

Using Technology to Connect Topics and Contrast Scenarios

Mike Larsen, July 24, 2003

Beyond the Formula

Four points

- I. Contrasts and connections are useful teaching tools.
- II. Contrasts and connections can be expanded and enhanced.
- III. Technology can be used to expand and enhance contrasts and connections.
- IV. Using technology to make contrasts and connections encourages stat'l thinking.

Contrast 1: Mean vs. median

Data set 1: 3, 5, 4, 8, 4

Sum = 24, Average = 4.8, Median = 4

Data set 2: 3, 5, 4, 28, 4

Sum = 44, Average = 8.8, Median = 4

Mean vs. median

How?

Blackboard, Graphing calculator,
Spread sheet, Statistical package,
Internet web site.

Active?

If students do computations, pick own
large outlier, yes.

Mean vs. median, cont.

Disadvantage?

Devoid of context,

Data set too small for other purposes

Alternative?

Larger data set,

Data on a topic of interest

Mean vs. median, cont.

Expanded contrasts with larger data:

Contrasts of percentiles

Contrasts of IQR vs. s (spread)

Enhanced contrasts with technology:

Graphs, Quick computations,

Experimentation.

Woodlawn, Chicago, July 23, 2003, real estate, chicagotribune.com

Sample 25 properties
(1000s, rounded)

Statistics:

Sum 4293

339, 299, 249, 239, 235

Mean 171.72

179, 179, 177, 175, 170

Median 159

159, 159, 159, 158, 158

142, 142, 142, 139, 136

SD 58.73

115, 112, 112, 110, 109

IQR 179-139=40

Woodlawn, Chicago, July 23, 2003, real estate, chicagotribune.com

Sample 25 properties
(1000s, rounded)

Statistics:

Sum 7344

3390, 299, 249, 239, 235

Mean 293.76

179, 179, 177, 175, 170

Median 159

159, 159, 159, 158, 158

142, 142, 142, 139, 136

SD 646.78

115, 112, 112, 110, 109

IQR 179-139=40

Contrast 2: Transformations

Data set:		F	mph
	Albany	75	9
Temp. in degrees F,	Buffalo	72	0
	Ithaca	70	9
Wind speed in mph	New York	73	7
	Niagra	70	5
wunderground.com,	Rochester	75	12
July 23, noon CST	Saranac Lake	66	5
	Syracuse	77	6
	Utica	77	6

Transformations

Linear Transformations:

Degrees F to degrees C: $C = (F-32)*5/9$

Degrees F to degrees K: $K = C - 273$

Mph to kph: $KPH = MPH/.62 = 1.61*MPH$

Mph to feet per minute:

$$FPM = MPH / (5280*60)$$

Transformations, cont.

