

# EGM101 – Skills Toolbox

Week 5, Part 4: Frequency

- With large datasets, want/need to **group** (categorize) data
- “How” depends on the type of data:
  - **Nominal data**
  - **Discrete data**
- Continuous data require something different

Responses to an airline survey on passenger complaints

Nature of complaint	Number of complaints
Inadequate leg room	719
Uncomfortable seats	914
Narrow aisles	146
Insufficient toilets	58
Other	384
<b>Total</b>	<b>2221</b>

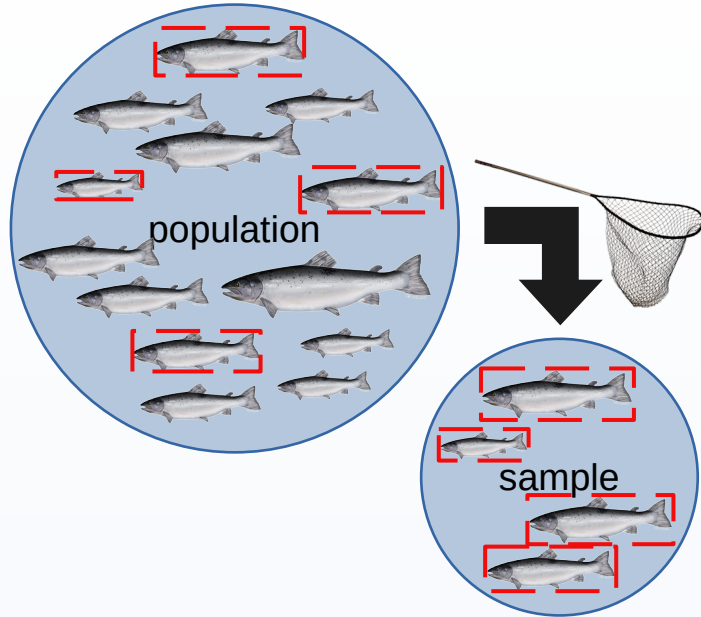
How often have you visited your local zoo in the past year?

Number of visits to local zoo	Number of persons
0	90
1	71
2	26
3	8
4	3
5	0
6	1
> 6	1
<b>Total</b>	<b>200</b>

- With discrete data, can create a tally chart
- For each distinct value, add a tally mark to table
  - Usually grouped by 5s
- Questions: what to do with:
  - Continuous data?
  - Discrete data with a large range?

5	6	2	6	4	4	0	6	10	8	1	5	0	0	6	4	3	1	5	6
5	1	2	7	6	2	5	5	2	7	4	3	5	4	3	4	3	3	7	9
8	0	0	6	6	4	1	4	8	1	6	4	4	3	6	4	0	0	2	9
6	1	5	10	4	2	8	4	3	6	2	8	5	6	10	8	6	4	4	3
1	8	1	5	7	3	2	3	6	6	5	5	1	5	6	4	0	5	6	4

Number of visits	Tally	Count
0		8
1		9
2		8
3		10
4		17
5		14
6		18
7		4
8		7
9		2
10		3
Total		100



Sample ( $n=25$ ) of salmon with lengths recorded in mm:

601	623	585	589	606
607	587	635	604	574
604	598	593	608	616
605	592	584	591	605
615	597	619	589	597

- Effective way of organizing numerical data
- Stem (on left)
  - All but last digit
- Leaf (on right)
  - Final digit

Stem	Leaf
57	4
58	4 5 7 9 9
59	1 2 3 7 7 8
60	1 4 4 5 5 6 7 8
61	5 6 9
62	3
63	5

601	623	585	589	606
607	587	635	604	574
604	598	593	608	616
605	592	584	591	605
615	597	619	589	597

# Grouping Continuous Data

- With continuous data, **classify** (group) values
  - Also discrete data with large ranges
- How to choose number/size of groups?
  - Too few: too much simplification (data loss)
  - Too many: too much detail
- In general:
  - Use equally-spaced intervals
  - Use between ~5 and ~20 classes
  - Choose “easy” boundaries

