Ch. 2 - Measurement & Stats

- [2-day version]

Measurement & Stats

- Day 1
  - Measurement Scales
  - Why numbers?
  - Distribution & Graphs : Histogram
  - Central Tendency
  - Mean, SoR, SSR, Variance, Standard Deviation
  - Start Exercise 1
- Day 2
  - Population vs. Sample
  - Precision vs. Accuracy
  - Logic and Logical Fallacies
  - Percentile Rank, Norms
  - Z-score exercise (in class)

Measurement Scales

- Nominal
- Ordinal
- Interval
- Ratio

Nominal Scale

- Nominal: Name or ID only
  - red, blue, green....
    john, tony, fred...
    Sci2-243, Sci2-245...
  - does not signify Ordering, Ranking, or More/Less
  - Gotcha: even if used with Numbers it may be still a Nominal.
  - Example: colors, names, room numbers, ID numbers

Ordinal Scale

- Ordinal : ordering
  - first, second, third….
    1, 2, 3...
    A, B, C...
  - signifies Order, but can't assume distance between items is the same, e.g. the difference between an A and a B may be much different than a B and a C
  - Example: Class Rank, Assignment Grade, Product Ratings

Interval Scale

- Interval: specifies orders AND inter-item distance
  - -3, -2, -1, 0, 1, 2, 3.... 100, 105, 115
  - the difference between two numbers IS the same, e.g. 100 to 105 should be the same amount as 105 to 110
  - Does NOT have an absolute zero.
  - Example: temperature in Degrees Farenheit

Ratio Scale

- Ratio: specifies orders AND inter-item distance and has absolute zero
  - 0, 1, 2, 3.... 100, 105, 115
  - the difference between two numbers IS the same, e.g. 100 to 105 should be the same amount as 105 to 110
  - Does have an absolute zero.
  - Example: temperature in Degrees Kelvin

Measurement Scales

	Magnitude	Equal Intervals	Absolute Zero
Nominal			
Ordinal	✓		
Interval	✓	✓	
Ratio	✓	✓	✓

Scales: Practical Info

- Nominal Scale: common
  - common stats: Count, Frequency, Mode
- Ordinal Scale: less common
  - stats: specialized "nonparametric" techniques required
- Ratio and Interval: common
  - Often can be treated identically with same statistical techniques

Basic Statistics

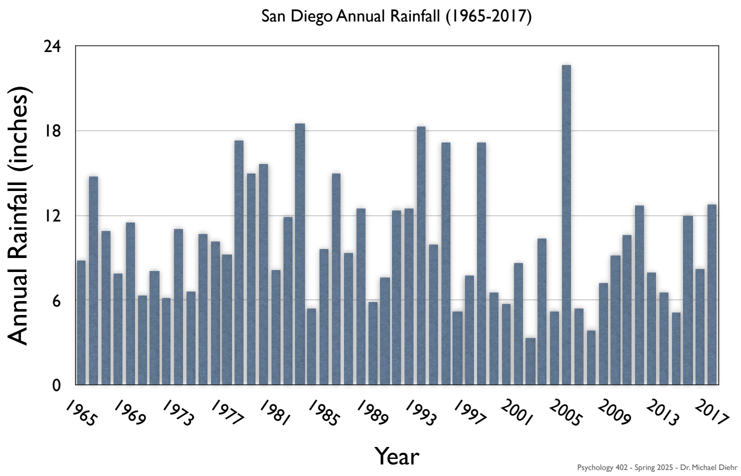
- Why use numbers?
  - Pros:
    - convenient, succinct
    - universal
    - well-defined, repeatable
  - Cons:
    - precision vs. accuracy
    - numerical fallacy

