Chapter 1

Key Terms

- Data: contains information and disinformation
- Statistics: the art and science of making sense out of data
- Population: all the individual data elements that matter for the question
- Census: all the data/information needed for the population
- Sample: subset of the population of interest

Scientific Method=Statistical Method (Fisher)

- Empirical approach to science
- Hypothesis—Design a Study—Data—Analyze—Conclusions—Back to Beginning

Correct Data can also be disinformation

More Key Terms

- Parameter: a numerical characteristic of a population; measure alongside a population
- Statistic: a numerical characteristic of a sample; measured alongside a sample

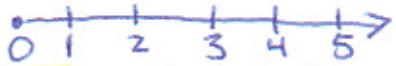
Determining in the same way, ex.: max, min, average

P-P S-S

-what is being measured, not necessarily the value

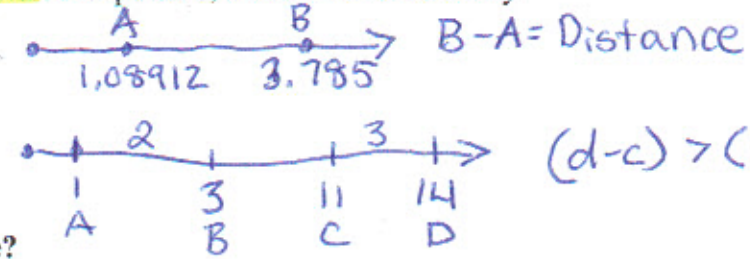
Types of Data

- Quantitative: can be measured; has amount and magnitude
  - Discrete: counts can be mapped to the integers, related to counting



  - Continuous: speed, velocity; hits everything; every value has a neighbor; can be measured

  - Interval: comparable, treated mathematically



Ratios make sense?

- If there is a natural zero, then rations make sense and the data are called ratio data

- Something that isn't ratio data is temperature

Interval vs. Ratio

- If quantitative is at least interval, then it is ration too if there is a natural zero

Goal of a Sample is to make it look like the population

Types of Studies

- Observational: get data without bothering the world; observe it in its natural state
- Experimental: get data from response you stimulate; treat and respond
- Retrospective/Case Control: look back in time and see how they got there; health history
- Prospective/Longitudinal: look at the future
- Cross Sectional: look at now

Experiment Designs

- Treatment is applied to the Experimental Unit
- Placebo (effect): example of control treatment (baseline); the benefit that comes from something that's not really doing anything; the actual treatment must be better
- Blinding: recipient of treatment doesn't know what they're getting
- Double blinding: person administering treatment doesn't know what the experimental unit is getting either
- Random: all alternatives are equally likely
- Replication: repetition of an experiment on more than one subject to legitimize the initial results
- Blocking: controlling confounding from happening
- Confounding: when 2 effects are mixed up to where you can't discern what's causing something
- Block: a group of experimental units that have something naturally in common

Sampling Techniques

- Simple Random Sample: ensures that every possible sample (of size n) and each experimental unit must be independent of every other one (individually picked on their own)
- Stratified (Random) Sample: separate the population of experimental units into groups, strata, blocks
    - Stratify, then random sample from each stratum
- Systematic Sampling: a system is devised that has a good chance of getting a representative sample
- Cluster sampling: once you get a cluster at random they are all in or all out; not independently selected
- Convenience sampling: by definition biased; self-selection

Chapter 2

Simple random also means each pick is independent of each other (independence of items involved)

Mean of frequency $\frac{\Sigma(f \cdot x)}{\Sigma f}$

Descriptive Statistics

- Distribution: the nature and shape of the spread of the data over the values

Midpoints
1.3
1.8
2.3
2.8

boundaries
1.55
2.05
2.55

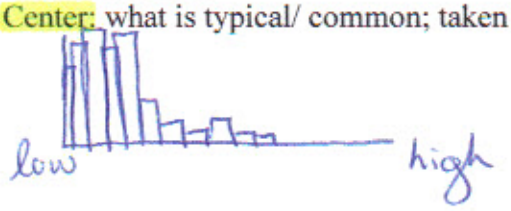
Class Limits Lower	Upper	Tally	Frequency	Relative Frequency	Cumm. Freq.	Cumm. Rel. Freq.
1.1	1.5	I	1	1/9	1	1/9
1.6	2.0	III	3	3/9	4	4/9
2.1	2.5	III	3	3/9	7	7/9
2.6	3.0	II	2	2/9	9	9/9
				Freq / n	Add Up	Cumm. / n

