STAT VOCAB

THE QUESTION PHASE

NULL HYPOTHESIS

The hypothesis that most people already think is true. Ex. Eating a good breakfast before a test will help you focus

Notation → H_o

ALTERNATIVE HYPOTHESIS

The hypothesis that goes against the current held belief. Ex. Eating a big breakfast does not help you focus before a test.

Notation → H_a

POPULATION

The entire group you want information about.

THE DESIGN PHASE

OBSERVATIONAL STUDY

A characteristic is observed and data is collected. No outside factors are placed on the population.

EXPERIMENTAL STUDY

A treatment is assignment to a group in order to study a specific characteristic. Data is then compared based on the results of the treatment.

TREATMENT

In an experimental study, the treatment is what is *changed* or *added* to the population or sample.

VARIABLE OF INTEREST

The variable of interest is what characteristic the surveyor wants to know about. NOTE- this is NOT the treatment

CONTROL GROUP

In an experimental study, this group has no treatment done on them.

PLACEBO

A treatment that has no effect, however, the subject(s) does not know this. Often times in medical studies, the placebo is a sugar pill.

PLACEBO EFFECT

This occurs when subjects on a placebo show change because they think they are on an actual treatment.

CATEGORICAL DATA

Data that cannot be arranged numerically. For example: hair color, gender, ethnicity, religious affiliation, etc.

QUANTITATIVE DATA

Data that can be arranged numerically. For example: height, age, income, number of children, etc.

CLOSED QUESTION

In a survey, this type of question requires the subject to choose an answer based on listed choices. (multiple choice)

OPEN QUESTION

In a survey, this question has no answer choices and leaves the subject to answer freely. “How do you feel about _____”

BIASED QUESTION

In a survey, a biased question is worded so that the subject would tend to answer in a certain way.

BIASED SAMPLING

When some members of a population are more likely to be selected than others.

NON-REPRESENTATIVE SAMPLING

A sample that does not reflect the makeup of the population (a biased sample). For example population: High School

Sample: Only Freshmen

RESPONSE BIAS

Responders do not give accurate answers. For example: The subject might be embarrassed or trying to impress.

NON-RESPONSE BIAS

Refusal to answer a question.

OBSERVER EFFECT

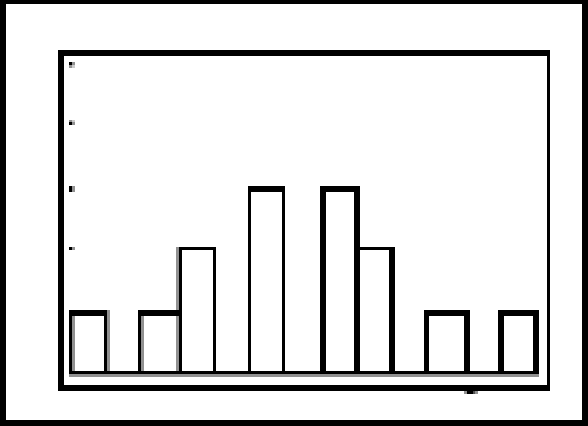
Similar to response bias. Subjects being observed act differently when they know they are being observed.

WORDING OF QUESTIONS

When the wording of questions makes a responder lean to a certain response. This could also occur in a closed question with not enough options.

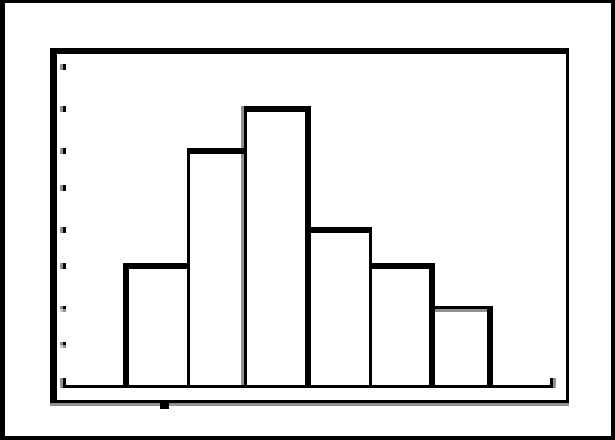
HIGH VARIABILITY

When data is spread over a wide range of values.