Tabular Data

Table 1:
San Diego Annual Rainfall (in Inches) by Year

Year	Rainfall (inches)
1965	8.81
1966	14.76
1967	10.86
1968	7.86
1969	11.48

Data Distributions

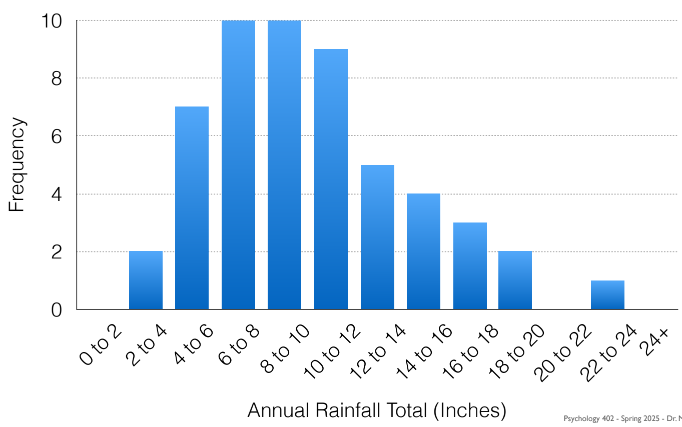


Histogram

- Frequency Distribution
- Invented by Karl Pearson
- Shows data from *one* variable only
- Data is (often) collected into groups ("bins")

Histogram

Figure 1
Frequency Distribution of San Diego Annual Rainfall



Describing Distributions

- Why? Large lists are inconvenient. Reduce many data points to a few numbers.
- Issue: Reducing data ("Degrees of freedom") : throws away data.
- We are modeling our data using a simplification.
- "All models are wrong, some models are useful"
- Simple vs. Simplistic?

Descriptive Statistics

- Statistical Assumptions: When these are not met, weird things happen.
- Joe Smith is 6 feet tall, his child is 1 foot tall. Thus, the average height in the Smith household is 3.5 feet.
- If you are sitting in bar, and Bill Gates walks in, suddenly everyone in the bar is (on average) a multi-millionaire.

Alternative Notations

- Square Root (x) sqrt(X) $\sqrt{X}$ $\sqrt[2]{X}$
- X-Squared X^2 X**2 X^2
- Sum(x) $x_1+x_2+x_3$... $\sum_{i=1}^{N} x_i$ $\sum x$
- Mean M $\frac{\sum x}{N}$ $\bar{X}$

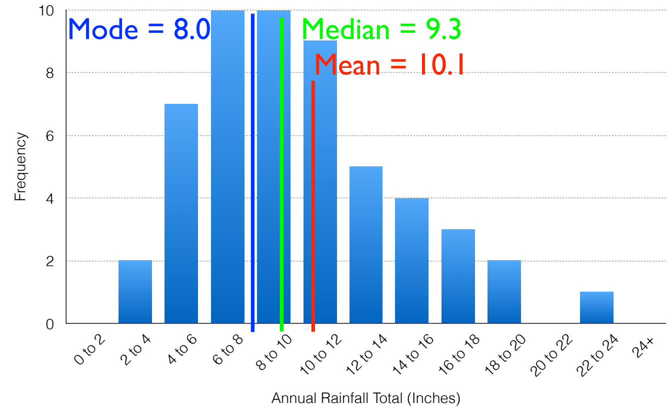
Central Tendencies

- Values tend to cluster around a point.
- **Mean** : most common statistic, commonly referred to as the "average". Formula ∑X / N
- **Mode**: the most common value in a set
  - rare to use in statistics
- **Median**: the middle-most value in a set
  - the value at which half are above and half are below. Aka the 50th percentile.