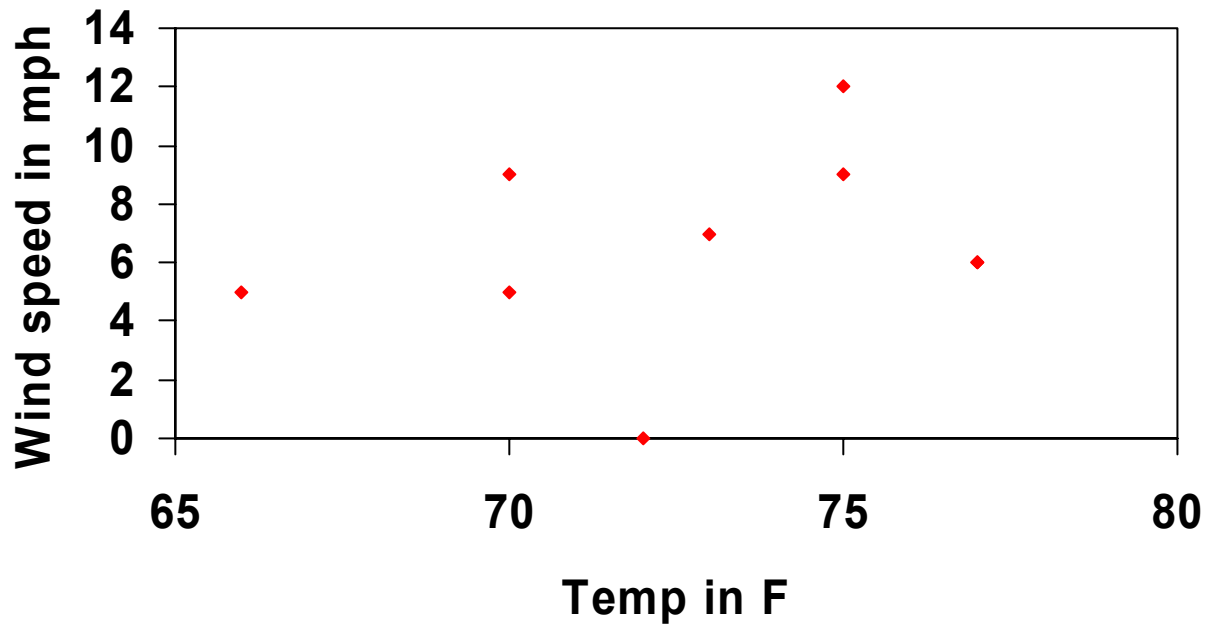
How?

Blackboard? Useful for illustrating a few calculations. Difficult for exploration.

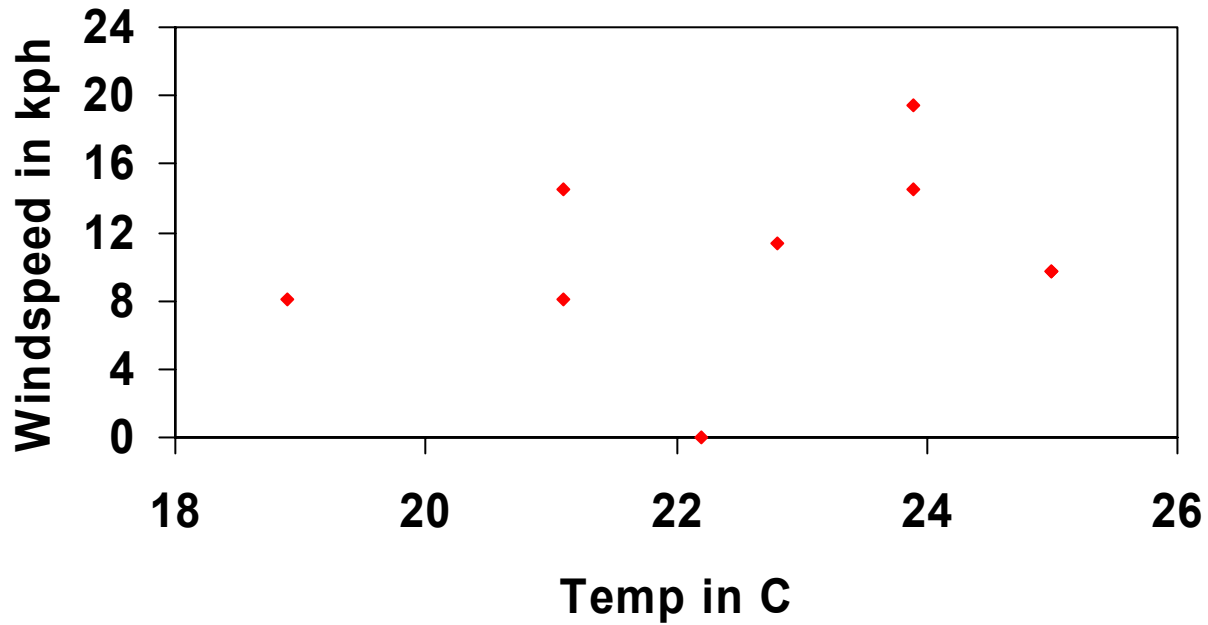
Graphing calculator and statistical package – Quick, exploration possible, answer multiple questions, larger data set possible.

