

PROB/STAT

Name _____

Unit 1 Introduction to Statistics

		8/30 First day "stuff"	8/31 <u>Intro Activity:</u> <i>Love is Blind Study</i> <i>Clock Buddies</i> <u>HW:</u> read and take notes section 1.1 including examples	9/1 Section 1.1 Intro to Stats notes review <u>HW:</u> pg 6 #1-18,21-31 odd, 34,35
9/4 Labor Day No school	9/5 Section 1.2 Data Classification notes <u>HW:</u> pg 13. #7-20,28 read and take notes section 1.3	9/6 no A <i>Come in for xtra help!</i>	9/7 Check HW Note quiz <i>Case Study: Rating TV shows (pg 15) (clock buddies)</i>	9/8 Quiz Review <u>On your own:</u> Read/notes pg 23 <i>Uses and Abuses</i> - anytime during week- have it read by Friday
9/11 Quiz 1.1,1.2 Section 1.3 Experimental Design* <u>HW:</u> Read polling article Pg 23	9/12 Finish Notes <u>Classwork/HW:</u> pg 21 #1-24 all,28	9/13 no A <i>Come in for xtra help!</i>	9/14 <i>Rolling Down the River Activity</i>	9/15 (Quiz 1.3) Whey Protein Article <i>"Hooked on a Feeling", "How does Gallup polling work"</i> <u>HW:</u> <i>Read given articles and answer questions</i>
9/18 <i>Paired Article Activity</i> (project grade- due at end of class)	9/19 <i>Finish paired activity</i> Begin Problem set (Project grade)	9/20 no A <i>Come in for xtra help!</i>	9/21 Work on Problem set	9/22 ½ day Problem Set Due Chapter Review
9/25 Unit Review	9/26 CH 1 TEST	9/27 No class	9/28 BEGIN UNIT 2	9/29 ½ day

Why Take Statistics?

The Philosophical:

Statistics deals with how we can learn about the world from observations when those observations are fallible. It teaches us to work with randomness and make it a tool for discovering the unknown rather than something to fear. The reasoning of statistics is the foundation of scientific reasoning. When examined closely, it is amazing that we actually can draw conclusions from a random sample or fallible experiment. If nature were just a bit more intransigent, we'd be unable to get anywhere. There is something quite amazing and beautiful in reasoning and methods that tread so close to the edge of the impossible and yet are so fundamental to scientific progress.

The Political:

Your parents may have told you that their statistics course was the worst experience of their academic lives, but things are different now. The AP course is a modern course that focuses on data rather than on probability theory and combinatorics. We use calculators and computers and deal with real-world problems. We use discovery learning methods and multimedia materials. I guarantee that your experience will be nothing like what your parents suffered through.

The Practical:

Statistics is used by charlatans to fool the masses. Statistical literacy is good self-defense in a world of advertising and political claims that only seem to be based on data. This course will arm you with the necessary understanding so you can smell a statistically-disguised lie a mile away. And think of the money you'll save not playing the lottery.

