Please go to Update page and Course page to see information about class and to keep up to date. The links are in the main page of Canvas and also on http://nature-lover.net/math. You can see old notes from spring 25 etc at this website. It will help you prepare for class.

————————————————————————————————-

Today: More statistics: variance and standard distribution

————————————————————————————————-

Quiz 7 wed April 8: mean, median, mode, std deviation, percentile,

Normal distribution

————————————————————————————————

Review: Mean, median and mode:

MEASURES OF CENTRAL TENDENCY

Helps in understanding what the data is telling us.

Mean (average), Median, and the Mode. (*for numerical data).*

Average = Total of all values / Number of values

Median = value at half-way mark: (arrange in ascending order, pick the middle number; if there are two values in middle, take average)

Mode: most frequent value or item (more than one mode is possible; if all are same value, then there is no mode).

Examples:

Sports (batting averages, scoring averages, QB ratings, etc.,)

Grade point average

How average global temperature anomalies are calculated:

https://www.encyclopedie-environnement.org/en/climate/average-temperature-earth/

Video showing how this anomaly has changed (for the worse)

https://www.youtube.com/watch?v=Z4bSxb5THm4

Actually, the graphs on climate change look at the anomaly between the global average temperature and a base point temperature.

CDC data on average number of flu deaths from 2010 to 2020

https://www.cdc.gov/flu/about/burden/index.html

————————————————————————————————

Average (mean) is NOT same as the median, but they could be same in some cases

Example: 5 6 7 8 9 : average = median = 7.

Average when not all values are given:

Simple example: Find the total value of 10 items in range 2 to 4 approximately.

Approximate total value = 10 x average of 2 and 4 = 10 x 3 = 30.

So it is as if we suppose that all the items are of price 3, and so we are just adding 3 repeatedly, for a total of 10 times 3.

This will be a good approximation if prices are random, because then the values will sort of “average out” to be the same.

Average = total value / number of values.

So Total value: Number of values x average

Example: Average is 5 and there are 10 values and total value is 50.

VARIANCE AND STANDARD DEVIATION

Help to measure how much the data is **dispersed.**

They help us to go deeper into how the data is distributed.

Question: Are there a lot of values much above average or much below average?

Suppose average incomes of different countries are given, we can compare their relative prosperity. But if you want to know how people are doing on the whole, we have to know how the values differ from the average. That will give us an idea how the wealth is distributed. Are a few people making a lot of income or are most people in same range?

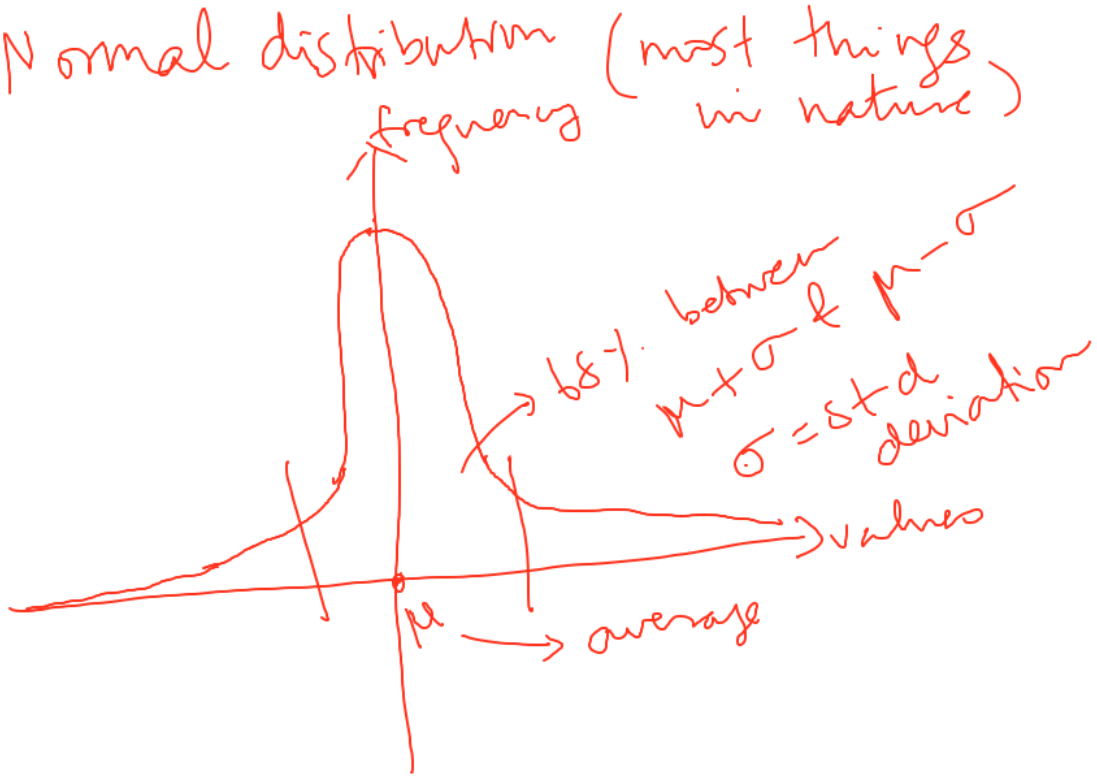
Key point: What is range in which most people’s income is located?

