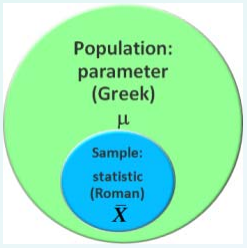
INTRODUCTION TO BIOSTATISTICS
FOR GRADUATE AND MEDICAL STUDENTS

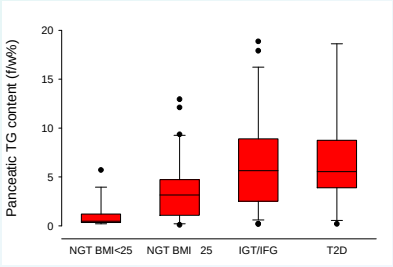


- Introduce fundamental statistical principles
- Cover a variety of topics used in biomedical publications
  - Design of studies
  - Analysis of data
- Focus on interpretation of statistical tests
  - Less focus on mathematical formulas

June 25, 2013

INTRODUCTION TO BIOSTATISTICS
FOR GRADUATE AND MEDICAL STUDENTS

Descriptive Statistics and Graphically Visualizing Data



Beverley Adams Huet, MS
Assistant Professor
Department of Clinical Sciences, Division of Biostatistics

June 25, 2013

Files for today (June 25)

- Lecture and handout (2 files)
  - Biostat_Huet1_25Jun2013.pdf (PPT presentation)
  - Biostat_handout_Altman_BMJ2006.pdf (Read article)
- Homework -- either handwritten paper or email OK
  - To be assigned Thursday

Contact information

beverley.huet@utsouthwestern.edu
Office E5.506
Phone 214-648-2788

"The best thing about being a statistician is that you get to play in everyone else's backyard."
John Tukey, Princeton University

Today's Outline

- Introduction
  - Statistics in medical research
- Types of data
  - Categorical
  - Continuous
  - Censored
- Descriptive statistics
  - Measures of Central Tendency

June 25, 2013

Statistics Information/Explanations

- The Little Handbook of Statistical Practice by Gerard E. Dallal, Ph.D
  http://www.tufts.edu/~gdallal/LHSP.HTM
- WISE: Web Interface for Statistical Education
  http://wise.cgu.edu/index.html
- New view of statistics
  http://www.sportsci.org/resource/stats/index.html

June 25, 2013

Links to on-line statistical calculators

For online (e.g., t-tests or chi-sq):

- GraphPad quick calcs
  http://www.graphpad.com/quickcalcs/
- OpenEpi
  http://www.openepi.com/OE2.3/Menu/OpenEpiMenu.htm
- SISA General simple statistics & sample size
  http://www.quantitativeskills.com/sisa/

June 25, 2013

Statistical and Graphics software (download at UTSW IR)

http://www.utsouthwestern.net/intranet/administration/information-resources/

Statistics and graphics software GraphPad Prism and SigmaPlot can be downloaded from the UTSW **Information Resources** INTRAnet

- **GraphPad Prism** (Mac and Windows)
- **SigmaPlot** (Windows)

June 25, 2013

Statistics in the medical literature

"Medical papers now frequently contain statistical analyses, and sometimes these analyses are correct, but the writers violate quite as often as before, the fundamental principles of statistical or of general logical reasoning."

Greenwood M. (1932) *Lancet*, I, 1269-70.

June 25, 2013

Statistics

"Statistics may be defined as a body of methods for making wise decisions in the face of uncertainty."

(W.A. Wallis)

Use data from **sample** to make **inferences** about a **population**

- Statistics is not just an extension of mathematics
  - Not akin to a cookbook.
  - Involves logic and judgment.
- Key concepts
  - **variability**
  - **bias**

June 25, 2013

Sources of Bias

- Wrong sample size
- Selection of study participants
- Non-responders
- Withdrawal
- Missing data
- Compliance
- Repeated peeks at accumulating data

Steps in a research study

Planning

Design

Execution (data collection)

Data management & processing

Data analysis

Presentation

Interpretation

Publication

Biostatistics

Applicable to

- Clinical research
- Basic science and laboratory research
- Epidemiological research

June 25, 2013

Role of a Biostatistician when planning a study

- Assess study design integrity, validity, biases, blinding
  - Is it analyzable?
- Power and sample size estimates
- Randomization schemas
- Analysis plans
- Data safety and monitoring
  - Interim analyses, stopping rules?

June 25, 2013

When to choose the statistical test? When to contact a Biostatistician?

BEFORE data is collected

The study design, sample size, and statistical analysis must be able to properly evaluate the research hypothesis set forth by the investigator

Why learn statistics?

Myth

"You can prove anything with statistics"

Fact

You cannot PROVE anything with statistics, just put limits on uncertainty

Why learn statistics?

Statistics pervades the medical literature (Colton, 1974).