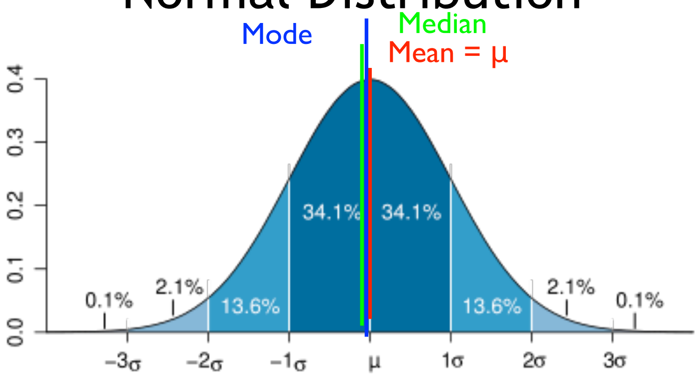
Histogram

Figure 1

Frequency Distribution of San Diego Annual Rainfall

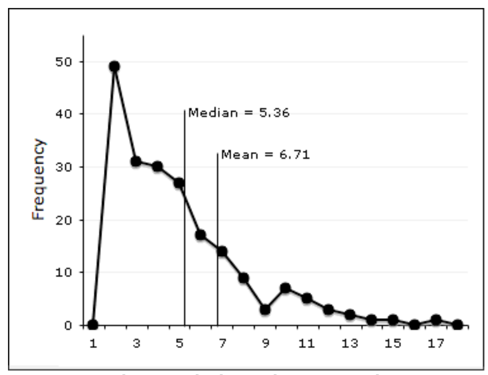


Normal Distribution



In a normal distribution, the mean, mode, and median are all the same

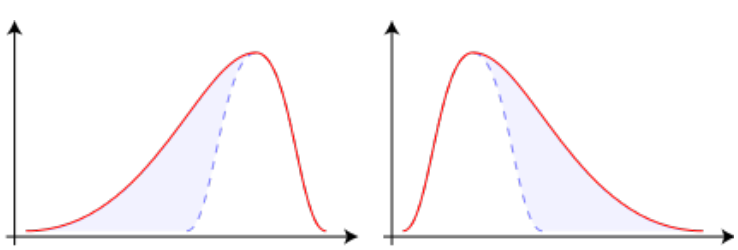
Skewed Distribution



In a skewed distribution, the mean, mode, and median are all often different

Skew

- negative skew : fatter tail on the left
- positive skew : fatter tail on the right



Negative Skew
"Left Skew"

Positive Skew
"Right Skew"

Measures of Central Tendency 1

	Description	Algorithm	Formula
Mean	the "average"	sum values, divide by N	$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$
Median	the "middle-most value"	sort values, find middle value	50th percentile
Mode	the "most common" value	find most frequent value	…

Measures of Central Tendency 2

Behavior:	Normal Distribution	Skewed Distribution
Mean	same	overly affected by outliers
Median	same	fairly resistant to outliers
Mode	same	resistant to outliers

Measures of Dispersion 1

- Compare each measured value to the average
- "for a typical value, how far away is it from the mean"
- "Difference score" or "residual" can be calculated as the difference between the actual score and the mean. In other words, $d_i = x_i - \bar{X}$
- Take the average (mean) of the difference scores.
- Average difference score = Sum(d) / N

Average Difference Score

	Score (x)	Mean ($\bar{X}$)	Difference d = (x - $\bar{X}$)
	2	6	-4
	3	6	-3
	9	6	3
	11	6	5
	14	6	8
	1	6	-5
	6	6	0
	4	6	-2
	5	6	-1
	5	6	-1
Sum	**60**	**60**	**0**
Mean	**6**	**6**	**0**

Sum of Residuals

- Given N samples of x : $x_1, x_2, x_3 \dots x_N$
- mean of x $$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$
- residuals $$d_i = x_i - \bar{x}$$
- Sum of Residuals is always zero

$$\sum_{i=1}^{N} d_i = 0$$

Sum of Residuals

- The "average difference score" score will *always* equal zero
- Solution:
  - Square the residuals *before* adding: removes the negative values.
  - "SSR" or Sum of Squared Residuals

SSR: Sum of Squared Residuals

- Given N samples of X: $x_1, x_2, x_3 \dots x_N$
- mean of x $$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$
- residuals $$d_i = x_i - \bar{x}$$
- Sum of Squared Residuals (SSR)

$$SSR = \sum_{i=1}^{N} (d_i)^2$$

Problems with SSR

- SSR depends on units of measurement:
  - a meter is 1000 millimeters, so SSR will be 1000*1000 = one million times higher when using meters vs. millimeters
- SSR depends on N (# of samples)
  - Doubling N will cause SSR to double (roughly)
- Therefore, SSR is hard to understand:
  - is SSR = 0.00342 high or low?
  - is SSR = 2343249 high or low?