- Class boundaries: separate one class from the other; halfway between the upper class limit and the next lower class limit; *equal to or lower then stay down, bigger then go up*
- Class Width: the distance from one boundary to the next, from one upper limit to the next, or from one lower limit to the next
- Class Midpoint: (Upper limit plus lower limit)/2

Chapter 3

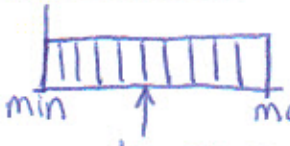
Measures of the Center

- Center: what is typical/ common; taken by mean, median, mode, and midrange



Skewed to the right

- Uniform distribution



midrange good here since it's not skewed

- Mode: value that occurs most often; can be multiple

Most Common ①

- Mean: average, add all numbers together and divide by how many numbers there are in the set

$$\frac{\Sigma x}{n}$$

$$\{x: x_1, x_2, x_3, \ldots x_n\}$$

$$\{x: x_{(1)}, x_{(2)}, x_{(3)}, \ldots x_{(n)}\}$$ ordered

min max

2nd Most Common Homes ②

- Median: the value in the middle when the data are in sorted order
- Midrange: use when desperate, (min plus max)/2

$$\frac{(\text{min}+\text{max})}{2}$$

Measures of Variation

- Variation: the flipside of uniformity; gets in the way of our understanding; want to clarify or understand why it's happening
- Range: not very useful, max minus min max-min
- Standard Deviation: most common measure of variation, the mean distance of the values from the center; how far the data is spread out from the middle (mean)

① Most Common

statistic $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

parameter σ

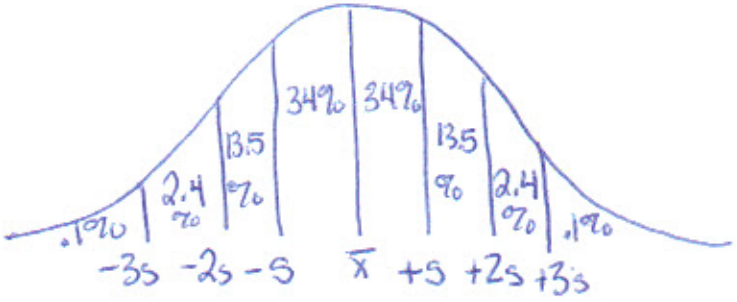
- Variance: standard deviation squared s^2 or σ^2 $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

What is Standard Deviation Good For?

1. Range Rule

$$s \approx \frac{\text{Range}}{4} = \frac{\text{max-min}}{4}$$

2. Empirical Rule (Only for Bell Curves) pg 106



a. 1 standard deviation contains 68% of the data
b. 2 contains 95% of data
c. 3 contains 99.7%

Measures of Relative Standing (position)

- Z scores
  - Concept of unusual
  - Unusual when the absolute value of the score if bigger than 2

statistic $\frac{x - \bar{x}}{s}$ parameter $\frac{x - \mu}{\sigma}$

  - Largest absolute value is most unusual

- Percentiles

- Z scores tell position above or below the means as measured in standard deviation units, percentiles tell you where in the order of the data the value falls (from smallest to largest)

$P_K = X$; K^{th} percentile = value X

Types of Percentile Problems

- You know the value of x, want to know k

$$K = \left[\frac{\#\text{ of values} < X}{\text{total \# of values}}\right] \times 100$$

Write as P_K = Value (Given)

- You know k and want to know x (what is the location, has decimal, round up)

$$\text{location} = \left(\frac{K}{100}\right) n$$

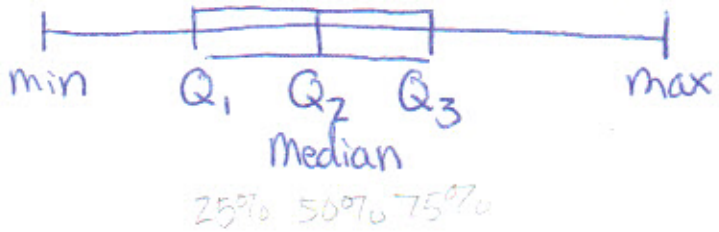
write as $P_{\text{Given K}} = $ location

no decimal, always that + next 1 up

What does a percentile do? The 17th percentile separates the lower 17 percent from the rest

Box Plots

- Use 5 values, found with the same percentile formulas as above



Chapter 4

Probability Terms

- Or=Addition Rule; And=Multiplication Rule
- Event: something that might happen

Coin Flip: H or T ; Dice Roll: 1 → 6 etc.