- For properly conducting your own research
- Evaluate others' research
  - Many statistical design flaws and errors are still found in the medical literature

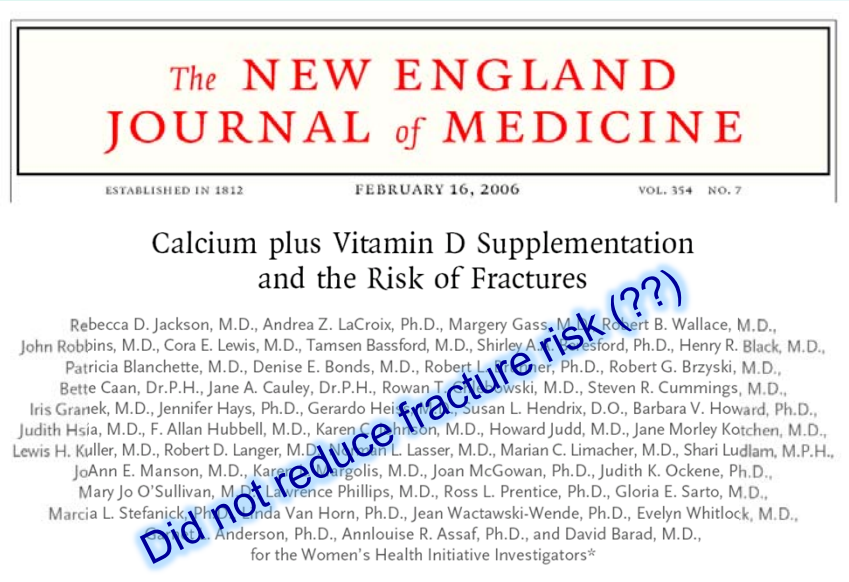
June 25, 2013

Clinical Trials: WHI

WOMEN'S HEALTH INITIATIVE

- 15 year $735 million study sponsored by the NIH
- 161,000 women ages 50-79, and is one of the largest programs of research on women's health ever undertaken in the U.S.

June 25, 2013



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 FEBRUARY 16, 2006 VOL. 354 NO. 7

Calcium plus Vitamin D Supplementation and the Risk of Fractures

Rebecca D. Jackson, M.D., Andrea Z. LaCroix, Ph.D., Margery Gass, M.[illegible] Robert B. Wallace, M.D., John Robbins, M.D., Cora E. Lewis, M.D., Tamsen Bassford, M.D., Shirley A[illegible]esford, Ph.D., Henry R. Black, M.D., Patricia Blanchette, M.D., Denise E. Bonds, M.D., Robert L[illegible]er, Ph.D., Robert G. Brzyski, M.D., Bette Caan, Dr.P.H., Jane A. Cauley, Dr.P.H., Rowan T[illegible]wski, M.D., Steven R. Cummings, M.D., Iris Granek, M.D., Jennifer Hays, Ph.D., Gerardo Hei[illegible], Susan L. Hendrix, D.O., Barbara V. Howard, Ph.D., Judith Hsia, M.D., F. Allan Hubbell, M.D., Karen C[illegible]hnson, M.D., Howard Judd, M.D., Jane Morley Kotchen, M.D., Lewis H. Kuller, M.D., Robert D. Langer, M.D.[illegible]n L. Lasser, M.D., Marian C. Limacher, M.D., Shari Ludlam, M.P.H., JoAnn E. Manson, M.D., Kare[illegible]rgolis, M.D., Joan McGowan, Ph.D., Judith K. Ockene, Ph.D., Mary Jo O'Sullivan, M.D., Lawrence Phillips, M.D., Ross L. Prentice, Ph.D., Gloria E. Sarto, M.D., Marcia L. Stefanick, Ph.[illegible]nda Van Horn, Ph.D., Jean Wactawski-Wende, Ph.D., Evelyn Whitlock, M.D., [illegible] Anderson, Ph.D., Annlouise R. Assaf, Ph.D., and David Barad, M.D., for the Women's Health Initiative Investigators*

Did not reduce fracture risk (??)

June 25, 2013

WHI (Women's Health Initiative)

15 year, $735 million study sponsored by the NIH

Calcium plus Vitamin D Supplementation and the Risk of Fractures. NEJM 2006;354:669-83

Inadequate design left many questions unanswered

- Significant limitations to the study including*
  - low dose of vitamin D
  - allowance of calcium and vitamin D supplements, and anti-osteoporotic medications *(Study of calcium and vitamin D versus MORE Calcium and vitamin D?)*
- The women enrolled were not at risk for fracture!!
  - Lower rate (about half) of hip fractures than expected and this decreased study power to <50% to show a significant finding.
    - low rates could be due to a number of factors
      - high BMD and BMI of participants
      - inclusion of relatively few women age > 70 years
      - many participants were already using calcium & vit D supplements, or were on HRT

* Courtesy of Naim Maalouf, MD, Dept Internal Medicine, UT Southwestern Medical Center

WHI (Women's Health Initiative)

