	48	0
89	2	75	1	61	0	47	0
88	1	74	2	60	1	46	1
87	2	73	1	59	0		
86	2	72	3	58	1		

Grouped Frequency Distributions

table 3.3 Scores from Table 3.1 grouped into class intervals of different widths

Class Interval (width = 2)	f	Class Interval (width = 19)	f
98–99	1	95–113	4
96–97	2	76–94	38
94–95	2	57–75	23
92–93	4	38–56	5
90–91	1		$N = 70$
88–89	3		
86–87	4		
84–85	1		
82–83	6		
80–81	3		
78–79	5		
76–77	10		
74–75	3		
72–73	4		

Grouped Frequency Distributions

- Communication tool that help simplify our data, but…
  - Too few ”groups” can oversimplify
  - Too many groups are difficult to process

- How do we choose the right level of simplicity?
  - **Ultimate goal**: Let humans understand…
    - Choose intrinsically meaningful intervals when possible
    - 5-10
    - Square root of *N*

table 3.7 Relative frequency, cumulative frequency, and cumulative percentage distributions for the grouped scores in Table 3.4

Class Interval	*f*	Relative *f*	Cumulative *f*	Cumulative %
95–99	4	0.06	70	100
90–94	6	0.09	66	94.29
85–89	7	0.10	60	85.71
80–84	10	0.14	53	75.71
75–79	16	0.23	43	61.43
70–74	9	0.13	27	38.57
65–69	7	0.10	18	25.71
60–64	4	0.06	11	15.71
55–59	4	0.06	7	10.00
50–54	2	0.03	3	4.29
45–49	1	0.01	1	1.43
	70	1.00		

Translations

- *"f"* = frequency

- "Relative" just means 'frequency divided by *N*'
  - Proportion of all observations in one group

- "Cumulative" means frequency of observations at or below this groups *upper real limit*

Class Interval	f
95–99	4
90–94	6
85–89	7
80–84	10
75–79	16
70–74	9
65–69	7
60–64	4
55–59	4
50–54	2
45–49	1
	70

1. What was the most frequent observation?

2. What is the ***relative frequency*** of scores between 70 and 79?

3. What is the ***cumulative frequency*** for the group 65-69?

Figures or Tables?

- Either can work, but try both & see what human readers intuit most easily!!
- Types of Figures:
  - Bar graph
  - Histogram
  - Frequency polygon
  - Smoothed frequency distribution

Bar Graph (pp 63)