Remove the influence of N

- The Sum of a set of values depends on the number (N) of values:
- $\sum_{i=1}^{N} x_i$
- Take the average (mean)
  - this divides by N
  - removes the influence of N

Variance

- Problem: SSR depends on N
- Solution: Take the average of SSR to remove the influence of N
- The average of the squared residuals is called Variance (S^2)

Variance

- Variance = SSR/N
- Variance = mean of squared residuals

$$S^2 = \frac{\sum_{i=1}^{N} (d_i)^2}{N}$$

$$S^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$

Problems with Variance

- Units are squared:
  - measuring height in meters?
    variance is meters^2
  - measuring # of cupcakes eaten?
    variance is (# of cupcakes eaten)^2
- Won't someone rid me of these meddlesome squared units?

Standard Deviation

- Improving on Variance:
- The square root of Variance (S^2) gives S, which is called "Standard Deviation".
- Also abbreviated SD, StdDev or σ (Greek letter sigma), or sometimes just "S"
- SD : easier to understand because it's in the same units as your measurement.
- SD is a unique property of the normal distribution -- given a mean and a SD you have uniquely specified the distribution

Standard Deviation

- SD = Square root of Variance

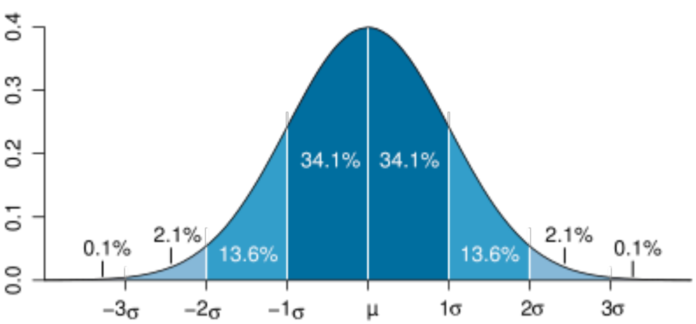
$$S = \sqrt{\frac{\sum_{i=1}^{N} (d_i)^2}{N}}$$

$$S = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}}$$

Standard Deviation

- can be thought of as the "average deviation"
- (but it's not literally average deviation, since we showed earlier the average difference score is always Zero)
- Technically:
  - (in a normal distribution) scores will be within plus or minus 1 SD about 68% of the time

Normal Distribution

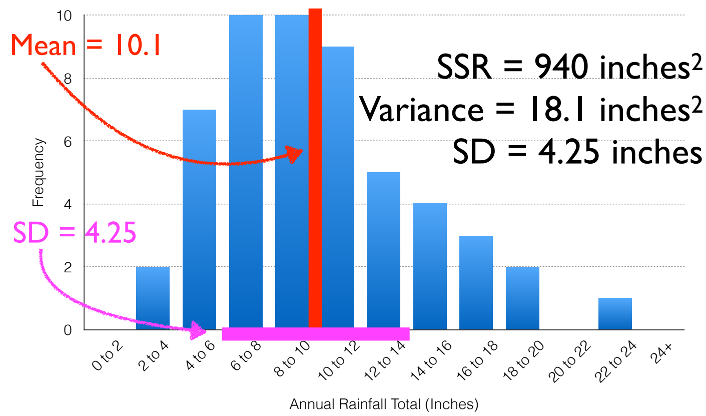


In a normal distribution, about 68.2% of values fall within ±1 SD

SSR, Variance and SD

Figure 1

Frequency Distribution of San Diego Annual Rainfall



Central Limit Theorem

- No matter the shape of the Population distribution, if you take enough (*) samples of the mean, the distribution of your samples of the mean will have a Normal distribution
- Central Limit Theorem Exercise (Javascript)
- This fact makes our life easy: Many statistics assume a normal distribution. The CLT provides us a normal distribution in most cases, even when the population data is skewed

Exercise: normal distribution

- Roll one 10-sided die 10 times and record the results
- Prediction
  - Your Distribution: Uniform (flat)
  - Mean : 4.5
  - Class Distribution: ???
- hint: What is N? # die rolls, # of students?
- List and Graph results
- Does the distribution look normal?
  - if so, why?