Untangling Results of Women's Health Study

- Newspapers Examine Confusion Over Results Of Recent Women's Health Initiative Studies
- "toss out the calcium pills"
- "The Worrisome Calcium Lie…"

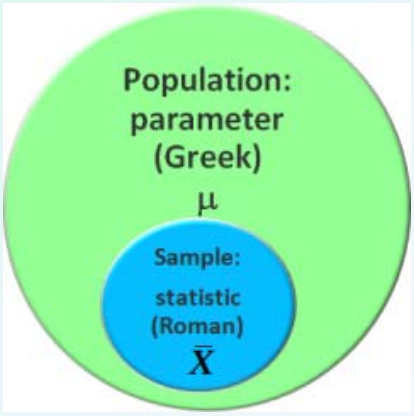
June 25, 2013

Statistics in the medical literature

- Errors in design and execution
- Errors in analysis
- Errors in presentation
- Errors in interpretation
- Errors in omission

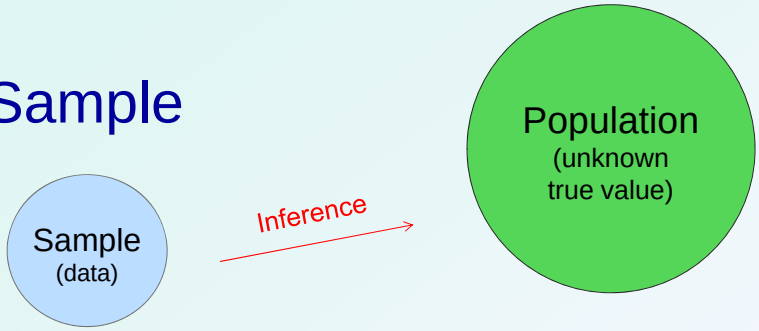
June 25, 2013

Statistics - notation



June 25, 2013

Sample

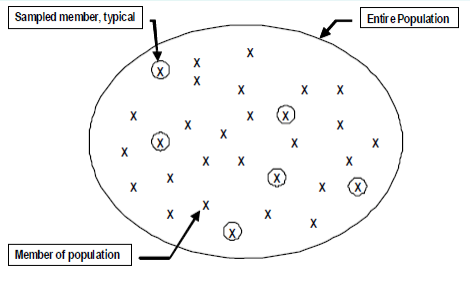


We use data from **sample** to make **inferences** about a **population**

June 25, 2013

Statistics

A sample is a set of observations drawn from a larger population.



- The **sample** is the numbers (data) collected.
- The **population** is the larger set from which the sample was taken; contains all the subjects of interest.

June 25, 2013

Types of Statistics

Descriptive statistics

Summary statistics used to organize and describe the data

Inferential statistics

Making decisions in the face of uncertainty

June 25, 2013

Types of Statistics

Descriptive statistics

Inferential statistics

Results From baseline to 18 weeks, dark chocolate intake reduced mean (SD) systolic BP by –2.9 (1.6) mm Hg (*P* < .001) and diastolic BP by –1.9 (1.0) mm Hg (*P* < .001)

JAMA. 2007;298:49-60.

Types of Statistics

Descriptive statistics

- Which summary statistics to use to organize and describe the data?
  - Proportion, mean, median, SD, percentiles
- Descriptive statistics do not generalize beyond the available data

Types of Statistics

Inferential statistics

- Generalize from the sample.
  - Hypothesis testing, confidence intervals
    - t-test, Fisher's Exact, ANOVA, survival analysis
    - Bayesian approaches
- Making decisions in the face of uncertainty

Types of Data

Variable – anything that varies within a set of data

- Mortality rates
- Survival time
- LDL cholesterol
- Surgery type
- Biopsy stage
- Compliance
- Marital status
- Age
- Weight
- Smoking status
- Adverse drug reaction
- Energy intake
- Parity
- Drug dose

Types of Data

Important in deciding which analysis methods will be appropriate

Categorical (qualitative) variables

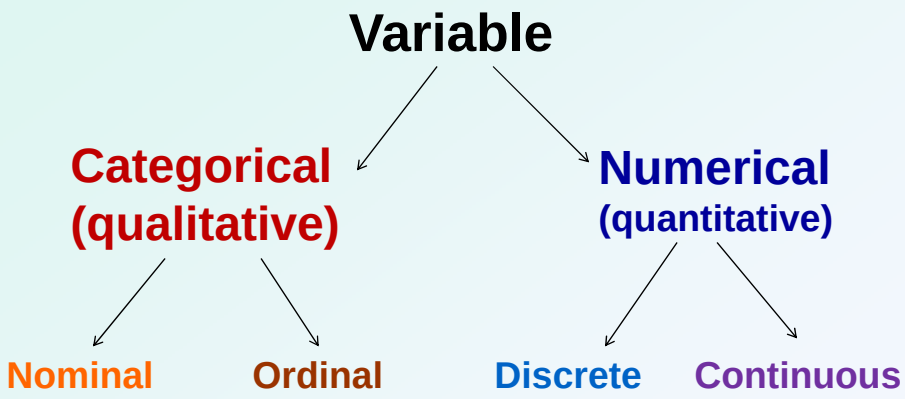
- Sex, ethnicity, smoker/non-smoker, blood type