If incomes are distributed randomly according to a commonly occurring distribution called normal distribution, this would be the case.

The graph of the normal distribution is the Bell Curve.

An example is the distribution of heights of people.

If the standard deviation of heights of a population is 1 inch, it means most people are within 1 inch of the average.



For example, If average height is 6 feet, and std dev is 6 inches, then most people would be within 5 ft 6 in to 6 ft 6 in

Example for finding standard deviation:

Predicted high temperatures in next week:

73 63 58 60 61 69 69

Average = 64.714 approximately. This is denoted by μ.

If the values are denoted by x, then the standard deviation σ is given by the formula : $\sigma = \sqrt{\left(\frac{\sum(x-\mu)^2}{n}\right)}$ The table below shows how it is calculated. The steps are described below the table.

FINDING STANDARD DEVIATION

73 63 58 60 61 69 69	DIFFERENCE FROM THE AVERAGE	SQUARES OF DIFFERENCES		
73	8.2857143	68.6530614612245		
63	-1.7142857	2.93877546122449		
58	-6.7142857	45.0816324612245		
60	-4.7142857	22.2244896612245		
61	-3.7142857	13.7959182612245		
69	4.2857143	18.3673470612245		
69	4.2857143	18.3673470612245		
64.7142857142857	0.0000001	189.428571428571	SUM OF SQUARES OF DIFFERENCES	
	ACTUALLY THIS SUM IS ALWAYS 0	27.0612244897959	AVERAGE OF SQUARES OF DIFF	AKA VARIANCE
		5.20204041600946	STANDARD DEVIATION!	(SQUARE ROOT OF VARIANCE)
		ON THE AVERAGE THE NUMBERS DIFFER BY 5 APPROX	MOST OF THEM ARE WITHIN 1 STD DEV OF AVERAGE	
			THAT IS, BETWEEN AVERAGE + SD AND AVERAGE - SD	

NOTE: In the computer program, choose “population standard deviation”

What did we do here? First find the average, then find the differences of each date point from the average, then square the differences, then

find average of the squares of the differences aka variance and then (population) standard deviation by square rooting the variance.

To find how the data deviate from average, cannot just add differences.

Sum of the differences from the average is always zero. (Prove it to yourself!) You can see in the second column of table above the differences add to zero (although it says 0.0000001 it is really zero).

Example: average of 5, 6, 7, 8, 9 is 7 and sum of differences is

-2+-1+0+1+2 = 0.

VARIANCE = average of squares of differences (sum of squares/number of values)

STD Deviation = square root of variance

NOTE: When there are too many pieces and you can only use a sample, then there is the **sample standard deviation**. In Excel, if you are using only small number of values, choose **population standard deviation.**

Std deviation doesn't tell us *what the average is.*

It tells you how things differ from the average, on the average.

Are all of them close to the average, or are many of them far from average?

PRACTICE QUESTIONS FOR TODAY:

Problem 1: Find the standard deviation for the following temperature values by excel or other program:

55, 57, 61, 63, 69, 65, 61, 66, 70, 71.

Problem 2: Find the standard deviation for the same temperature values by hand, showing each step : calculate average first, find squares of the differences, then the average of the differences squared (variance), then the square root of the variance which is the standard deviation.

Problem 3: (Watch video about six sigma first) What percentage of the values are within one standard deviation of the average? What percentage are within two standard deviations? What would be six standard deviations from the average?