- Simple Event: an outcome or event that can't be broken down; can happen in only one exact way; roll dice and get sum=12
- Compound Event: can happen in more than 1 way; flip coin twice and get at least 1 H
- Sample Space: the collection of all possible simple events

Flip 2 Coins. What is the sample space

	H	T
H	HH	TH
T	HT	TT

or HH, TH, HT, TT

Notation

Events: A, B, C

Prob.: $P(A) = \frac{\#\text{ ways it can happen}}{\#\text{ ways anything}}$

p = probability

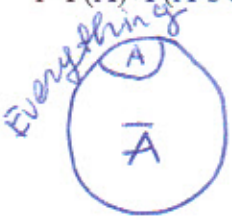
$0 \le p \le 1$ 0 = impossible, 1 = certain

1. **Relative frequency** approach to Probability
   a. You have all the data but inconvenient to use precise data
   b. Reggie Miller makes 94% of his free throws historically, what is the likelihood that he makes his next one. 94%
   c. Uses the recent past to estimate the probability in the near future

2. **Classical Approach** to Probability
   a. Requires equally likely outcomes

Addition Rule: P(_or_) – P (Both)
P (A or B) means if A happens then true, if B happens then true, if A and B happen then true
Two events are disjoint if there is no overlap

Compliment Rule
1=P(A)+P(A's compliment, everything that is not A)



$\bar{A}$ (A Compliment)
- Easier to solve problems

Multiplication Rule
Key word 'and'

P(Head & Head)

	H	T
H	HH	TH
T	HT	TT

$P(H_1) = 1/2$

$P(H_2) = 1/2$

$1/2 \cdot 1/2 = 1/4$

Conditional Probability

P(B given A) restricts the world to when A must happen

$$P(B|A) = \frac{Prob(A \& B)}{P(A)}$$

Independence, statistically speaking or probabilistically speaking

Age	Root Beer	Soft	Water	
≤ 10	85	10	5	100
> 10	20	78	2	100

$P(RB|\leq 10) = \frac{85}{100}$

$P(RB|>10) = \frac{20}{100}$ Dependent on age

Conditional Probability Formula

$$P(B|A) = \frac{P(B \& A)}{P(A)}$$

Ways to Count

- Approach probability through relative frequency
- Count all ways for A=Success (s)
- Count all ways=n
- Probability of Even A=s/n

1. **Fundamental Counting rule**
   a. Event A can happen in m ways
   b. Event B can happen in n ways
   c. Then A and B can happen in (m)(n) ways

Coin Flip (2 ways) x Dice Roll (6 ways)

D L L L D D D

10 2 6 2 6 2 6 10 10 10

2. **Factorial Rule**
   a. The number of ways to arrange n distinct items using all of them, is n!
   b. Sequences, Put them in Order, Arrangements, Permutations

8 Books arranged on shelf

$8! = 40,320$

3. **Permutations Rule**
   a. N distinct items with t number of ways to choose the n items and arrange them
   b. Care about the order within the arrangement 1,2,3 is not the same a 2,3,1

$_nP_r$ (n: # of items, r: # ways)

10 books, shelf holds 6 = $_{10}P_r = 151,200$

4. **Combinations Rule**
   a. How many different groups, combinations, sets, can you make
   b. Order doesn't matter, 1,2,3 is the same as 2,3,1

$_nC_r$

8 books, choose 6

$_8C_6 = 28$

Chapter 5

Discrete Probability Distributions

- Discrete: x is a counting number, only positive integers
- Random variable: governed by probability; reason for different values is random by probability

- Proper Probability Distribution: the probability of all possible outcomes is always equal to and never exceeds 1

X	P(x)	x*P(x)	P(x) $(x-\mu)^2$
1	.11	0.11	.677
2	.17	0.34	.372
3	.21	0.63	.348
4	.26	1.04	.070
5	.14	0.70	.323
6	.11	0.66	.698
$\Sigma P(x)=1$ ✓		$\Sigma = 3.48 = \mu$	$\Sigma = 2.188 = \sigma^2$

$\sigma = \sqrt{2.188}$
1.480

$\mu = x \cdot P(x)$ $\quad \sigma^2 = P(x)(x-\mu)^2$ $\quad \sigma = \sqrt{\sigma^2}$

Binomial Distribution

- Can only get 1 of 2 outcomes (gender at birth, coin flip, right/wrong)
- Will be implied in the problem

Success $P(S) = p$

Failure $P(F) = (1-p) = q$

n = # trial
x = # of successes
P(x) = prob of exactly x successes in n trials

⇒ Quiz 7 Problem 3

Exam w/ 10 multiple choice (4 possibles) Prob. exact 4 right

(x) $P(X=4)$

$${}_{10}C_4 (.25)^4 (.75)^6 = .146$$

$\mu_x = np$

$\sigma_x = \sqrt{npq}$