MAT 121: Probability and Statistics

Summer 2018
Lecturer: Caleb McWhorter
Syracuse University

Course Webpage: `http://coffeeintotheorems.com/courses/2019-2/summer/mat-121/`

Chapter 1

How to Collect Data
Types of Data & Measurement

What is Statistics?

Definition (Statistics)

Science of collecting, organizing, analyzing, interpreting, and presenting data.

What is Statistics Good For?

Explain Patterns

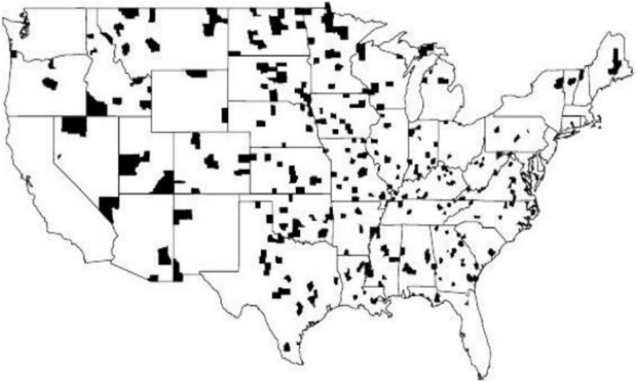


Figure 1: Counties with the highest 10% age-standardized death rates of kidney/ureter cancer in males, 1980–1989.

A. Gelman and D. Nolan (2017). *Teaching Statistics: A Bag of Tricks*. Oxford University Press.

Make Data Driven Decisions

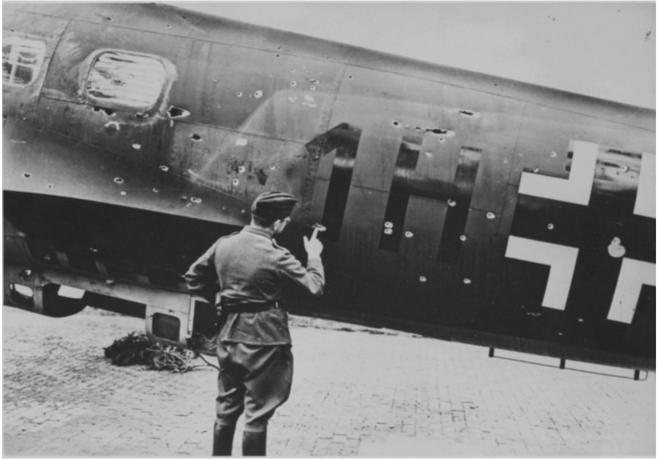


Figure 2: Soldier examing bullet holes on a bomber.

Evgeniy. “The German officer is studying bullet holes on the fuselage of the bomber He.111, returning from combat sortie.” War Thunder, 01/25/2016. http://waralbum.ru/274297

Reveal Human Nature

Figure 3: Monty Hall in 'Let's Make a Deal'.

Frauenfelder, Mark. "Monty Hall." BoingBoing, 10/01/2017. https://boingboing.net/2017/10/01/monty-hall-1921-2017.html. Accessed 05/20/2018

Describe the Universe

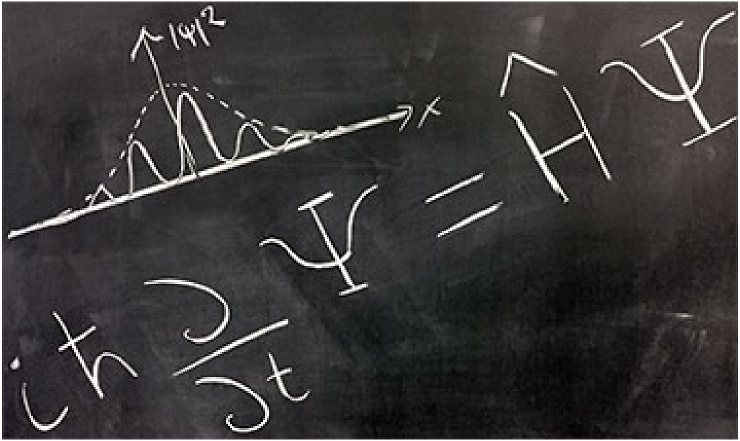


Figure 4: Quantum Mechanics on a blackboard.

edX. "Quantum Mechanics: Quantum physics in 1D Potentials." Online video clip. https://www.youtube.com/watch?time_continue=2&v=0jmW2PeQ-oQ. YouTube. Youtube, 10/02/2017. Web 05/20/2018

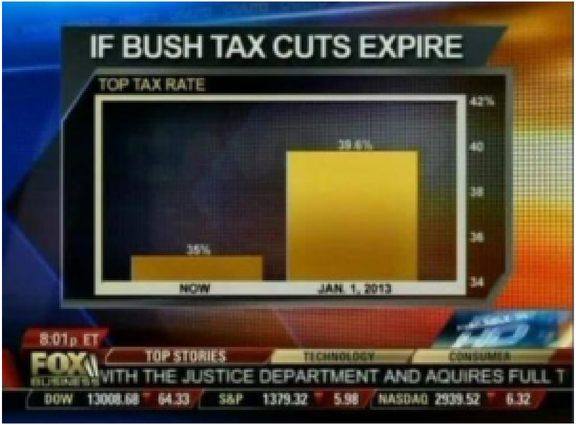


Figure 5: If Bush tax cuts expire.

Robbins, Naomi. "Another Misleading Graph of Romney's Tax Plan." 08/04/2012. https://www.forbes.com/sites/naomirobbins/2012/08/04/another-misleading-graph-of-romneys-tax-plan/#b71ed4133b89. Accessed
05/20/2018

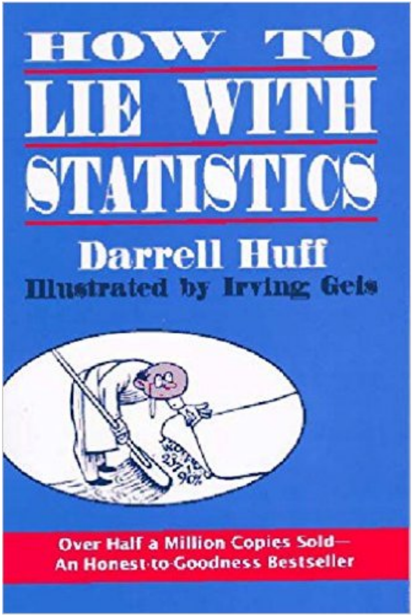

HOW TO
LIE WITH
STATISTICS
Darrell Huff
Illustrated by Irving Geis
PROFITS
23%
90%
Over Half a Million Copies Sold—
An Honest-to-Goodness Bestseller

Figure 6: How to lie with Statistics.

Detecting Fraud

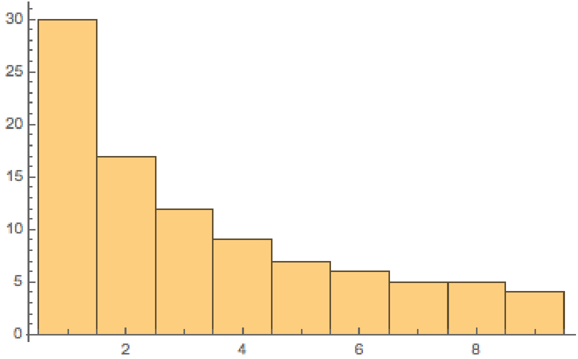


Figure 7: Distribution of of percentage of leading digits.

Two Main Types of Statistics

Descriptive

&

Inferential

Two Main Types of Statistics

Definition (≈Descriptive Statistics)

Describes data, e.g. how one presents data or interesting characteristics: mean, median, min/max, standard deviation, skew, kurtosis, etc..

Definition (Inferential Statistics)

Draws inferences from data, e.g. z-statistic, t-statistic, χ^2, F-ratio, etc..

How does one approach Statistics?

“Statistical Method”

1. Ask a question.
2. Collect data.
3. Analyze the data.
4. Interpret the results.
5. Present the conclusions

“Statistical Method”

1. Ask a question.
2. Collect data.
3. Analyze the data.
4. Interpret the results.
5. Present the conclusions

VS

Scientific Method

1. Observation
2. Question
3. Hypothesis
4. Experiment
5. Analysis
6. Conclusion

Collecting Data

Basic Definitions

Definition (Data)

Collection of observations, e.g. measurements, colors, responses, etc..

Basic Definitions

Definition (Data)

Collection of observations, e.g. measurements, colors, responses, etc..

Definition (Population)

Collection of *all* measurements or data, e.g. data from group or object of study.

Basic Definitions

Definition (Data)

Collection of observations, e.g. measurements, colors, responses, etc..

Definition (Population)

Collection of *all* measurements or data, e.g. data from group or object of study.

Definition (Census)

Collection of data from *every* member of a population.

Basic Definitions

Definition (Data)

Collection of observations, e.g. measurements, colors, responses, etc..

Definition (Population)

Collection of *all* measurements or data, e.g. data from group or object of study.

Definition (Census)

Collection of data from *every* member of a population.

Definition (Sample)

Subcollection of members selected from a population.

Basic Definitions

Definition (Experiment)

The process of applying a procedure and observing its effects.

Basic Definitions

Definition (Experiment)

The process of applying a procedure and observing its effects.

Definition (Experimental Units/Subjects)

The objects of study in an experiment. When the object is an individual/group of individuals, these are called subjects.

Basic Definitions

Definition (Experiment)

The process of applying a procedure and observing its effects.

Definition (Experimental Units/Subjects)

The objects of study in an experiment. When the object is an individual/group of individuals, these are called subjects.

Definition (Observational Study)

An experiment where one observed and measures characteristics but does not attempt to modify the experimental units/subjects.

Types of Sampling Response

Voluntary

&

Involuntary

Definition (Voluntary Response Sample)

A sample where participants decide whether to be included or not. Also called self-selected sample.

Definition (Voluntary Response Sample)

A sample where participants decide whether to be included or not. Also called self-selected sample.

Definition (Involuntary Response Sample)

A sample where the participants were forced to be included.

Example (Voluntary Response Sample)

1. Internet poll
2. Mail-in poll
3. Telephone call-in poll

Example (Voluntary Response Sample)

1. Internet poll
2. Mail-in poll
3. Telephone call-in poll

Example (Involuntary Response Sample)

“Turn in the course evaluation with the final exam or it will not be accepted.”

Types of Sampling

Definition (Random Sample)

A sample where each individual has the same probability of being selected.

Definition (Random Sample)

A sample where each individual has the same probability of being selected.

Definition (Simple Random Sample (SRS))

A sample of a population where every sample of size n has the same probability of being selected.

Random Sample

vs

Simple Random Sample

- Random Samples are about *individuals*.
- Simple Random Samples are about *groups*.

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Random Sample? Simple Random Sample?

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Pawnee:

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee	*Gotham*
Scranton	*Metropolis*

Probability of Pawnee: $\frac{1}{2}$

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee	*Gotham*

Scranton	*Metropolis*

Probability of Pawnee: $\frac{1}{2}$

Probability of Gotham:

Example: Random Sample vs SRS

Flip a coin:

Heads	—	Choose the top row of cities
Tails	—	Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Pawnee: $\frac{1}{2}$

Probability of Gotham: $\frac{1}{2}$

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee	*Gotham*
Scranton	*Metropolis*

Probability of Pawnee:	$\frac{1}{2}$
Probability of Gotham:	$\frac{1}{2}$
Probability of Scranton:	$\frac{1}{2}$
Probability of Metropolis:	$\frac{1}{2}$

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Pawnee: $\frac{1}{2}$

Probability of Gotham: $\frac{1}{2}$

Probability of Scranton: $\frac{1}{2}$

Probability of Metropolis: $\frac{1}{2}$

Therefore, this is a random sample

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Top Row: $\frac{1}{2}$

Probability of Bottom Row: $\frac{1}{2}$

Example: Random Sample vs SRS

Flip a coin:

Heads	—	Choose the top row of cities
Tails	—	Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Pawnee & Gotham: $\frac{1}{2}$

Probability of Scranton & Metropolis: $\frac{1}{2}$

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Pawnee & Gotham: $\frac{1}{2}$

Probability of Scranton & Metropolis: $\frac{1}{2}$

Probability of Scranton & Pawnee: 0

Example: Random Sample vs SRS

Flip a coin:

Heads — Choose the top row of cities
Tails — Choose the bottom row of cities

Pawnee *Gotham*

Scranton *Metropolis*

Probability of Pawnee & Gotham: $\frac{1}{2}$

Probability of Scranton & Metropolis: $\frac{1}{2}$

Probability of Scranton & Pawnee: 0

Therefore, this is not a SRS.

Types of Sampling

Definition (Systematic Sampling)

A sampling where one selects the start point and every k^{th} sample is chosen.

Example (Systematic Sampling)

1. Call every fifth number in the telephone book.
2. Request a response from every 100th visitor to a website.
3. Examine every other car produced in a factory.

Definition (Stratified Sampling)

A sampling where one divides the population into at least two different subgroups (or strata), so that the subjects within each subgroup share some characteristic, and sampling within these subgroups.

Example (Stratified Sampling)

1. Collecting the GPA of students by academic major.
2. Surveying random individuals in six different towns in a county.
3. Polling individuals broken into age brackets.
4. Record the average yearly temperature in each European country.

Definition (Cluster Sampling)

A sampling where one divides the population into clusters (or sections) and then randomly select from those clusters, choosing *all* members from the selected clusters.

Example (Cluster Sampling)

1. Surveying every individual from random households in a city.
2. Quizzing every person in a side of a classroom.
3. Examine all the plants in a collection of 50 ft x 50 ft squares in a large field.

The key difference between stratified and cluster sampling is the subgroup composition. In each case, you want to judge the population as a whole by breaking the population into subgroups and examining a few of these subgroups. You want the final sample to 'look like' the larger population.

Stratified sampling does this by choosing different subgroups, each subgroup consisting of individuals with something feature in common. Though the members within any subgroup will be the same, they will be different from individuals from other subgroups. Putting the samples from each chosen subgroup together, you hopefully have a sample resembling the original population.

Cluster sampling also takes the population and breaks it into groups. But these subgroups should consist of lots of ‘different’ individuals. You choose a few of these subgroups, and sample as many individuals possible (hopefully all) from these chosen subgroups. When you put these samples together, you hopefully have a sample which resembles the original population.

In summation, stratified sampling depends on breaking the population into subgroups so that each of the individuals in any given subgroup are all similar, whereas cluster sampling depends on breaking the population into subgroups where ‘all’ the individuals in any given subgroup are different from each other.

Definition (Convenience Sampling)

Using data that is simple to collect.

Example (Convenience Sampling)

1. Polling each person you pass on the street.
2. Sampling volunteers for a study at a university.
3. Recording the height of all trees on your street.

Definition (Multistage)

A method of data collection that utilizes a variety of sampling methods at various stages.

When is Data Collected?

Definition (Retrospective Study)

Data is collected from past data collections.

Definition (Cross-Sectional Study)

Data is observed, measured, and collected at one point in time.

Definition (Prospective Study)

Data is collected in the future from groups (that share common factors).

Types of Data

Types of Data

Definition (Quantitative Data)

Data consisting of numbers representing counts or measurements. Also called numerical data.

Definition (Qualitative Data)

Data consisting of names or labels. Also called categorical data.

Example (Quantitative Data)

1. The price of gasoline.
2. The average temperature of a region.
3. The number of foxes in a region.

Example (Quantitative Data)

1. The price of gasoline.
2. The average temperature of a region.
3. The number of foxes in a region.

Example (Qualitative Data)

1. The sex of participants in a study.
2. The names of countries.
3. The color of objects.

Remark

If it cannot be represented by a number, then it *must* be qualitative. However, being numerical does not mean that data is qualitative. The numerical values must hold meaning relative to object being measured.

Types of Quantitative Data

Definition (Discrete Data)

Numerical data that is finite or “countable”.

Definition (Continuous Data)

Numerical data such that every number between every two possible data values is possible.

Example (Discrete Data)

1. The number of serves in a tennis match.
2. The number of subscribers to a newspaper.
3. The number of particles in the universe.

Example (Discrete Data)

1. The number of serves in a tennis match.
2. The number of subscribers to a newspaper.
3. The number of particles in the universe.

Example (Continuous Data)

1. Distance measured in meters.
2. Wavelength of light.
3. Total watts of electricity used.

Remark

Whether data is discrete or continuous can depend on the scale or method of measurement.

Remark

Whether data is discrete or continuous can depend on the scale or method of measurement.

Example

Temperature: Measured on a mercury thermometer (continuous) vs measured digital thermometer (discrete).

Remark

Whether data is discrete or continuous can depend on the scale or method of measurement.

Example

Temperature: Measured on a mercury thermometer (continuous) vs measured digital thermometer (discrete).

Example

Money: Average hourly income (discrete) vs GDP of a country (continuous).

Levels of Measurement

Definition (Nominal Data)

Data can only be arranged into categories. Order does not make sense (or is arbitrary relative to the object being ‘measured’).

Example (Nominal Data)

1. Eye color
2. Sex
3. Social Security number
4. Yes/No/Undecided
5. Coded Survey Responses

Definition (Ordinal Data)

Data which can be arranged in an order but differences are meaningless.

Example (Ordinal Data)

1. High school class ranking
2. Time of day (morning, midday, afternoon, night)
3. Socioeconomic status
4. Ranks of colleges
5. Political orientation
6. Likert Scale: strongly disagree, disagree, neutral, agree, strongly agree

Definition (Interval Data)

Data which can be ordered and differences have meaning. There is no natural zero.