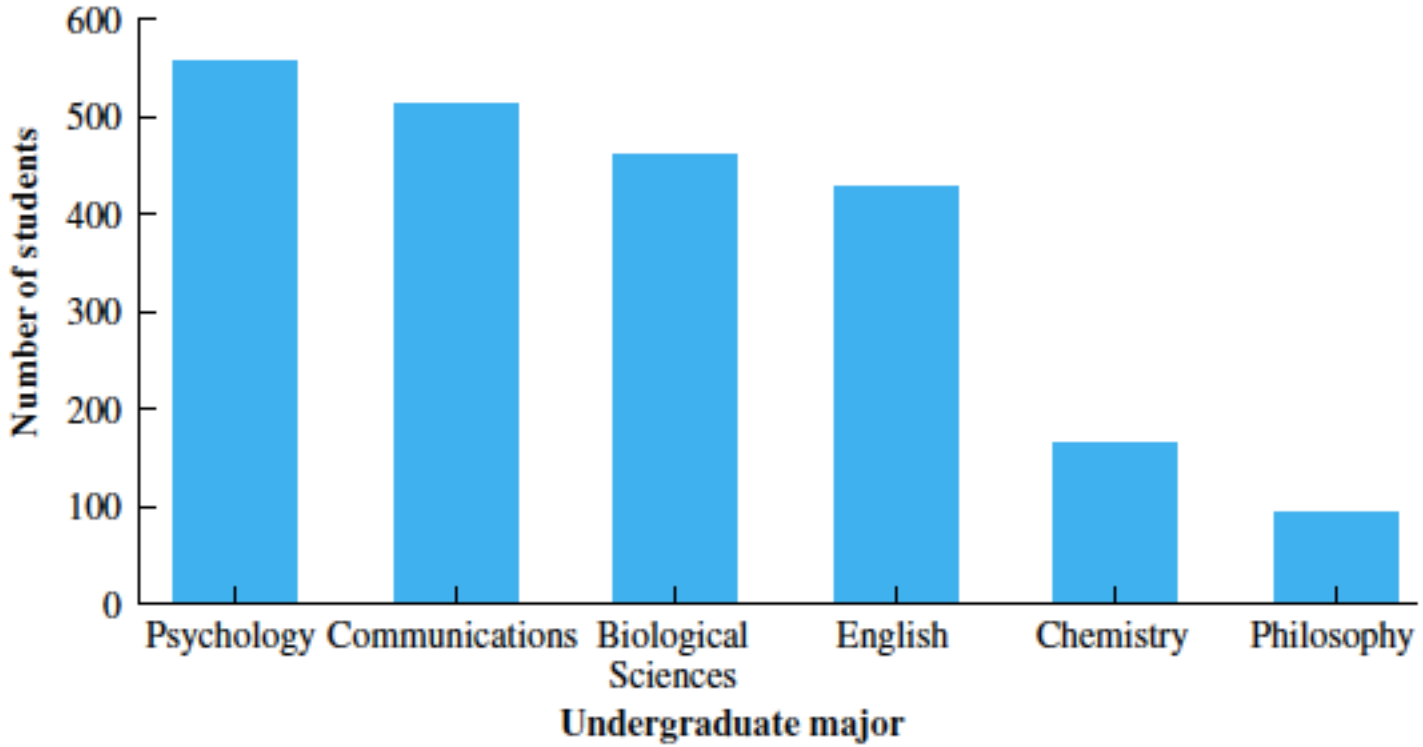


figure 3.3 Bar graph: Students enrolled in various undergraduate majors in a college of arts and sciences.

- **Note:** What is implied by disconnected bars?

Histogram (pp 64)

