



Computing Statistics

ID1050– Quantitative & Qualitative Reasoning

Single-variable Statistics

- We will be considering six statistics of a data set
- Three measures of the middle
 - Mean, median, and mode
- Two measures of spread
 - Variance and standard deviation
- One measure of symmetry
 - Skewness
- We can compute these values for either discrete or continuous data.

Mean or Average

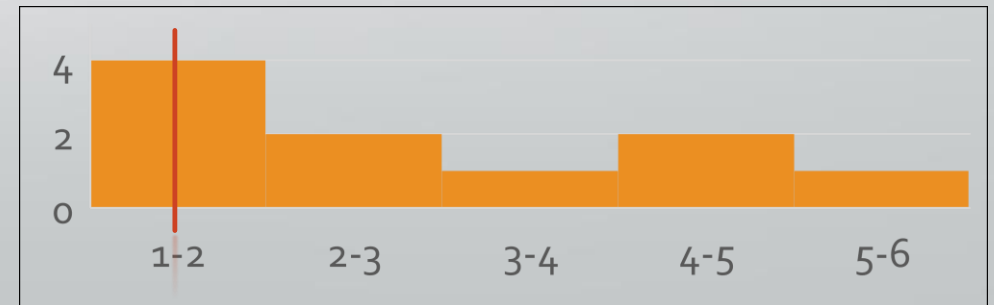
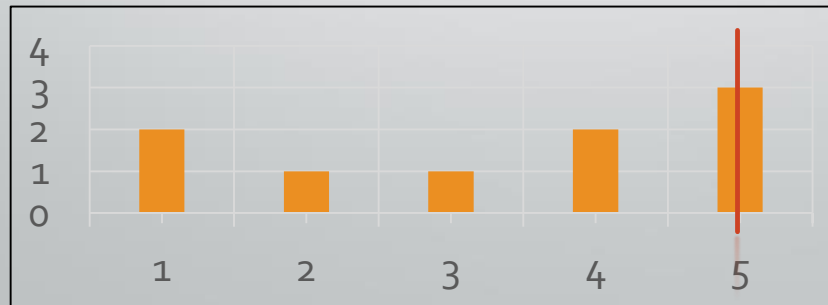
- The mean is defined as the **sum of the data divided by the number of data**
 - The variable often used is μ , the Greek 'mu', or $\bar{x}$. Often μ is associated with a population and $\bar{x}$ is associated with a sample.
 - Symbolically, $\bar{x} = \frac{\sum x}{n}$, where $\sum x = x_1 + x_2 + \cdots + x_n$, and n is the number of data values. (The capital letter sigma, Σ , represents summation.)
 - Example: Data is (1, 2, 3, 4, 5). The sum is $1+2+3+4+5=15$. There are 5 data values, so the average is $15/5=3$.
- Many calculators have a 'statistics' mode. The way the manufacturer chooses to implement statistical calculation varies widely. There are tutorials for this course's standard calculator, the TI-30Xa, for [entering data](#) and [computing statistics](#). If you have a different brand or model, consult your calculator's user's manual or website for details how to work with statistics.

Median

- The median is **the middle number when the data is listed in order**. If there is an even number of data points, the median is the average of the two middle values.
 - Example: Data is (1, 2, 3, 4, 5). The median is 3
 - Example: Data is (1, 2, 3, 4, 5, 6). The median is $(3+4) / 2 = 3.5$
- Why is this quantity useful?
 - The median ignores outlying values. What if our data had been (1, 2, 3, 4, 1000)?
 - The mean is 202, which is not characteristic of any of the actual values.
 - The median is 3, which is more typical of most of the values.
 - The median is helpful when looking for a house to buy. The median house price is the typical price you'd pay, even though the millionaire's house at the corner of the block raises the mean of the house prices above the value most people paid for theirs.

Mode

- The mode represents the **most populated class**, or the group with the most members. This is yet another reasonable way of finding the middle of the data.
- Determining the mode is different for discrete data than it is for continuous data.
 - For discrete data, the mode is simply the number that appears the most times.
 - Data is (1, 1, 2, 3, 4, 4, 5, 5, 5). **The mode is 5.**
 - For continuous data, the mode is the center of the range of the class that has the most members in it.
 - Data is (1.1, 1.2, 1.3, 1.8, 2.0, 2.6, 3.1, 4.6, 4.8, 5.1). The class from 1-2 has the most members. The center of this range is 1.5, so the **mode is 1.5.** (Note: 1.5 does not even appear in the data.)
- In both cases, the mode can be quickly determined from the graph. The mode is the x-value that is at **the center of the tallest bar** in either the bar graph (discrete data) or histogram (continuous data).
- Data can have two modes (*bi-modal*), but if there are more, we usually say it is *amodal* (no distinct mode).