Example (Interval Data)

1. Time of day (hours and minutes)
2. Temperature (in °F)
3. The calendar year

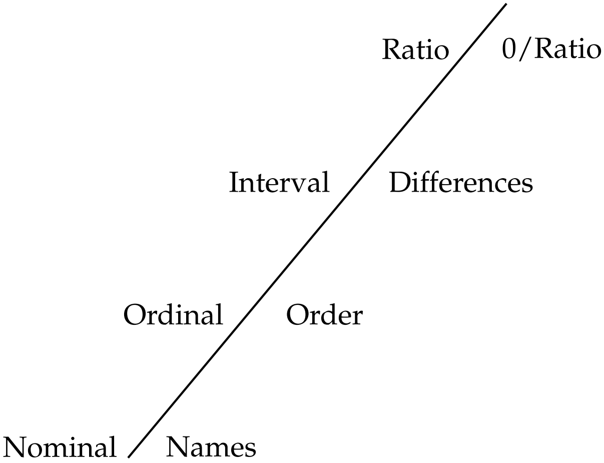
Definition (Ratio)

Data which can be ordered with a natural zero and differences/ratios ‘make sense.’

Example (Ratio Data)

1. Heights
2. Distances
3. Volumes
4. Temperature (Kelvin)

Hierarchy of Measurements



Ratio 0/Ratio

Interval Differences

Ordinal Order

Nominal Names

Why is Sampling & Design so Hard?

Question Design is Difficult

Example (Loaded Questions)

1. Do you believe the President's choice of inexperienced candidate should be confirmed?
2. Do you believe there should be locations to inject illegal drugs?
3. "Should a smack as part of good parental correction be a criminal offence in New Zealand?"[1]

[1] Newshub. "Anti-smacking debate goes to referendum.". 06/15/2009. Accessed 05/21/2018.

Question Design is Difficult

Example (Leading Questions)

1. Do you believe responsible parents should discipline their children?
2. How good was the new film?
3. Do you have problems with your employer?

Question Design is Difficult

Example (Double Barreled Questions)

1. Are you satisfied or dissatisfied are you with your current level of salary and benefits?
2. When was the last time you showered or shaved?
3. Is the new textbook useful for students and teachers?

Question Design is Difficult

Example (Absolutes)

1. Do you believe the President ever lies?
2. Do you always shower in the morning?
3. Should one never terminate a pregnancy?

Question Design is Difficult

Example (Language/Knowledge/Experience Considerations)

1. Should CPS receive more state and/or federal funding?
2. Do you believe individuals convicted of illicit treatment of animals have predilections which may indicate they will participate in future in other felonious behavior?
3. Which suite of cards is your favorite?

Question Design is Difficult

Example (Poor Phrasing/Question Design)

1. Does it seem possible or impossible that the Nazi extermination of the Jews never happened?
2. Did you grow up with a family with a mother and father or only a mother and father?
3. In the past 30 days, when you might have consumed chewing gum, did you share it with anyone?

Costs

- Pay people to design, distribute, collect, and analyze survey/experiment.
- Experiment could involve large and expensive apparatuses.
- Does one offer incentives? [This could create other issues.]

Practical vs Statistical Significance

Definition (Statistical Significance)

When a result is unlikely to have occurred by chance.

Example

ProCare Industries offered Gender Choice to increase chances a baby would be born with a desired sex. It increased the chance to 52% of girls when desired. [Study on 10,000 couples, 5200 had girls. 0.003% chance of this happening.] But this is not practically significant. [Note: actual probability of girl 48.8%.]

Other Important Considerations

- The survey/experiment should actually text the question at hand without bias.

- Replication: The experiment/survey should be able to be repeated.

- Blinding: Should the subjects know the point of the experiment or whether they are receiving treatment? One must be aware of the *placebo effect*. Whenever possible (especially in medical studies), double-blinding should be used.

Double-blinding is when *both* the subjects and the experimenters do not know who is or is not receiving treatment.

Other Important Considerations

- Randomization: Where individuals are assigned to different groups through a process of random selection.
- Confusing Responses/Non-response: What is a survey response is (partially) illegible, confusing, or otherwise problematic? What if some of the responses are missing or incomplete?
- Timeframe: How long will the data collection take? Will it be useful/relevant by then? What are the costs with the timeframe? What if participants drop out or disappear? Will short term data be useful?

Bad Experiment/Survey = Bad Data

Good Experiment/Survey $\neq$ Good Data

Sampling Matters!

Example (Poor Sampling)

1. A liberal news website asks whether visitors are satisfied with the current president.
2. A survey about a women's rights issue that mostly surveyed male individuals or a drug is given to only female subjects and not male subjects.
3. A otherwise well designed survey where participants can see the results as they come in.
4. An internet poll about computer usage.
5. A survey about health where individuals are asked about their weight.

Remark

Even in a well designed and executed study, the data and analysis may not be sufficient.

Definition (Confounding Variable)

A variable that influences the dependent and independent variable, causing an association. Also called a lurking or hidden variable.

Example

1. Birth order and Presence of Down's Syndrome: Here a hidden variable is the mother's age.
2. Exercise and Weight Gain by sex: Possible confounding variables are age, amount eaten, occupation, etc..
3. Murder rate and ice cream consumption: Possible confounding variable is weather.

They are also important for understanding jokes…

`https://www.youtube.com/watch?v=e8StK-vIkBs`

Remark

One must be careful when examining and interpreting data.

Correlation $\neq$ Causation

EXAMPLE

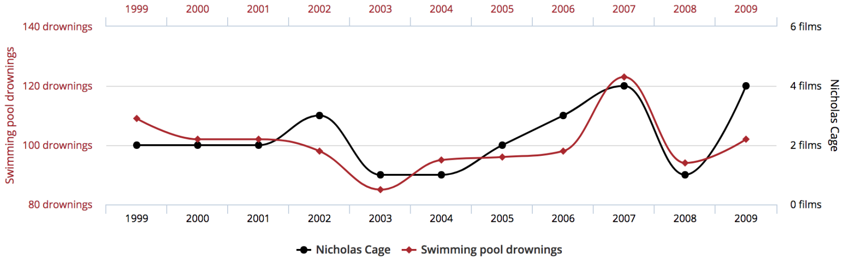


Figure 8: Number of people who drowned by falling into a pool correlates with films Nicholas Cage appeared in ($r = 0.666004$).

T. Vigen. *Spurious Correlations*. http://www.tylervigen.com/spurious-correlations. Accessed 05/21/2018

Example

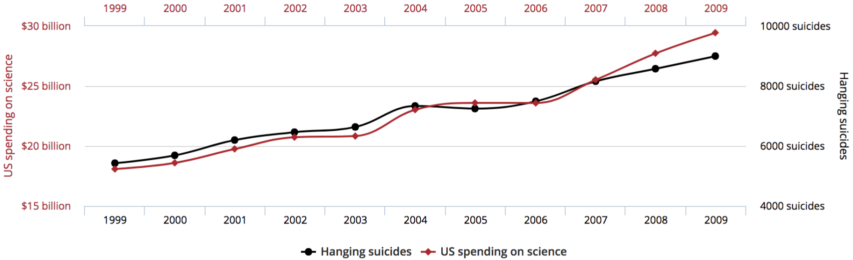


Figure 9: US spending on science, space, and technology correlates with suicides by hanging, strangulation, and suffocation ($r = 0.99789126$).

T. Vigen. *Spurious Correlations*. http://www.tylervigen.com/spurious-correlations. Accessed 05/21/2018

EXAMPLE

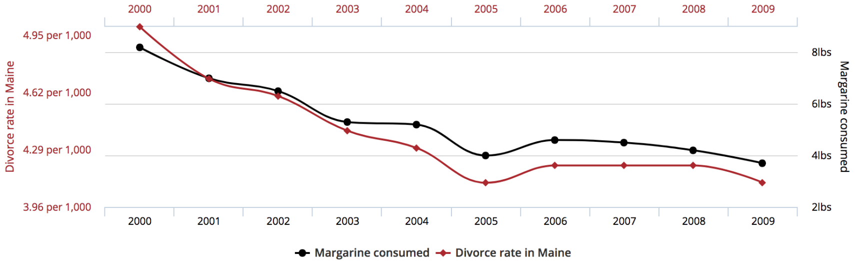


Figure 10: Divorce rate in Maine correlates with per capita consumption of margarine ($r = 0.992558$).

T. Vigen. *Spurious Correlations*. http://www.tylervigen.com/spurious-correlations. Accessed 05/21/2018

EXAMPLE

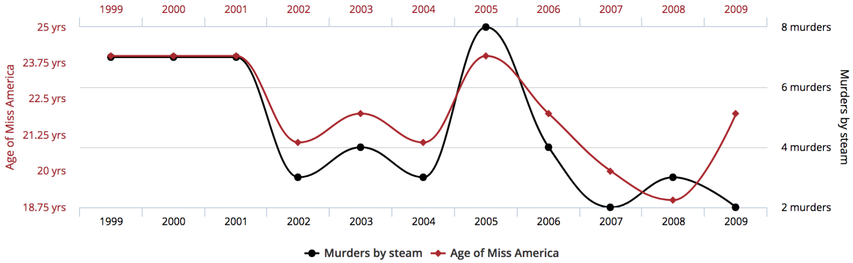


Figure 11: Age of Miss America correlates with murders by steam, hot vapors, and hot objects ($r = 0.870127$).

T. Vigen. *Spurious Correlations*. http://www.tylervigen.com/spurious-correlations. Accessed 05/21/2018

Example

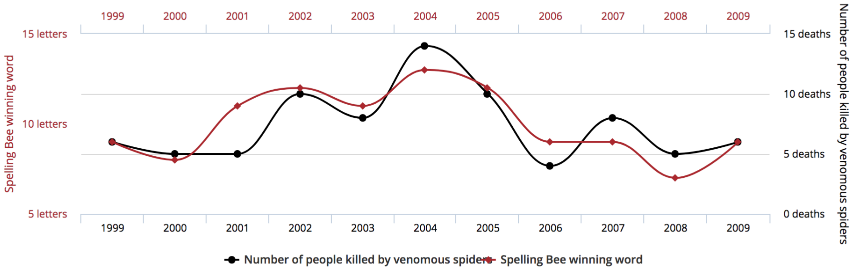


Figure 12: Letters in winning word of Scripps National Spelling Bee correlates with number of people killed by venomous spiders ($r = 0.8057$).

T. Vigen. *Spurious Correlations*. http://www.tylervigen.com/spurious-correlations. Accessed 05/21/2018

Example (More Seriously)

Example

At young ages, there is a strong correlation between IQ and astrological sign. [This correlation does not last into later childhood. Why?]

Chapter 2

Frequency Distributions & Histograms
Graphing Data & Regressions

Characteristics of Data

What characteristics of data are we interested in?

- Center: The “middle” of the data.

Characteristics of Data

What characteristics of data are we interested in?

- Center: The “middle” of the data.
- Variation: How “different” are the data values?

Characteristics of Data

What characteristics of data are we interested in?

- Center: The “middle” of the data.
- Variation: How “different” are the data values?
- Distribution: What is the “shape” of the data?

Characteristics of Data

What characteristics of data are we interested in?

- Center: The “middle” of the data.
- Variation: How “different” are the data values?
- Distribution: What is the “shape” of the data?
- Outliers: Are there data values which are “very different.”

Characteristics of Data

What characteristics of data are we interested in?

- Center: The “middle” of the data.
- Variation: How “different” are the data values?
- Distribution: What is the “shape” of the data?
- Outliers: Are there data values which are “very different.”
- Time: How does the data vary with time?

In order to investigate center, variation, distribution, outliers, and time factors of data, we will need to organize and visualize the data.

The first step will be to group data values.

Definition (Frequency Distribution)

A frequency distribution shows how data are partitioned among several categories (or classes) by listing the categories along with the number (frequency) of data values in each of them.

Example

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Frequency Distribution Terms

Definition (Lower Class Limit)

The smallest number in each class.

Definition (Upper Class Limit)

The largest number in each class.

Definition (Class Boundaries)

The “middle” number between classes.

Definition (Class Midpoints)

The middle value of each class.

Definition (Class Width)

The length of each class, i.e. the difference between the upper and lower class boundaries.

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Classes	90–100	80–89	70–79	60–69	50–59	0–49

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Classes	90–100	80–89	70–79	60–69	50–59	0–49
Lower Class Limit	90	80	70	60	50	0

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Classes	90–100	80–89	70–79	60–69	50–59	0–49
Lower Class Limit	90	80	70	60	50	0
Upper Class Limit	100	89	79	69	59	49

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Classes	90–100	80–89	70–79	60–69	50–59	0–49
Lower Class Limit	90	80	70	60	50	0
Upper Class Limit	100	89	79	69	59	49
Class Midpoints	95	84.5	74.5	64.5	54.5	24.5

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Classes	90–100	80–89	70–79	60–69	50–59	0–49
Lower Class Limit	90	80	70	60	50	0
Upper Class Limit	100	89	79	69	59	49
Class Midpoints	95	84.5	74.5	64.5	54.5	24.5
Class Boundaries	89.5, 100.5	79.5, 89.5	69.5, 79.5	59.5, 69.5	49.5, 59.5	−0.5, 49.5

Table 1: Exam scores on a Calculus II exam

Grade	Frequency
90–100	4
80–89	3
70–79	6
60–69	3
50–59	2
0–49	0

Classes	90–100	80–89	70–79	60–69	50–59	0–49
Lower Class Limit	90	80	70	60	50	0
Upper Class Limit	100	89	79	69	59	49
Class Midpoints	95	84.5	74.5	64.5	54.5	24.5
Class Boundaries	89.5, 100.5	79.5, 89.5	69.5, 79.5	59.5, 69.5	49.5, 59.5	−0.5, 49.5
Class Width	11	10	10	10	10	50

A Few Cautions

Remark

The class width is *not* the difference of the upper and lower class limits. It is the difference of the upper and lower class boundaries!

A Few Cautions

Remark

If the adjacent class limits are equal, then the boundaries are the same. [There is a problem with this. Why?]

A Few Cautions

Remark

If the adjacent class limits are equal, then the boundaries are the same. [There is a problem with this. Why?]

Example

Value	100–150	150–200	200–250
Frequency	5	19	7

Class	100–150	150–200	200–250
Class Boundaries	100, 150	150, 200	200, 250
Class Width	?	?	?

How to Construct a Frequency Distribution

How to Construct a Frequency Distribution

1. Select the number of classes — usually between 5 and 20.[2]

2. Calculate the class width $= \dfrac{\text{Max Value } - \text{ Min Value}}{\text{Number of classes}}$. [Rounding up, if necessary.]

3. Choose a value for first lower class limit, either using the min or another ‘appropriate’ value.

4. Construct the other class limits.

5. Calculate the frequency in each class.

[2] “Sturges” Guideline: Number of classes $= 1 + \log_2(n)$, where n is the number of data values.

Example

Data Values

59	827	920
217	1112	616
601	700	980
299	827	946

Example

1. **Decide on the number of classes:** We shall use 4 classes.

Example

1. **Decide on the number of classes:** We shall use 4 classes.
2. **Calculate the class width:** Max Value=1112, Min Value=59

Example

1. **Decide on the number of classes:** We shall use 4 classes.

2. **Calculate the class width:** Max Value=1112, Min Value=59

$$\text{Class width} \approx \frac{1112 - 59}{4} = \frac{1053}{4} = 263.25 \approx 300$$

   We shall then use a class width of 300.

Example

1. **Decide on the number of classes:** We shall use 4 classes.

2. **Calculate the class width:** Max Value=1112, Min Value=59

$$\text{Class width} \approx \frac{1112 - 59}{4} = \frac{1053}{4} = 263.25 \approx 300$$

   We shall then use a class width of 300.

3. **Choose a starting point:** A starting value of 0 seems appropriate. We shall use 0.

Example

1. **Decide on the number of classes:** We shall use 4 classes.
2. **Calculate the class width:** Max Value=1112, Min Value=59

   $$\text{Class width} \approx \frac{1112 - 59}{4} = \frac{1053}{4} = 263.25 \approx 300$$

   We shall then use a class width of 300.
3. **Choose a starting point:** A starting value of 0 seems appropriate. We shall use 0.
4. **Calculate the lower class limits:** 0, 300, 600, 900

5. **Count the frequencies and construct the table:**

Classes	0–	300–	600–	900–
Frequency				

5. **Count the frequencies and construct the table:**

Classes	0–299	300–599	600–899	900–1199

5. **Count the frequencies and construct the table:**

Classes	0–299	300–599	600–899	900–1199
Frequency	3	0	5	4

But what if the data is not quantitative?

Definition (Relative Frequency)

$$\text{Relative Frequency } = \frac{\text{frequency for class}}{\text{sum of frequencies}}$$

Example

Table: Samples of 23 species of gilled mushrooms in the Agaricus and Lepoita Family.

Cap Shape	Bell	Conical	Convex	Flat	Knobbed	Sunken	Total
Frequency	452	4	3656	3152	828	32	8124

Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Example

What is the relative frequency for convex capped mushrooms?

Convex: 3656
Total: 8124

Example

What is the relative frequency for convex capped mushrooms?

Convex: 3656
Total: 8124

$$\text{Convex Relative Frequency } = \frac{3656}{8124} \approx 0.450$$

Example

What is the relative frequency for convex capped mushrooms?

Convex: 3656
Total: 8124

$$\text{Convex Relative Frequency } = \frac{3656}{8124} \approx 0.450$$

What is the relative frequency for bell capped mushrooms?