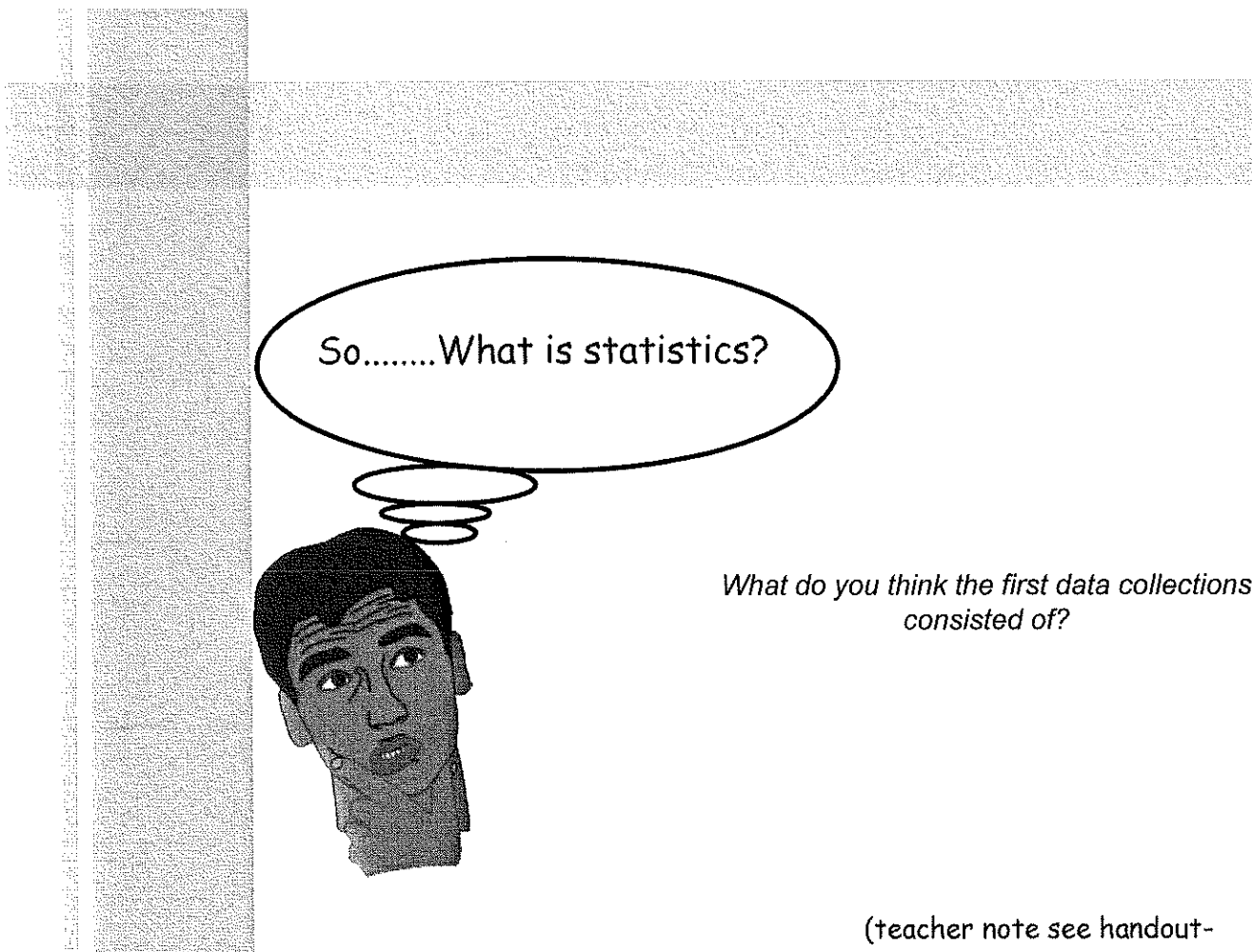
The Crass:

1) Statistics is one of those courses that will get you a job and/or a promotion. When we survey the companies that hire our graduates (I teach at Cornell) they tell us that they'd like our students to know more statistics and more computing. When we survey our recent graduates (5-years out), they tell us that statistics was one of the most important courses to them in getting their current positions.

2) Statistics is required by virtually every social science major, engineering, pre-med, and many others. It is now required for most law degrees. AP Statistics credit is therefore a useful commodity -- it will save you time and money in college.

--Paul Velleman

From a post to the AP Stat List serve.



So.....What is statistics?



What do you think the first data collections consisted of?

(teacher note see handout-pdf)

What it entails:

- collecting data
- taking surveys
- describing populations
- analyzing data- in a VARIETY of ways

Can be difficult and costly. How?

