

Statistics 1

Terminology and types of data

Temporary Room Change

- 9/2/09-9/23/09 Meet in **Mendocino 2003**
(Computer Lab)
- Except:
 - 9-9:50 class 9/18 meet in BRH-104
 - 1-1:50 class 9/14 meet in RVR-1004
- Onecard to print, USB memory stick

Important websites

- Class info: <http://www.csus.edu/indiv/n/norrisa/>
- Homework: www.webassign.net
- R statistical Software www.r-project.org/ also add on the R package Rcmdr (optional)

Working Definition of Statistics

- Statistics: tools for making sense out of data
Statistician's motto: "Got Data?"
- Data may be from a population or a sample
 - Population consists of *all objects of interest* in a study
 - Sample is a *subset* of the population

Example: Popularity Study

Datafile Name: Popular Kids

Datafile Subjects: [Psychology](#) , [Social science](#)

Story Names: [Students' Goals](#) , [What Makes Kids Popular](#)

Reference: Chase, M. A., and Dummer, G. M. (1992), "The Role of Sports as a Social Determinant for Children," *Research Quarterly for Exercise and Sport*, 63, 418-424

Authorization: Contact authors

Description: Subjects were students in grades 4-6 from three school districts in Ingham and Clinton Counties, Michigan. Chase and Dummer stratified their sample, selecting students from urban, suburban and rural school districts with approximately 1/3 of their sample coming from each district. Students indicated whether good grades, athletic ability, or popularity was most important to them. They also ranked four factors: grades, sports, looks, and money, in order of their importance for popularity. The questionnaire also asked for gender, grade level, and other demographic information.

Number of cases: 478

Variable Names:

1. Gender: Boy or girl
2. Grade: 4, 5 or 6
3. Age: Age in years
4. Race: White, Other
5. Urban/Rural: Rural, Suburban, or Urban school district
6. School: Brentwood Elementary, Brentwood Middle, Ridge, Sand, Eureka, Brown, Main, Portage, Westdale Middle
7. Goals: Student's choice in the personal goals question where options were 1 = Make Good Grades, 2 = Be Popular, 3 = Be Good in Sports
8. Grades: Rank of "make good grades" (1=most important for popularity, 4=least important)
9. Sports: Rank of "being good at sports" (1=most important for popularity, 4=least important)
10. Looks: Rank of "being handsome or pretty" (1=most important for popularity, 4=least important)
11. Money: Rank of "having lots of money" (1=most important for popularity, 4=least important)

