



**Using Computational Software to Teach Statistical Concepts**  
 An interactive workshop  
 Robert H. Carver  
  
 July 24, 2003  
 10:45 am - 12:00 noon  
  
**Beyond the Formula**  
 Statistics Conference  
 at Stonehill College, North Andover, MA

---

---

---

---

---

---

---

---

**Workshop Goals**

- Identify strategies for extending use of available software for *conceptual development* in introductory courses
- Illustrate such use of software
- Engage participant expertise in developing course-specific applications

 July 24, 2003

---

---

---

---

---


---

---

---

**Agenda**

10:45	Introductions, brief overview
11:00	Identify participants' problems, constraints, opportunities
11:10	Develop strategies to apply software as part of the solution
11:45	Share results
12:00	Adjourn

 July 24, 2003

---

---

---

---

---

---

---

---

Beyond the Formula  
Student Learning  
Outcomes

### Context

<ul style="list-style-type: none"> <li>+ Intro Course                             <ul style="list-style-type: none"> <li>• Emphasis on reasoning</li> <li>• Service course vs. intro to major                                     <ul style="list-style-type: none"> <li>• First course?</li> <li>• Last course?</li> </ul> </li> <li>• Student preparation to read statistics texts</li> <li>• Beyond the formula</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>+ Constraints                             <ul style="list-style-type: none"> <li>• Loosening                                     <ul style="list-style-type: none"> <li>• Student computing</li> <li>• Classroom technology</li> </ul> </li> <li>• Tightening Fixed                                     <ul style="list-style-type: none"> <li>• 12 weeks</li> <li>• Budgets for software</li> <li>• Cognitive budget: stats vs software</li> </ul> </li> </ul> </li> </ul>
---	--

July 24, 2003
4

---

---

---

---

---

---

---

---

Beyond the Formula  
Student Learning  
Outcomes

### Computational Software

- + In the classroom & workplace
- + Permits emphasis on interpretation & statistical reasoning rather than computation
- + Vehicle for constructing concepts
- + Ease of replication and simulation – permitting illustrative comparison
- + Individualized results
- + Integration with texts & other technologies
- + *Ad hoc* interactivity—teachable moments

July 24, 2003
5

---

---

---

---

---

---

---

---

Beyond the Formula  
Student Learning  
Outcomes

### Some Introductory Concepts

<ul style="list-style-type: none"> <li>+ Variation                             <ul style="list-style-type: none"> <li>• Shape</li> <li>• Center</li> <li>• Spread</li> </ul> </li> <li>+ Randomness</li> <li>+</li> <li>+</li> </ul>	<ul style="list-style-type: none"> <li>+ Random Sampling                             <ul style="list-style-type: none"> <li>• Bias</li> <li>• Null hypothesis</li> <li>• “Error”</li> </ul> </li> <li>+ Linear (&amp; non-linear relationships)</li> <li>+</li> </ul>
--	---

July 24, 2003
6

---

---

---

---

---

---

---

---

Beyond the Formula  
Statistics 100  
Fall 2002

### Variation & Distribution

- + Annual student survey
- + Some Variables:
  - + Weight & Ideal Weight
  - + Study hours
  - + Outside work hours

STONEHILL COLLEGE  
JUL 24 2003

---

---

---

---

---

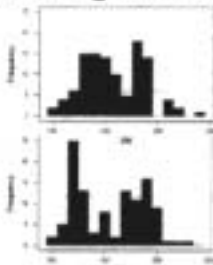
---

---

---

Beyond the Formula  
Statistics 100  
Fall 2002

### Visualizing Variation



What's the conversation about  
Shape?  
Center?  
Spread?

← What does this comparison add?

- + What did last year's graphs look like?
- + What might explain this?
- + What do you want to look at next?

STONEHILL COLLEGE  
JUL 24 2003

---

---

---

---

---

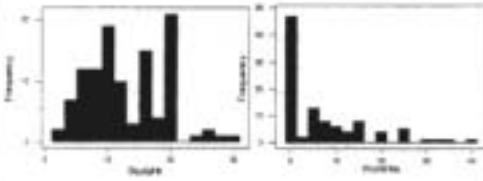
---

---

---

Beyond the Formula  
Statistics 100  
Fall 2002

### What does shape tell us?



STONEHILL COLLEGE  
JUL 24 2003

---

---

---

---

---

---

---

---


Beyond the Formula  
Statistics Curriculum

### Random Sampling

"Random Rectangles" from  
Activity-Based Statistics  
(Scheaffer, et al.)

Exercise: Select 3 samples  
 $n = 5$ : compute mean

- "Random"
- "Representative"
- Simple Random



STONHILL COLLEGE  
July 24, 2003

---

---

---

---

---

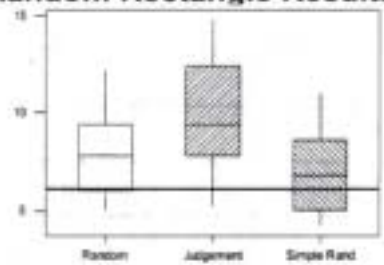
---

---

---

Beyond the Formula  
Statistics Curriculum

### Random Rectangle Results



Random Judgement Simple Rand

STONHILL COLLEGE  
July 24, 2003

---

---

---

---

---

---

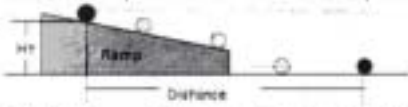
---

---

Beyond the Formula  
Statistics Curriculum

### Linearity

- Galileo's experiments with inclined plane



- What is the relationship between height and distance?
- Which is the response variable?