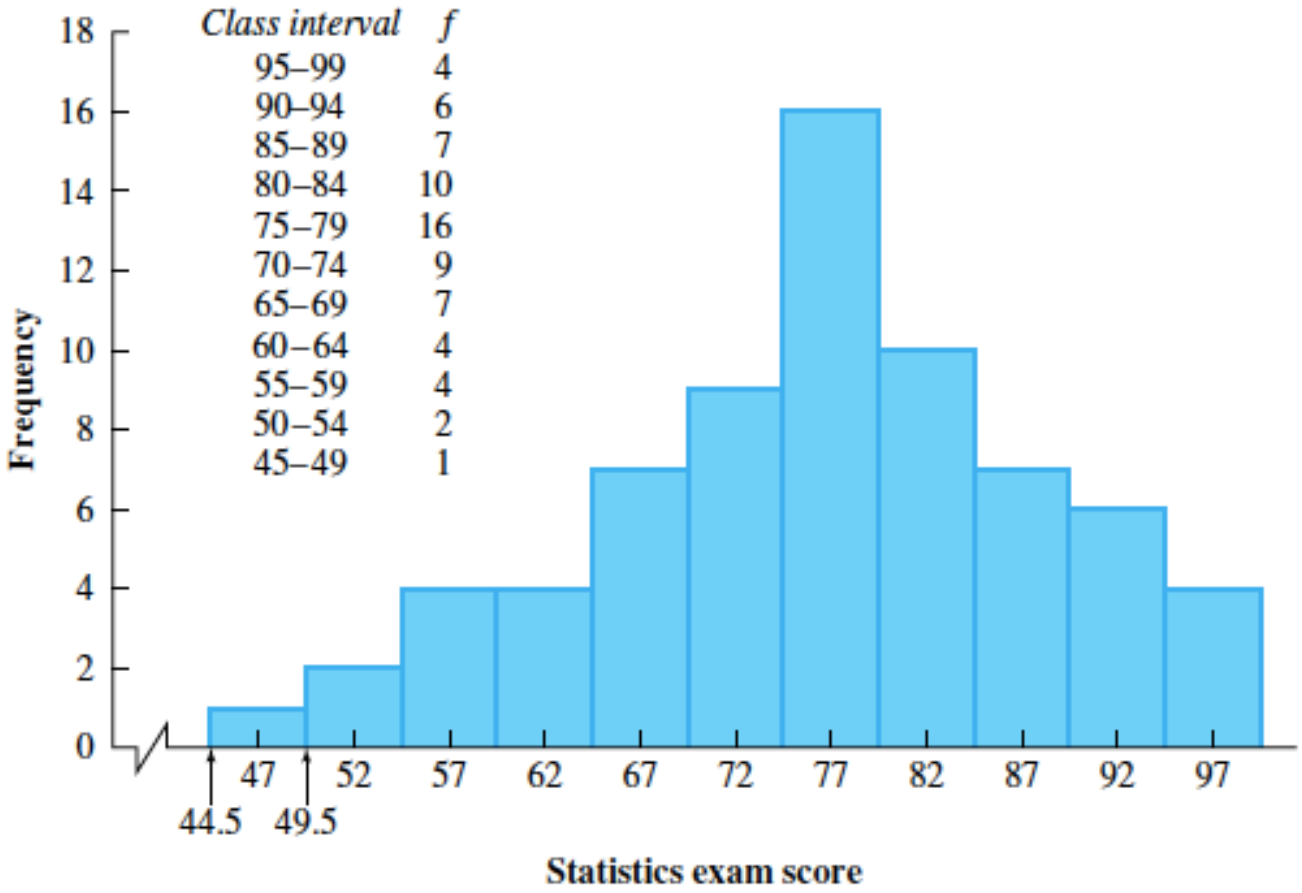


figure 3.4 Histogram: Statistics exam scores of Table 3.4.

- **Note:** 'Squiggle' on the left side indicates that the *x*-axis has been stretched

Frequency Polygon (pp 65)