LOW VARIABILITY

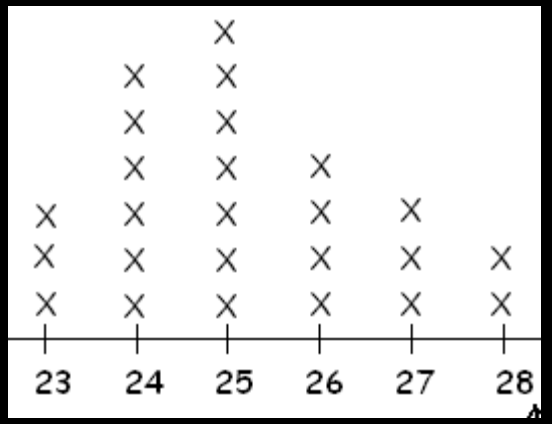
When data is not spread out (condensed).



HIGH STATISTICAL BIAS

When the trial outcome is NOT close to the population mean.

SAMPLE RESULTS →



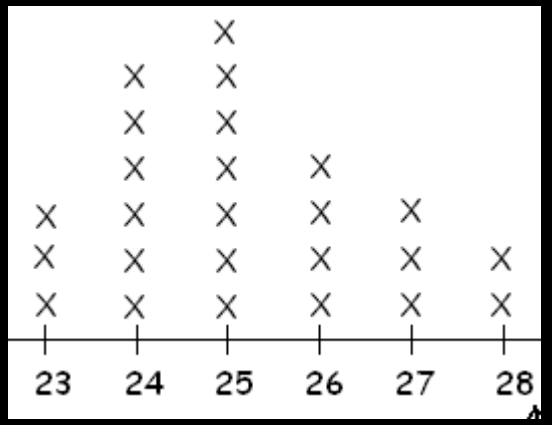
Population mean: 32

LOW STATISTICAL BIAS

When the trial outcome is close to the population mean.

SAMPLE RESULTS →

Population mean: 24.5



UNDERCOVERAGE

Not enough members of the population are being studied.

POPULATION: Sprayberry High School

Sample: 10 students

THE COLLECT PHASE

CENSUS

A study in which the ENTIRE population is sampled.

SAMPLING METHOD

How a sample is chosen in a statistical study. If the sample is not picked correctly, all of the data collected could be inaccurate.

SIMPLE RANDOM SAMPLE

Every subject receives a number, and numbers are randomly selected either through a random drawing or random number generator.

STRATIFIED RANDOM SAMPLE

Population is divided into distinct groups. Then a simple random sample is done for each group.

Ex. Split into boys + girls- then draw numbers out of hat.

SYSTEMATIC SAMPLE

A rule is used to select members of a population. For example: Every third person who comes through the door will be selected.

CLUSTER SAMPLE

Clusters are formed which are small scale representations of the population. The entire cluster is studied.

Example: D lunch (has members of all parts of the school)

CONVENIENCE SAMPLE

Only members of the population who are easily accessible are sampled. This generally results in poor samples.

THE ANALYZE PHASE

POPULATION MEAN

The average of EVERY member of the population. This is generally impossible, depending on what your population is.

Notation → μ

SAMPLE MEAN

The mean of a sample of the population.

Notation → $\bar{x}$

UNIVARIATE DATA

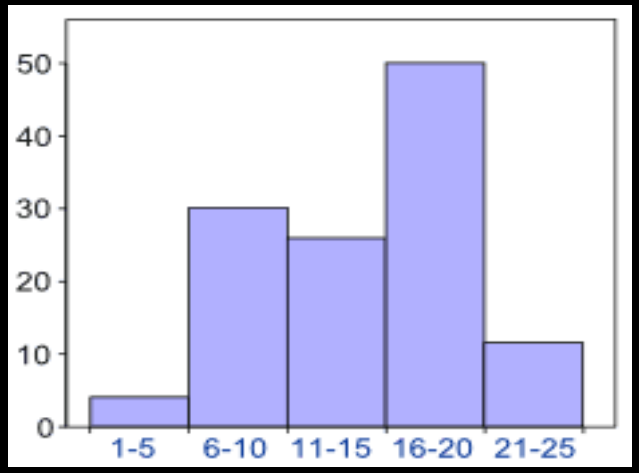
When you conduct a study that only looks at one variable, you are working with univariate data.

BIVARIATE DATA

When you conduct a study that looks at the relationship between two variables, you are working with bivariate data.