Bell: 452
Total: 8124

Example

What is the relative frequency for convex capped mushrooms?

Convex: 3656
Total: 8124

$$\text{Convex Relative Frequency } = \frac{3656}{8124} \approx 0.450$$

What is the relative frequency for bell capped mushrooms?

Bell: 452
Total: 8124

$$\text{Bell Relative Frequency } = \frac{452}{8124} \approx 0.056$$

Example

Table: Samples of 23 species of gilled mushrooms in the Agaricus and Lepoita Family.

Cap Shape	Bell	Conical	Convex	Flat	Knobbed	Sunken	Total
Frequency	452	4	3656	3152	828	32	8124
Relative Frequency	0.056	0.000	0.450	0.388	0.102	0.004	1.000
Percentage	5.6%	0%	45%	38.8%	10.2%	0.4%	100%

Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

This can also be done with quantitative data.

Table 2: Frequency of heights for 100 U.S. males between 20–29

Height	Number	Relative Frequency
Under 5′3″	1	1%
5′4	3	3%
5′5″	4	4%
5′6″	6	6%
5′7″	7	7%
5′8″	12	12%
5′9″	17	17%
5′10″	17	17%
5′11″	12	12 %
6′	7	7%
6′1″	6	6%
6′2″	4	4%
6′3″	3	3%
6′4″ and Up	1	1%
Total	100	100%

Remark

The sum of the relative frequencies should be “close” to 100% but need not be exactly 100% (due to rounding).

Definition (Cumulative Frequency Distribution)

A frequency distribution in which the frequency in each class is the sum of the frequencies for the previous classes and that class.

Table: Frequency of heights for 100 U.S. males between 20–29

Height	Number	Relative Frequency	Cumulative Frequency
Under 5′3″	1	1%	1%
5′4	3	3%	4%
5′5″	4	4%	8%
5′6″	6	6%	14%
5′7″	7	7%	21%
5′8″	12	12%	33%
5′9″	17	17%	50%
5′10″	17	17%	67%
5′11″	12	12 %	79%
6′	7	7%	86%
6′1″	6	6%	92%
6′2″	4	4%	96%
6′3″	3	3%	99%
6′4″ and Up	1	1%	100%
Total	100	100%	100%

Frequencies were made to be visualized!

Remark

All data (whenever possible) was meant to be visualized. One should *always* try to plot data before performing any statistical analyses.

Definition (Histogram)

A graph consisting of bars of equal width, drawn adjacent to each other (unless there are gaps in the data). The horizontal scale represents the classes or quantitative data and the vertical axes the frequency.

Example

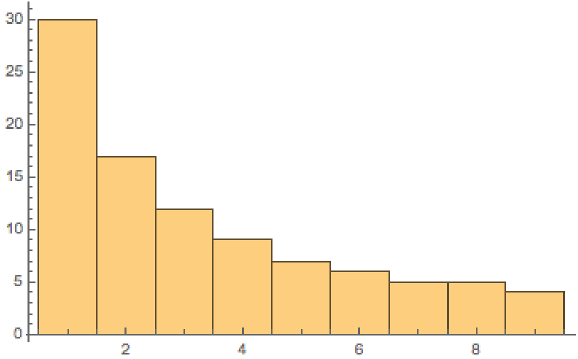


Figure 13: Distribution of of percentage of leading digits.

Advantages of a Histogram

- Visually displays the *shape/distribution* of the data.

Advantages of a Histogram

- Visually displays the *shape/distribution* of the data.
- Visually displays the *center* of the data.

Advantages of a Histogram

- Visually displays the *shape/distribution* of the data.
- Visually displays the *center* of the data.
- Visually displays the *spread* of the data.

Advantages of a Histogram

- Visually displays the *shape/distribution* of the data.
- Visually displays the *center* of the data.
- Visually displays the *spread* of the data.
- Visually displays any *outliers*.

Common Distribution Shapes

Uniform Distribution

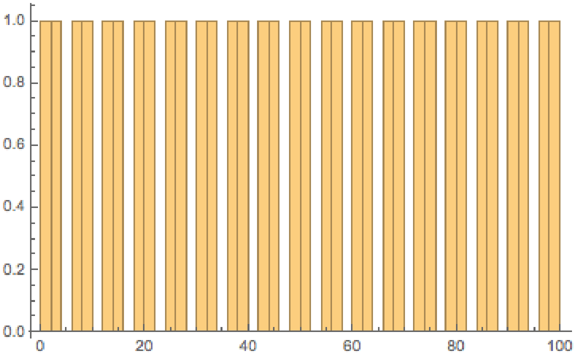


Figure 14: Uniform Distribution.

Uniform Distribution

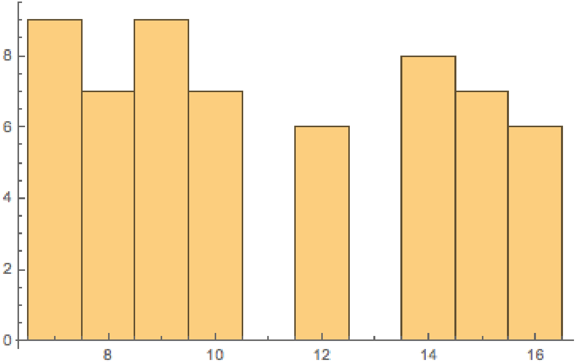


Figure 15: Uniform Distribution.

Normal (Bell/Gaussian) Distribution

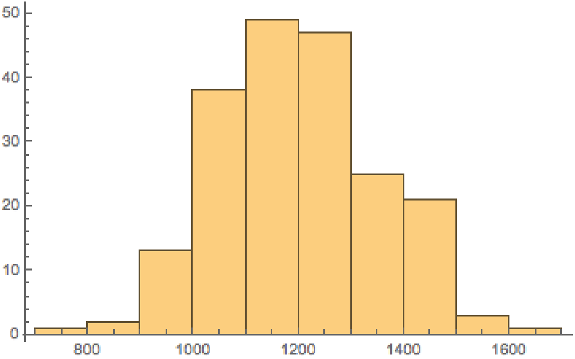


Figure 16: Normal Distribution.

Normal (Bell/Gaussian) Distribution

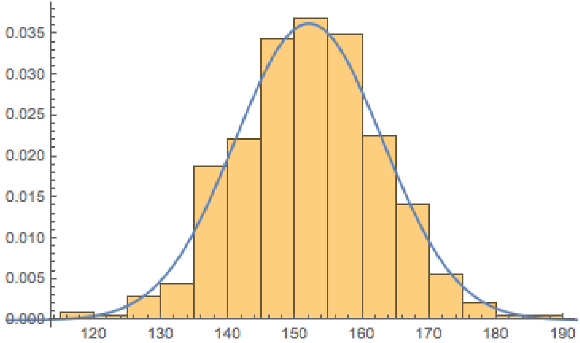


Figure 17: Normal Distribution.

Right Skew

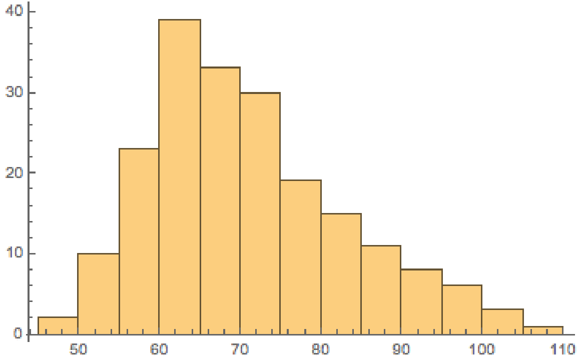


Figure 18: Right Skewed Distribution.

Left Skew

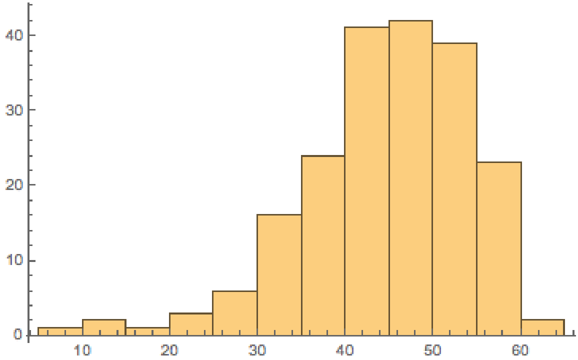


Figure 19: Left Skewed Distribution.

Some Distributions are Naturally ‘Skewed’

Chi-Squared (χ^2) Distribution

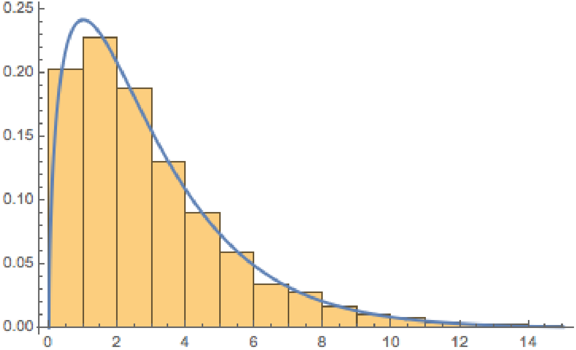


Figure 20: Chi Squared Distribtion.

Beta Distribution

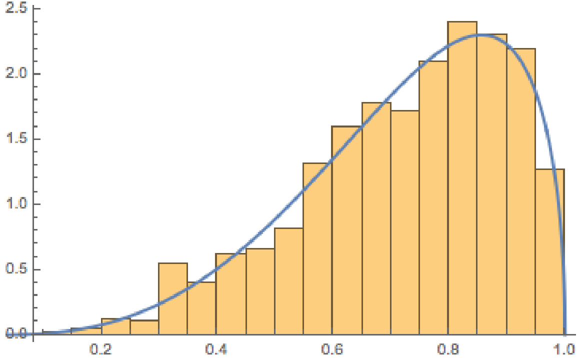
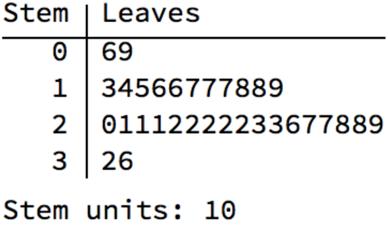


Figure 21: Beta Distribution.

Other Plots

Stem-and-Leaf Plot



```
Stem | Leaves
-----+-------------------
   0 | 69
   1 | 34566777889
   2 | 01112222233677889
   3 | 26

Stem units: 10
```

Figure 22: Stem-and-Leaf Plot.

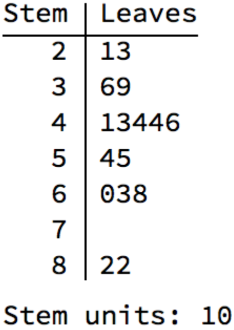
Stem-and-Leaf Plot

1. Choose a stem.
2. Write out the leaves in order. Values which repeat should be repeat in the plot.
3. If a stem should have no leaves, the line should be left blank but included.

Construct a Stem-and-Leaf Plot

41	43	55	44
46	21	63	60
44	36	23	54
82	39	68	82

Stem-and-Leaf Plot



Stem	Leaves
2	13
3	69
4	13446
5	45
6	038
7	
8	22

Stem units: 10

Figure 23: Stem-and-Leaf Plot.

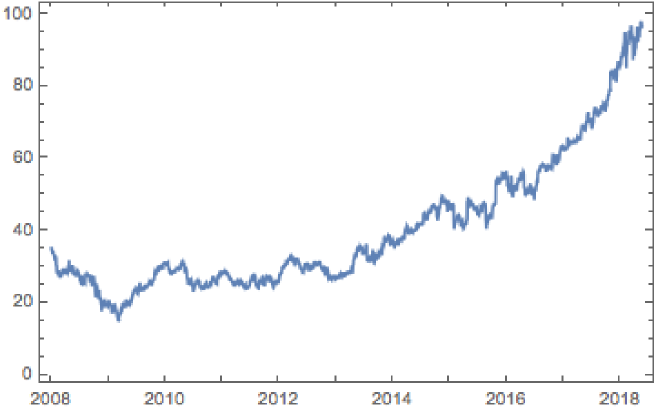


Figure 24: Time Series for Microsoft's Stock since 2008.

Pie Chart

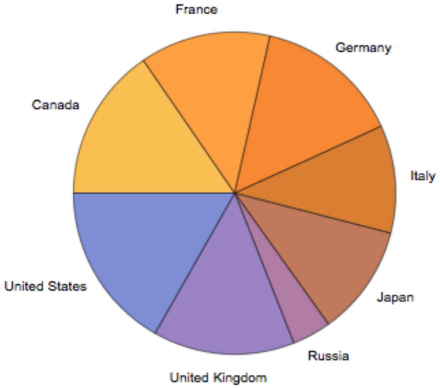


Figure 25: GDP per capita.

SCATTERPLOT

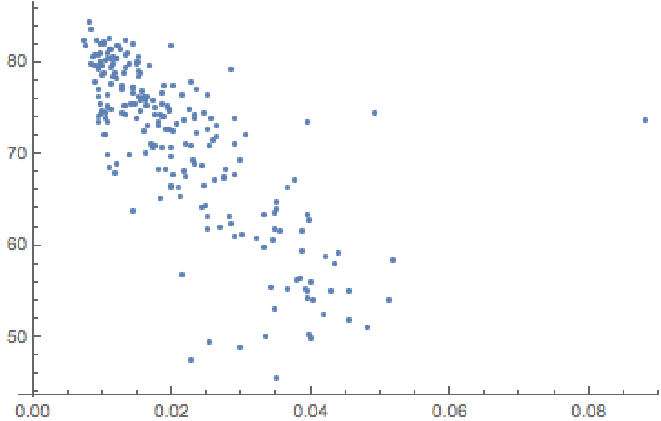


Figure 26: Life expectancy against birth rates.

Scatterplots actually tell you more!

Definition (Correlation)

When two more more variables are somehow associated with one another.

Definition (Correlation)

When two more more variables are somehow associated with one another.

Definition (Linear Correlation)

When two variables are correlated and the plotted pairs of points result in a pattern that can be approximated by a straight line.

CORRELATION DOES NOT IMPLY CAUSALITY!

Linear Regression

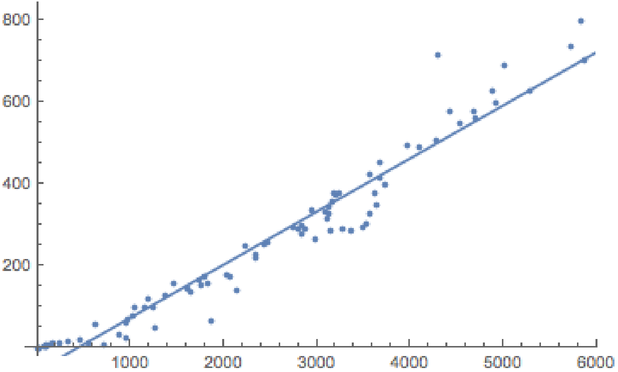


Figure 27: Vaporization heat plotted against enthalpy.

Definition (Correlation Coefficient)

r, measures the strength and direction of a linear relationship.

Remark

- Negative correlation, negative slope
- Positive correlation, positive slope
- $-1 \leq r \leq 1$
- The closer to -1 or 1, the stronger the linear relationship
- r^2 measures the "percent linearity."

Definition (p-value)

The probability that getting a paired sample with a linear correlation coefficient r that is at least as extreme as the one obtained from the paired sample.

Linear Regression

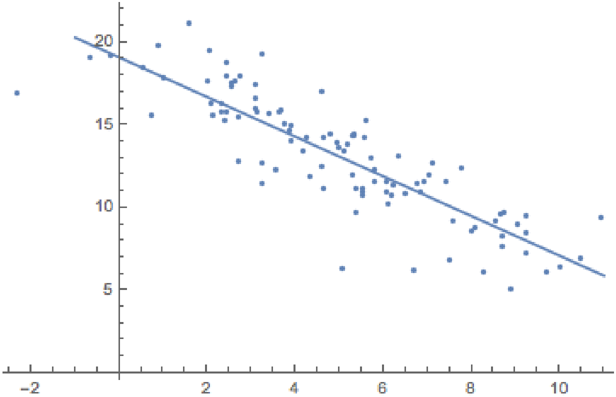


Figure 28: $r = -0.892$, $r^2 = 0.796$, $p \approx 0$

Linear Regression

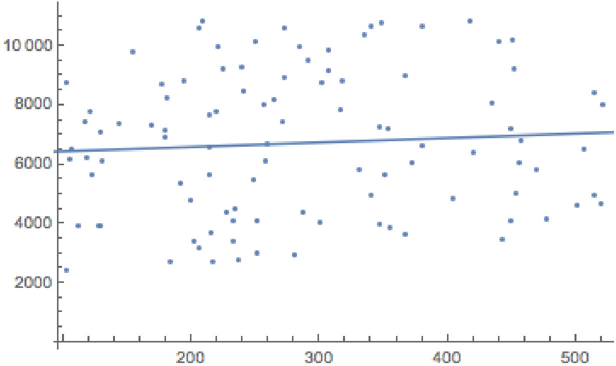


Figure 29: $r = 0.0734$, $r^2 = 0.005$, $p \approx 0.467$

Graphs Should Lead, Not Mislead

Example

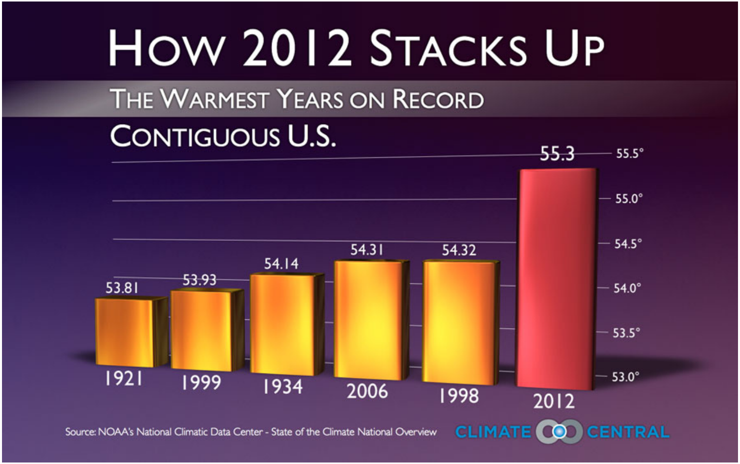


Figure 30: Climate data.