Numerical (quantitative) variables are measured

- Age, weight, parity, triglycerides, tumor size

June 25, 2013

Types of variables



June 25, 2013

Categorical variables

Sex, race, compliance, adverse events, family history of diabetes, hypertension diagnosis, genotype

- Summarized as
  - Frequency counts, fractions, proportions, and/or percentages
- Graphically displayed as
  - Bar charts

Categorical variable

Nominal data - no natural ordering

- Gender
- Race/ethnicity
- Religion
- Yes/no
- Zip code, SSN

Summarizing categorical variables

Bar Graph

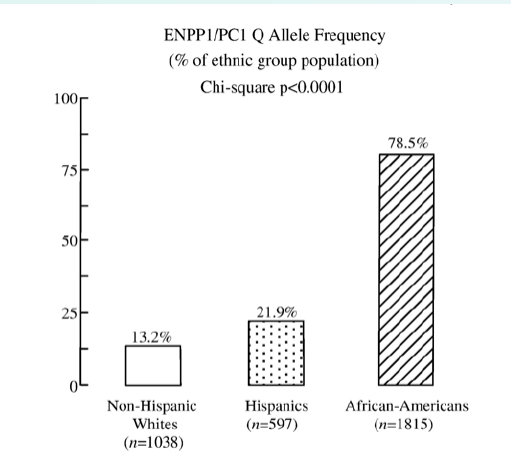


Fig. 1. ENPP1 121Q allele frequency for non-Hispanic Whites, Hispanics, and African-Americans.

M. Chandalia et al. / Journal of Diabetes and Its Complications 21 (2007) 143–148

June 25, 2013

Ordered categorical variable

Ordinal data – can be ranked

- Attitudes (strongly disagree, disagree, neutral, agree, strongly agree)
- Education (grade school, high school, college)
- Cancer stage I, II, III, IV
- Coffee – tall, grande, venti

June 25, 2013

Summarizing categorical variables

Table 1. Characteristics of the Participants in the Calcium with Vitamin D Trial at the Time of the WHI Screening, According to Randomly Assigned Group.*

Characteristic	Calcium + Vitamin D (N = 18,176)	Placebo (N = 18,106)
Age at screening		
Mean — yr	62.4±7.0	62.4±6.9
50 to 59 yr — no. (%)	6,728 (37.0)	6,694 (37.0)
60 to 69 yr — no. (%)	8,275 (45.5)	8,245 (45.5)
70 to 79 yr — no. (%)	3,173 (17.5)	3,167 (17.5)
Race or ethnic group — no. (%)†		
White	15,047 (82.8)	15,106 (83.4)
Black	1,682 (9.3)	1,635 (9.0)
Hispanic	789 (4.3)	718 (4.0)
American Indian or Native American	77 (0.4)	72 (0.4)
Asian or Pacific Islander	369 (2.0)	353 (1.9)
Unknown or not identified	212 (1.2)	222 (1.2)
Family history of fracture after 40 yr of age — no. (%)	6,835 (37.6)	6,692 (37.0)
History of fracture — no. (%)		
At any age	6,311 (34.7)	6,228 (34.4)
At age ≥55 yr	1,948 (10.7)	1,968 (10.9)
No. of falls in previous 12 mo — no. (%)		
None	11,193 (61.6)	11,200 (61.9)
1	3,421 (18.8)	3,386 (18.7)
2	1,462 (8.0)	1,426 (7.9)
≥3	732 (4.0)	701 (3.9)

Don't forget to report the denominators!

Frequency

Percent

Calcium plus Vitamin D Supplementation and the Risk of Fractures. NEJM 2006;354:669-83

June 25, 2013

Categorical data
Software output from SAS program

The FREQ Procedure

Sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
M	39	48.15	39	48.15
F	42	51.85	81	100.00

RaceEth	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AmerInd	2	2.47	2	2.47
Black	25	30.86	27	33.33
Hisp	42	51.85	69	85.19
White-NH	12	14.81	81	100.00

Cross tabulation

Frequency Percent Row Pct Col Pct	Table of Sex by RaceEth				
	RaceEth				
Sex	AmerInd	Black	Hisp	White-NH	Total
M	2 2.47 5.13 100.00	12 14.81 30.77 48.00	19 23.46 48.72 45.24	6 7.41 15.38 50.00	39 48.15
F	0 0.00 0.00 0.00	13 16.05 30.95 52.00	23 28.40 54.76 54.76	6 7.41 14.29 50.00	42 51.85
Total	2 2.47	25 30.86	42 51.85	12 14.81	81 100.00