Exercise 1: Die Rolls

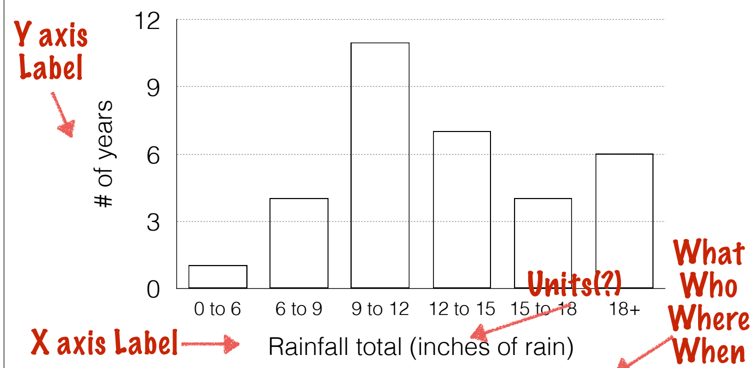
Trial #	X_i	M	d = $(x_i - M)$	$(\text{residual})^2$
1	2	3.5	-1.5	2.25
2	4	3.5	0.5	0.25
3	6	3.5	2.5	6.25
4	2	3.5	-1.5	2.25
Sum:	14			

N # of rolls	$M = \bar{X}$		Σ Residuals	$\Sigma (\text{residual}^2)$	$\frac{\Sigma (\text{residual}^2)}{N-1}$	$\sqrt{S^2}$
4	3.5		0.0	11.0	3.67	1.92

Example of APA-7 style Histogram

Figure 1

Frequency Distribution of Annual Rainfall in San Diego



Note. Annual rainfall for the years 1970-2002, measured at Lindberg Airport, by water year (October-September). Note the somewhat bi-modal distribution, with both 9 to 12 and 18 or more inches being most common.

Exercise: normal distribution 2

- Compute Mean ($\bar{X}$) - is it near 4.5?
- Compute residuals
- Compute sum of residuals -- do they add to zero?
- Compute squared residuals
- Compute Sum of squared residuals (SSR)
- Divide SSR by (N-1) - this is Variance or (S^2)
- Take square root of variance - this is S or Standard Deviation
- For this exercise, SD should be near 2.8

Ch. 2 - Part 2

Population vs. Sample

- Ideally, measure *everyone* to get the exact value (*Population parameter)*
- Practically, this is impossible.
- Take samples instead, and calculate the *Sample statistic*.
- The "Law of Large Numbers", "Sampling Theory", "Central Limit Theorem" makes life easier
- Central Limit Theorem Exercise (Javascript)
- Some formulas differ for *Population* vs. *Sample* (divide by N or divide by N-1 ?)

Population v. Sample

	Population	**Sample**
Definition	the entire set of items	the actual subset you measured
Descriptives	"Parameters"	"Statistics"
Symbols	Greek	Roman
Mean	μ	$\bar{X}$
Std. Deviation	σ	S
Variance	σ^2	S^2
Divide by	N	N-1

Law of Large Numbers

- If you take enough* samples, the sample mean approaches the population mean.
- Example: a coin has two sides. If heads=1 and tails = 0, then the average expected result is exactly 50% Heads (0.5) in the long run.
- However, if you flip a coin just a few times, getting exactly 0.5 is not likely.
- The LLN states that you will if you take enough samples.

* what is "enough"? Rule of thumb : 100.

LLN Demonstration

- Law of Large Numbers
- Demonstration with Coin Flips

Logical Fallacies

Logical Arguments

- Logical arguments or inferences generally have several components:
  - Premises
  - Conclusions
- Example:
  - Premise: All English people are musicians
  - Premise: John Lennon was English
  - Conclusion: John Lennon was a musician

Logical Arguments 2

- An Inference can be either Valid or Invalid -- this refers to the Structure of the argument (not the Facts themselves)
  - All A are B
    All C are A
    All C are B
- A Valid inference can still come to a false conclusion, and vice-versa

Logical Fallacies

- A Logical Fallacy generally means that your inference is *Invalid*
- In addition, your facts may or may not be true, but the flaw in reasoning has occurred before you even apply facts.

