BS – Important Concepts Central Tendency– Lessons 10, 11, 12

1. Ideally, one number to describe data. How choose one number or a small range

of numbers in a data-set?

-> measures of center because they describe the center of a distribution

-> this allows us to summarize all our data;

-> multiple numbers to describe the whole data set vs interval estimate

1. Mode:

-> the value at which frequency is highest, the most common value;

-> of data set = a number;

-> of distribution a range/bin that occurred with highest frequency

-> uniform distribution = no mode;

-> multimodal = two or more clear trends, bimodal distribution ;

-> The data doesn't need to be numeric to find a mode: we can also compute the mode for categorical data as well! (the mode of a categorical data set: the preferred M&M flavor of 8,000 Udacity students);

-> the mode occurs on the X-axis, so you are looking for whatever value has the highest frequency;

-> not all scores in the data set affect the mode;

-> there isn't an equation for the mode;

-> the mode is not the same in all the samples;

-> it is not affected by extreme scores (outliers);

1. Median:

-> value in the middle;

-> exact middle of any distribution ;

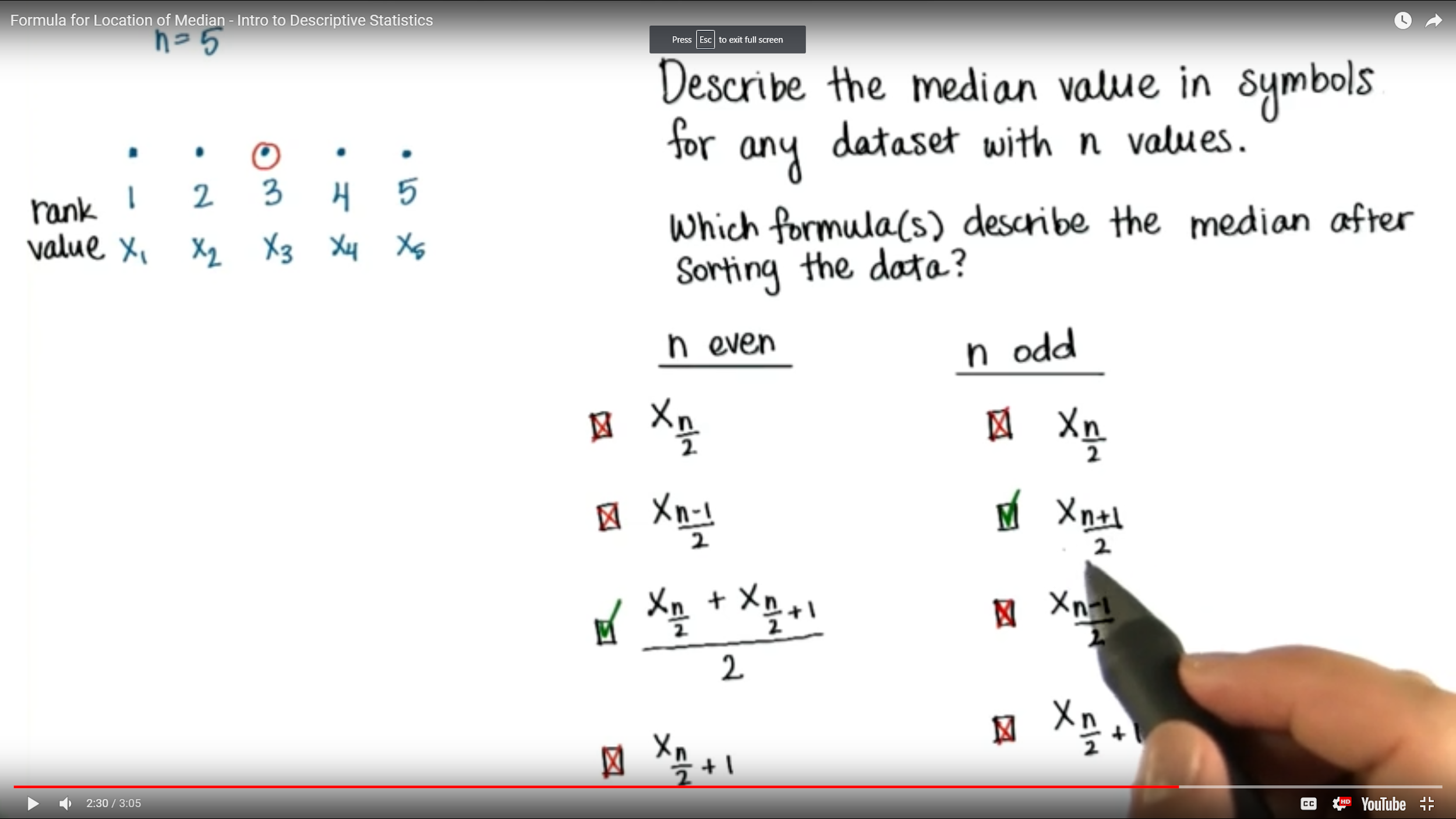
-> to make it a useful statistic we have to put the data in order, least to greatest or greatest to least (doesn't matter) => in general sort from least to greatest;

