

Measures of Central Tendency

Definition - the three kinds of numbers that can be used to "best describe" the data:

MEAN - the average number of the data. Used when the mode is not majority and there are **NO** outliers.

To find the mean, divide the sum of the data by the quantity of the data.

MEDIAN - the middle number of the data. Used when the mode is not majority and there **ARE** outliers.

To find the median, move equally from lowest and highest numbers towards the middle.

- An even number of data always has two medians. If they are different, the median is the average of the two.
- An odd number of data has one median.

MODE - the most common number in the data. Checked first, and used when it **IS** the majority (more than half).

To find the mode, count the quantity of each unique data.

- There may not be a mode (must have at least 2).
- There may be more than one mode. (Two modes is called "bimodal"; three modes is called "trimodal"; 4 or more modes is called "multi-modal.")

Statistics Vocabulary

KEY ITEMS OF DATA:

MINIMUM - the **smallest** number in a group of data

MAXIMUM - the **largest** number in a group of data

RANGE - the spread (**difference**) between the highest and lowest number in a group of data (maximum minus minimum).

STANDARD DEVIATION - how far apart the data is spread from the mean

OUTLIER - data that is **beyond** two standard deviations from the mean **OR** more than 1.5 times the IQR (inter-quartile range) from the median

GRAPHS AND TABLES:

FREQUENCY TABLE – a 3-column table that groups frequencies into intervals

HISTOGRAM – a vertical, connected bar graph of frequencies

STEM-AND-LEAF PLOT - a graph that groups data by placing stems (larger place-value) on the left and leaves (smaller place-value) on the right

BOX-AND-WHISKERS PLOT - a graph that shows data in quartiles with minimum, maximum, median, and any outliers

Statistics Vocabulary, cont.

population - all; the whole

sample - part of the whole

\bar{x} (x-bar) - mean (average) of the sample

μ (mu) - mean of the population

S - standard deviation of the sample

σ (sigma) - standard deviation of the population

variance - expectation of squared data from the mean

Z-score - how many standard deviations data is from the mean

Empirical rule - in normal distribution, almost all data falls within 3 standard deviations of the mean

random sampling - choosing a sample by chance

Census - a numbering of the population

survey - collecting data randomly by asking questions

bias - preference or prejudice

Venn Diagram - a diagram of circles with data as quantities

Data Collection Methods

Quantitative:

Experiments/Trials

Management Information Systems (MIS)

Observations/Counting/Reporting

Surveys (Close-ended questions)

Interviews

Questionnaires/Census

Qualitative:

In-depth Interview

Surveys (open-ended questions)

Observations

Document Review

Focus Groups

Sampling

20. 65% of the voters of a particular county are registered democrats. In a poll of 1,000 citizens, half of which were democrats, 350 voted “yes” and 50 non-democrats voted “yes.” Estimate the percent of the whole voting population who would vote “yes.”

Democrats:

percentage of whole * percentage of sample

$$.65 * 350/500 =$$

$$.455$$

Non-Democrats:

percentage of whole * percentage of sample

$$.35 * 50/500 =$$

$$.035$$

$$.455 + .035 = .49$$

49% of population would vote "yes"

Data Distribution

EVEN DISTRIBUTION:

All data is represented **equally**. A flat-topped rectangle.

NORMAL DISTRIBUTION:

Median and mean are together. A "bell" curve.

1st Standard Deviation - 68.27% [**68.27%**] of the data (34.13% on each side)

2nd Standard Deviation - 27.18% [**95.45%**] of the data (13.59% on each side)

3rd Standard Deviation - 4.28% [**99.73%**] of the data (2.14% on each side)

Tails - .27% [100%] of the data (.13% on each side)

SKEWED DISTRIBUTION:

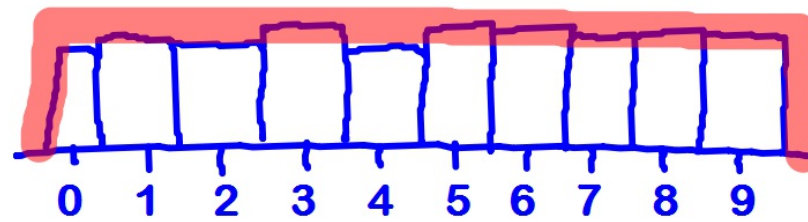
When median and mean are apart, the distribution becomes skewed.