Active? Yes, if student manipulates data.

Wind speed and temp



Wind speed and temp



Transformations, cont.

Expanded comparisons:

Graphs, mean, median,

Min, max, 1Q, 3Q, IQR, s , s^2

Correlation coefficient r

Estimated regression slope and intercept

Explicitly compare graphs and statistics.

-- Also nonlinear (log, square root)

Contrast 3: Binomial distributions

X is a random variable that records the number of successes in n independent trials that each have success probability p

$X=0, 1, \dots, n$

$$P(X=k) = {}_n C_k p^k (1-p)^{n-k}$$

Do students understand this distribution?

Binomial distributions

Contrast two or more baseball players:
mlb.com, AAA Iowa Cubs, stats

Trenidad Hubbard, batting avg 0.322
Augie Ojeda, batting avg 0.240

Imagine 16 at bats (about 4 games)
Assume the binomial model is OK

How do the players compare?

Binomial distribution, cont.

Compute $P(X=k)$ for two players, various k .

How?

Blackboard? Explains formula, but not interesting.

Table look up? Not all values of p and n .

Graphing calculator, statistical package, or Internet: Allow graphing and more extensive comparison (larger n , more values of p).

Active? Yes, if students perform calculations and make graphs and summarize comparisons.

Binomial distributions, cont.

Web sites concerning the Binomial

-- Input parameters and compare in real time

- <http://www-stat.stanford.edu/~naras/jsm/example5.html>
- <http://www.stat.sc.edu/~west/applets/binomialdemo.html>
- <http://stat-www.berkeley.edu/~stark/Java/BinHist.htm>
- http://www.ruf.rice.edu/~lane/stat_sim/normal_approx/

Binomial distribution, cont.

Expand the range of comparisons and questions:

1. $P(X=k)$, $P(X>k)$, $P(X\geq k)$ for various k
2. Shape – which is more skew?
3. Which has higher mean? Where does the probability histogram balance?
4. Which has a greater IQR and standard deviation?
5. Which has more probability 1 SD beyond the mean?

Expand the contrasts:

Change n , change p ($p<.5$, $p=.5$, $p>.5$)

Four points, reviewed

- I. Contrasts and connections are useful teaching tools.
- II. Contrasts and connections can be expanded and enhanced.
- III. Technology can be used to expand and enhance contrasts and connections.
- IV. Using technology to make contrasts and connections encourages *stat'l thinking*.

Contrast 4: Central limit theorem

What affects the number of observations needed before the CLT is effective?

How?

Physical simulations (e.g., age of coins, pulling tokens out of a box): good, but time consuming (few reps) and few scenarios contrasted.

CLT

Showing class pictures: More scenarios but not active (students are skeptical).

Simulations using Internet web pages, graphing calculators, and statistical packages:

Many repetitions

Contrast scenarios

Active

Can look at original distributions

CLT, cont.

Uniform random numbers/integers

Set number of observations to 1000

Randomly generate values into 16 columns

Average first 4, first 8, then first 16 columns

Summarize (mean, sd, graph) 1000 reps.

Repeat for another range of numbers and/or another distribution.

CLT, cont.

Transformations: If u is uniform(0,1), then

$x = (b-a)*u+a$ is uniform(a,b).

$y = \text{int}(b*u)+a$ is uniform on $(a, a+1, \dots, b)$,
where int is the integer part of $b*u$.

See floor, round.

$r = -\ln(u)/b$ is Exponential(b), mean b , $r>0$

$s = u^{a-1}$ is Beta(a,1) $0<s<1$

$t = 1-u^{b-1}$ is Beta(1,b) $0<t<1$

CLT, cont.

Contrast

Shape of original distribution, extremes,
mean and sd of original distribution.