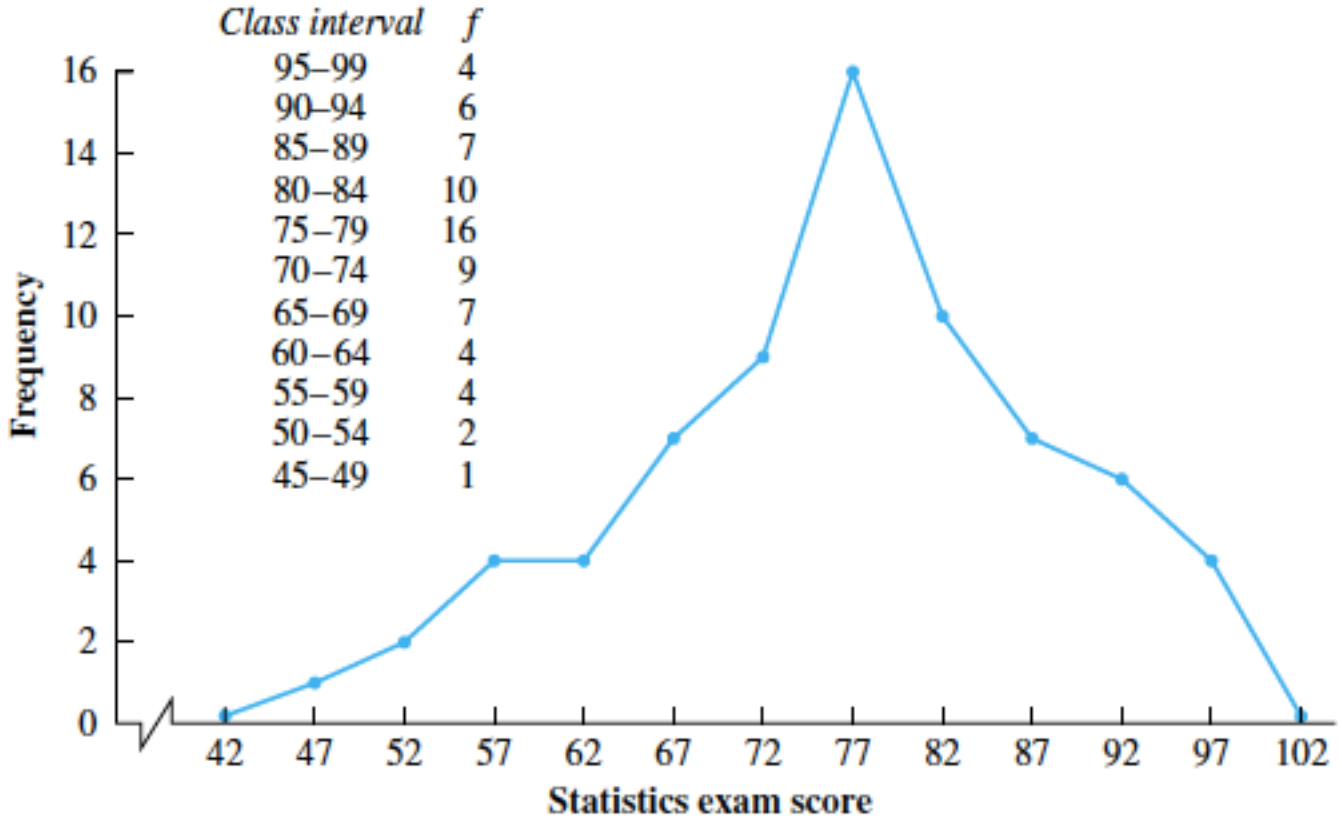
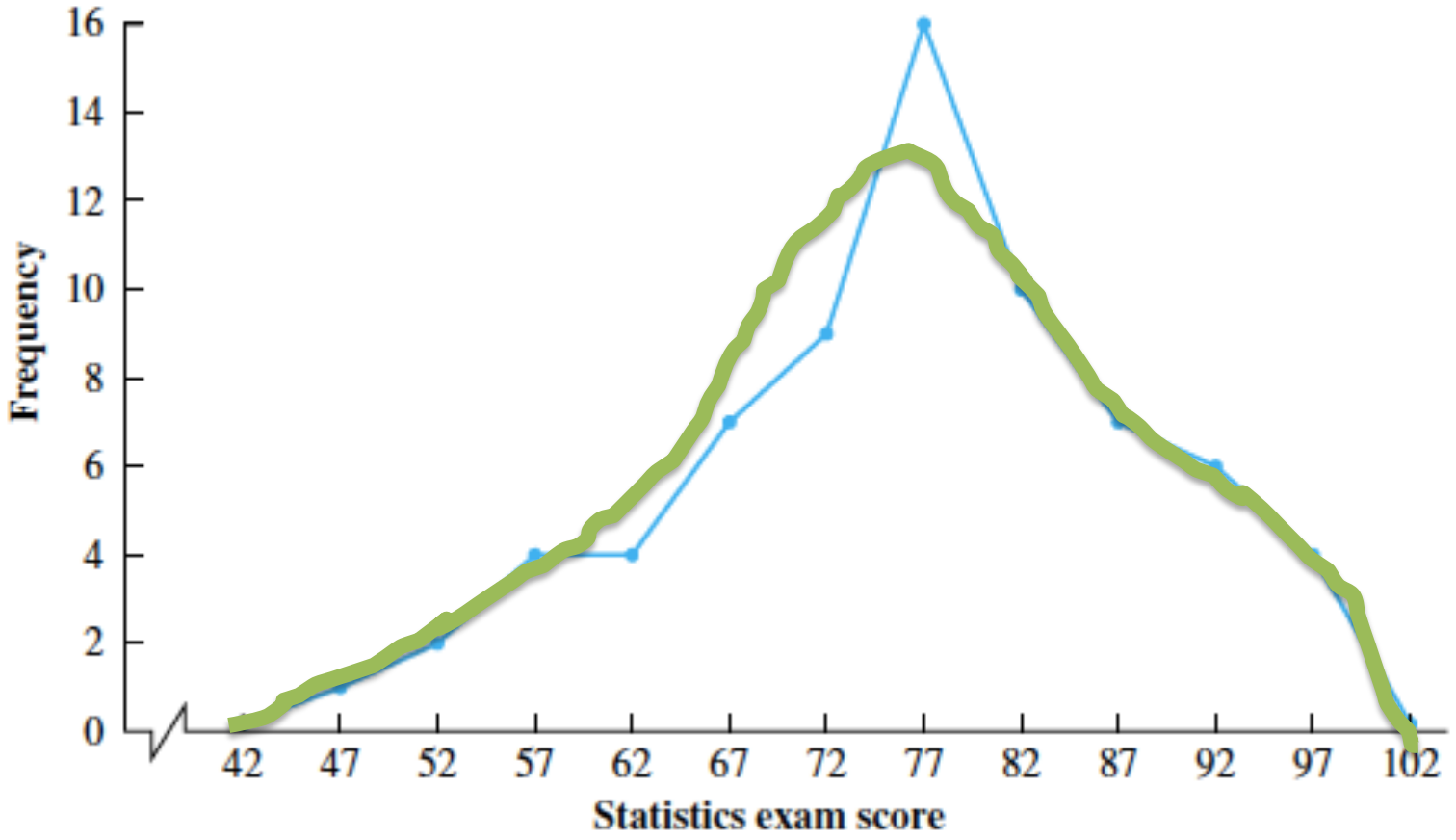