Skewed Right (positive skew) - positive tail (right side) is longer

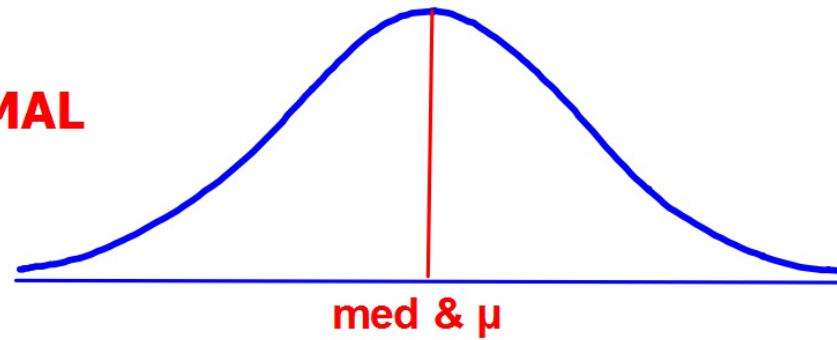
Skewed Left (negative skew) - negative tail (left side) is longer

Types of Distribution

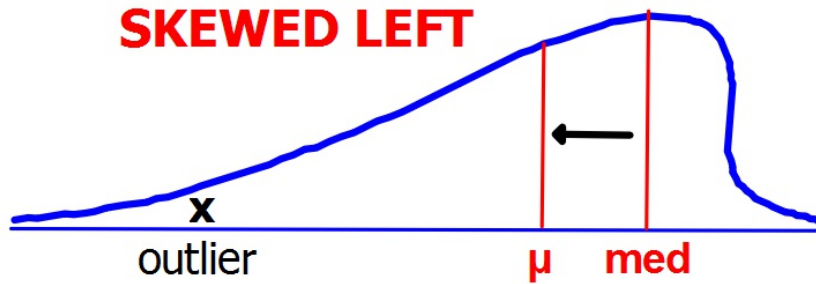
EVEN



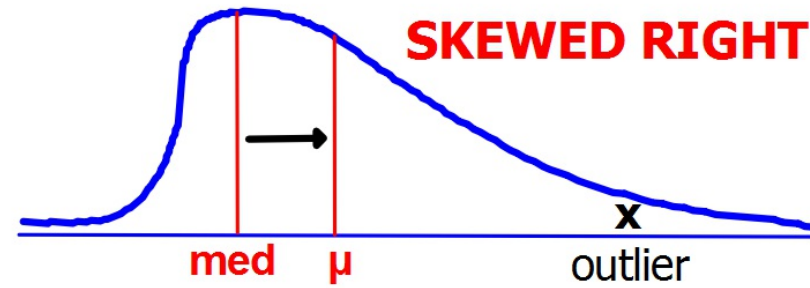
NORMAL



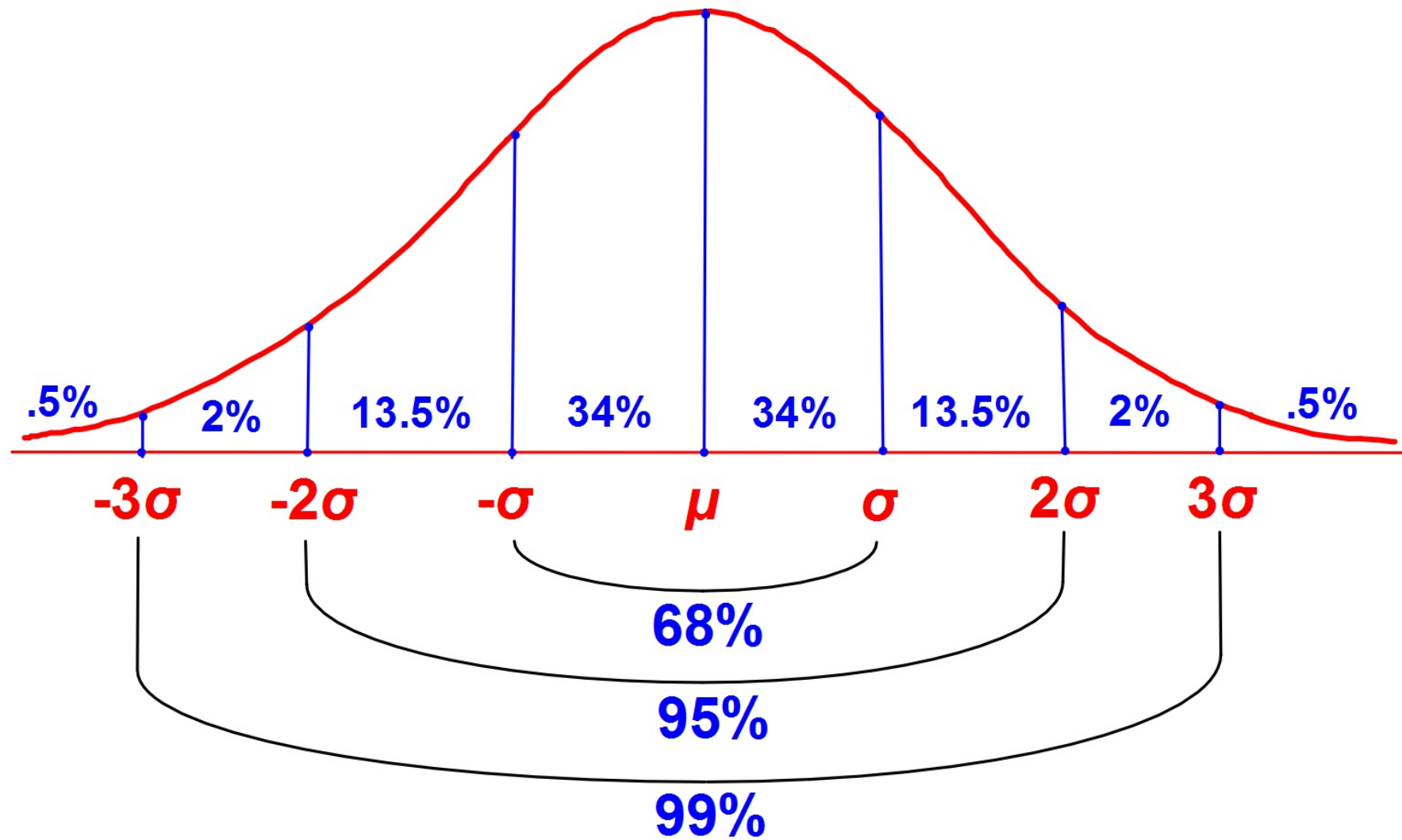
SKEWED LEFT



SKEWED RIGHT



Normal Distribution Percentages



Distribution Calculator Commands

Z-Test - Finds z-scores and p-values of samples compared to mean using either data or stats.

Type under Z-test: μ , σ , \bar{x} , n , and $<$ or $>$. Select calculate or draw.

normalcdf - normal cumulative distribution function; used to find percentage of area under the curve (p-values).

Type: "normalcdf(lower bound, upper bound, mean, standard deviation)"

invnorm - inverse of normal; used to find z-scores or actual x-values. Always calculates from left so subtract percentile from 1 to calculate from right.

To find z-scores, type: "invnorm(percentile from left, mean of 0, standard deviation of 1)"

To find x-values, type: "invnorm(percentile from left, mean, standard deviation)"