- Example: Affirming the consequent
- If P, then Q — bank owners are rich
  Q is true — Bill Gates is rich
  Therefore P — Bill Gates owns a bank

Biased Sample

- Every individual x that we have seen from sample X has characteristic Z
  Therefore ALL X have characteristic Z

- Every student I talk to in this class is interested in Psychology
  Therefore, ALL students are interested in Psychology

Nominal and Numeric

- Nominal Fallacy: The tendency to believe that something has a name or identification, it exists or has special meaning.
  “I am sleepy” vs. “I am suffering from activity-induced-rest-reduction-performance-impairment syndrome”
- Numerical Fallacy: belief that something has been measured and assigned a number, it actually exists. “I’m really sad” vs. “I scored a 32 on the Beck Depression Inventory”

Reification Fallacy

- To Reify - to make something more concrete or real

- Examples:
  - “An A student”
  - “High IQ”
  - “Top of the class”
  - “A F Grade”

Ranking Fallacy

- Reducing a complex phenomenon (e.g. intelligence), giving it a single number (reification) and then ordering based on that number

- Examples:
  - A IQ of 93 is better than an IQ of 90
  - An income of $50,000 is better than $45,000

Correlation = Causation

Other Fallacies

- Begging the question -- circular argument
- Post hoc ergo propter hoc (*after* this, therefore *because* of this)
- Appeal to Authority
- False Dilemma (Black & White thinking)

Significant Figures

- “digits of precision” or “sig. fig.”
- 1.3 has 2 sig. fig.
- 1.3455 has 5 sig. fig.

- More significant figures —> more precision

Precision vs. Accuracy

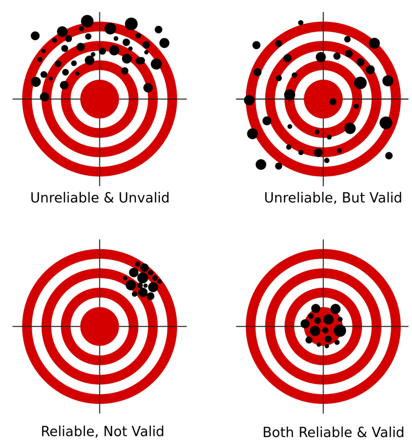
- Precision : the level of detail a measurement is made with, often specified with an error-range
  - “about 6 feet plus or minus 1 foot” vs. “6 foot 11 inches plus or minus 1 inch”
- Accuracy: how close the measured value is to the actual value, does it “hit the target”
- Think arrow vs. shotgun
- A number can be precise and accurate, precise but inaccurate, or accurate but imprecise.

Precision Fallacy

- A number that is *precise* may seem to be *accurate* when it is not
- A measurement that is *reliable* may seem to have *validity* when it does not

Precision vs. Accuracy

- Target shooting analogy
- Similar to Reliability vs. Validity



Fallacies re: Probability

- Classical
- Gambler’s Fallacy
- Bayesian Reasoning

9 Heads in a row

- You are flipping a coin, and get 9 heads in a row
  H H H H H H H H H
- What is the % chance the next flip will be a H ?

- Three common answers:
  - 50/50
  - more likely Heads
  - more likely Tails

9 Heads: Classical Inference

- Coin flips are independent 50/50 events, therefore 50% : Logical/Statistical
- This is the *correct* answer *for a fair coin*

9 Heads: Gambler’s Fallacy

- Coin flips are independent 50/50 events, but we just saw 9/10 heads, therefore a Tail is “due”
- This is the “Gambler’s Fallacy” and one reason Casinos make tons of money. The reasoning is faulty.
- Note: when dealing with draws w/o replacement, this logic is *correct*. For example, a single-card blackjack deck -- if no face cards have come up after 30 cards, then face cards are due