figure 3.5 Frequency polygon: Statistics exam scores of Table 3.4.

- **Note:** What is tacitly implied by this plot?

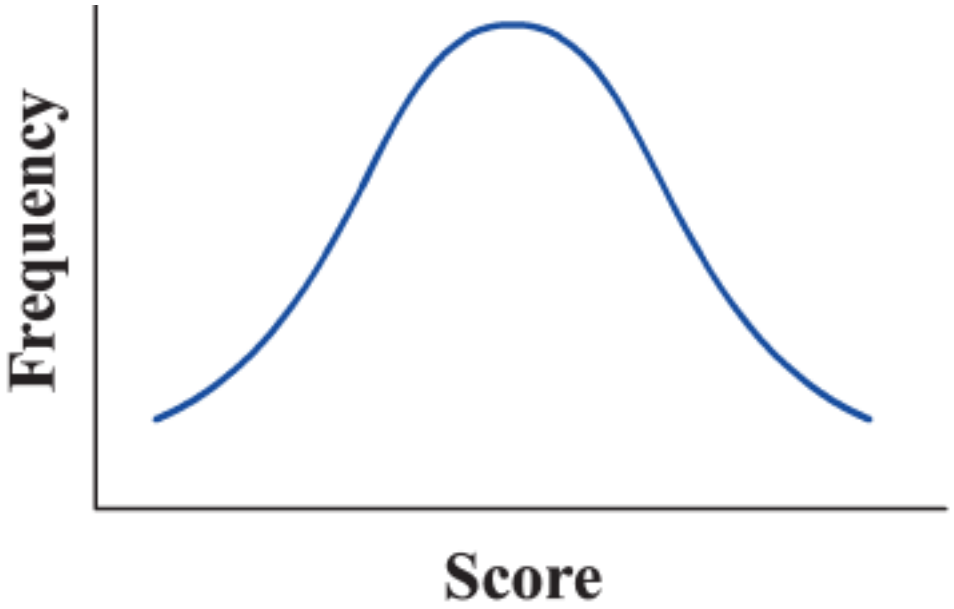
Me? I'm a ***smooth*** dude…



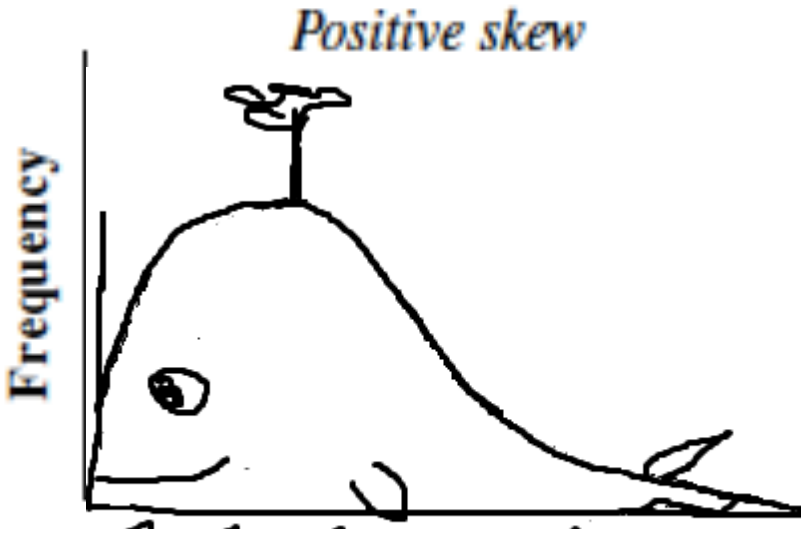
- **Note:** Smoothing increases simplicity, but can reduce precision if overdone