Definition: *Statistics* is the science of collecting, organizing, analyzing, and interpreting data

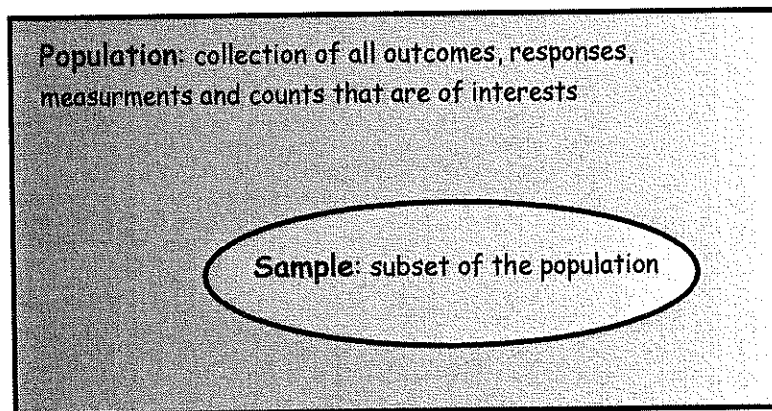
Definition: *Data* consists of information coming from observations, counts, measurements, or responses. (*datum*)

Data is presented in all kinds of ways

BTW: data collection=census



There are two types of data sets you will be studying: populations and samples



Telling the difference can **SOMETIMES** be tricky- at least for me.....it gets tricky when I try to **OVERTHINK** it...don't be me

Example:

In a recent survey, 4210 adults in the U.S. were asked if they own an ipad. 1234 of these people said yes. What is the populations and what is the sample?

Population:

Sample:

Data set:

Example:

In a recent survey, 235 students at HK were asked if they own a graphing calculator. 102 of them said yes.

Population:

Sample:

Data set:

Parameters and Statistics.... What's the diff?

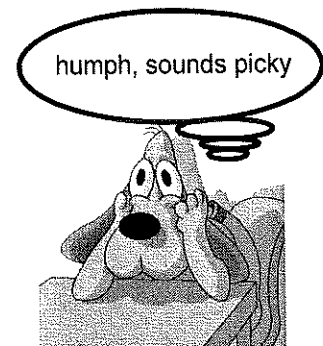
A *parameter* is a numerical description of a *population* characteristic.

A *statistic* is a numerical description of the *sample* characteristic.

Examples:

Tell if a P or an S

1. A recent survey of a sample of MBAs reports the average starting salary is less than \$65,000
2. Starting salaries for 667 MBA grads from Univ Chicago increased by 8.5% from prev year.



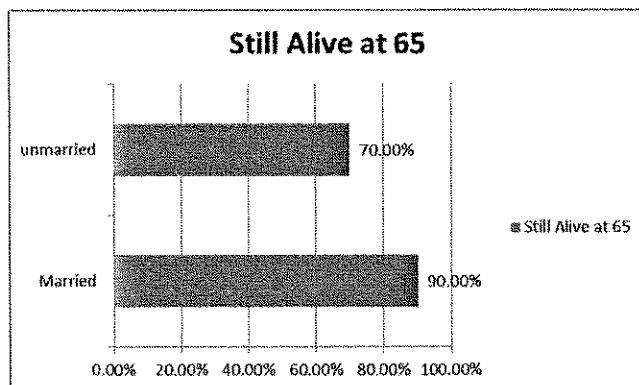
Branches of Statistics

THESE TWO BRANCHES CO-EXIST WITHIN STUDIES

Descriptive Statistics: organization, summarization, and display of data

Inferential Statistics: using samples to draw conclusions about populations (*one tool is probability*)

<http://www.nbcnews.com/id/44122528/ns/health-behavior/t/single-people-may-die-younger-new-study-finds/>



A large sample of men, aged 48, was studied for 18 years, Interpret the chart and answer questions.

What is the population?
 What is the Sample?
 What is the descriptive part of the study?
 What inferences could be made?

*What could be some extraneous variables?

Can inferences be 100% sure?

Example:
Click link

<http://www.forbes.com/sites/rickferri/2013/01/10/ts-official-gurus-cant-accurately-predict-ma>

Descriptive Branch?
Inferences?
Population?
Sample?