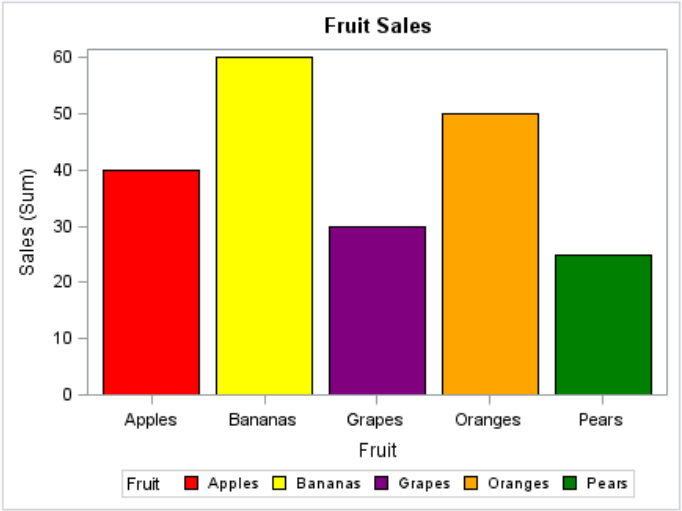
HISTOGRAM

A diagram consisting of rectangles whose height represents the frequency of a variable, and width represents an interval. There is no spacing between the rectangles.



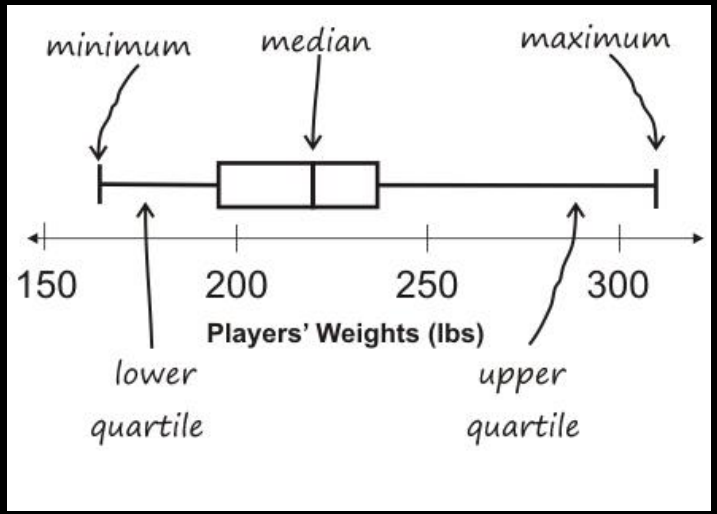
BAR GRAPH

A diagram in which the numerical values of variables are represented by the height of rectangles with equal width. There is spacing between the rectangles.



BOX & WHISKER PLOT

A graphic way to display the median, quartiles, and extremes of a data set on a number line to show the distribution of the data. Sometimes called a boxplot.

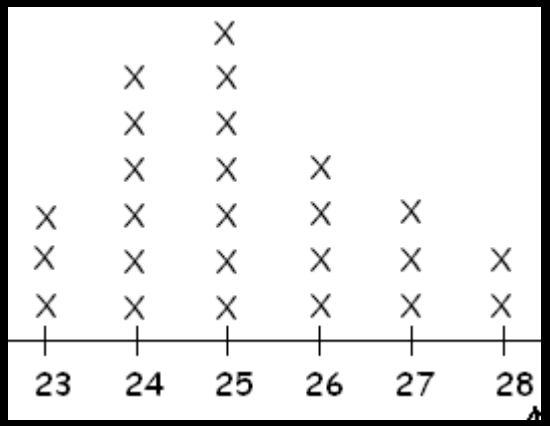


PIE CHART

A type of graph in which a circle is divided into sectors that each represent a proportion of the whole.

DOT PLOT

The representation of data by using dots over a number line. The number of dots over each increment on the number lines represents the frequency of the data. (the dots can be Xs as well)



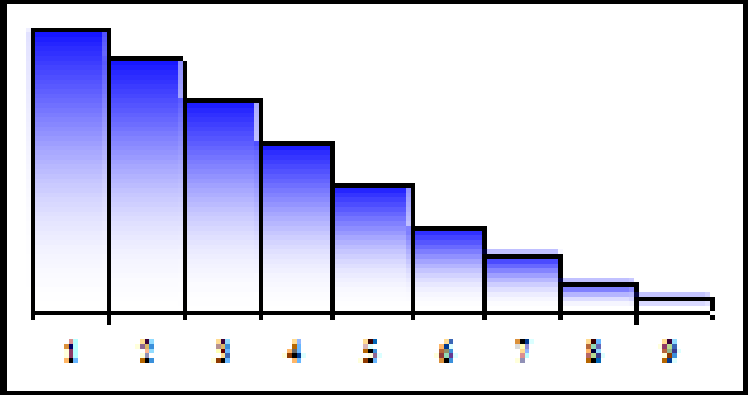
FREQUENCY TABLE

A table displaying how often each value in a set of data occurs.