Variance

- *Variance* (*var.* or σ^2 or s^2) is a measure of the spread of data about the average. We don't care which direction the difference is, so we will be ignoring the sign of the difference. In words, the variance is **the sum of the squares of the differences divided by one less than the number of data values**.
 - The equation is $var. = \frac{\sum(x-\bar{x})^2}{n-1}$
- Example: Data is (1, 2, 3, 4, 5) and mean ($\bar{x}$) is 3.
 - Variance is $10/(5-1)=2.5$
- If you are using a calculator, it is most likely that the calculator will compute the standard deviation (σ) instead. To get the variance from the standard deviation, simply find **the square of the standard deviation**:
 - $var = \sigma^2$

x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
1	3	-2	4
2	3	-1	1
3	3	0	0
4	3	1	1
5	3	2	4
			10

Standard Deviation

- *Standard deviation* (*std. dev.* or σ or s) is a measure of the spread of data about the average. We don't care which direction the difference is, so we will be ignoring the sign of the difference. In words, the standard deviation is the square root of (the sum of the squares of the differences divided by one less than the number of data values).
- The equation is $std. dev. = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{var.}$
- Example (from previous slide): Data is (1, 2, 3, 4, 5), mean ($\bar{x}$) is 3, and we previously found that the variance is $var. = 2.5$
- Since the standard deviation is the square root of variance,
 - Standard deviation is $\sigma = \sqrt{2.5} = 1.58$
- If you are using a calculator, it is most likely that the calculator will compute the standard deviation (σ) as part of its normal statistical function. There is a tutorial for using this course's standard calculator, the TI-30Xa, to [calculate standard deviation](#).
- Question: Since standard deviation and variance differ by one keystroke, why do we need both?
 - The units of standard deviation are the same as the data. Variance has other direct uses (e.g. Analysis of Variance) and is also more easily computed.

Skewness

- The distribution of a set of data may have symmetry about the mean, or it may have a longer 'tail' to one side or the other.
- Imagine draping a sheet over the graph of the data. The side of the sheet that is **least steep** is the side that has the **longer tail**.
 - If the **tail points to the right** (toward positive x values), the **skewness will be a positive number**.
 - If the **tail points to the left**, **skewness will be negative**.
 - Zero skewness indicates symmetric tails to both sides.
- It is sometimes difficult to estimate from the graph what the skewness will be, but there is a formula for calculating skewness in all cases:
 - **Skewness = (mean-mode)/(standard deviation)**

Data is (1.1, 1.2,
1.3, 1.8, 2.0,
2.6, 3.1, 4.6,
4.8, 5.1).