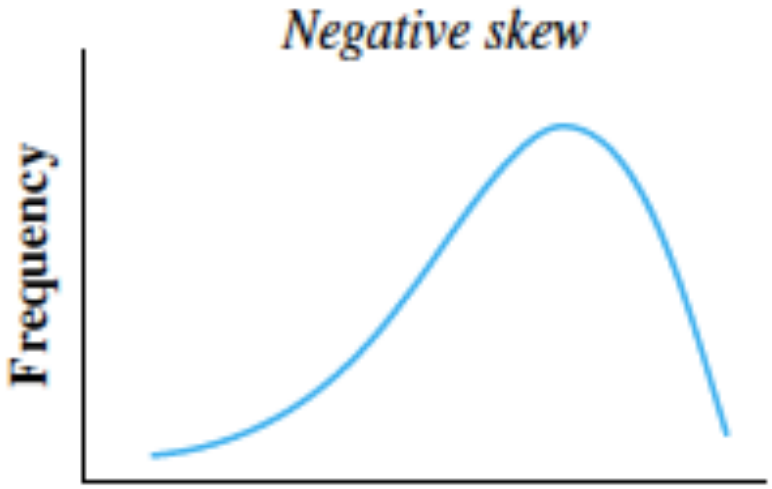
Distribution shapes

- The *shape* of distributions often reveals important information
  - Many statistics require certain shapes
    - Ideal distribution: Bell curve w/symmetry & unimodal



Skew





Or, statisticians are weird, so
“just think the opposite”

Mathematics Review

Order of operations:	Examples:
1. Parentheses	( ) [ ] { }
2. Exponents	x^2, x^2, or $\sqrt{x}$
3. Division	x/2 or $\frac{x}{2}$
4. Multiplication	x*2 or $x \times 2$
5. Summation	$\sum \boldsymbol{x}$
6. Addition	$x + 2$
7. Subtraction	$x - 2$

Summation

- Summation is very common operation in statistics

  "Find the sum of scores for variable *X*, from *i* = 1 to *i* = *N*"

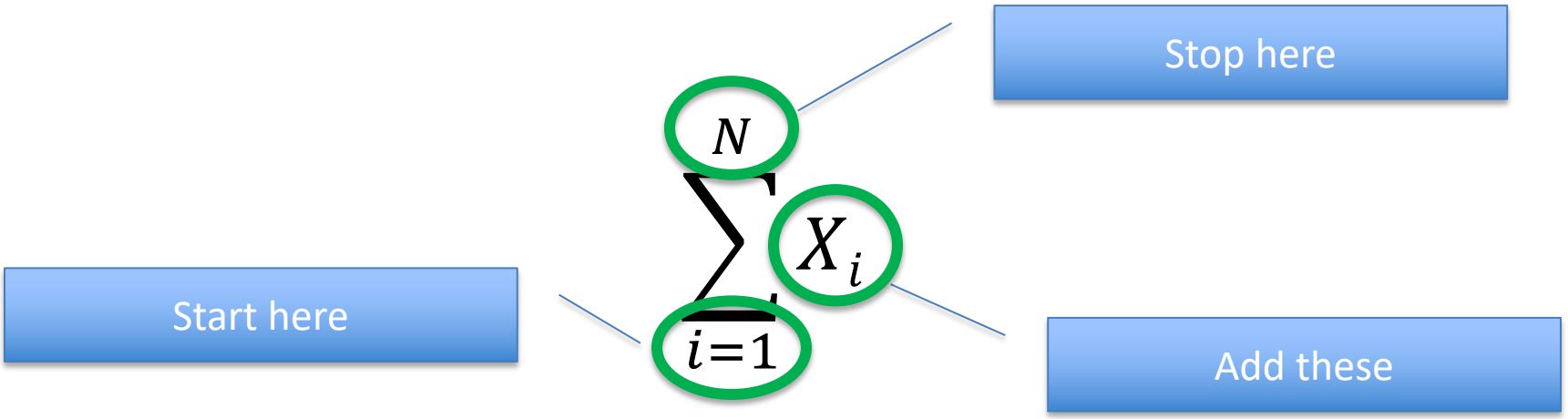
$$\sum_{i=1}^{N} X_i$$

$$X_1 + X_2 + X_3 + X_4 + \ldots X_N$$

Summation

- Summation is very common operation in statistics

"Find the sum of scores for variable *X*, from *i* = 1 to *i* = *N*"