A. Gelman, "How 2012 stacks up: The worst graph on record?". 01/08/2013. http://themonkeycage.org/2013/01/how-2012-stacks-up-the-worst-graph-on-record/. Accessed 05/22/2018

Example

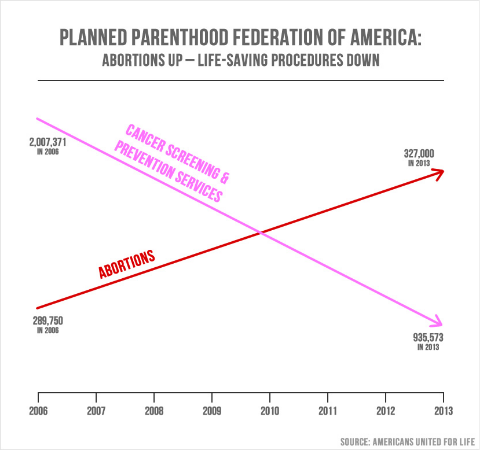


Figure 31: Planned Parenthood services.

L. Qiu, “Chart shown at Planned Parenthood hearing is misleading and ’ethically wrong”’. 10/01/2015. http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/. Accessed
05/22/2018

Example (Corrected)

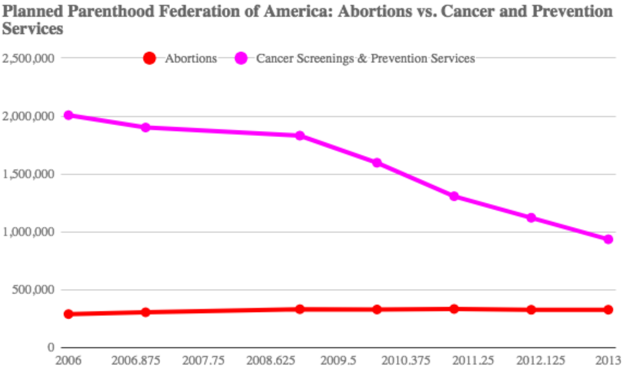


Figure 32: Planned Parenthood services.

L. Qiu, "Chart shown at Planned Parenthood hearing is misleading and 'ethically wrong"'. 10/01/2015. http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/. Accessed 05/22/2018

Example

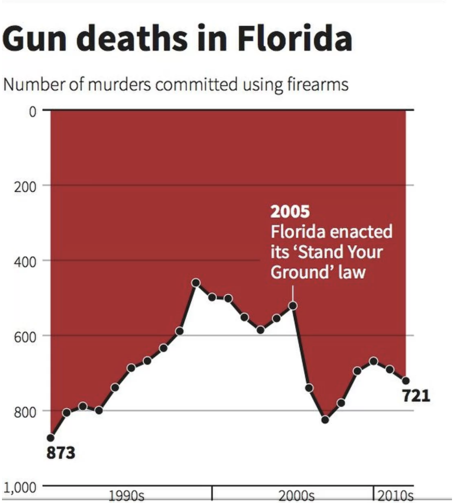


Figure 33: Stand your ground (Source: Florida Department of Law Enforcement).

M. Lallanilla. "Misleading Gun-Death Chart Draws Fire". 04/23/2014. https://www.livescience.com/45083-misleading-gun-death-chart.html. Accessed 05/22/2018

EXAMPLE (CORRECTED)

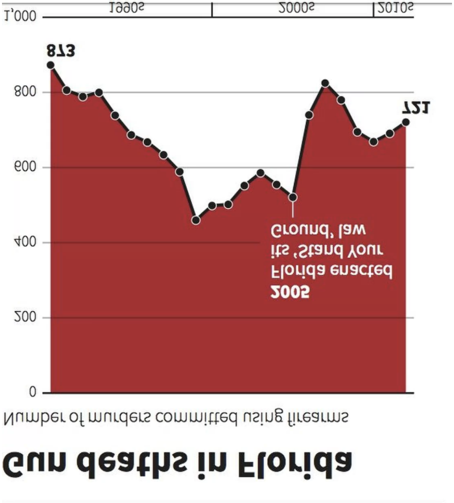
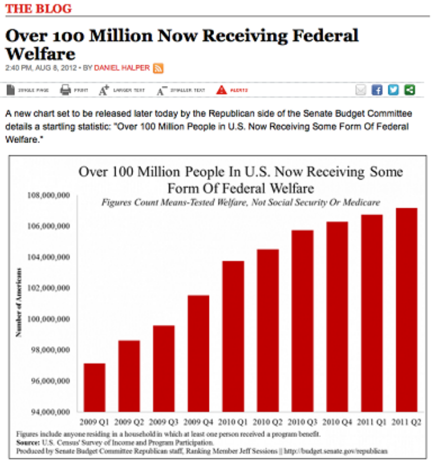


Figure 34: Stand your ground (Source: Florida Department of Law Enforcement).

M. Lallanilla. “Misleading Gun-Death Chart Draws Fire”. 04/23/2014. https://www.livescience.com/45083-misleading-gun-death-chart.html. Accessed 05/22/2018

Example



THE BLOG

Over 100 Million Now Receiving Federal Welfare

2:40 PM, AUG 8, 2012 • BY DANIEL HALPER

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."

Over 100 Million People In U.S. Now Receiving Some Form Of Federal Welfare

Figures Count Means-Tested Welfare, Not Social Security Or Medicare

Figures include anyone residing in a household in which at least one person received a program benefit.

Source: U.S. Census' Survey of Income and Program Participation.

Produced by Senate Budget Committee Republican staff, Ranking Member Jeff Sessions || http://budget.senate.gov/republican

Figure 35: Welfare.

Stephanie. "Misleading Graphs". 01/24/2014. http://www.statisticshowto.com/misleading-graphs/. Accessed 05/22/2018.

EXAMPLE

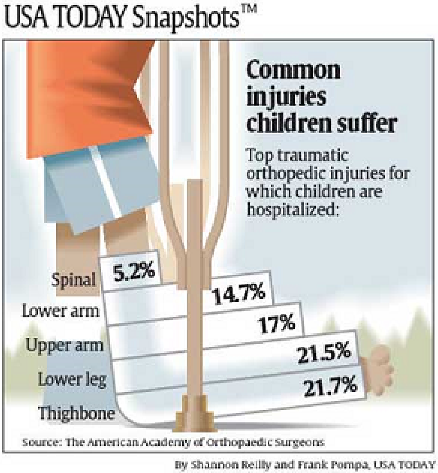


Figure 36: Common injuries.

Stephanie. “Misleading Graphs”. 01/24/2014. http://www.statisticshowto.com/misleading-graphs/. Accessed 05/22/2018.

Example

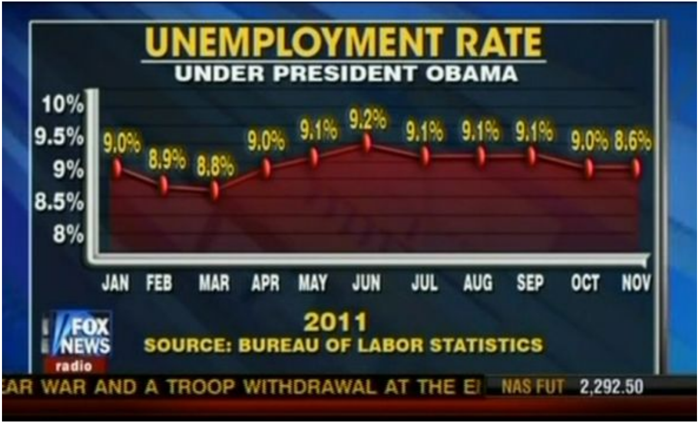


Figure 37: Unemployment rate.

Stephanie. “Misleading Graphs”. 01/24/2014. http://www.statisticshowto.com/misleading-graphs/. Accessed 05/22/2018.

Example (Corrected)

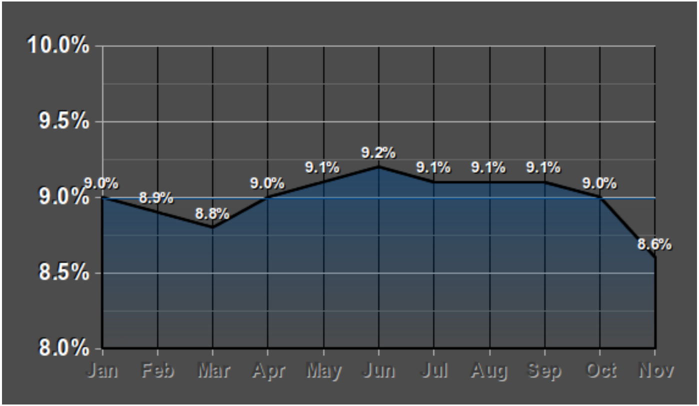


Figure 38: Unemployment rate.

Stephanie. "Misleading Graphs". 01/24/2014. http://www.statisticshowto.com/misleading-graphs/. Accessed 05/22/2018.

Example

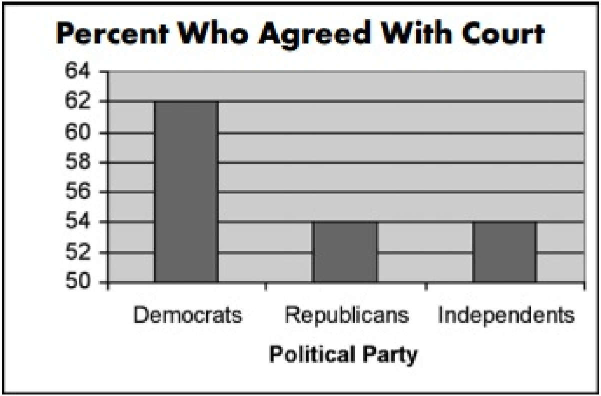


Figure 39: Should Terry Schiavo be removed from life support?

Stephanie. “Misleading Graphs”. 01/24/2014. http://www.statisticshowto.com/misleading-graphs/. Accessed 05/22/2018.

Example

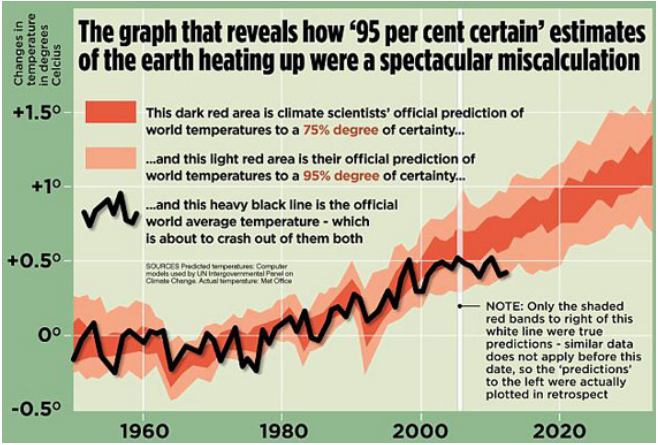


Figure 40: Global Warming.

Stephanie. "Misleading Graphs". 01/24/2014. http://www.statisticshowto.com/misleading-graphs/. Accessed 05/22/2018.

Chapter 3

Measures of Center & Variability
Boxplots

We want to explore ‘center’ and ‘variation’ in data

Notions of center

Definition (Midrange)

A value midway between the minimum and maximum value of a data set: $\dfrac{\text{min} + \text{max}}{2}$

Definition (Median)

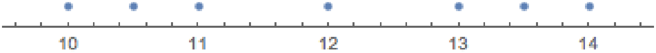
The ‘middle’ of a data set.

Definition (Mean)

The ‘average’ value of a data set: $\dfrac{\sum x_i}{n}$

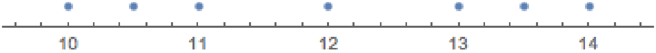
Midrange

10, 10.5, 11, 12, 13, 13.5, 14,



Midrange

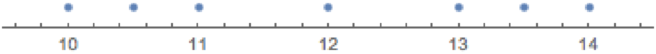
10, 10.5, 11, 12, 13, 13.5, 14,



Midrange =

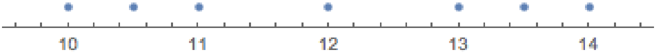
Midrange

10, 10.5, 11, 12, 13, 13.5, 14,



$$\text{Midrange } = \frac{10+14}{2} = \frac{24}{2} = 12$$

MIDRANGE



Notice for this dataset, the midrange does indeed give us the center we would imagine.

Midrange

However, there is a problem with midrange!

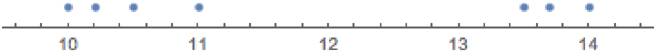
Midrange

However, there is a problem with midrange!

The midrange only consider the highest & lowest data points. Thus, the midrange 'ignored' how the data is actually distributed. This could lead to bad measures of center.

Midrange

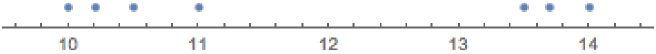
10, 10.2, 10.5, 11, 13.5, 13.7, 14



Midrange =

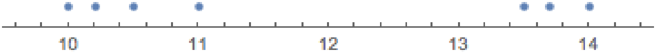
Midrange

10, 10.2, 10.5, 11, 13.5, 13.7, 14



$$\text{Midrange} = \frac{10+14}{2} = \frac{24}{2} = 12$$

Midrange



Notice for this dataset, the midrange does not produce a 'good' notion of center. There are no values in the data set anywhere near the midrange of 12! [But it still is in the center…take that how you will.]

Midrange

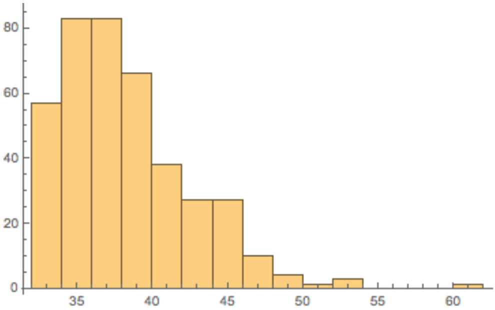
In the previous example, the midrange gave a ‘bad’ measure of center — there were no data values near the midrange value.

Midrange

In the previous example, the midrange gave a ‘bad’ measure of center — there were no data values near the midrange value.

There are other issues when we pair this with variability or measure outliers, as we shall discuss. However, even if there are data values near the midrange, the midrange can still be a ‘bad’ measure of center.

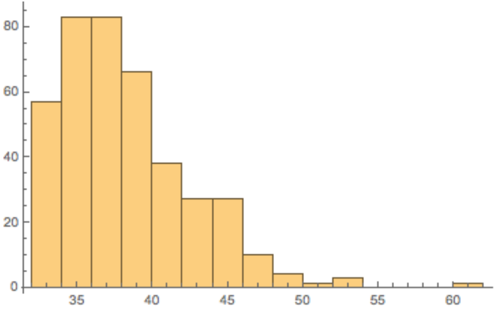
Midrange



Min: 32.2 Max: 61.8

$$\text{Midrange} = \frac{32.2 + 61.8}{2} = 47.0$$

Midrange



Here, the midrange is 47.0, which is too far right.

Midrange

So what is meant by center?

Midrange

So what is meant by center?

We want the word ‘center’ for us to really indicate more than just the ‘middle’ of the data but also tell us about what a ‘typical’ value in the dataset is, i.e. the ‘average’ (not the same thing as mean) data value.

Midrange

So what is meant by center?

We want the word ‘center’ for us to really indicate more than just the ‘middle’ of the data but also tell us about what a ‘typical’ value in the dataset is, i.e. the ‘average’ (not the same thing as mean) data value.

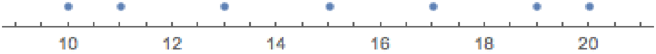
The midrange certain fails us in this second aspect.

Median

The median will produce a better notion of center for us.

Median

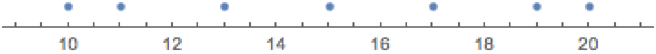
10, 11, 13, 15, 17, 19, 20



Median:

Median

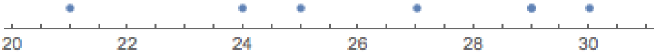
10, 11, 13, 15, 17, 19, 20



Median: 15

Median

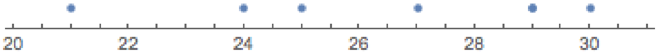
21, 24, 25, 27, 29, 29, 30



Median:

Median

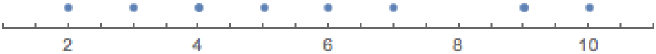
$$21, 24, 25, 27, 29, 29, 30$$



Median: 27

Median

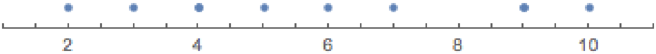
$$2, 3, 4, 5, 6, 7, 9, 9, 10$$



Median:

MEDIAN

$$2, 3, 4, 5, 6, 7, 9, 9, 10$$



Median: $\dfrac{5+6}{2} = 5.5$

Median

Notice the general rule here to find which number is the median. If the total amount of numbers is…

…even, then average the $n/2$th number with the $n/2+1$th number.