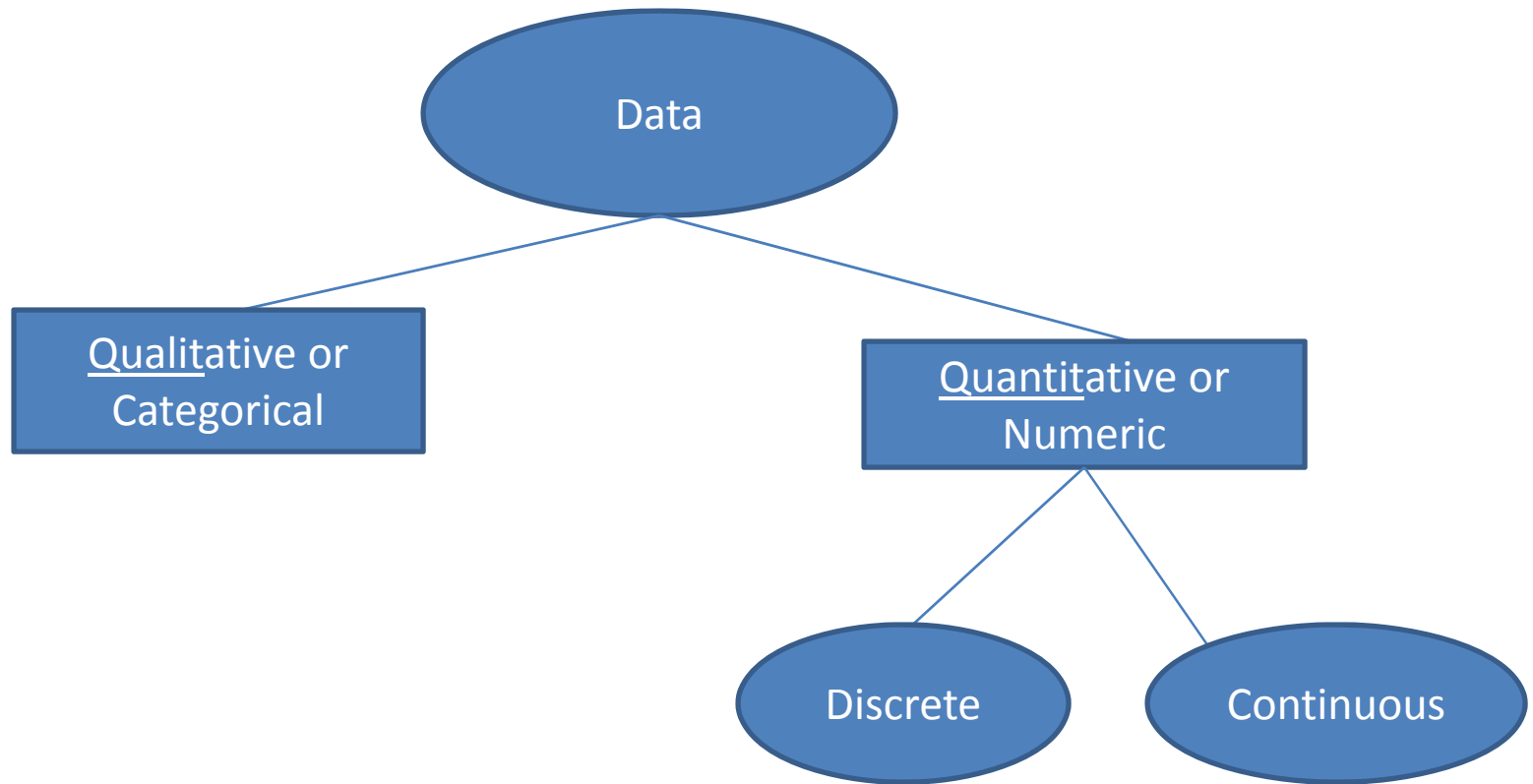
The Data:

Gender	Grade	Age	Race	Urban/Rural	School	Goals	Grades	Sports	Looks	Money
boy	5	11	White	Rural	Elm	Sports	1	2	4	3
boy	5	10	White	Rural	Elm	Popular	2	1	4	3
girl	5	11	White	Rural	Elm	Popular	4	3	1	2
girl	5	11	White	Rural	Elm	Popular	2	3	4	1
girl	5	10	White	Rural	Elm	Popular	4	2	1	3
girl	5	11	White	Rural	Elm	Popular	4	2	1	3

Example: Popularity Study

- What is the population of interest?
- What is the sample?
- What variables are observed for each student?

Types of Data



Types of Data

Determine whether each variable is qualitative, quantitative-discrete or quantitative-continuous

Variable Names:

1. Gender: Boy or girl
2. Grade: 4, 5 or 6
3. Age: Age in years
4. Race: White, Other
5. Urban/Rural: Rural, Suburban, or Urban school district
6. School: Brentwood Elementary, Brentwood Middle, Ridge, Sand, Eureka, Brown, Main, Portage, Westdale Middle
7. Goals: Student's choice in the personal goals question where options were 1 = Make Good Grades, 2 = Be Popular, 3 = Be Good in Sports
8. Grades: Rank of "make good grades" (1=most important for popularity, 4=least important)
9. Sports: Rank of "being good at sports" (1=most important for popularity, 4=least important)
10. Looks: Rank of "being handsome or pretty" (1=most important for popularity, 4=least important)
11. Money: Rank of "having lots of money" (1=most important for popularity, 4=least important)

Statistics: Definition and Branches

Statistics: methods for organizing, summarizing and drawing conclusions from data

Three branches of Statistics:

- Descriptive Statistics: using tables, graphs and numbers to organize and summarize data
- Probability: calculating the likelihood of chance events
- Inferential Statistics: methods for drawing conclusions about a population based on sample data

Examples

- Descriptive: summarize students goals from Popularity study

	Grades	Popular	Sports
Percent of Respondents	52%	29%	19%

- Probability: When you toss a fair coin, the probability of heads is $\frac{1}{2}$
- Inferential Statistics (Demo):
Estimate the proportion of students in this class who have posted a message on Facebook. Use a sample of size 10.