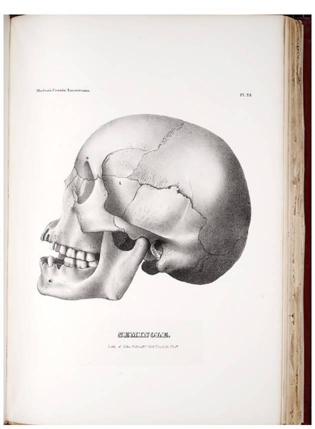
9 Heads: Bayesian Statistics

- Coin flips are supposed to be 50/50 events, but we just saw 9/10 heads, therefore the data is telling us that perhaps this is not a fair coin.
- Bayes’ theorem suggests you evaluate the prior probabilities in determining future behavior
- In this case, you’d conclude that Head is <u>more</u> likely on the next flip

Cognitive Biases of Discrimination

- Which cognitive biases (logical fallacies) are involved in racism, sexism and other bigoted beliefs?

Crania Americana



Samuel George Morton
1839

Morton’s Data as printed

Race	N	Cranial Volume Mean
Caucasian	52	87
Mongolian	10	83
American	144	82
Malay	18	81
Ethiopian	29	78

Data, corrected

Race	Mean (Morton)	Mean (corrected)
Caucasian	87	87
Mongolian	83	87
American	82	86
Malay	81	85
Ethiopian	78	83

Seed vs. Shot

Race	Difference (seed - shot)
Caucasian	1.8
Mongolian	n/a
American	2.2
Malay	n/a
Ethiopian	5.4

Morton's errors

- Fundamental arithmetic errors
- Data selection errors
- Failure to measure or control for external variables (biological sex, body size, etc.)
- Basic Statistical errors (averaging measurements from unequal size subgroups)
- The racist thumb press?

- Is he a liar? Conscious or subconscious?

Morton's Model

Internal vs. External Validity

- Internal Validity - how did it work?
  - were the methods good
  - did the IV cause the DV

- External Validity - what does it mean?
  - does skull size indicate IQ?
  - does IQ indicate personal worth?

Comparing Scores

- Compare a single score to the population
- One way: difference scores

- Problem: Is a difference of "3" big or little? On a 100 point test it's not very large, but on a 10 point test it's the difference between an A and a C

Comparing Scores

- Desire a system independent of the raw score units (just like letter grades)
- Two methods:
  - Ranks & Percentile Ranks...
  - Standard Scores...

Ranks, Percentiles

- Given a distribution of scores, and a single score
- **Rank** = the item # of the single score when sorted high to low
- **Percentile Rank** = the % of scores which are lower than the given score
- **Percentile** = the score at which a given percent of scores are below a given score
- Note: “**Percentile**” often used informally to mean “**Percentile Rank**”

Rank & Percentile

- Coronavirus Deaths
- Total deaths, per million people, as of September 2020
- Sort low to high

Country	Score
Mozambique	0.9
China	3.0
Ethiopia	8.0
Japan	11.0
Zambia	16.0
Colombia	424.0
France	471.0
Sweden	577.0
USA	584.0
Bolivia	599.0

Rank & Percentile

- Determine Rank #

Country	Score	Rank
Mozambique	0.9	1
China	3.0	2
Ethiopia	8.0	3
Japan	11.0	4
Zambia	16.0	5
Colombia	424.0	6
France	471.0	7
Sweden	577.0	8
USA	584.0	9
Bolivia	599.0	10