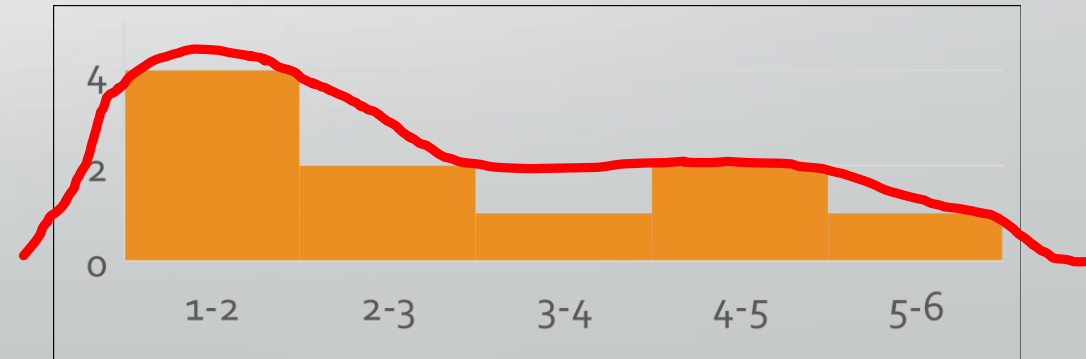
Mean is 2.76

Mode is 1.5

Std. Dev. is 1.56

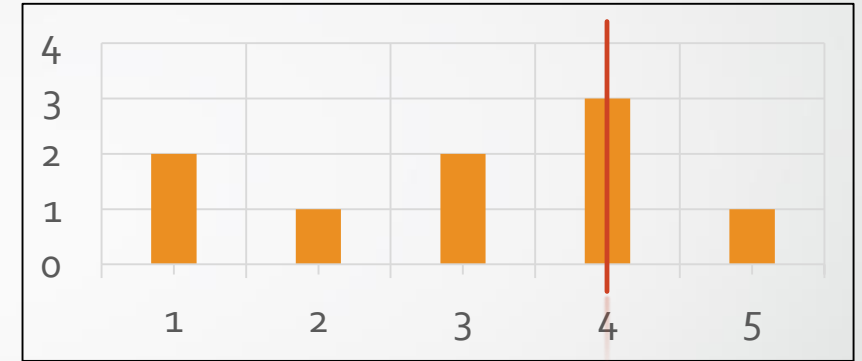
$$\text{Skewness} = \frac{(2.76 - 1.5)}{1.56}$$

= 0.81
(tail to the right)



Example: Discrete Data

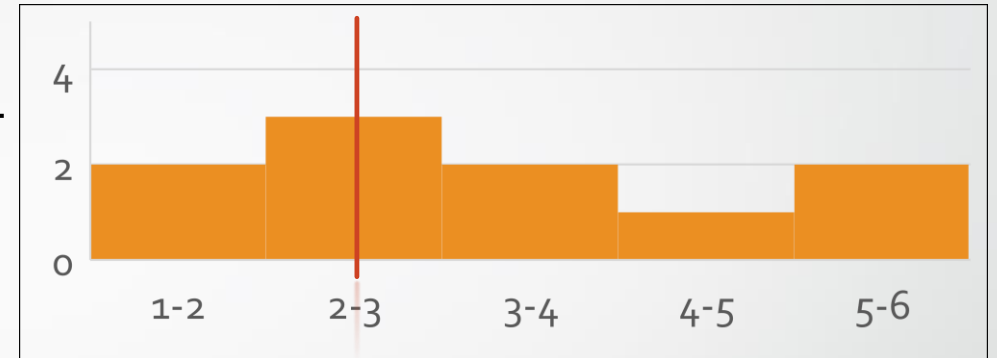
- Data: 1, 1, 2, 3, 3, 4, 4, 4, 5
- N: 9
- Graph:
- Mean: 3
- Median: 3
- Mode: 4
- Variance: 2
- Standard Deviation: 1.41
- Skewness: -0.71



x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
1	3	-2	4
1	3	-2	4
2	3	-1	1
3	3	0	0
3	3	0	0
4	3	1	1
4	3	1	1
4	3	1	1
5	3	2	4
27			16

Example: Continuous Data

- Data: 1.5, 1.7, 2.4, 2.5, 2.7, 3.5, 3.8, 4.7, 5.1, 5.1
- N: 10
- Graph:
- Mean: 3.3
- Median: 3.1
- Mode: 2.5
- Variance: 1.81
- Standard Deviation: 1.35
- Skewness: 0.6



x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
1.5	3.3	-1.8	3.24
1.7	3.3	-1.6	2.56
2.4	3.3	-0.9	0.81
2.5	3.3	-0.8	0.64
2.7	3.3	-0.6	0.36
3.5	3.3	0.2	0.04
3.8	3.3	0.5	0.25
4.7	3.3	1.4	1.96
5.1	3.3	1.8	3.24
5.1	3.3	1.8	3.24
33			16.34

Conclusion

- We can answer a great deal of statistical questions by examining the graph and six standard statistical variables for the data:
 - Bar graph or histogram
 - Measures of the middle
 - Mean (can be done on a calculator)
 - Median (obtained from the sorted list of data)
 - Mode (obtained from the graph)
 - Measures of the spread
 - Variance (calculated using a tabular method) [or the square of the std. dev.]
 - Standard Deviation (obtained from calculator's statistics mode) [or the square root of the variance]
 - Measure of symmetry
 - Skewness (calculated from the above values Mean, Mode, and Std. Dev.)