…odd, then the median is the $(n+1)/2$th number.

Median

Notice the general rule here to find which number is the median. If the total amount of numbers is…

…even, then average the $n/2$th number with the $n/2+1$th number.

…odd, then the median is the $(n+1)/2$th number.

You can always derive these formulas by considering $n = 2$ with the numbers 1, 2, 3 and $n = 4$ with the numbers 1, 2, 3, 4.

The median produces a better measure of center. Moreover, it is a resistant measure of center.

Definition (Resistant)

We say a measure is resistant if the presence of outliers does not change the measure ‘much’.

The midrange is *not* a resistant measure of center.

EXAMPLE

$$1, 2, 3, 4, 5$$

Midrange: $\dfrac{1+5}{2} = \dfrac{6}{2} = 3$

Example

$$1, 2, 3, 4, 5$$

Midrange: $\dfrac{1+5}{2} = \dfrac{6}{2} = 3$

Now we add an outlier of 11.

$$1, 2, 3, 4, 5, 11$$

Midrange: $\dfrac{1+11}{2} = \dfrac{12}{2} = 6$

Example

$$1, 2, 3, 4, 5$$

Midrange: $\dfrac{1+5}{2} = \dfrac{6}{2} = 3$

Now we add an outlier of 11.

$$1, 2, 3, 4, 5, 11$$

Midrange: $\dfrac{1+11}{2} = \dfrac{12}{2} = 6$

The midrange has nearly doubled, even though only a single outlier was added.

The median *is* a resistant measure of center.

The median *is* a resistant measure of center. In fact, it is very resistant. [The more data values, the more resistant it is!]

Example

Original Data Set

1, 2, 3, 4, 5, 6   Median: 3.5

Example

Original Data Set

1, 2, 3, 4, 5, 6   Median: 3.5

Data Set	Median
1,2,3,4,5,6, 10 billion	4
1,2,3,4,5,6, 10 billion, 11 billion	4.5
1,2,3,4,5,6, 10 billion, 11 billion, 12 billion	5

Example

Original Data Set

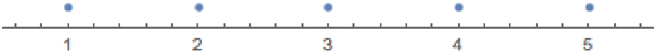
1, 2, 3, 4, 5, 6   Median: 3.5

Data Set	Median
1,2,3,4,5,6, 10 billion	4
1,2,3,4,5,6, 10 billion, 11 billion	4.5
1,2,3,4,5,6, 10 billion, 11 billion, 12 billion	5

Despite adding 3 ‘stupidly’ big outliers, the median has barely changed! In fact, we would have to do this 3 more times to see a large change!

Mean

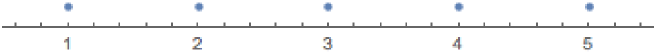
$$1, 2, 3, 4, 5$$



Mean:

Mean

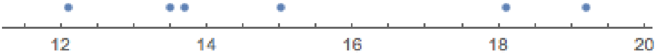
$$1, 2, 3, 4, 5$$



Mean: $\dfrac{1+2+3+4+5}{5} = \dfrac{15}{5} = 3$

Example

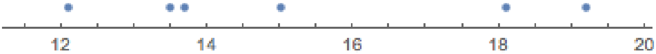
12.1, 13.5, 13.7, 15.0, 18.1, 19.2



Mean:

Example

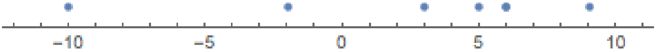
$$12.1, 13.5, 13.7, 15.0, 18.1, 19.2$$



$$\text{Mean: } \frac{12.1 + 13.5 + 13.7 + 15.0 + 18.1 + 19.2}{6} = \frac{91.6}{6} = 15.27$$

Example

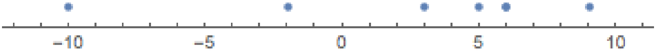
$$-10, -2, 3, 5, 6, 6, 9$$



Mean:

Example

$$-10, -2, 3, 5, 6, 6, 9$$



$$\text{Mean: } \frac{-10 + (-2) + 3 + 5 + 6 + 6 + 9}{7} = 2.43$$

Remark

Always use the word mean, not average. Average can mean many things, e.g. ‘average’ meaning typical. To avoid confusion, always use ‘mean’.

The mean is *not* a resistant measure of center.

Example

$$1, 2, 3, 4, 5$$

Mean: 3

Example

$$1, 2, 3, 4, 5$$

Mean: 3

$$1, 2, 3, 4, 5, 25$$

Mean: $\dfrac{1+2+3+4+5+25}{6} = \dfrac{40}{6} = 6.67$

There is only one outlier yet the mean has nearly doubled.

Resistant Measures

Measure	Resistant
Midrange	✗
Median	✓
Mean	✗

Remark

Note that if we want to use the mean as our measure of center, we actually want to be able to calculate the mean. We cannot do this with a frequency table alone. However, we can still estimate the mean.

Example

Class	Frequency
1–5	3
6–10	1
11–15	2
16–20	2

Example

Class	Frequency
1–5	3
6–10	1
11–15	2
16–20	2

We do not know the values, just the number of values which occur in certain intervals. But we can estimate the mean by estimating that each number in the interval is approximately the class midpoint.

Example

Class	Frequency
1–5	3
6–10	1
11–15	2
16–20	2

That is, we pretend the first class consists of the numbers 3, 3, 3, the second class of the number 8, the third class of the numbers 13, 13, and the last class of the numbers 17, 17.

EXAMPLE

Class	Frequency
1–5	3
6–10	1
11–15	2
16–20	2

This gives us mean:

$$\begin{aligned}
\text{Mean } &\approx \frac{3+3+3+8+13+13+17+17}{8} \\
&= \frac{3(3)+1(8)+2(13)+2(17)}{8} \\
&= \frac{77}{8} \\
&= 9.625
\end{aligned}$$

Remark

This is *not* the actual mean, just an approximation of it.

Remark

Note all we did was multiply the class midpoint by the number of elements of the class, added these values, and divided by total number of data values. This gives a formula for the approximation:

$$\text{Mean } \approx \frac{\sum \text{ freq } \cdot \text{midpoint}}{n}$$

Example

Class	Frequency
10–12	1
13–20	2
21–25	1
25–40	4

Example

Class	Frequency
10–12	1
13–20	2
21–25	1
25–40	4

The class midpoints are: 11, 16.5, 23, 32.5.

Example

Class	Frequency
10–12	1
13–20	2
21–25	1
25–40	4

The class midpoints are: 11, 16.5, 23, 32.5.

There are $1 + 2 + 1 + 4 = 8$ total data points.

Example

Class	Frequency
10–12	1
13–20	2
21–25	1
25–40	4

The class midpoints are: 11, 16.5, 23, 32.5.

There are $1 + 2 + 1 + 4 = 8$ total data points.

$$\text{Mean} \approx \frac{1(11) + 2(16.5) + 1(23) + 4(32.5)}{8} = \frac{197}{8} = 24.625$$

Remark

This is exactly how course grades and GPAs are calculated. However instead of midpoints, a percentage or ‘weight’ is attached to the frequency (in this case value). This gives a weighted mean, e.g. grade average or GPA

$$\text{Weighted mean } = \frac{\sum \text{ weight } \cdot \text{ value}}{n}$$

Grade Average

Homework: 20%	Your Homework Grade: 95%
Quizzes: 30%	Your Quiz Grade: 80%
Midterm: 20%	Your Midterm Grade: 75%
Final: 30%	Your Final Grade: 87%

Grade Average

Homework: 20%	Your Homework Grade: 95%
Quizzes: 30%	Your Quiz Grade: 80%
Midterm: 20%	Your Midterm Grade: 75%
Final: 30%	Your Final Grade: 87%

The percentage worth are the weights, i.e. 20% (that is 0.20) is the weight for homework. The 'total' grade is 100.

This gives weighted mean (your course grade) of…

$$\begin{aligned}\text{Grade} &= \frac{0.20(95) + 0.30(80) + 0.20(75) + 0.30(87)}{100} \\ &= \frac{84.1}{100} \\ &= 0.841 = 84.1\%\end{aligned}$$

GPA

GPA is an example of a weighted average (by credit hour). Different letter grades are worth different amounts. For example, on a 4.0 rather than a 4.3 scale:

Grade	Weight
A	4.0
A−	3.66
B+	3.33
B	3.0
B−	2.66
C+	2.33
C	2.0
C−	1.66
D	1.0
F	0.0

$$\text{GPA} = \frac{\sum \text{weight} \cdot \text{grade value}}{\text{credit hours}}$$

Example

You take 16 credits one semester.

Class	Credits	Grade
HST 409	3	A−
MAT 121	4	A
GER 367	3	B−
ANT 348	3	C+
ECN 313	3	B

EXAMPLE

You take 16 credits one semester.

Class	Credits	Grade
HST 409	3	A−
MAT 121	4	A
GER 367	3	B−
ANT 348	3	C+
ECN 313	3	B

Then your GPA this semester is...

$$\text{GPA} = \frac{3(3.66) + 4(4.0) + 3(2.66) + 3(2.0) + 3(3.0)}{3+4+3+3+3} = \frac{49.96}{16} = 3.12$$

Now that we have a measure of center, we want to see how the different data values ‘vary’ about this center and identity ‘weird’ values, i.e. we want to measure spread and identify outliers.

Measuring Variation

Range

The simplest measure of spread is the range:

$$\text{Range} = \text{Max value} - \text{Min Value}$$

Range

The simplest measure of spread is the range:

$$\text{Range} \;=\; \text{Max value} \;-\; \text{Min Value}$$

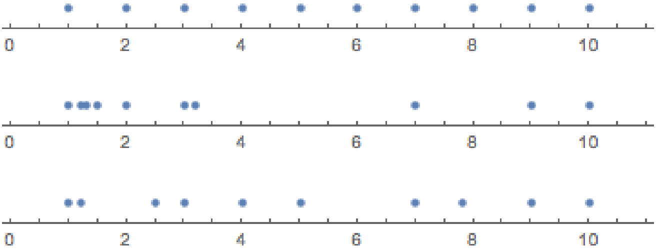
Example:

$$1, 3, 5, 5, 6, 11, 13$$

$$\text{Range} \;= 13 - 1 = 12$$

Remark

The range only measures total spread. We want to somehow measure both the total spread and the spread between the individual data values. For example, the three following data sets all have the same range but are spread out in different ways, which we want our measure to detect.



Measuring Variation about Median

Variation about Median

Before discussing variation about the median, we will need to discuss proportions.

Variation about Median

We can measure spread about the median using percentiles:

$$\text{Percentile Data Value } = \frac{\text{Number Values Less than } x}{\text{Total Number Values}} \cdot 100$$

Note: We denote the ith percentile by P_i. For example, the 25th percentile is P_{25}, the 62nd percentile is P_{62}, ect..

Example

$$1, \quad 1.5, \quad 3, \quad 3.5, \quad 4, \quad 5.6, \quad 8.1, \quad 9.6$$

What percentile is the number 5.6?

Example

$$1, \quad 1.5, \quad 3, \quad 3.5, \quad 4, \quad 5.6, \quad 8.1, \quad 9.6$$

What percentile is the number 5.6?

It is the $\frac{5}{8} = 0.625$, the 62.5 percentile, i.e. $P_{62.5}$. This means 5.6 is larger than slightly more than half the total values in the data set. In fact, it is larger than 62.5% of the data values.

There is a problem with repeated values…

Example

$$1, 5, 5, 5, 5, 5, 5, 5, 5, 6$$

What percentile is the number 5?

Example

$$1, 5, 5, 5, 5, 5, 5, 5, 5, 6$$

What percentile is the number 5?

It is the $\frac{1}{10} = 0.10 = 10$th percentile, i.e. P_{10}.

Example

$$1, 5, 5, 5, 5, 5, 5, 5, 5, 6$$

You do not ‘want’ 5 to be the 10th percentile. But you do not ‘want’ it to be the 90th percentile (using the last five).

With repeated values, use

$$\frac{\text{number values less than } x + 0.5 \text{ \# } x\text{'s appearing}}{n} \cdot 100$$

Example

$$1, 5, 5, 5, 5, 5, 5, 5, 5, 6$$

So what percentile is 5?

Example

$$1, 5, 5, 5, 5, 5, 5, 5, 5, 6$$

So what percentile is 5?

$$\frac{1 + 0.5 \cdot 8}{10} \cdot 100 = \frac{5}{10} \cdot 100 = 50$$

So 5 is the 50th percentile.

Finding Proportions

We can also go the other direction, though it involves an extra 'step':

$$L^{\text{th}} \text{ Percentile } = \frac{k}{100} \cdot n.$$

This gives the location of the L^{th} percentile, up to one exception.

Example

Given the data set

$$1, \quad 3, \quad 5, \quad 7, \quad 9, \quad 10, \quad 13, \quad 15, \quad 17$$

Find P_{30}, the 30th percentile, i.e. the value such that 30% of all data set values are at most this value.

Example

Given the data set

$$1, \quad 3, \quad 5, \quad 7, \quad 9, \quad 10, \quad 13, \quad 15, \quad 17$$

Find P_{30}, the 30th percentile, i.e. the value such that 30% of all data set values are at most this value.

$L = \frac{30}{100} \cdot 9 = 2.70 \rightsquigarrow$ 3rd data set number. Therefore, P_{30} is 5 for this dataset.

Remark

The rule to find the percentile value is to *always* round up. So if you find $L = 7.00001$, then you take the 8th data set value. Make sure the data set is in order!

Remark

The rule to find the percentile value is to *always* round up. So if you find $L = 7.00001$, then you take the 8th data set value. Make sure the data set is in order!

Now the 'extra step' comes in for the special case where L is *exactly* an integer (which is 'rare'). If this happens, you take the average of that value and the next value.

Example

For the data set

1.1, 2.3, 5.5, 7.8, 9.9, 10.3, 11.7, 12.9, 13.5, 14.8

Find P_{80}, the 80th percentile.

Example

For the data set

1.1, 2.3, 5.5, 7.8, 9.9, 10.3, 11.7, 12.9, 13.5, 14.8

Find P_{80}, the 80th percentile.

$L = \frac{80}{100} \cdot 10 = 8$. Then we take the average of the 8th and 9th value:

$$P_{80} = \frac{12.9 + 13.5}{2} = \frac{26.4}{2} = 13.2$$

Variation about Median: 5-number summary

We can now measure variation about the median by breaking our data into fourths:

Min: The minimum Value.
Q_1: The 25th percentile (the first quarter), i.e. P_{25}
Q_2/median: The 50th percentile (the second quarter, i.e. P_{50})
Q_3: The 75th percentile (the third quarter), i.e. P_{75}.
Max: The maximum value.

This is called the 5-number summary.

Definition (5-Number Summary)

An (ordered) list of numbers arising from a dataset consisting of minimum, Q_1, median, Q_3, and maximum.

Definition (Boxplot)

A plot visually showing the five numbers in the 5-number summary. The upper/lower line correspond to the max/min. The lower line in the middle box to Q_1, the middle line to the median, and the upper line of the center box to Q_3.

Example

2, 2, 5, 6, 8, 9, 12, 13, 17

Example

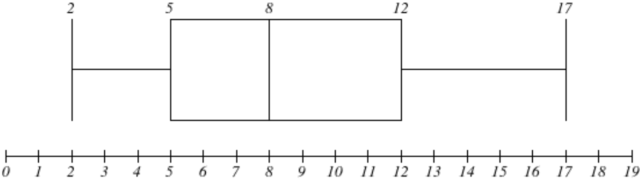
$$2, \quad 2, \quad 5, \quad 6, \quad 8, \quad 9, \quad 12, \quad 13, \quad 17$$

Min: 2, Max: 17
Median: 8
Q_1: 5, $\quad Q_3$: 12

Example

2,   2,   5,   6,   8,   9,   12,   13,   17

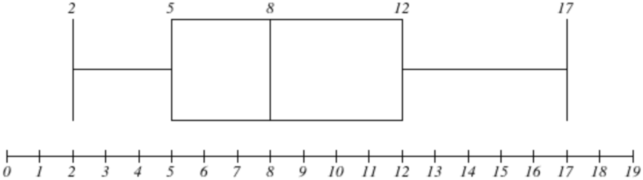
Min: 2, Max: 17
Median: 8
Q_1: 5, Q_3: 12



Example

$$2, \quad 2, \quad 5, \quad 6, \quad 8, \quad 9, \quad 12, \quad 13, \quad 17$$

Min: 2, Max: 17
Median: 8
Q_1: 5, Q_3: 12



Note: A boxplot shows us the skewness of the distribution. The distribution is perfectly symmetric if each segment is the same length.

Remark

Box plots can be drawn horizontally or vertically.

These quartiles now give a way of measuring how far a data value has to be from the center (in this case median) to be considered ‘weird’, i.e. an outlier.

Definition (IQR)

$\text{IQR} = Q_3 - Q_1$

We consider a value an outlier (in terms of the median) if the value is at least 1.5 x IQR from the median.