1.2 TYPES OF DATA LEVELS OF MEASUREMENT

So we have.....

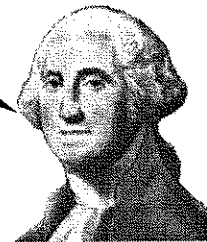
Qualitative Data

attributes, labels, non numerical entities



Quantitative Data

numerical measurements or counts



How about an example of what we are taking about:

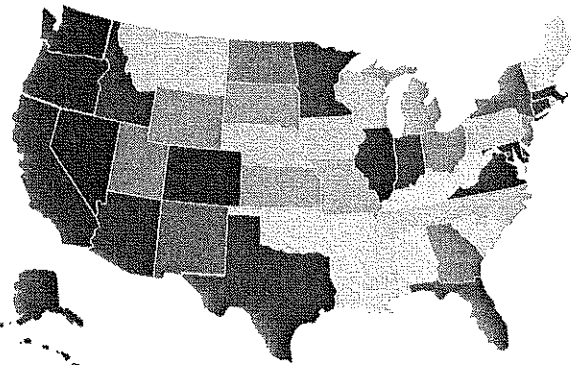
State	population
AL	4,822,023
CA	38,041,430
CT	3,590,347
FL	19,317,568

Quantitative?

Qualitative?

A map of which U.S. states have the most number of Starbucks (per-capita)

- #1 Washington
- #2 Nevada
- #3 Colorado
- #4 Oregon
- #5 California
- #6 Hawaii
- #7 Arizona
- #8 Alaska
- #9 Illinois
- #10 Virginia
- #11 Maryland
- #12 Idaho
- #13 Texas
- #14 Massachusetts
- #15 Minnesota
- #16 Indiana
- #17 Connecticut
- #18 Florida
- #19 New York
- #20 New Mexico
- #21 Delaware
- #22 North Dakota
- #23 Georgia
- #24 Utah
- #25 Wyoming



- #26 Ohio
- #27 South Dakota
- #28 New Jersey
- #29 Kansas
- #30 Michigan
- #31 Missouri
- #32 North Carolina
- #33 Wisconsin
- #34 Pennsylvania
- #35 Tennessee
- #36 Nebraska
- #37 Montana
- #38 Iowa
- #39 Rhode Island
- #40 Maine
- #41 South Carolina
- #42 Kentucky
- #43 Oklahoma
- #44 Louisiana
- #45 Alabama
- #46 New Hampshire
- #47 West Virginia
- #48 Mississippi
- #49 Arkansas
- #50 Arkansas

Created by [ilovecoffee.jp](http://en.ilovecoffee.jp)

<http://en.ilovecoffee.jp/>

LEVELS OF MEASUREMENT

nominal level of measurement are qualitative ***only***- NO mathematical computations can be made. (its more "informational")

ordinal level of measurement are qualitative and quantitative: data can be arranged in order but differences are not meaningful between entries.

five top ranked TV sitcoms 2012

5. *Parks and Recreation*
4. *Girls*
3. *Louie*
2. *Mad Men*
1. *Breaking Bad*

Interval level of measurement are quantitative. Data can be ordered, and calculate meaningful differences between data entries. Zero entry simply represents a position on a scale- *not an inherent zero*

Ratio level of measurement are similar to interval with added property that a zero entry IS an *inherent zero*. A ratio of two data values can be formed so one data value can be expressed as a multiple of another.

Inherent zero.....what is that?
its a zero that implies NONE.

the difference between these requires some application common sense to discern between them- refer to chart pag 12 in text

Heres some help for us!

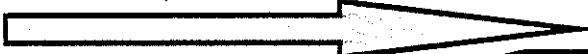
<http://www2.webster.edu/~woolfm/statwhatis.html>

<http://www.socialresearchmethods.net/kb/measlev1.php>

1.3 Experimental Design

Though you may never have to develop a statistical study, you will likely have to interpret results of one.