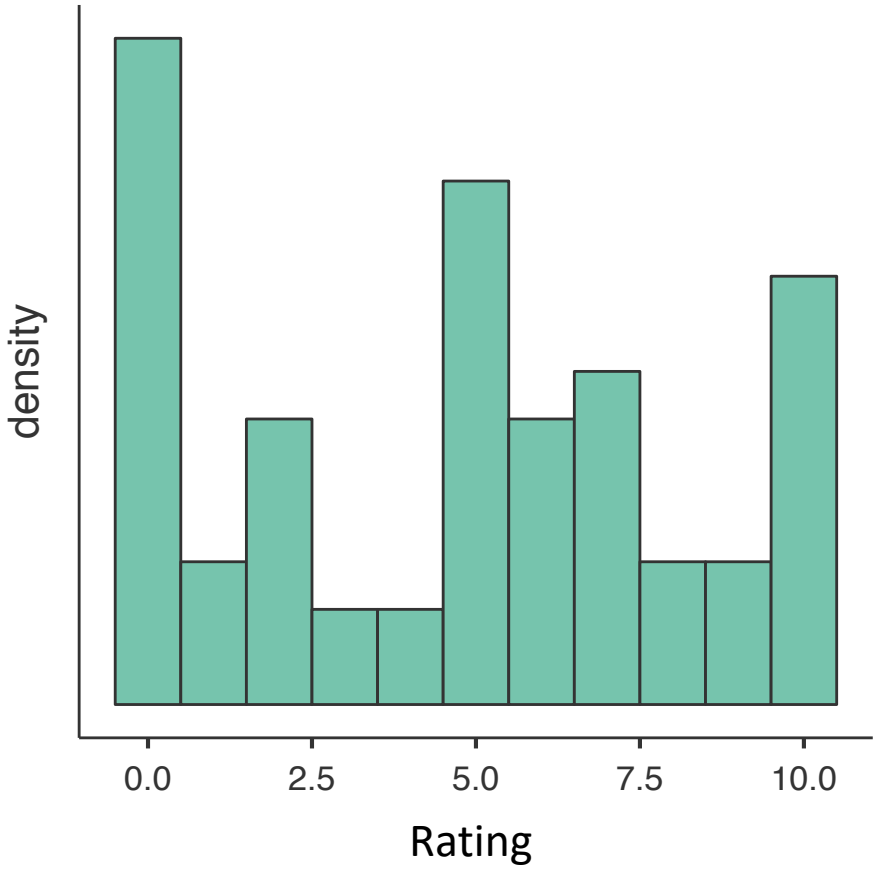
$$X_1 + X_2 + X_3 + X_4 + \ldots X_N$$

…the Greek letter ”sigma” makes me feel…

1 (‘at ease’) – 10 (‘very anxious’)

N = 66

- Sample of 5 participants for math review



Example…*the letter "sigma" makes me feel…*

$X_1 = 9$

$X_2 = 5$

…

What is: $\sum_{i=1}^{N} X_i$

How about: $\sum_{i=2}^{3} X_i$

Participant ID	DV
1	9
2	5
3	10
4	5
5	10

Fortunately, we'll mostly use ΣX_i

Sum of Squares

Only 1 change:

$$\Sigma X_i^2$$

Remember order of operations!!

Participant ID	DV (x)	x^2
1	9	81
2	5	25
3	10	100
4	5	25
5	10	100

Answer: 81 + 25 + 100 + 25 + 100 = 331

Now, calculate $(\Sigma X_i)^2$

Answer: $(9 + 5 + 10 + 5 + 10)^2 = 39^2 = 1521$