Class interval x(weight in kg)	Tally	Frequency f
40 - 44	II	2
45 - 49	IIII	4
50 - 54	𝍸	5
55 - 59	𝍸 III	8
60 - 64	𝍸	5
65 - 69	IIII	4
70 - 74	II	2
		30

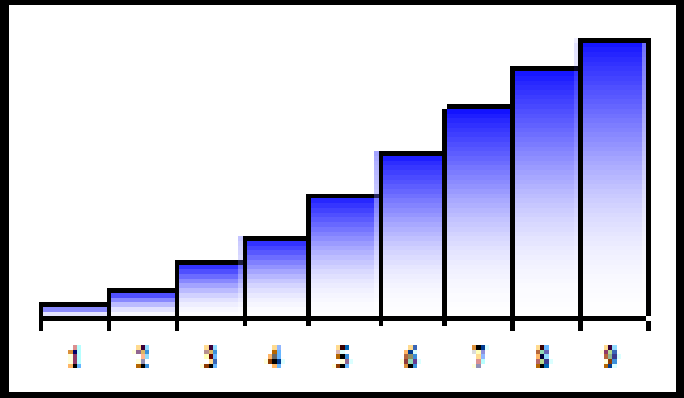
SKEWED RIGHT

A distribution with most of the observations on the left. (It has a right tail)



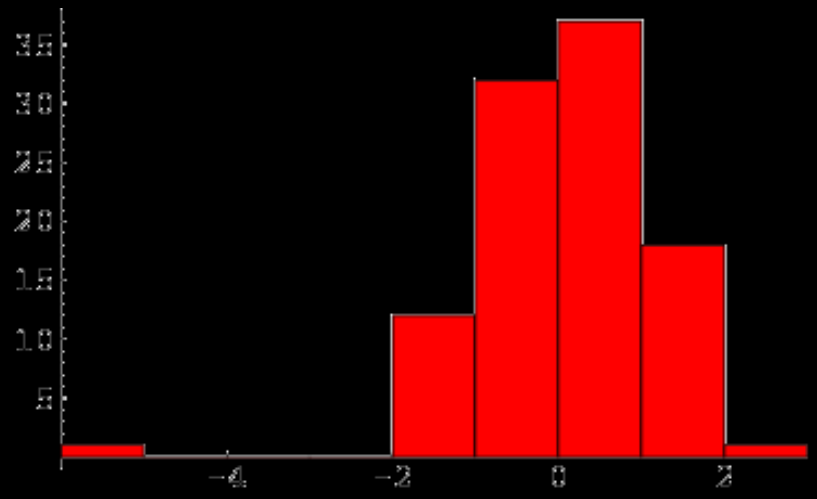
SKEWED LEFT

A distribution with most of the observations on the right. (It has a left tail)



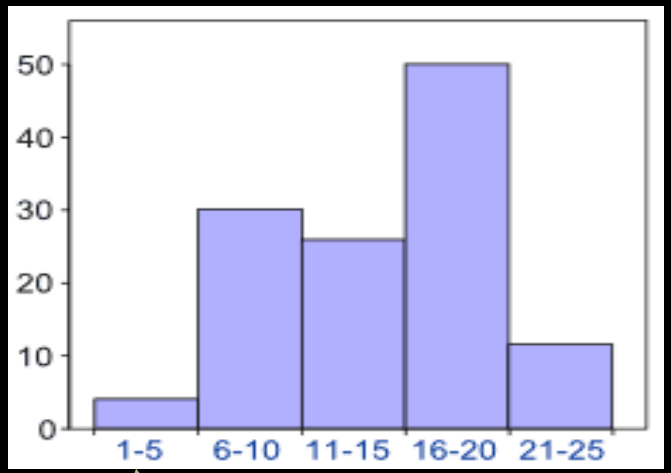
OUTLIER

A data value that is far away from the rest of the distribution on either side.



BIN SIZE

In a histogram, this refers to the size of the intervals used at the bottom of the chart.



Bin size is in 4s