Designing Process:

1. Identify the variables of interest and population
2. Develop plan for collecting data.
- If using a sample, what is important?
3. Collect the data. 
4. Describe the data
5. interpret the data and make decisions
6. Identify possible errors

Gallup poll handout

Data Collection

- Observational study
- Perform experiment
- Simulation
- Survey

OBSERVATIONAL STUDY: researcher observes and measures characteristics of interest of PART of a specific populations

EXPERIMENT: a treatment is applied to part of a population (people or things), response to treatment is observed and recorded.

SIMULATION: mathematic or physical model used to reproduce conditions of a situation or process, collect data from this.

SURVEY: can be interview, mailer, telephone, Internet, etc.....
(this can introduce bias.)

A bit more about BIAS.

what will influence bias?

EX. Does either question introduce bias? If so, how?

A)The governor has appropriated transportation funds well in order to build CT's economy?

Agree or Disagree

OR

B)The governors appropriated spending of 3 Billion dollars on a bus line connecting New Britain to Hartford, instead of education, is a good way to build CT economy?

Agree or Disagree

Let's break out the
"Gallop Poll" and "How Polls are Conducted"
that you read for HW.

Sampling Techniques:

Random Sample: every member of population has an equal chance of being selected.

- **Simple Random Sample (SRS):** every possible sample of same size has same chance of being selected.

Using the table of Random Numbers in your book (appendix), How would you select a random sample of 12 students in our school population of 672 (this number is made up)

How do we do this with the calc? (MATH>PRB>RandInt(____,____,____))

What is a
biased sample?

Appendix B

Table 1—Random Numbers

92630	78240	19267	95457	53497	23894	37708	79862	76471	66418
79445	78735	71549	44843	26104	67318	00701	34986	66751	99723
59654	71966	27386	50004	05358	94031	29281	18544	52429	06080
31524	49587	76612	39789	13537	48086	59483	60680	84675	53014
06348	76938	90379	51392	55887	71015	09209	79157	24440	30244
28703	51709	94456	48396	73780	06436	86641	69239	57662	80181
68108	89266	94730	95761	75023	48464	65544	96583	18911	16391
99938	90704	93621	66330	33393	95261	95349	51769	91616	33238
91543	73196	34449	63513	83834	99411	58826	40456	69268	48562
42103	02781	73920	56297	72678	12249	25270	36678	21313	75767
17138	27584	25296	28387	51350	61664	37893	05363	44143	42677
28297	14280	54524	21618	95320	38174	60579	08089	94999	78460
09331	56712	51333	06289	75345	08811	82711	57392	25252	30333
31295	04204	93712	51287	05754	79396	87399	51773	33075	97061
36146	15560	27592	42089	99281	59640	15221	96079	09961	05371
29553	18432	13630	05529	02791	81017	49027	79031	50912	09399
23501	22642	63081	08191	89420	67800	55137	54707	32945	64522
57888	85846	67967	07835	11314	01545	48535	17142	08552	67457
55336	71264	88472	04334	63919	36394	11196	92470	70543	29776
10087	10072	55980	64688	68239	20461	89381	93809	00796	95945
34101	81277	66090	88872	37818	72142	67140	50785	21380	16703
53362	44940	60430	22834	14130	96593	23298	56203	92671	15925
82975	66158	84731	19436	55790	69229	28661	13675	99318	76873
54827	84673	22898	08094	14326	87038	42892	21127	30712	48489
25464	59098	27436	89421	80754	89924	19097	67737	80368	08795
67609	60214	41475	84950	40133	02546	09570	45682	50165	15609
44921	70924	61295	51137	47596	86735	35561	76649	18217	63446
33170	30972	98130	95828	49786	13301	36081	80761	33985	68621
84687	85445	06208	17654	51333	02878	35010	67578	61574	20749
71886	56450	36567	09395	96951	35507	17555	35212	69106	01679
00475	02224	74722	14721	40215	21351	08596	45625	83981	63748
25993	38881	68361	59560	41274	69742	40703	37993	03435	18873
92882	53178	99195	93803	56985	53089	15305	50522	55900	43026
25138	26810	07093	15677	60688	04410	24505	37890	67186	62829
84631	71882	12991	83028	82484	90339	91950	74579	03539	90122
34003	92326	12793	61453	48121	74271	28363	66561	75220	35908
53775	45749	05734	86169	42762	70175	97310	73894	88606	19994
59316	97885	72807	54966	60859	11932	35265	71601	55577	67715
20479	66557	50705	26999	09854	52591	14063	30214	19890	19292
86180	84931	25455	26044	02227	52015	21820	50599	51671	65411
21451	68001	72710	40261	61281	13172	63819	48970	51732	54113
98062	68375	80089	24135	72355	95428	11808	29740	81644	86610
01788	64429	14430	94575	75153	94576	61393	96192	03227	32258
62465	04841	43272	68702	01274	05437	22953	18946	99053	41690
94324	31089	84159	92933	99989	89500	91586	02802	69471	68274
05797	43984	21575	09908	70221	19791	51578	36432	33494	79888
10395	14289	52185	09721	25789	38562	54794	04897	59012	89251
35177	56986	25549	59730	64718	52630	31100	62384	49483	11409
25633	89619	75882	98256	02126	72099	57183	55887	09320	73463
16464	48280	94254	45777	45150	68865	11382	11782	22695	41988