June 25, 2013

Numerical data
Discrete numerical variables

Discrete - cannot take on all values within the limits of the variable

- Parity, gravidity (0, 1, 2, …)
- Number of deaths
- Number of abnormal cells

Numerical data
Continuous variables

Usually a *measurement*

- Age, weight, BMI, %body fat
- Cholesterol, glucose, insulin
- Prices, $
- Time of day or time of sample collection
- Temperature
  - In degrees Kelvin – ratio scale
  - in C or F – interval scale

Types of Data

ID	Sex	Ethnicity	Age_yrs	Height_ cm	Wt_kg	BMI	Heart Rate	Pain	Pain code
62401	F	Hisp	32	162.56	56.82	21.50	71	Mild	1
62402	F	AA	45	182.88	90.91	27.18	74	Moderate	2
62403	F	NHW	29	149.86	81.82	36.43	86	Severe	3
62404	M	AA	36	139.70	47.73	24.46	86	Severe	3
62405	M	NHW	41	187.96	88.64	25.09	62	Mild	1
62406	M	Hisp	52	180.34	106.82	32.84	76	Moderate	2

Nominal Nominal Nominal Continuous* Continuous Discrete* Ordinal Ordinal

*Though age at last birthday is discrete, treat age as a continuous variable

*analyze as if continuous

Continuous variables

Data entry note - height

ID	Height		Height_in	Height_cm
101	5'4"		64.00	162.56
102	6'		72.00	182.88
103	5'9"		59.00	149.86
104	5'5		55.00	139.70
105	6 2		74.00	187.96
106	5'11"		71.00	180.34
n			6	6
Mean			65.83	167.22
SD			7.73	19.64

Continuous variables

Data entry note

ID	Height_in	Height_cm	Wt_lb	Wt_kg	BMI
101	64.00	162.56	125.00	56.82	21.50
102	72.00	182.88	200.00	90.91	27.18
103	59.00	149.86	180.00	81.82	36.43
104	55.00	139.70	105.00	47.73	24.46
105	74.00	187.96	195.00	88.64	25.09
106	71.00	180.34	235.00	106.82	32.84
n	6	6	6	6	6
Mean	65.83	167.22	173.33	78.79	27.92
SD	7.73	19.64	49.06	22.30	5.63

BMI (body mass index) = weight (kg) / height (m^2)

June 25, 2013

Continuous variables

Data entry note – blood pressure

ID	BP	SBP	DBP
101	130/90	130	90
102	145/98	145	98
103	110/70	110	70
104	120/80	120	80
105	116/82	116	82
106	128/85	128	85
n	0	6	6
Mean	#DIV/0!	124.83	84.17
SD	#DIV/0!	12.37	9.47

June 25, 2013

Continuous variables

Use the actual data, avoid reducing continuous data to categorical data

Always record the actual value not a category

- Example
  record age 26 instead of a category such as
  - ❑ 20 – 30 years

Statistical analysis with continuous data is more powerful and often easier

June 25, 2013

Comparing two groups: BMI analyzed two ways

BMI_Group A	BMI_Group B
33.4867	30.1023
32.1351	38.2888
28.3923	32.9024
27.2876	33.9424
25.5880	34.6334
38.3914	29.4910
22.9572	37.7789
21.7224	40.3879
20.9584	21.5714
38.4195	28.5903
40.6966	29.6120
30.6242	34.0294
39.7852	34.2624
26.5991	38.7278
27.0852	44.0202
27.4631	34.7421
30.4258	37.1738
38.4931	24.7027
30.0664	40.0076
29.4561	32.3284
40.1199	29.4166
33.0703	40.3387
29.3968	39.6101
24.7864	
n	**24**	**23**
Mean	**30.7**	**34.2**
SD	**6.0**	**5.5**

T-test (comparing means)
p-value =0.044

Dichotomize: "Obese" BMI >30 kg/m^2

=12/24 =17/23

0.50 0.74

or 50% vs 74%

Fisher's Exact test

p-value= 0.135 *Less powerful analysis!*

Note: Do not round numbers until the final presentation

June 25, 2013

Continuous variables

Use the actual data, avoid reducing continuous data to categorical data

- Information is lost when a continuous variable is reduced to a categorical (dichotomous or ordinal)

See handout:

Douglas G Altman and Patrick Royston.
The cost of dichotomising continuous variables.
BMJ, May 2006; 332:1080.