Example

1, 7, 36, 62, 79, 81, 88, 91, 95, 101, 117, 119, 153

Example

1, 7, 36, 62, 79, 81, 88, 91, 95, 101, 117, 119, 153

Median: 88
$Q_1 = 62$
$Q_3 = 101$

IQR $= Q_3 - Q_1 = 101 - 62 = 6$
Median $+$ 1.5 x IQR $= 146.5$
Median $-$ 1.5 x IQR $= 29.5$

Example

$$1, \;\; 7, \;\; 36, \;\; 62, \;\; 79, \;\; 81, \;\; 88, \;\; 91, \;\; 95, \;\; 101, \;\; 117, \;\; 119, \;\; 153$$

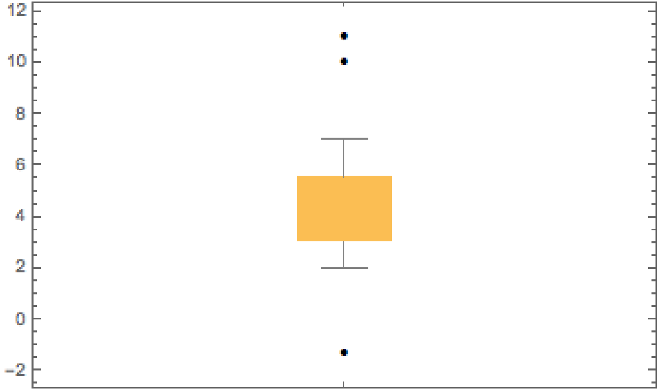
Median: 88
$Q_1 = 62$
$Q_3 = 101$

IQR $= Q_3 - Q_1 = 101 - 62 = 6$
Median $+ 1.5$ x IQR $= 146.5$
Median $- 1.5$ x IQR $= 29.5$

Therefore, the outliers are 1, 7 and 153.

Remark

One can also construct box plots differently to reflect outliers. Instead of having the outermost lines go to the max/min, have them go to the values of 1.5 x IQR ± Median. Then for values outside this range, use a dot. The dots are the outliers. These are called modified box plots.



Variation about Mean

Variation about the Mean

We need to measure variation about the mean. We could add up the differences between the values and the mean:

$$\sum (x - \overline{x})$$

Variation about the Mean

We need to measure variation about the mean. We could add up the differences between the values and the mean:

$$\sum (x - \overline{x})$$

But this is the problem with range. We do not want a total difference, but somehow an 'average' deviation. So we can take the average:

$$\frac{1}{n}\sum (x - \overline{x})$$

Variation about the Mean

However, there is a problem with this! Take the following dataset:

$$1, \quad 2, \quad 3, \quad 4, \quad 5$$

Variation about the Mean

However, there is a problem with this! Take the following dataset:

$$1,\ \ 2,\ \ 3,\ \ 4,\ \ 5$$

Mean, $\bar{x} = 3$

Variation about the Mean

However, there is a problem with this! Take the following dataset:

$$1,\ \ 2,\ \ 3,\ \ 4,\ \ 5$$

Mean, $\overline{x} = 3$

x_i	$x_i - \overline{x}$
1	-2
2	-1
3	0
4	1
5	2
Total:	0

Variation about the Mean

However, there is a problem with this! Take the following dataset:

$$1,\quad 2,\quad 3,\quad 4,\quad 5$$

Mean, $\overline{x} = 3$

x_i	$x_i - \overline{x}$
1	-2
2	-1
3	0
4	1
5	2
Total:	0

$$\frac{1}{n}\sum(x - \overline{x}) = \frac{1}{5} \cdot 0 = 0$$

Remark

This is *always* the case. The mean is the ‘middle’ of a dataset in the sense that for every value of a certain distance to the right of the mean, there is a corresponding value that distance to the left of the mean. Then the differences cancel — summing to zero.

We fix this by squaring the differences: $(x - \overline{x})^2$.

We fix this by squaring the differences: $(x - \overline{x})^2$.

Now the differences are always positive or zero, never negative, so there is no cancellation. We take the average of these squares:

$$\frac{1}{n}\sum(x - \overline{x})^2$$

We fix this by squaring the differences: $(x - \overline{x})^2$.

Now the differences are always positive or zero, never negative, so there is no cancellation. We take the average of these squares:

$$\frac{1}{n} \sum (x - \overline{x})^2$$

But now we have made the differences much bigger. For example, if $x = 6$ and $\overline{x} = 2$, then $x - \overline{x} = 6 - 2 = 4$. So the distance is 4 but we square this to get 16. We do not want to make the differences large, so we will square root to undo this:

$$\sqrt{\frac{1}{n} \sum (x - \overline{x})^2}$$

This is our measure of spread about the mean — the standard deviation:

$$s = \sqrt{\frac{1}{n} \sum (x - \overline{x})^2}$$

This is our measure of spread about the mean — the standard deviation:

$$s = \sqrt{\frac{1}{n}\sum(x-\overline{x})^2}$$

We will often calculate the variance first:

$$\text{Variance } = s^2 = \frac{1}{n}\sum(x-\overline{x})^2$$

This is our measure of spread about the mean — the standard deviation:

$$s = \sqrt{\frac{1}{n} \sum (x - \overline{x})^2}$$

We will often calculate the variance first:

$$\text{Variance } = s^2 = \frac{1}{n} \sum (x - \overline{x})^2$$

Note: We also use σ for the standard deviation and σ^2 for the variance.

Note this is *not* what we want. We actually want

$$s = \sqrt{\frac{1}{n-1}\sum(x-\overline{x})^2}$$

$$\text{Variance} = s^2 = \frac{1}{n-1}\sum(x-\overline{x})^2$$

Note this is *not* what we want. We actually want

$$s = \sqrt{\frac{1}{n-1}\sum(x-\overline{x})^2}$$

$$\text{Variance} = s^2 = \frac{1}{n-1}\sum(x-\overline{x})^2$$

This has to do with the fact that if you fix $n-1$ of the values, the last value can be solved for from the given equations. So you have '$n-1$ degrees of freedom', i.e. freedom of choice before the final value becomes forced. You want to average over the degrees of freedom. Moreover, you want to force the formula to 'breakdown' in the case $n = 1$. [Why?]

Standard Deviation & Variance

$$\text{Standard Deviation: } \sigma = \sqrt{\frac{1}{n-1}\sum(x-\overline{x})^2}$$

$$\text{Variance: } \sigma^2 = \frac{1}{n-1}\sum(x-\overline{x})^2$$

Standard Deviation & Variance

$$\text{Standard Deviation: } \sigma = \sqrt{\frac{1}{n-1}\sum(x-\overline{x})^2}$$

$$\text{Variance: } \sigma^2 = \frac{1}{n-1}\sum(x-\overline{x})^2$$

Note that we often calculate the variance first. Then take the square root to find the standard deviation.

Remark

We use Greek letters μ (for the mean) and σ (for the standard deviation) for the *whole population*. We often do not know these exactly (unless we took a census), so we estimate them by finding the mean and standard deviation of a sample (which we hope is a good approximation). For these samples, we use English letters $\overline{x}$ for the sample mean and s for the sample standard deviation. How 'good' these are as approximations is one of the points of inferential statistics.

Example

$$1, \quad 2, \quad 3, \quad 4, \quad 5$$

Mean: 3

Example

$$1, \quad 2, \quad 3, \quad 4, \quad 5$$

Mean: 3

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4
	Total:	10

$$s = \sqrt{\frac{1}{n-1}\sum(x-\overline{x})^2} = \sqrt{\frac{1}{4}\cdot 10} = \sqrt{2.5} \approx 1.58$$

Example

6.1, 7.9, 8.0, 9.2, 9.7, 10.6, 11.3

Mean: 8.97

Example

$$6.1, \quad 7.9, \quad 8.0, \quad 9.2, \quad 9.7, \quad 10.6, \quad 11.3$$

Mean: 8.97

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
6.1	−2.87	8.24
7.9	−1.07	1.14
8.0	−0.97	0.94
9.2	0.23	0.05
9.7	0.73	0.53
10.6	1.63	2.66
11.3	2.33	5.43
	Total:	18.99

Example

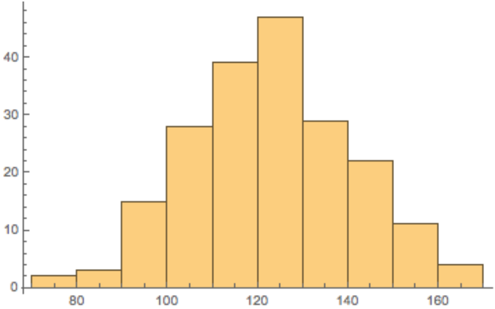
$$6.1,\quad 7.9,\quad 8.0,\quad 9.2,\quad 9.7,\quad 10.6,\quad 11.3$$

Mean: 8.97

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
6.1	−2.87	8.24
7.9	−1.07	1.14
8.0	−0.97	0.94
9.2	0.23	0.05
9.7	0.73	0.53
10.6	1.63	2.66
11.3	2.33	5.43
	Total:	18.99

$$s = \sqrt{\frac{1}{n-1}\sum(x-\overline{x})^2} = \sqrt{\frac{1}{6}\cdot 18.99} = \sqrt{3.165} \approx 1.78$$

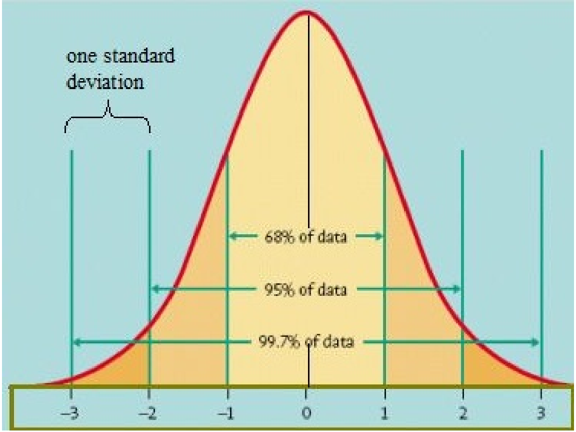
Note that when we do not have the data, we can still estimate the standard deviation using the rough estimate: $s \approx \dfrac{\text{Range}}{4}$



$$s = \frac{165 - 75}{4} \approx 22.50$$

When the underlying distribution is normal, there is a special relationship between percent of data values between given values of a distribution and standard deviations.

68–95–99.7 Rule

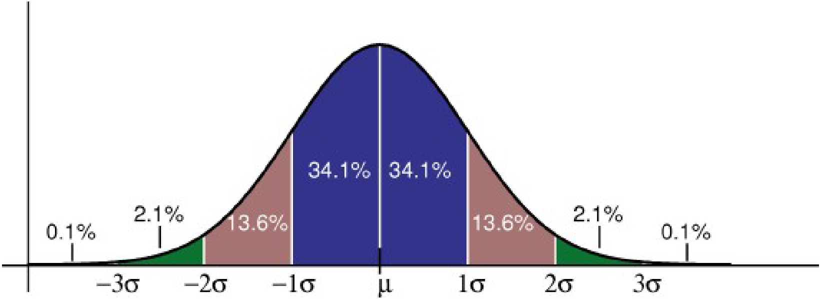


68%: Lie within 1 standard deviation of the mean
95%: Lie within 2 standard deviations of the mean
99.7%: Lie within 3 standard deviations of the mean

Stephanie. "68 95 99.7 Rule in Statistics." 09/15/2013. Accessed 05/26/2018.

68–95–99.7 Rule

Since the normal distribution is symmetric, we can actually be more specific. For example, since 68% lies within one standard deviation, 34% must lie 1 standard deviation above and 34% one below.



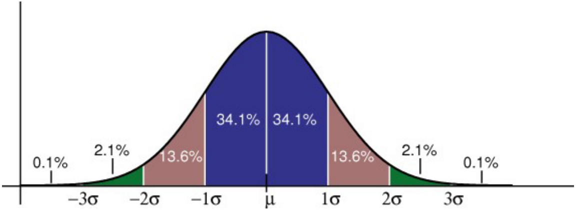
Stephanie. "68 95 99.7 Rule in Statistics." 09/15/2013. Accessed 05/26/2018.

Remark

The distance from the center of a normal distribution to where the curve changes how it ‘bends’ is the value of the standard deviation for that normal distribution. [If you know Calculus, this is the point of inflection for the curve.]

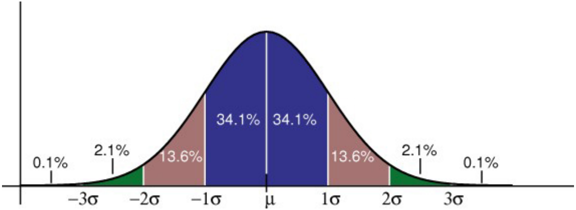
Example

An exam is approximately normally distributed with mean 72 and standard deviation 8.



EXAMPLE

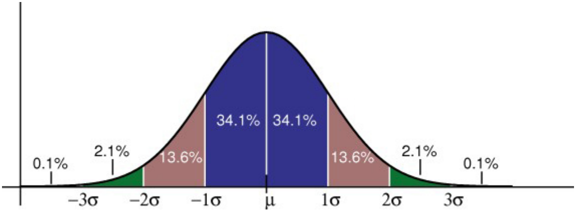
An exam is approximately normally distributed with mean 72 and standard deviation 8.



Standard Deviation Above/Below:	-3	-2	-1	0	1	2	3
Score:	48	56	64	72	80	88	96

EXAMPLE

An exam is approximately normally distributed with mean 72 and standard deviation 8.

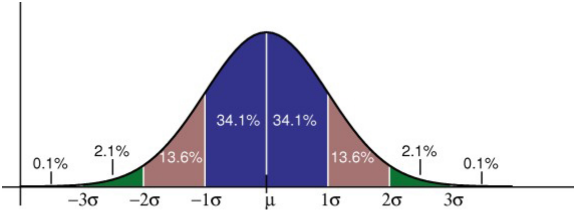


Standard Deviation Above/Below:	−3	−2	−1	0	1	2	3
Score:	48	56	64	72	80	88	96

% between 56–80: $13.6 + 34.1 + 34.1 = 81.8\%$

Example

An exam is approximately normally distributed with mean 72 and standard deviation 8.



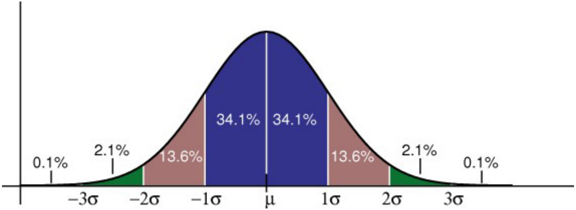
Standard Deviation Above/Below:	−3	−2	−1	0	1	2	3
Score:	48	56	64	72	80	88	96

% between 56–80: $13.6 + 34.1 + 34.1 = 81.8\%$

% between 80–88: 13.6%

Example

An exam is approximately normally distributed with mean 72 and standard deviation 8.



Standard Deviation Above/Below:	-3	-2	-1	0	1	2	3
Score:	48	56	64	72	80	88	96

% between 56–80: $13.6 + 34.1 + 34.1 = 81.8\%$

% between 80–88: 13.6%

% between 72–96: $34.1 + 13.6 + 2.1 = 49.8\%$

Even better, this gives a way of measuring how unusual a value is. The further from the mean, the more ‘unusual’ a value is. The more standard deviations you are from the mean, the further you are from the mean and vise versa. So we want to measure number of standard deviations from the mean.

z–SCORE

$$z = \underbrace{\frac{\overbrace{x - \overline{x}}^{\text{distance from mean}}}{s}}_{\text{S.D. from mean}}$$

z-score: Measures the number of standard deviations from the mean.

z–SCORE

$$z = \underbrace{\frac{\overbrace{x - \overline{x}}^{\text{distance from mean}}}{s}}_{\text{S.D. from mean}}$$

z-score: Measures the number of standard deviations from the mean.

So the larger the z-score (in absolute value) the more ‘unusual’ the score. [Unusual in the sense of the probability of seeing such an ‘extreme’ score.]

EXAMPLE

$$\mu = 12, \sigma = 3, x = 15, \qquad z_{15} = \frac{15 - 12}{3} = \frac{3}{3} = 1.0$$

$$\mu = 156, \sigma = 9, x = 141, \qquad z_{141} = \frac{141 - 156}{9} = -1.67$$

$$\mu = 87, \sigma = 7, x = 100, \qquad z_{100} = \frac{100 - 87}{7} = 1.86$$

Example

$$\mu = 12, \sigma = 3, x = 15, \qquad z_{15} = \frac{15 - 12}{3} = \frac{3}{3} = 1.0$$

$$\mu = 156, \sigma = 9, x = 141, \qquad z_{141} = \frac{141 - 156}{9} = -1.67$$

$$\mu = 87, \sigma = 7, x = 100, \qquad z_{100} = \frac{100 - 87}{7} = 1.86$$

Note: This also gives a way of determining if a data value is an outlier. We say a value is an outlier if it more than two standard deviations from the mean, i.e. $z \geq 2$ or $z \leq -2$.

Example

Susie took the ACT and scored a 26. Johnny took the SAT and scored 1240. If the ACT has mean 20.8 and standard deviation 4.8, while the SAT has a mean of 1060 and standard deviation 195, who did better?

Example

Susie took the ACT and scored a 26. Johnny took the SAT and scored 1240. If the ACT has mean 20.8 and standard deviation 4.8, while the SAT has a mean of 1060 and standard deviation 195, who did better?

$$z_{\text{Susie}} = \frac{26 - 20.8}{4.8} = 1.08$$

$$z_{\text{Johnny}} = \frac{1240 - 1060}{195} = 0.92$$

Therefore, Susie did better on her exam than Johnny did on his.