-> even number of scores=> median is the average of the two middle numbers;

-> robust/strong tendency, not affected much by departures from the norm, an outlier;

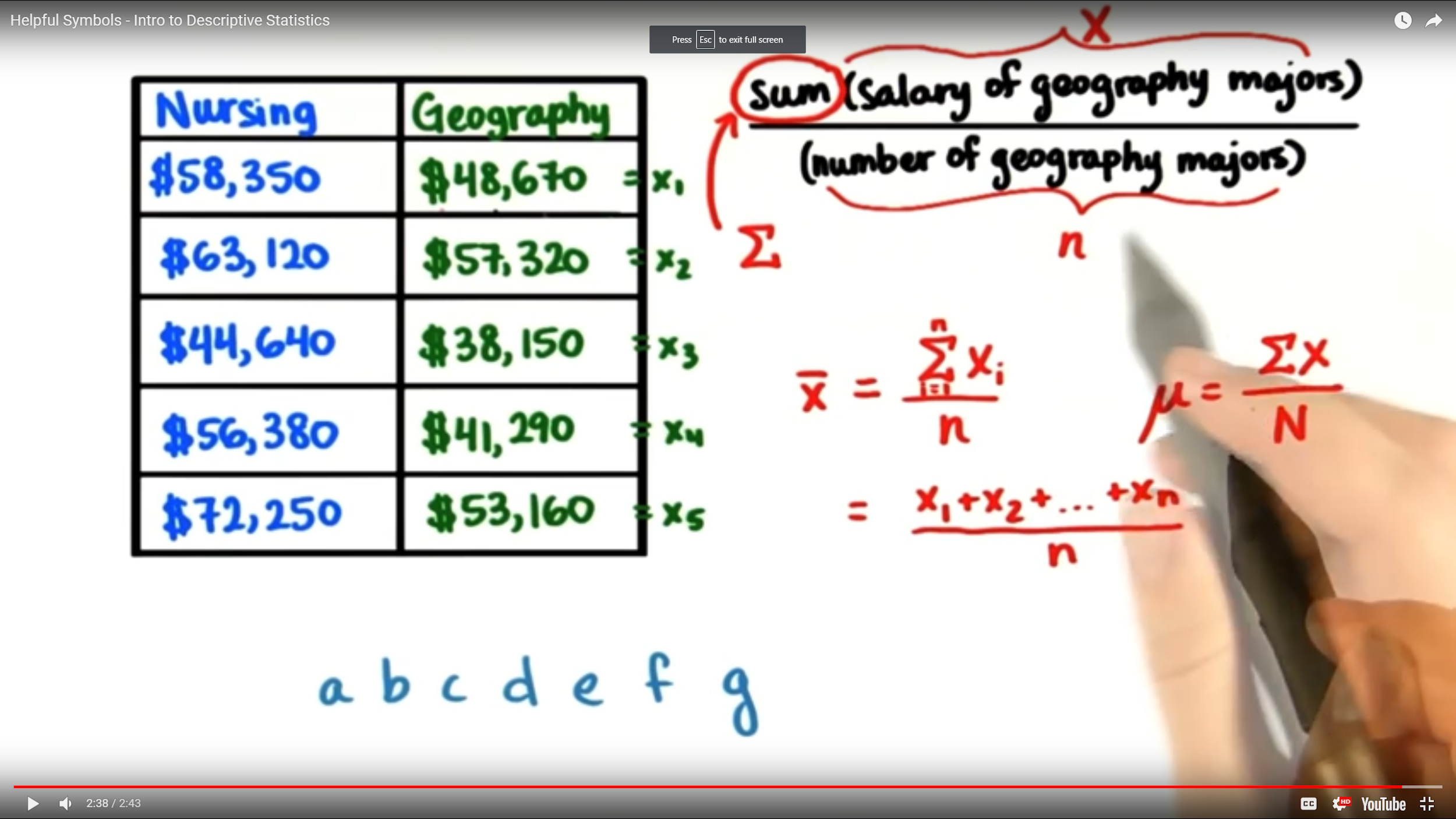
-> better statistic to determine starting salary, for highly skewed distributions;

-> formula for location of median;



1. Mean/average: X bar for sample and mu is for population

-> statistic that rests at a specific spot in the middle of the distribution;

-> takes all the values into account because we add them all up, then divide by how many values there are-> procedure for finding mean = there is an equation for the mean

1. Properties of the Mean:

-> all scores in the distribution affect the mean

-> the mean can be described with a formula

-> many samples from the same population will have similar means

-> the mean of a sample can be used to make inferences about the population it came from

-> the mean will change if we add an extreme value to the data set

1. Outlier:

-> values that are unexpectedly different from the other observed values

1. Mean with outlier:

-> the average is misleading, less representative of the middle of the data

-> create skewed distributions by pulling the mean toward the outlier

1. Orders Measures of Center:

-> positively skewed distributions: mode < median < mean

-> normal distribution: mode = median = mean

1. Measures of center:

-> make good decisions

-> make comparisons (ex Facebook friends)