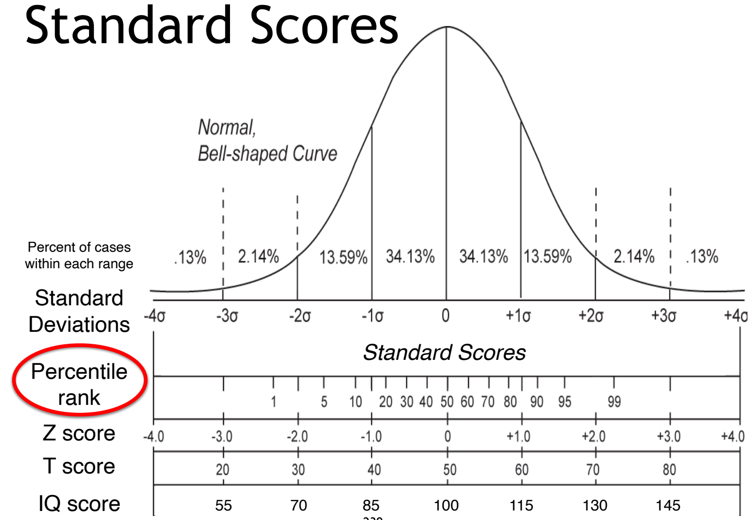
Percentile Rank

Percentile Rank = # of cases with worse value divided by # of cases

e.g. France is 7th of 10 (it has 3 cases with worse values)
3 / 10 = 30 % percentile rank

Country	Score	Rank	%ile Rank
Mozambique	0.9	1	90
China	3.0	2	80
Ethiopia	8.0	3	70
Japan	11.0	4	60
Zambia	16.0	5	50
Colombia	424.0	6	40
France	471.0	7	30
Sweden	577.0	8	20
USA	584.0	9	10
Bolivia	599.0	10	0

Standard Scores



Standard Scores 2

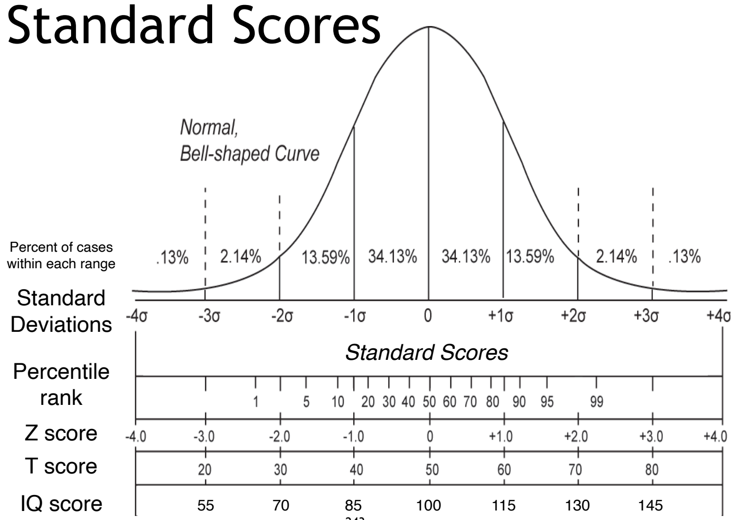
- Use the mean and standard deviation as points of reference.
- Standard score : distance from the mean, scaled by standard deviation
- Not affected by raw score units.
- Different standard scores mean the same thing, but are expressed differently.
  - just like how 1.0 and 100% mean the same thing
- Unfortunately, there are several different Standard Score systems!

Z-score

- A Z score is the # of standard deviations above (+) or below (-) the mean of a single measurement.
- Algorithm: given a single score (X_i), subtract the mean M, divide by the standard deviation S
- Formula
  - Z = (X-M)/SD

$$Z_i = \frac{X_i - \overline{X}}{S}$$

Standard Scores



Standard Scores: Z, T, IQ

	Z scores	T scores	IQ scores
Mean	0	50	100
SD	1	10	15
Example: top 3%			
Example: top 1%			
Formula: from Z Score	Z	(Z*10)+50	(Z*15)+100

z-Score (X-X̄)/s	T-Score 10z + 50	Wechsler IQ (15z + 100)	Percentile Rank
3.0	80	145	99.9
2.9	79	144	99.8
2.8	78	142	99.7
2.7	77	141	99.6
2.6	76	139	99.5
2.5	75	138	99.4
2.4	74	136	99.2
2.3	73	135	98.9
2.2	72	133	98.6
2.1	71	132	98.2
2.0	70	130	97.7
1.9	69	129	97.1

Norms 1

- Standard Scores provide us with a way of describing how a particular score relates to others in the population.
- Describing how an individual score relates to the population, which we assume are "normal".
- Terms "normative data" and "norms"
- Key questions: What is the normative group? What features or factors of the group may affect scores?

Z-score Exercise

- This is for practice, not graded for points
- PDF is on class website