Describing Continuous variables

- Summarize with
  - Means, medians, ranges, percentiles, standard deviation
- Numerous graphical approaches
  - Scatterplots, dot plots, box and whisker plots

HDL-C in control subjects and subjects with Type 2 diabetes (raw data)

ID	Group	HDL	ID	Group	HDL
732001	Control	51	732033	DM	42
732002	Control	46	732034	DM	40
732003	Control	47	732035	DM	44
732004	Control	48	732036	DM	45
732005	Control	54	732037	DM	38
732006	Control	47	732038	DM	41
732007	Control	45	732039	DM	40
732008	Control	52	732040	DM	43
732009	Control	50	732041	DM	36
732010	Control	52	732042	DM	41
732011	Control	46	732043	DM	38
732012	Control	42	732044	DM	40
732013	Control	50	732045	DM	35
732014	Control	47	732046	DM	38
732015	Control	44	732047	DM	41
732016	Control	40	732048	DM	40
732017	Control	49	732049	DM	42
732018	Control	40	732050	DM	36
732019	Control	45	732051	DM	40
732020	Control	45	732052	DM	38
732021	Control	45	732053	DM	33
732022	Control	42	732054	DM	36
732023	Control	46	732055	DM	37
732024	Control	40	732056	DM	37
732025	Control	37	732057	DM	33
732026	Control	43	732058	DM	32
732027	Control	35	732059	DM	35
732028	Control	40	732060	DM	29
732029	Control	39	732061	DM	35
732030	Control	43	732062	DM	33
732031	Control	35	732063	DM	29
732032	Control	37	732064	DM	27
			732065	DM	32

SAS code for descriptive statistics

```
proc means n mean std median min max maxdec=5
     data= BIOSTAT.ancova ;
title3 'Descriptive statistics';
class group;
var  hdl;
run;
```

June 25, 2013

Descriptive statistics

Two groups: control subjects and subjects with Type 2 diabetes
Endpoint: HDL-C

Descriptive statistics

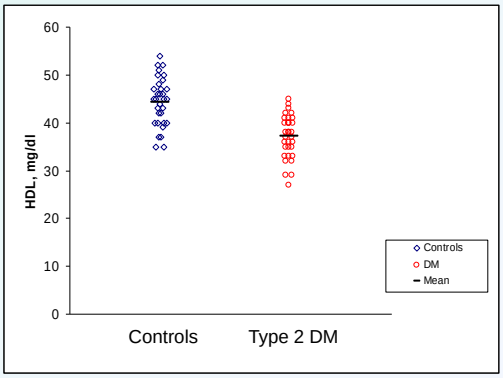
The MEANS Procedure

Analysis Variable : HDL							
Group	N Obs	N	Mean	Std Dev	Median	Minimum	Maximum
Controls	32	32	44.43750	5.03496	45.00000	35.00000	54.00000
DM	33	33	37.15152	4.45899	38.00000	27.00000	45.00000

June 25, 2013

Present the individual data whenever possible

HDL-C in control subjects and subjects with Type 2 diabetes
Endpoint: HDL-C

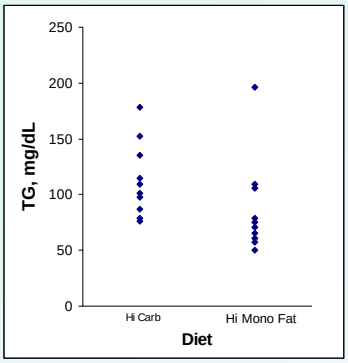


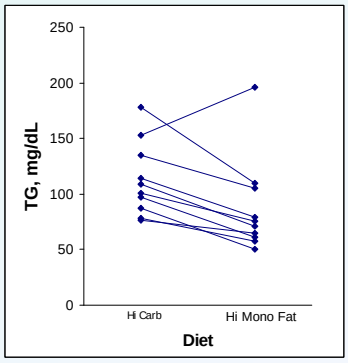
June 25, 2013

High Carbohydrate Diet Versus High Mono Fat Diet
Endpoint: Triglycerides

Design is a crossover study - each subject was given both diets in a randomized order

Graph paired data so that the relationship between pairs is preserved

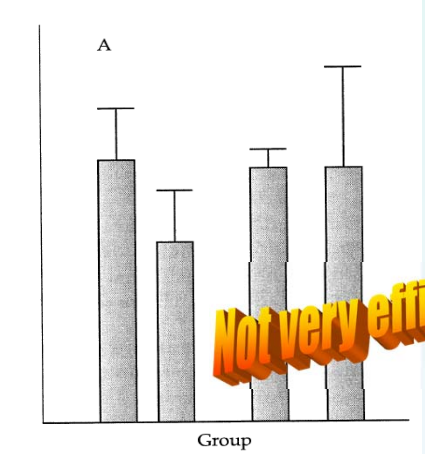




Data adapted from Garg et. al., NEJM 319:829-834, 1988.

June 25, 2013

Bar graphs for continuous data?