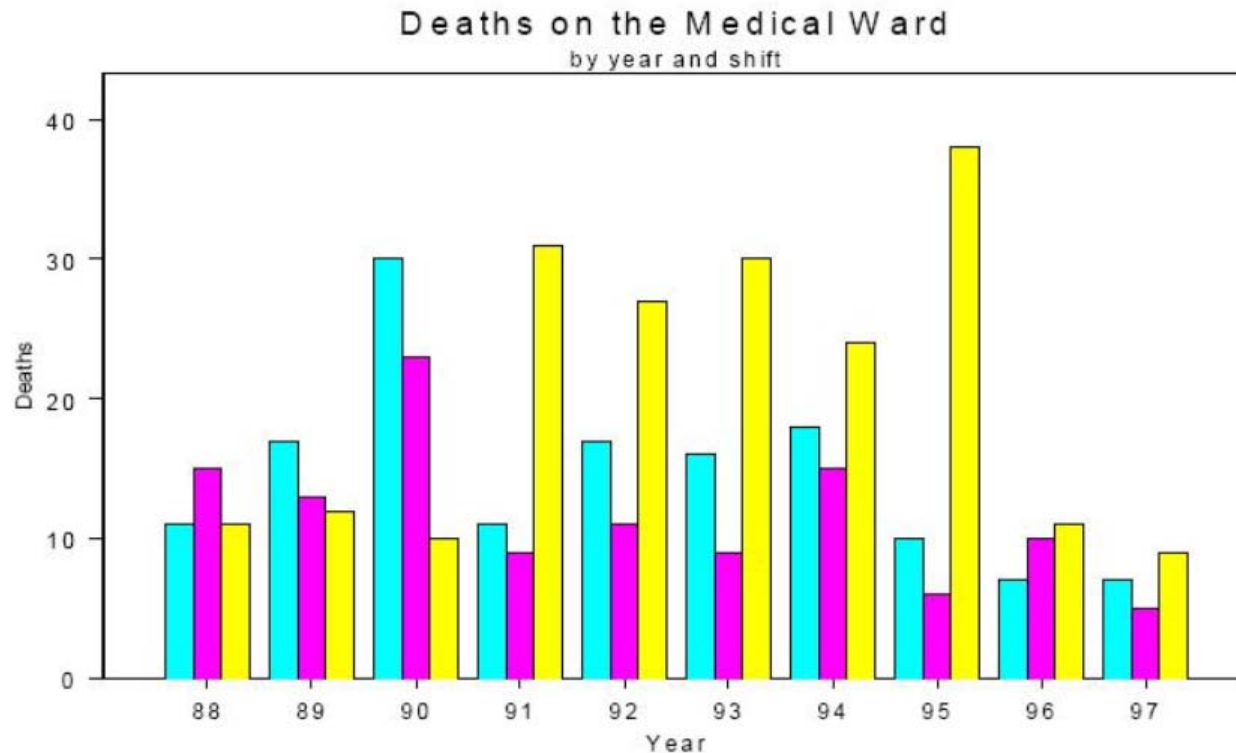
Applications of Statistics: Is this nurse a serial murderer?

- Kristen Gilbert was a nurse who worked in the Veterans hospital in Northampton, Massachusetts
- She was respected for her skill in crisis situations
- Was often first to sound “code blue” for cardiac emergency and skill at administering the epinephrine shot
- Suspicions by coworkers – too many “code blues” during her shift, many patients died
- Not mercy killings – patients were expected to survive

US vs Gilbert

- What evidence would you look for to decide if something “fishy” was going on?
- The evidence for the indictment was statistical (Dr. Gehlbach, Univ. of Mass.)!

United State vs. Kristen Gilbert



Display 1: The pattern of deaths, by year and by shift
In each set of three bars, left = night (midnight - 8 am), middle = day (8 am - 4 pm), right = evening (4 pm - midnight).