Skewness/symmetry of sampling distr'n.

Parameters (or estimates based on
simulation) of mean and SD of
sampling distribution.

Sampling distributions of other statistics:
medians, 3rd quartiles, maximums, etc.

CLT, cont.

<http://www.randomizer.org/form.htm>

100 sets of 5 uniform integers (w/r) (1 to 10)

Output to Excel (5 by 100 format)

100 means, 100 medians

100 means: mean is 5.274, sd is 1.34

100 medians: mean is 5.2, sd is 1.98

Means are less variable than medians!

Compare to graph

Speed of Presentation and Simulations

A concern was raised at the conference that students can't follow what happens in simulations because simulations go too fast. A few suggestions were given:

1. Slow down – explain very clearly what is being simulated.
2. Look at a couple of examples first – show how results are obtained based on one sample, or two samples, using the actual numbers.
3. Vary one dimension of the problem at a time.
4. Summarize what was learned.

Contrast 5: Confidence intervals

What is the meaning of level of confidence?

Level of confidence has constant meaning across comparisons, but assumptions of interval procedure have to be met.

What affects the confidence interval?

There are several factors, including the underlying distribution of the data.

Confidence Intervals

How?

Blackboard: Show the formula and alter the numbers – n , s , \bar{x} , level of confidence. Doesn't illustrate coverage in repeated sampling very well.

Coordinated work: As above, but with each student computing based on different numbers. Better for coverage illustration and more active, but requires time for synthesizing results.

Conf. Intervals, cont.

Internet simulations:

<http://www.stat.sc.edu/~west/javahtml/ConfidenceInterval.html>

http://www.ruf.rice.edu/~lane/stat_sim/normal_approx_conf/

Many replications, but not much to contrast.

Also, a lot of the details of intervals are hidden.

Simulations using graphing calculators and stat'l packages:

More work to design and implement, but potentially more 'active' and amenable to contrasts.

What to do? Combine technology and traditional.

Contrast 6: Hypothesis Tests and Power

What does power mean?

What is type I error?

In what sense is the t distr'n robust?

On-line Simulation:

http://www.ruf.rice.edu/~lane/stat_sim/robustness/

Vary one factor at a time.

Technology allows rapid contrast of scenarios.

Contrast using formulas still useful!

Contrast 7: Probability Trees

Probability trees are a basic tool for studying conditional probability and Bayes' theorem.

How?

Blackboard: Words, diagrams, calculations
– all very good. Contrast requires repeating calculations.

Probability Trees

Simulation with graphing calculator and statistical package:

Generate Binomial(10000, p), where $p=P(A)$.

Let n_A be the number of successes.

Generate Binomial(n_A , p_2), where $p_2=P(B|A)$.

Let n_{AB} be the number of successes and
 n_{AnotB} be the number of failures ($n_A - n_{AB}$).

Generate Binomial($10000 - n_A$, p_3), where $p_3=P(B|not A)$.

Let n_{notAB} be the number of successes and
 $n_{notAnotB}$ be the number of failures.

Make a two-way table of counts and answer questions.

Contrast requires repeating simulations.

Also – On-line calculators for probability trees

Four points and last comment

- I. Contrasts and connections are useful teaching tools.
- II. Contrasts and connections can be expanded and enhanced.
- III. Technology can be used to expand and enhance contrasts and connections.
- IV. Using technology to make contrasts and connections encourages stat'l thinking.

Contrasts and connections can form the basis of writing assignments.

Contact Information

larsen@galton.uchicago.edu

<http://galton.uchicago.edu/~larsen>

<http://galton.uchicago.edu/~larsen/btf03>

Moving to Iowa State in fall, 2003

<http://www.stat.iastate.edu>

See also “Internet Companion for Statistics”

<http://larsen.duxbury.com>

About Writing Assignments

The usefulness of writing assignments was discussed at the conference. Suggestions included

1. Assign them regularly – they are very valuable if students are going to learn to write about statistical ideas.
2. Make assignments focused – this makes them easier to grade and of higher quality.
3. Consider limiting the length of writing assignments – 25 words or less.
4. Consider making some problems (plug and chug computations) optional work. This offsets some of the grading time when writing assignments are given.