- A column is not needed to describe a mean
- These error bars imply the variability is only in one direction

From Lang and Secic, **How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers** (Paperback), 2006

June 25, 2013

Censored data

Cannot be measured beyond some limit

- Left censoring
- Right censoring

June 25, 2013

Left Censored data

Cannot be measured beyond some limit

- Lab data – “undetectable”, “below lower limit”
  - Example CRP “< 0.2 mg/dL”

Censored at the limit of detectability

Subject	CRP
001	0.7
002	1.6
003	**<0.2**
004	3.8

June 25, 2013

Right Censored data

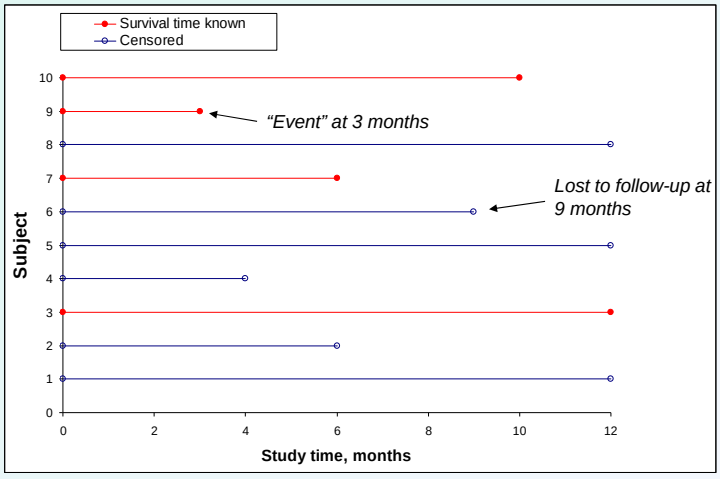
Cannot be measured beyond some limit

- Right censoring
  - “Survival” data – the period of observation was cut off before the event of interest occurred.

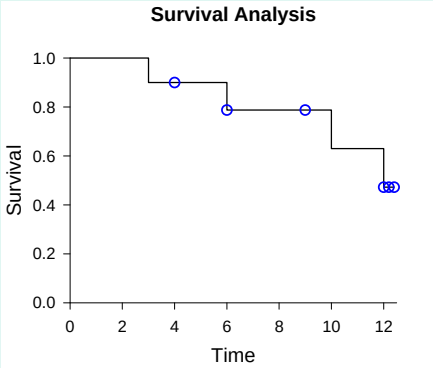
Note – an event in a ‘survival’ analysis may be infection, fracture , transplant , metastasis

June 25, 2013

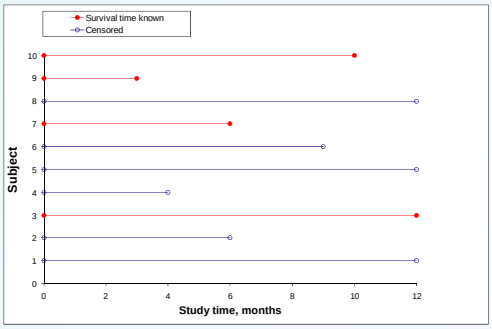
Right censored survival data



June 25, 2013



Right censored survival data



June 25, 2013

Descriptive statistics

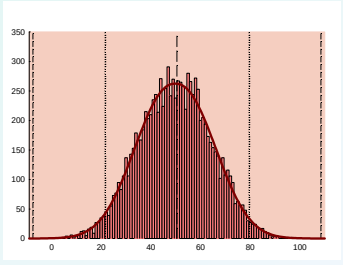
- Measures of Central Tendency
- Measures of Dispersion

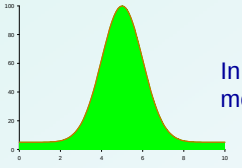
June 25, 2013

Measures of Central Tendency*

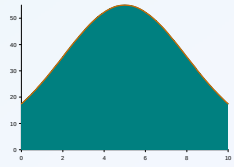
*or Measures of Location

- Mean
- Median
- *Geometric mean*
- Mode





In a symmetric distribution, the median, mode and mean will have the same value.



June 25, 2013

Measures of Central Tendency*

*or Measures of Location

- Mean
  - Arithmetic average or balance point
  - Discrete/continuous data; symmetric distribution
  - May be sensitive to outliers
  - Sample mean symbol is denoted as ‘x-bar’

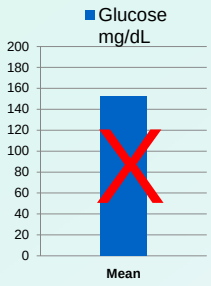
$$\bar{X} = \frac{\sum X}{N}$$

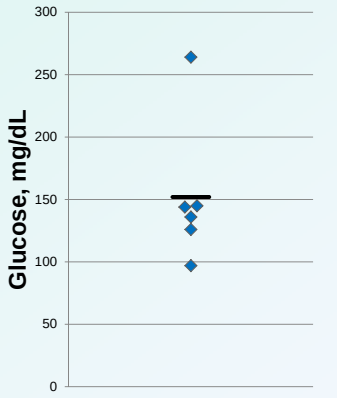
Fasting plasma glucose, n=6

SubjectID	Glucose mg/dL
0204	145
0205	126
0206	136
0210	97
0211	264
0212	144
Mean	152

June 25, 2013

Fasting plasma glucose, n=6





SubjectID	Glucose mg/dL
0204	145
0205	126
0206	136
0210	97
0211	264
0212	144
Mean	152
Median	140

What about other measures of central tendency?