Lecture 4

Chapter 4

Probability Vocabulary
Addition/Multiplication Rules
Counting & Probability Theorems

Why Probability?

Probability measures the likelihood of events. This is central to Statistics. Statistics wants to analyze data and make decisions based on samples.

But to understand what data samples are ‘saying’, you need to understand the probability of seeing similar samples.

In summary, Statistics seeks to reject explanations/claims with low probabilities of occurring in favor of more probable (in the literal sense) explanations.

Definition (Event)

A collection of results or outcomes of a procedure.

Definition (Simple Event)

An event which cannot be broken down into simpler components.

Definition (Sample Space)

The collection of all simple events.

Definition (Compound Event)

A combination of simple events.

Example

Procedure: Collected birth sex.

Simple Event: Girl, Boy

Sample Space: { Girl, Boy }

Example

Procedure: Collect birth sexes after woman's third birth.

Simple Event: Girl, Boy

Sample Space: { GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB }

Notation

P will denote probability

English letters, such as A, B, C, will denote a sample space

$P(A)$ will denote the probability of event A (or events from A) occurring

Example

Consider keeping track of the sex of births after a woman's third child.

Sample Space: { GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB }

Let A denote the collection of events where a woman has at least two girls.

$A =$ { GGG, GGB, GBG, BGG }

$P(A) =$ the probability that a woman has at least two girls.

This is part of the beauty of Mathematics. Mathematics is really a language of logic using mathematical systems.

This is part of the beauty of Mathematics. Mathematics is really a language of logic using mathematical systems.

So Mathematics can turn lexical/etymological questions of interest, e.g. what is the probability that someone will commit a crime, where do we invest money to have the greatest effect, etc., into mathematical expressions that can be evaluated.

This is part of the beauty of Mathematics. Mathematics is really a language of logic using mathematical systems.

So Mathematics can turn lexical/etymological questions of interest, e.g. what is the probability that someone will commit a crime, where do we invest money to have the greatest effect, etc., into mathematical expressions that can be evaluated.

Vise versa, examining mathematical formulations, what can these tell us about human nature, behaviors and properties of systems, and natural structures?

Remark

Note that for any event or collection of events A, $0 \leq P(A) \leq 1$; that is, the probability of something between 0 (impossible, or statistically impossible) and 1 (certain, or statistically certain).

Remark

Note that for any event or collection of events A, $0 \leq P(A) \leq 1$; that is, the probability of something between 0 (impossible, or statistically impossible) and 1 (certain, or statistically certain).

Remark

Note that $P(A) = 0$ or $P(A) = 1$ does not mean impossible/certain but rather *statistically* impossible/certain. For example, it is statistically impossible that tomorrow I will win the lottery only to later that day be struck by lightening and fall into a shark tank. It is also statistically certain that I will die. But we will have to wait until my death (should it ever arrive) to know for sure. But even then…

Walking Dead, AMC

Probability [for Finite Sets]

This is the ‘Fundamental Law’ of Probability:

Theorem

$$P(A) = \frac{\textit{number of times A occured}}{\textit{number of times the procedure was repeated}}$$

This is the traditional ‘rule of thumb’, i.e.

$$P(A) = \frac{\text{number of ways } A \text{ can happen}}{\text{Total outcomes}}$$

Example 1

Flip a coin: What is the probability of heads?

$$\frac{\text{heads}}{\text{head or tail}} = \frac{1}{2}$$

Example 1

Flip a coin: What is the probability of heads?

$$\frac{\text{heads}}{\text{head or tail}} = \frac{1}{2}$$

This only works with events that are *equally* likely.

Example 2

Consider keeping track of the sex of births after a woman's third child.

Sample Space: { GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB }

Let A denote the collection of events where a woman has at least two girls.

$A = \{$ GGG, GGB, GBG, BGG $\}$

$P(A) =$ the probability that a woman has at least two girls.

Note that each event (having a boy/girl) are equally likely. So each of the combinations of birth sex are equally likely. [Note: in reality, having a girl is *slightly* more likely, 50–52%.]

Example 3

{ GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB }

$A = \{$ GGG, GGB, GBG, BGG $\}$

$$P(A) = \frac{4}{8} = \frac{1}{2} = 0.5 = 50\%$$

Example 4

{ GGG, GGB, GBG, GBB, BGG, BGB, BBG, BBB }

$A =$ { GGG, GGB, GBG, GBB, BGG, BGB, BBG } (Probability of at least one girl.)

$$P(A) = \frac{7}{8} \approx 0.875 = 87.5\%$$

Non-Example

I buy a lottery ticket. What is the probability that I win?

$$P(\text{win}) = \frac{\text{win}}{\text{win or lose}} = \frac{1}{2}$$

This is clearly wrong. What is wrong with this? As another example of this type of fallacy:

Non-Example

What is the probability that I will die tomorrow?

$$P(\text{death}) = \frac{\text{die tomorrow}}{\text{die tomorrow or live tomorrow}} = \frac{1}{2}$$

Again, this clearly cannot be the case. What is wrong with the logic/Math above?

Theorem (Law of Large Numbers)

As a procedure is repeated, the probability of an event approaches the actual probability.

Remark

This *does not say anything* about a particular event. Rather, it says what should (indeed must) happen if you repeated the experiment a large number of times. Moreover, it does not say what should happen over a series of events.

Remark

This *does not say anything* about a particular event. Rather, it says what should (indeed must) happen if you repeated the experiment a large number of times. Moreover, it does not say what should happen over a series of events. Worse yet, it does not say what is ‘enough’, i.e. how many times the procedure should be repeated.

What is Probable?

According to the FAA (Federal Aviation Administration):

What is Probable?

According to the FAA (Federal Aviation Administration):

Probable: $0.00001 \leq P(A)$, i.e. 1 in 100,000.

Improbable: $0.000000001 \leq P(A) \leq 0.00001$, i.e. between 1 in 100,000 to 1 in a billion.

Extremely Improbable: $P(A) \leq 0.000000001$, i.e. 1 in a billion or less.

Remark

Humans are *really* bad at estimating probabilities. Moreover, humans prescribe to ‘actors’ what is better explained by chance.

Example

I flip a coin several times, which combination is most likely?

(i) TTHTHTTT

(ii) HTTHTTTT

(iii) THHHHHHH

(iv) HTHTHTTH

(v) HHHTTTHH

Example

I flip a coin several times, which combination is most likely?

(i) TTHTHTTT

(ii) HTTHTTTT

(iii) THHHHHHH

(iv) HTHTHTTH

(v) HHHTTTHH

They are all equally likely!!

Example

On a roulette wheel, black has come up the last 4 times. Is red or black more likely on the next spin given the last four were black? What about the next two times?

Example

On a roulette wheel, black has come up the last 4 times. Is red or black more likely on the next spin given the last four were black? What about the next two times?

The chances are still 50-50.

Examples

1. Which is more safe, cars or planes?
2. Which is more deadly, shark attacks or vending machines?
3. How likely are stocks to go up/down?
4. What products are likely to become popular?
5. How likely is it that you are ‘watched’?

Law of Truly Large Numbers

"With a large enough sample, any outrageous thing is likely to happen (Diaconis and Mosteller 1989). Littlewood (1986) considered an event which occurs one in a million times to be "surprising."

Taking this definition, close to 100,000 surprising events are "expected" each year in the United States alone and, in the world at large, "we can be absolutely sure that we will see incredibly remarkable events" (Diaconis and Mosteller 1989)."

Weisstein, Eric W. "Law of Truly Large Numbers." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/LawofTrulyLargeNumbers.html

Law of Truly Large Numbers

"With a large enough sample, any outrageous thing is likely to happen (Diaconis and Mosteller 1989). Littlewood (1986) considered an event which occurs one in a million times to be "surprising."

Taking this definition, close to 100,000 surprising events are "expected" each year in the United States alone and, in the world at large, "we can be absolutely sure that we will see incredibly remarkable events" (Diaconis and Mosteller 1989)."

That is, 'miracles' happen.

Weisstein, Eric W. "Law of Truly Large Numbers." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/LawofTrulyLargeNumbers.html

…back to Probability.

Definition (Complement)

The complement of an event(s) A, denoted $\overline{A}$ and pronounced ‘A complement’ or A bar’, is the set of all outcomes in which the event(s) A do not occur.

Notation

There is also other notation we use:

Union: $A \cup B$ (also written A or B) is the collection of all objects in A or B. This is read "A union B."

Intersection: $A \cap B$ (also written A and B) is the collection of all objects in A *and* B. This is read "A intersect B."

$P(\overline{A})$ is the probability that something in A *does not* happen.

Because the language is often used, we will define a few additional terms:

Odds Against: $P(\overline{A})/P(A)$

Odds in Favor: $P(A)/P(\overline{A})$

Payoff Odds: Net Profit : Bet Amount

Example

Assume $P(\text{win}) = \dfrac{1}{16}$.

Example

Assume $P(\text{win}) = \dfrac{1}{16}$.

So $P(\text{loss}) = \dfrac{15}{16}$. Then…

Example

Assume $P(\text{win}) = \dfrac{1}{16}$.

So $P(\text{loss}) = \dfrac{15}{16}$. Then…

Odds Against$= \dfrac{15/16}{1/16} = \dfrac{15}{1} = 15 : 1$.

Example

Assume $P(\text{win}) = \dfrac{1}{16}$.

So $P(\text{loss}) = \dfrac{15}{16}$. Then…

Odds Against$= \dfrac{15/16}{1/16} = \dfrac{15}{1} = 15 : 1$.

Odds in Favor$=$ 1/Odds Against $= 1 : 15$.

Example

Assume $P(\text{win}) = \dfrac{1}{16}$.

So $P(\text{loss}) = \dfrac{15}{16}$. Then…

Odds Against$= \dfrac{15/16}{1/16} = \dfrac{15}{1} = 15 : 1$.

Odds in Favor$=$ 1/Odds Against $= 1 : 15$.

Assume you bet \$20 (and won) and the payoff odds were 11:1. So

$$\frac{11}{1} \cdot \frac{20}{20} = \frac{220}{20}$$

you win \$220, i.e. $11 \times \$20 = \$220$.

How to Calculate Probabilities?

Remark

Before beginning, note that people use ‘or’ differently than mathematicians. When you’re asked if you want cake or pie, people generally mean cake or pie but not both. But ‘or’ in Mathematics always means cake or pie or both.

Remark

Before beginning, note that people use ‘or’ differently than mathematicians. When you’re asked if you want cake or pie, people generally mean cake or pie but not both. But ‘or’ in Mathematics always means cake or pie or both.

In Mathematics, you really can have your cake and eat it too!

Example

Consider the following table (non)-drug users and their test results.

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

Example

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$P(\text{Positive Test and Uses Drugs}) =$

Example

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive Test and Uses Drugs}) = \frac{80}{220} = \frac{4}{11} \approx 36.4\%$$

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$P(\text{Positive Test}) =$

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive Test}) = \frac{80+15}{220} = \frac{95}{220} = \frac{19}{44} \approx 43.2\%$$

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

P(Positive Test or Uses Drugs) $=$

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive Test or Uses Drugs}) = \frac{80 + 15 + 5}{220} = \frac{100}{220} = \frac{5}{11} \approx 45.5\%$$

Often, we calculate probabilities by breaking them down into simpler pieces, which are easier to handle, then put the probabilities ‘back together’.

One way of doing this is with OR/AND events.

One way of doing this is with OR/AND events.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Remark

This is really just counting. To count the people in A or B, you count the people in A, plus the count of people in B, but then you need to subtract off anyone you counted twice, i.e. the people in *A and B*.

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive Test or Uses Drugs}) = P(\text{Positive}) + P(\text{Drugs}) - P(\text{Positive and Drugs})$$

$$= \frac{95}{220} + \frac{85}{220} - \frac{90}{220} = \frac{100}{220} = \frac{5}{11} \approx 45.5\%$$

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$
\begin{aligned}
P(\text{Positive Test or Uses Drugs}) &= P(\text{Positive}) + P(\text{Drugs}) - P(\text{Positive and Drugs}) \\
&= \frac{95}{220} + \frac{85}{220} - \frac{90}{220} = \frac{100}{220} = \frac{5}{11} \approx 45.5\%
\end{aligned}
$$

Yay! Mathematics is internally self-consistent!

There is a special case when the formula is a bit simpler.

Definition (Disjoint)

Events A and B are disjoint, also called mutually exclusive, if the events cannot occur at the same time.

Example

1. A: Take the car to work.
   B: Take the bus to work.

2. A: Win the game.
   B: Lose the game.

3. A: Die rolls 5 or 6.
   B: Die rolls 3 or less.

If A and B are disjoint, then $P(A \text{ and } B) = 0$ because they cannot occur simultaneously. But then

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ &= P(A) + P(B) - 0 \\ &= P(A) + P(B) \end{aligned}$$

Testing Disjointness

This works both ways. Meaning two events are disjoint if and only if $P(A \text{ and } B) = 0$. This gives a way of testing if events are disjoint, i.e. see if $P(A \text{ and } B) = 0$. Equivalently, test if $P(A \text{ or } B) = P(A) + P(B)$.

Testing Disjointness

This works both ways. Meaning two events are disjoint if and only if $P(A \text{ and } B) = 0$. This gives a way of testing if events are disjoint, i.e. see if $P(A \text{ and } B) = 0$. Equivalently, test if $P(A \text{ or } B) = P(A) + P(B)$.

Otherwise, one has to think about whether there are events that satisfy both the conditions of A and B or not.

Another special case:

$$\begin{aligned} 1 = P(A \text{ or } \overline{A}) &= P(A) + P(\overline{A}) - P(A \text{ or } \overline{A}) \\ &= P(A) + P(\overline{A}) - 0 \\ &= P(A) + P(\overline{A}) \end{aligned}$$

Therefore, $P(A) + P(\overline{A}) = 1$.

Because $P(A) + P(\overline{A}) = 1$, we also know

$$P(\overline{A}) = 1 - P(A)$$
$$P(A) = 1 - P(\overline{A})$$

So knowing the probability of an event is ‘the same’ as knowing the probability of its opposite.

Definition (Conditional Probability)

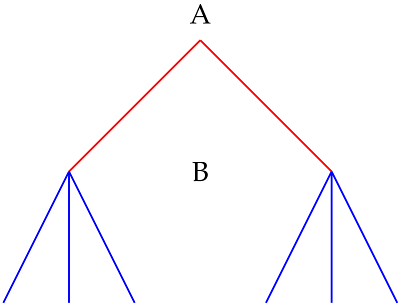
The probability of A happening given that an event B occurred is denoted $P(A \mid B)$, and is read ‘probability of A given B’.

$$\begin{aligned} P(A \text{ and } B) &= P(A) \cdot P(B \mid A) \\ &= P(B) \cdot P(A \mid B) \end{aligned}$$

There is a simpler formula which works in special cases (which most are familiar with):

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

What is the logic of this ‘multiplication rule’?



There are $2 \cdot 3$ total ways A and B can both happen.

Example

Rolling two die, what is the probability of getting a 1 and 6?

$$P(1 \text{ and } 6) = P(1) \cdot P(6) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Example

Flipping a coin twice, what is the probability of getting a H and then a T?

$$P(H \text{ then } T) = P(H) \cdot P(T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Example

Flipping a coin twice (or two coins once), what is the probability of a head and a tail (in total)?

$$\begin{aligned}
P(H \text{ and } T) &= P(H \text{ then } T) + P(T \text{ then } H) \\
&= P(H) \cdot P(T) + P(T) \cdot P(H) \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \\
&= \frac{1}{4} + \frac{1}{4} \\
&= \frac{2}{4} = \frac{1}{2}
\end{aligned}$$

Non-Example

Assume the following:

$P(\text{Fail}) = 75\%$
$P(\text{Pass}) = 25\%$

Non-Example

Assume the following:

$P(\text{Fail}) = 75\%$
$P(\text{Pass}) = 25\%$

$$P(\text{Fail and Pass}) = P(\text{Fail}) \cdot P(\text{Pass}) = 0.75 \cdot 0.25 = 0.1875$$

Non-Example

Assume the following:

$P(\text{Fail}) = 75\%$
$P(\text{Pass}) = 25\%$

$$P(\text{Fail and Pass}) = P(\text{Fail}) \cdot P(\text{Pass}) = 0.75 \cdot 0.25 = 0.1875$$

But we know that $P(\text{Fail and Pass}) = 0$ because it is impossible to both pass and fail! What went wrong is we used the 'special' case of the formula. Recall…

$$
\begin{aligned}
P(A \text{ and } B) &= P(A) \cdot P(B \mid A) \\
&= P(B) \cdot P(A \mid B)
\end{aligned}
$$

Definition (Independent)

Events A and B are called independent if the occurrence of one does not affect the probability of the occurrence of the other, i.e. the probability of either is the same regardless of whether the other occurred or not.