Internet Sites and Their Use in Introductory Statistics

Mike Larsen, July 25, 2003

Beyond the Formula

Outline

1. Challenges of using Internet sites
2. Where in course to use Internet sites?
3. Questions/projects – what to do?
4. Sites that illustrate methods
5. Constructing data sets
6. Simulation sites
7. General sites

Challenges of using Internet sites

Many sites are

- not organized by course topic
- not accompanied by questions
- of variable quality
- include additional material
- not linked to general sites or indexes

Comments by Audience

PROBLEM: Other sites are not stable – they change links rapidly (or just before class).

SUGGESTION: Government sites, archives and repositories, sites dedicated to teaching are likely to be more stable. Some commercial sites and sites related to college courses change rapidly (or at least by season/semester).

PROBLEM: Access in high schools – not all students have computers and (fast) Internet access at home. Computer labs and available lab time might be limited outside of class.

Addressing general challenges

An instructor must:

1. select sites for various topics
2. write questions related to the site
(and write instructions)
3. asses quality of sites
4. explain or avoid material on additional topics, or direct students to relevant part
5. find the sites in the first place

Where to use sites in course

In-class illustrations

- Need computer, Internet connection,
TV or projector/screen

- Throughout class vs. at end of class
(depends on physical classroom)

- Timing is an issue – how long will it take?

- Can students participate?

One idea is for students type and explain results in their own words.

Where to use in course, cont.

Self study/optional study

At home or independently in lab

Some sites are prepared already for self study

Other sites will need questions, directions

Homework assignments

Take home – instructions are important here

Computer laboratory

Other challenges are still relevant

Questions/projects – what to do?

What should students do with the sites?

1. Report on site
2. Comment on statistical presentations
3. Discuss possible alternatives
4. Verify and then modify computations
5. Manipulate sites if interactive
6. Search – find data, find examples, etc.
7. Print the screen – print a graph to record work

Illustrations

DASL: <http://lib.stat.cmu.edu/DASL/>

All methods, box plots, Cuckoos

NIST: <http://www.itl.nist.gov/div898/handbook/>

Section 1.1.6

CHANCE and CHANCE News

<http://www.dartmouth.edu/~chance/>

Other places on the web

Smithsonian budget, sections 7 and 8

<http://www.si.edu/about/budget/>

Percentiles

Salary data

<http://www.amstat.org/profession/index.html>

College Board

<http://www.collegeboard.com/counselors/hs/sat/scorereport/scoredata.html>

General research articles:

<http://www.collegeboard.com/research/home/>

CDC growth charts

<http://www.cdc.gov/growthcharts/>

Collections of data sets

DASL for small data sets

<http://lib.stat.cmu.edu/DASL/Stories/cuckoo.html>

Exploring Data Web site

<http://exploringata.cqu.edu.au>

Comments on Data Set Construction/Choice

Understand measurements,

be able to explain context and problem.

Population vs. sample data – which is it?

Sample size -- small, medium, large

If too small, consider alternative or simulation.

If too large, consider sampling.

Correlated measurements – be aware of relationships. In sports, for example, data are given by team, by game, by season, etc.

Summary versus individual measurements

Data Set Construction

1. Is there a sample from a population?
Yes – great, can you summarize the population for comparison?
No – use for description.
2. Can you record multiple variables?
Yes – do so, but don't create too big of a burden.
No – rethink. Are there categories or qualitative variables? Can you match the sample units to another table or source?
3. Is the sample size too small?
Yes – limits usefulness.
No – if its too big consider sampling.