June 25, 2013

Measures of Central Tendency

Median

- Middle value when the data are ranked in order (if the sample size is an even number then the median is the average of the two middle values)
- 50^{th} percentile
- Ordinal/discrete/continuous data
- Useful with highly skewed discrete or continuous data
- Relatively insensitive to outliers

June 25, 2013

Measures of Central Tendency

The median of 13, 11, 17 is 13
The median of 13, 11, 568 is 13
The median of 14, 12, 11, 568 is 13

June 25, 2013

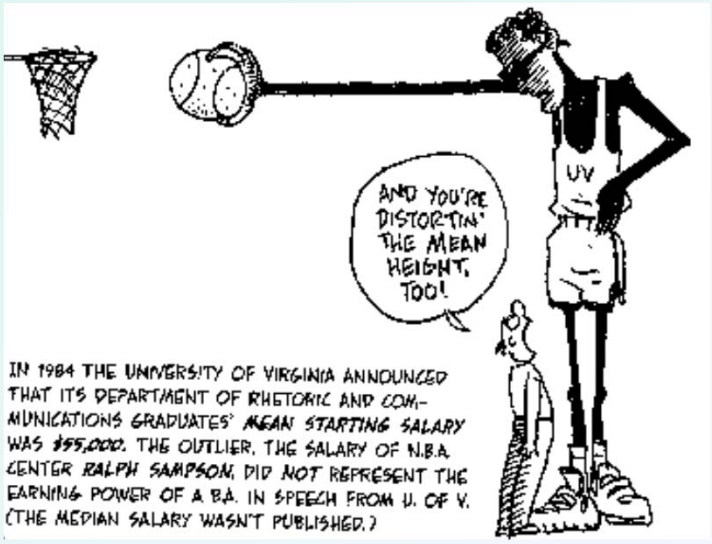
Measures of Central Tendency

SubjectID	Glucose mg/dL
0204	145
0205	126
0206	136
0210	97
0211	264
0212	144
Mean	152
Median	140

Order the glucose values from smallest to largest

SubjectID	Glucose mg/dL
0210	97
0205	126
0206	136
0212	144
0204	145
0211	264

The median is often better than the mean for describing the center of the data



Gonick & Smith (1993) The Cartoon Guide to Statistics.

Geometric mean

Log transformed data

SubjectID	Glucose mg/dL	ln(Glucose)
0204	145	4.976734
0205	126	4.836282
0206	136	4.912655
0210	97	4.574711
0211	264	5.575949
0212	144	4.969813
Mean	152	4.9743573
SD	57.644	0.330
Median	140	4.941234093

Geometric mean

Take the antilog of the mean

exp(4.974357) = 144.6558278

Geometric mean:

Back-transform (antilog) the mean of the log transformed data

June 25, 2013

Measures of Central Tendency

Mode

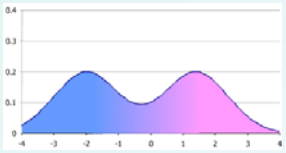
- Most frequently occurring value in the distribution
- Nominal/ordinal/discrete/continuous data

The mode of 13, 11, 22, 11, 17 is 11

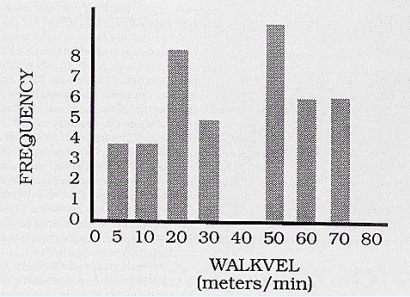
June 25, 2013

Measures of Central Tendency (Mode)

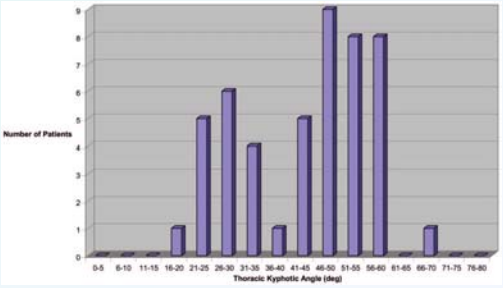
Bimodal distribution



The mode is not necessarily unique



Lunsford BR (1993) **JPO** 5(4), 125-130.



Bartynski et al. (2005) AJNR 26 (8): 2077.

June 25, 2013

Next class – Thursday, June 27

Room D1.602

- Describing data
  - Descriptive statistics – measures of dispersion
    - Variance, standard deviation
  - Other statistics
    - Coefficient of variation
    - Standard error of the mean
- Histograms and other graphs
- Transformations

June 25, 2013