Reprinted from *A Million Random Digits with 100,000 Normal Deviates* by the Rand Corporation (New York: The Free Press, 1955). Copyright 1955 and 1983 by the Rand Corporation. Used by permission.

Stratified Sample-

(strata=homogeneous grps-broken into common grps that are similar)- then put together-reflect whole pop.

similar characteristics: ethnicity, age, gender, grade level, socio-economic, political, etc

A simple random sample from each strata (don't have to be equal- proportional if some strata are bigger than others)

Cluster Sample-

when population falls into naturally occurring subgroups having similar characteristics of the population then this would be appropriate.

all clusters must have similar characteristics

what problems could arise?

Systematic Sample-

need all population available

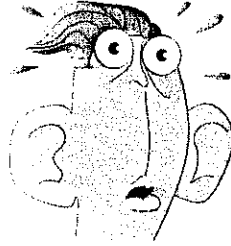
taking every "nth" person after picking a random beginning

Convenience Sample-

using only what's easily or readily available

why could this be an issue?

ACTIVITY TIME!!



River Farming
handout- work in
pairs

Types of Bias

Sampling Bias:

- *Selection Bias*- how was sample selected
- *Voluntary response Bias*-mass survey (no idea amount went out, responses are just from volunteer and uncontrolled "sample")
- *Non-response Bias*-know amount surveys that went out and know how many returned not returned

Non Sampling Bias:

- *Question wording* (may slant results)
- *Response*-lie, answer wrong, round off

Other important vocab not found in your book

Blind and double blind study

Blind: subjects (experimental units) do not know what treatment they are receiving or technician

whey protein
study

Double Blind: experimental unit as well as those administering do not know who is getting what treatments (coding)

placebo
article

As applied in EXPERIMENTS

Experimental Unit: *smallest unit for which a treatment is applied.*

Explanatory variables/Factors: *variables that have value and controlled by the experimenter. The thing that is done in the experiment.*

Response Variables: *measures an outcome of the experiment. variables that are not controlled by the experimenter but may be impacting the explanatory variables. (what other things could have caused x to occur in the experiment)*

Extraneous Variables: *may effect response but balanced out over the experiment (been accounted for)*

Confounding Variable: *variable that effects response and may effect overall results (not balanced throughout experiment)*

Blocking Variable: *may control outside variables known to be issue. may decrease bias*

Using the Whey Protein Article:

What are the experimental units?

What are the explanatory vars/factor(s)?

What is/are the response variable(s)?

Extraneous variable(s)?

Confounding Variable(s)?

Blocking Variable(s)