Data by State

U.S. Statistical Abstract – Data by State

<http://www.census.gov/statab/www/>

Bureau of Labor Statistics

<http://www.bls.gov/bls/blswage.htm>

County agriculture

<http://www.nass.usda.gov>

See historical data, statistics database,
select corn, Nebraska, 2001-2001, to screen

Federal Statistics sites in general

<http://www.fedstats.gov>

Some Examples of Data Sets

Weather

<http://wunderground.com>

<http://weather.chicagotribune.com>

House sales

<http://www.chicagotribune.com> (real estate)

Distance and prices of cheapest airline flights

<http://www.washingtonpost.com>

<http://www.travelnotes.org>

Animals and Fish

Google search on “fishing records” and your state

Look for animal sizes at a zoo site

<http://www.zooregon.org/Cards/cardindex.htm>

Sports data sets

<http://www.athletics.harvard.edu/>

Look under M-Heavy weight crew, roster

<http://sportsillustrated.cnn.com/golfonline/golfstats/lpgaSearch.html>

Search for a player by name (Sorenstam), select the 2002 season

<http://www.mlb.com>

Look for individual players and historical records

<http://www.wtatour.com>

Look at Players – various career summary info.

What are some of your favorite sites/sports?

Simulation sites

RVLS

<http://www.ruf.rice.edu/~lane/rvls.html>

http://www.ruf.rice.edu/~lane/stat_sim/descriptive/index.htm

Webster West and Todd Ogden

<http://www.stat.sc.edu/~west/applets/>

<http://www.stat.sc.edu/~west/javahtml>

Shodor Education Foundation

<http://www.shodor.org/interactivate/>

AP Statistics links, Chance links

Comments on simulation sites

If only summary results are given on the site, then there is less opportunity to report results in multiple ways.

The more options for changing parameters and conditions, the more interesting it can be, but also the more confusing to the student. Clear directions are needed.

Simulation/Display/Calculations of Distributions

Normal, T, Chi-square, F, Binomial

What to do?

1. Guide student through several computations
2. Have students write about findings
3. Print screen/print graph to report results

<http://stat-www.berkeley.edu/~stark/Java/BinHist.htm>

http://ruf.rice.edu/~lane/stat_sim/sampling_dist/index.html

General sites

AP Statistics Listserv

<http://a-s.clayton.edu/apstatfaq>

College Board AP statistics

<http://apcentral.collegeboard.com>

AP Statistics Web guide

Rice Virtual Lab and Hyper Stat

<http://www.ruf.rice.edu/~lane/rvls.html>

CHANCE

<http://www.dartmouth.edu/~chance/>

NIST Engineering Statistics Handbook

<http://www.itl.nist.gov/div898/handbook/>

About General Sites

Look up topics, Links to other sites
Watch for new additions

Some useful, some not as good, some redundancy
More general information than you can use/need

Also Ginger Holmes Rowell's (rowell@mtsu.edu)
talk (7/24/03) and <http://causeweb.org> (future)

Google searches

A lot of material is available

<http://google.com> Example: search on Poisson

Material can be not there, too technical,
poor quality for teaching, redundant.

Searches can produce useful results,
but consume time.

Conclusions

There is a lot of interesting and useful material for teaching and learning statistics on the Internet.

Preparing to use Internet sites in class and in assignments can be time consuming, so take care in planning.

Internet material cannot and should not replace effective traditional teaching methods.

Conclusions, cont.

The strengths of Internet sites include
interaction/active learning
context dependent
interesting examples
potential basis for writing assignments
exploration
simulation tools and source of data.

Contact Information

larsen@galton.uchicago.edu

<http://galton.uchicago.edu/~larsen>

<http://galton.uchicago.edu/~larsen/btf03>

Moving to Iowa State in fall, 2003

<http://www.stat.iastate.edu>

An effort of mine ...

See also “Internet Companion for Statistics: Guide and Activities for the Web,” first edition, <http://larsen.duxbury.com>

Problems are organized by course topics, linked to web sites, prepared with questions and a home web site.