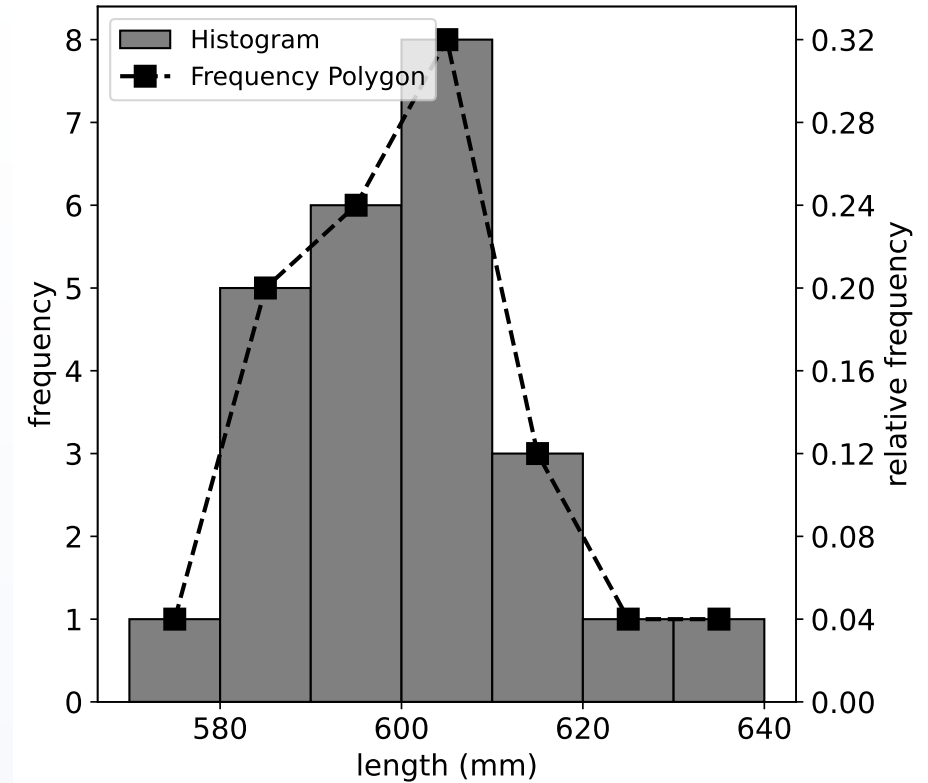
# Frequency and Relative Frequency

- **Frequency**: number of times value occurs
  - Always a whole number
- **Relative frequency**: proportion of times value occurs
  - Equal to frequency divided by number of values
  - Expressed as fraction or decimal
- **Cumulative relative frequency**: sum of previous relative frequencies

Frequency table of sampled salmon lengths ( $n=25$ )

Length (mm)	Frequency	Relative Frequency	Cumulative Relative Frequency
570 – 579.99	1	$1/25 = 0.04$	0.04
580 – 589.99	5	$5/25 = 0.20$	0.24
590 – 599.99	6	$6/25 = 0.24$	0.48
600 – 609.99	8	$8/25 = 0.32$	0.80
610 – 619.99	3	$3/25 = 0.12$	0.92
620 – 629.99	1	$1/25 = 0.04$	0.96
630 – 630.99	1	$1/25 = 0.04$	1.00
<b>Total</b>	<b>25</b>	<b><math>25/25 = 1.00</math></b>	<b>1.00</b>

- A graphical representation of frequency distribution
- Can use frequency or relative frequency
- Quickly shows shape, center, and spread of data
- Can also be plotted as a line graph (“frequency polygon”)
  - Markers plotted at center of intervals





- Goal: to summarize data
  - Easier to manage
  - Easier to see patterns
- With nominal, discrete data:
  - Count number of occurrences
- With continuous data (or large discrete datasets):
  - Group data by intervals
- Display data using tally charts, stem-and-leaf plots, or histograms

- Illowsky and Dean, Chapters 1.3, 2.1 – 2.2
- Caswell, Chapter 4
- Weiss, Chapter 2.3
- Manipulating Bin Sizes [[UW iSchool](#)]