So assuming A and B are independent…

$$P(A \text{ and } B) = P(A) \cdot P(B \mid A) = P(A) \cdot P(B)$$

$$P(A \text{ and } B) = P(B) \cdot P(A \mid B) = P(B) \cdot P(A)$$

Example

What is the probability that someone tests negative and is a drug user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$\begin{aligned} P(\text{Negative and Uses}) &= P(\text{Negative}) \cdot P(\text{Uses} \mid \text{Negative}) \\ &= \frac{125}{220} \cdot \frac{5}{125} \\ &= \frac{5}{220} = 0.023 \end{aligned}$$

Example

What is the probability that someone tests negative and is a drug user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$
\begin{aligned}
P(\text{Negative and Uses}) &= \\
&= \frac{125}{220} \cdot \frac{5}{125} \\
&= \frac{5}{220} = 0.023
\end{aligned}
$$

Example

What is the probability that someone tests negative and is a drug user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$
\begin{aligned}
P(\text{Negative and Uses}) &= P(\text{Negative}) \cdot P(\text{Uses} \mid \text{Negative}) \\
&= \frac{125}{220} \cdot \frac{5}{125} \\
&= \frac{5}{220} = 0.023
\end{aligned}
$$

Example (Again)

What is the probability that someone tests negative and is a drug user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$\begin{aligned} P(\text{Negative and Uses}) &= P(\text{Uses}) \cdot P(\text{Negative} \mid \text{Uses}) \\ &= \frac{85}{220} \cdot \frac{5}{85} \\ &= \frac{5}{220} = 0.023 \end{aligned}$$

Example (Again)

What is the probability that someone tests negative and is a drug user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$
\begin{aligned}
P(\text{Negative and Uses}) &= \\
&= \frac{85}{220} \cdot \frac{5}{85} \\
&= \frac{5}{220} = 0.023
\end{aligned}
$$

Example (Again)

What is the probability that someone tests negative and is a drug user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$\begin{aligned} P(\text{Negative and Uses}) &= P(\text{Uses}) \cdot P(\text{Negative} \mid \text{Uses}) \\ &= \frac{85}{220} \cdot \frac{5}{85} \\ &= \frac{5}{220} = 0.023 \end{aligned}$$

Disjoint and Independent

Can disjoint events be independent?

Disjoint and Independent

Disjoint events can *never* be independent!

Disjoint and Independent

Disjoint events can *never* be independent!

Why? If events are disjoint then both cannot happen at the same time. But then if one happens the other does not (hence its probability is 0). But unless the probability was 0 before, this changed the probability, so they cannot be independent.

Remark

Because disjoint events are *never* independent, if A and B are disjoint, it will *never* be the case that $P(A \text{ and } B) = P(A) \cdot P(B)$.

Example

What is the probability that two people are born on the same day of the week?

VS

What is the probability that two people are both born on a Monday?

Example

What is the probability that two people are born on the same day of the week?

What is the probability that two people are both born on a Monday?

$$P(\text{same day}) = \frac{7}{7} \cdot \frac{1}{7} = \frac{1}{7}$$

Example

What is the probability that two people are born on the same day of the week? $= 1/7$

What is the probability that two people are both born on a Monday?

$$P(\text{same day}) = P(\text{Monday}) \cdot P(\text{Monday}) = \frac{1}{7} \cdot \frac{1}{7} = \frac{1}{49}$$

Conditional Probability

Recall that the probability that B occurs given that A occurred is denote $P(B \mid A)$. We know that $P(A \text{ and } B) = P(A)\, P(B \mid A)$ so that…

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Conditional Probability

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Conditional Probability

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Remark

Because A has to occur, one can only consider events in A. So this becomes the 'new denominator' or the new 'n' or new total.

Example

What is the probability that someone tests positive given that they are a user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive} \mid \text{Uses}) =$$

Example

What is the probability that someone tests positive given that they are a user?

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$\begin{aligned} P(\text{Positive} \mid \text{Uses}) &= \frac{P(\text{Positive and Uses})}{P(\text{Uses})} \\ &= \frac{80/220}{85/220} \\ &= \frac{80}{220} \cdot \frac{220}{85} \\ &= \frac{80}{85} \end{aligned}$$

Example (Again)

What is the probability that someone tests positive given that they are a user? [Using this idea of 'new total' for the denominator.]

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive} \mid \text{Uses}) =$$

Example (Again)

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive} \mid \text{Uses}) = \frac{80}{85}$$

Example (Again)

Compare the probabilities:

	Positive Test	Negative Test	Total
Uses Drugs	80	5	85
Not Use Drugs	15	120	135
Total	95	125	220

$$P(\text{Positive}) = \frac{95}{220} \approx 43.2\%$$

$$P(\text{Positive} \mid \text{Uses}) = \frac{80}{85} \approx 94.1\%$$

Explain this difference.

Example

Passengers on the Titanic by class and survival.

	First Class	Second Class	Third Class	Crew	Total
Survived	199	119	174	214	706
Died	130	166	536	685	1517
Total	329	285	710	899	2223

$$P(\text{survived} \mid \text{first class}) = \frac{199}{329} = 0.605 = 60.5\%$$
$$P(\text{survived} \mid \text{third class}) = \frac{174}{710} = 0.245 = 24.5\%$$
$$P(\text{survived}) = \frac{706}{2223} = 0.318 = 31.8\%$$

EXAMPLE

Passengers on the Titanic by class and survival.

	First Class	Second Class	Third Class	Crew	Total
Survived	199	119	174	214	706
Died	130	166	536	685	1517
Total	329	285	710	899	2223

$$P(\text{third class} \mid \text{died}) = \frac{536}{1517} = 0.353 = 35.3\%$$

$$\begin{aligned} P(\text{third class} \mid \text{died}) &= \frac{P(\text{third class and died})}{P(\text{died})} = \frac{536/2223}{1517/2223} \\ &= \frac{536}{1517} = 0.353 = 35.3\% \end{aligned}$$

Another powerful tool is Bayes’ Theorem.

Bayes' Theorem (Short Form)

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(B)}$$

Bayes' Theorem (General Form)

If the sample space can be broken into disjoint events $\{A_i\}$, then

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_i P(A_i)P(B \mid A_i)}$$

It will come as no surprise that Probability Theory comes with a lot of counting. After all, our starting point for probability was…

$$P(\text{stuff we are looking at}) = \frac{\text{stuff we are looking at}}{\text{total stuff}}$$

So we need to learn to count.

Counting

Theorem (Multiplication Counting Rule)

If there are n_1 ways of performing a first action, n_2 ways to perform the second action,…, and n_r ways of performing the rth action, there are

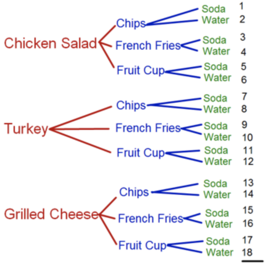
$$n_1 \cdot n_2 \cdot \ (\cdots) \ \cdot n_r$$

total ways of performing all the actions.

Remark

This is really just a tree diagram of counting the ways of doing something you may have seen in school before.

Example:



There are 3 ways of choosing a main course, $n_1 = 3$, 3 ways of choosing a side, $n_2 = 3$, and 2 ways of choosing a drink, $n_3 = 2$, then there are

$$n_1 \cdot n_2 \cdot n_3 = 3 \cdot 3 \cdot 2 = 18$$

total possible meals.

http://www.algebra-class.com/fundamental-counting-principle.html

Example

How many 5 digits passcodes are there using the digits 0–9?

Example

How many 5 digits passcodes are there using the digits 0–9?

Solution: Let n_i be the total possibilities for the ith digit, e.g. n_1 is the number of possible first digits. There are 10 possible choices for each digit, i.e. $n_i = 10$ for each digit, then the total number of passcodes is

$$n_1 \cdot n_2 \cdot n_3 \cdot n_4 \cdot n_5 = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 10^5 = 100,000$$

Example

How many 5 digits passcodes are there if the passcode must start with 1 and be even?

Example

How many 5 digits passcodes are there if the passcode must start with 1 and be even?

Solution: Let n_i be the total possibilities for the *i*th digit, e.g. n_1 is the number of possible first digits. Well, there is only one choice for the first digit, because it must be one. Therefore, $n_1 = 1$.

For the passcode to be even, the last digit must be even. The only choices for this last digit are then 0, 2, 4, 6, 8. Therefore, $n_5 = 5$. But the rest of the digits can be anything. Therefore, $n_2 = n_3 = n_4 = 10$. But then

$$n_1 \cdot n_2 \cdot n_3 \cdot n_4 \cdot n_5 = 1 \cdot 10 \cdot 10 \cdot 10 \cdot 5 = 5 \cdot 10^3 = 5 \cdot 1,000 = 5,000$$

total possible passcodes meeting these criterion.

In all the counting we have done so far, we did not care about the order. For example, when counting the passcodes, you could have chosen the first, then second,..., then fifth digit. But you could also have selected the last digit, then the next to last, and so forth.

In all the counting we have done so far, we did not care about the order. For example, when counting the passcodes, you could have chosen the first, then second,..., then fifth digit. But you could also have selected the last digit, then the next to last, and so forth.

But what if you wanted to count things where the order mattered?

Definition (Factorial)

The factorial, denoted !, is the product of the given nonnegative integer and all subsequent positive integers.

E.g.

$$\begin{aligned}
1! &= 1\\
3! &= 3 \cdot 2 \cdot 1 = 6\\
5! &= 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120\\
0! &:= 1
\end{aligned}$$

Note: 0! is defined to be 1. We shall be able to explain this in a moment.

Theorem

The number of ways of arranging n objects all in a particular order is $n!$.

Theorem

The number of ways of arranging n objects all in a particular order is $n!$.

Proof: There are n choices for which object comes first. Once placed, it cannot be placed again. Therefore, there are only $n-1$ choices for the second objects. Continue this process. Then the total number of ways of arranging all of these objects in a particular order is…

$$n \cdot (n-1) \cdot (n-2) \cdot (\cdots) \cdot 2 \cdot 1 =: n!$$

Example

How many ways are there of arrange Bill, Bob, and BillyJoBob in a line?

Note that order matters here: Bill, BillyJoBob, and Bob is a different lineup than BillyJoBob, Bill, and Bob.

$$3! = 3 \cdot 2 \cdot 1 = 6$$

Example

How many ‘words’ are there using the letters from the word ‘movie’?

Note that order here matters: mvoie is a different ‘word’ than oiemv.

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

We will need to differentiate counting objects where order does matter compared to counting objects where order *does not* matter.

Permutation: Order does matter.
Combination: Order does not matter.

A Mnemonic

Permutation:	**P**osition
Combination:	**C**ollection

Permutations

Permutation: Counting collections of objects where **order does matter**. The number of different ways of arranging a collection of k things from n objects (without replacement), where order does matter, is

$$ {}_nP_k := \frac{n!}{(n-k)!} $$

Permutations

Permutation: Counting collections of objects where **order does matter**. The number of different ways of arranging a collection of k things from n objects (without replacement), where order does matter, is

$$ {}_nP_k := \frac{n!}{(n-k)!} $$

Note: In the case where $n = k$, i.e. we are choosing all the objects, we obtain what we got before:

$$ {}_nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n! $$

The first example will show why this formula works.

Example

How many ways can you arrange 2 objects from 5 objects, i.e. what is ${}_5P_2$?

Example

How many ways can you arrange 2 objects from 5 objects, i.e. what is $_5P_2$?

There are 5 choices for the first spot and 4 for the second. So…

$$5 \cdot 4 = 20$$

Example (Again)

How many ways can you arrange 2 objects from 5 objects, i.e. what is $_5P_2$?

Example (Again)

How many ways can you arrange 2 objects from 5 objects, i.e. what is $_5P_2$?

$$_5P_2 = \frac{5!}{(5-2)!} = \frac{5!}{(3!} = \frac{5 \cdot 4 \cdot 3!}{3!} = \frac{5 \cdot 4 \cdot \cancel{3!}}{\cancel{3!}} = 5 \cdot 4 = 20$$

Permutation Formula

Why does the formula work? Suppose we want to arrange k things from n objects.

Permutation Formula

Why does the formula work? Suppose we want to arrange k things from n objects.

There are n choices for the first and $n-1$ choices for the second and…and $n-k+1$ choices for the kth objects. But then…

$$n\cdot(n-1)\cdot\dots\cdot(n-k+1) = n\cdot(n-1)\cdot\dots\cdot(n-k+1)\cdot\frac{(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$$

Example

In a race with 20 people, how many different possibilities are there for the top 3 finishers?

Example

In a race with 20 people, how many different possibilities are there for the top 3 finishers?

$$\begin{aligned} {}_{20}P_3 &= \frac{20!}{(20-3)!} \\ &= \frac{20!}{17!} \\ &= \frac{20 \cdot 19 \cdot 18 \cdot 17!}{17!} \\ &= \frac{20 \cdot 19 \cdot 18 \cdot \cancel{17!}}{\cancel{17!}} \\ &= 20 \cdot 19 \cdot 18 \\ &= 6,840 \end{aligned}$$

But what happens when there is repetition? For example, how many ways are there of arranging the letters of ‘peek’ to make new ‘words’?

One way is pkee. Then we can switch the two final e’s to obtain another, ‘pkee’. But this is the same ‘word’! We need a way of counting the number of combinations where order matters but there is repetition.

The number of ways of counting the number of arrangements from n total objects without replacement and there are repetitions is

$$\frac{n!}{n_1!n_2!\cdots n_k!},$$

where there are n_1 objects alike, n_2 objects alike, $\ldots$, n_k objects alike.

Example

How many ‘words’ can be made using the letters from the word ‘statistics’?

Example

How many 'words' can be made using the letters from the word 'statistics'?

There are 10 letters. There are 3 s's and t's, and also 2 i's. So here we have three repeating letters. Using the notation before, we need n_1, n_2, n_3, where $n_1 = 3$, $n_2 = 3$, and $n_3 = 2$. Then

$$\frac{10!}{3!3!2!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot \cancel{3!}}{3!\cancel{3!}2!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot \cancel{6}\ ^{2} \cdot 5 \cdot \cancel{4}}{(\cancel{3} \cdot \cancel{2}) \cdot \cancel{2}} = 50,400$$

Combination: Counting objects where order **does not matter**. The number of ways of arranging k things from n objects, where order does not matter, is

$${}_nC_k := \frac{n!}{(n-k)!\,k!}$$

Combination: Counting objects where order **does not matter**. The number of ways of arranging k things from n objects, where order does not matter, is

$${}_{n}C_{k} := \frac{n!}{(n-k)!\,k!}$$

Note: This is also denoted by what is called the *binomial coefficient*,

$$\binom{n}{k}$$

We say this "n choose k."

Example

How many committees can be made from a group of 5 people?

Example

How many committees can be made from a group of 5 people?

Notice that order does not matter. If a committee is made of Bob, Ashley, and Alice is the same committee as Ashley, Alice, and Bob.

$$\frac{5}{3} = {}_5C_3 = \frac{5!}{(5-3)!\,2!} = \frac{5 \cdot 4 \cdot 3 \cdot \cancel{2!}}{2!\,\cancel{2!}} = \frac{5 \cdot 4 \cdot 3}{2} = 5 \cdot 2 \cdot 3 = 30$$

Combination Formula

Why does this formula work? Suppose we want to choose k things from n objects, order not being important. We already know there are ${}_nP_k$ ways of choosing them if order mattered. This will choose the group of k objects for us. But we want to consider all the possible ways of arranging them as the same. There are $k!$ ways of arranging k objects (if order were to matter). So we must divide by this number, then we have

$$\frac{{}_nP_k}{k!} = \frac{n!}{(n-k!)\,k!} = {}_nC_k = \binom{n}{k}$$

Example

How many ways can you give 3 gifts to 8 people?

Example

How many ways can you give 3 gifts to 8 people?

Notice that order does not matter. Giving a gift to Bob, Ashley, and Alice is the same as giving the gift to Ashley, Alice, and Bob.

$$\binom{8}{3} = {}_8C_3 = \frac{8!}{(8-3)!\,3!} = \frac{8!}{5!\,3!} = 56$$

Note that problems can use both permutations and combinations!

Example

From 8 people, how many ways can you choose a committee of 4 people, designating one of them as president of the committee?

Example

From 8 people, how many ways can you choose a committee of 4 people, designating one of them as president of the committee?

There are many approaches. Let’s rephrase the problem as doing the following: choosing a committee of 4 people, then choosing a president among those 4. Then we are counting the number of ways of choosing a committee of 4 **AND** choosing a president from the 4.

$$ {}_8C_4 \cdot {}_4P_1 = \frac{4!}{(8-4)!\,4!} \cdot \frac{4!}{(4-1)!} = 70 \cdot 4 = 280 $$

Example (Again)

From 8 people, how many ways can you choose a committee of 4 people, designating one of them as president of the committee?

Example (Again)

From 8 people, how many ways can you choose a committee of 4 people, designating one of them as president of the committee?

Let's rephrase the problem as doing the following: choosing a president, then choose 3 other people to be on the committee along with the president. Then we are counting the number of ways of choosing a president **AND** choosing a committee of 3 people from the 7 remaining people.

$$ {}_8P_1 \cdot {}_7C_3 = \frac{8!}{(8-1)!} \cdot \frac{7!}{(7-3)!\,3!} = 8 \cdot 35 = 280 $$

The word emphasis shows a useful counting tool:

AND:	Multiply
OR:	Add

With the caveat that with OR you need to be sure that you are not double counting.

There will be ‘themes’ to counting problems that typically (at this level) force them into a few archetypes:

1. AND/OR Logic
2. Trees
3. Venn Diagrams
4. Tables

Summary

Order/Replacement	Without	With
Does (Permutation)	$\frac{n!}{(n-k)!}$	n^k
Does Not (Combination)	$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!\,k!}$

Summary

Order/Replacement	Without	With
Does (Permutation)	$\frac{n!}{(n-k)!}$	n^k
Does Not (Combination)	$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!\,k!}$

Note: We already saw how two of these formulas work (the without replacement). Can you explain why the other two (with replacement) are what they are?