STONHILL COLLEGE  
July 24, 2003

---

---

---

---

---

---

---

---

Beyond the Formula  
Student Copy

### Linear relationship

- + Explore distinction between positive relationship and linear relationship
  - + What happens when height = 0?
  - + What happens as height increases?
  - + What happens at infinite height?
- + Can the relationship be *linear*?

STONHILL COLLEGE  
JULY 24, 2003

---

---

---

---

---

---

---

---

Beyond the Formula  
Student Copy

### Visualizing the result

Regression Fit

Regression Fit

STONHILL COLLEGE  
JULY 24, 2003

---

---

---

---

---

---

---

---

Beyond the Formula  
Student Copy

### DIY

- + Which **concepts** are most elusive?
- + What's your "best approach"?
- + Which of these can we better approach experientially/interactively?
- + Role of illustrative data (raw/ cooked/ synthetic/ simulated)
- + How can a given approach exploit software at hand + data + student background?

STONHILL COLLEGE  
JULY 24, 2003

---

---

---

---

---

---

---

---

# Beyond the Formula

## Statistics Conference

at Monroe Community College, Rochester, NY

### Using Computational Software to Teach Statistical Concepts

Robert H. Carver, Stonehill College

Session S.2.1

July 24, 2003

#### Illustrative Data Set Descriptions

(further detail is available in the files using Project Manager)

#### File Name Source and Description/ [applications]

**Cholest** Minitab Inc. This files is among those shipping with Minitab.  
Cholesterol readings from a sample of heart attack patients at repeated intervals following heart attack (n=28) and a control group (n=30).  
[distributions, paired and independent observations, 1- and 2-sample t-tests.]

Column	Count	Missing	Name
C1	28	0	2-Day
C2	28	0	4-Day
C3	28	9	14-Day
C4	30	0	Control

**Florida Votes** Greg Adams and Chris Fastnow, "A Note on the Voting Irregularities in Palm Beach, Florida." <http://www.madison.hss.cmu.edu>  
County-by-county counts of votes cast during the 2000 Presidential election in the state of Florida. [linear regression, outliers, inference]

Column	Count	Name
T C1	67	County
C2	67	GORE
C3	67	BUSH
C4	67	BUCHANAN
C5	67	NADER
C6	67	BROWNE
C7	67	HAGELIN
C8	67	HARRIS
C9	67	MCREYNOLDS
C10	67	MOOREHEAD
C11	67	PHILLIPS
C12	67	Total

**Galileo** S.Drake, *\_Galileo at Work\_*, University of Chicago Press, 1978. D.A. Dickey and J.T. Arnold, "Teaching Statistics with Data of Historic Significance" *Journal of Statistics Education* 3(1), 1995.

Results of two classic experiments rolling balls down an inclined ramp. [linear and non-linear models]

Column	Count	Name
C1	7	DistRamp
C2	7	HtRamp
C3	5	DistShelf
C4	5	HtShelf

**File Name Source and Description/ [applications]**GSSGeneral "Collections \ Years Index: 1998a Variables" *General Social Survey, 1998.*<http://www.icpsr.umich.edu/GSS/rnd1998/merged/indx-mod/1998a.htm>

Selected responses from the 1998 General Social Survey. This file contains 6 variables: Marital status, age, sex, race, geographic region, and zodiac sign of respondent. [scales of measurement, distributions, independence, hypothesis development, various tests]

	Column	Count	Missing	Name
T	C1	1445	0	MARITAL
	C2	1445	2	AGE
T	C3	1445	0	SEX
T	C4	1445	0	RACE
T	C5	1445	0	REGION
T	C6	1445	22	ZODIAC

JFKLAX

US Department of Transportation, Bureau of Transportation Statistics, Office of Airline Information. March 2001.

Airline arrival and departure records for one randomly selected date for flights between JFK (NY) and LAX (CA) airports. [Probability distributions, variation, scales of measurement, operationalization, hypothesis development]

	Column	Count	Name
T	C1	52	Carrier
	C2	52	FlightNo
T	C3	52	Origin
T	C4	52	Destination
D	C5	52	DepSch
D	C6	52	DepAct
D	C7	52	ArrSch
D	C8	52	ArrAct
	C9	52	ArrDelay
	C10	52	DepDelay
T	C11	52	LateArr

New York

Bureau of Justice Statistics Crime & Justice Electronic Data Abstracts

<http://www.ojp.usdoj.gov/bjs/dtdata.htm>

Almanac data for the state of New York by county. [Linear relationships, scatterplots, distributions]

	Column	Count	Name
T	C1	62	COUNTY
	C2	62	POP1990
	C3	62	POP90_SQMI
	C4	62	MALES
	C5	62	FEMALES
	C6	62	AGE_UNDE5
	C7	62	AGE_5_17
	C8	62	AGE_18_64
	C9	62	AGE_65_UP
	C10	62	SQMI
	C11	62	PCTFEM

Rectangles

Author.

Results of in-class demonstration of random and judgement sampling. [random sampling, sampling methods, bias, variation]

	Column	Count	Name
	C1	19	Random
	C2	19	Judgement
	C3	19	Simple Rand
	C4	57	Means
T	C5	57	Method

**File Name** **Source and Description/ [applications]**

StateTrans Various Federal data sources.

Transportation-related data for 50 states and Washington DC. [distributions, outliers, linear relationships, data transformations, hypothesis development]

	Column	Count	Name
T	C1	51	State
	C2	51	Pop
	C3	51	Area
	C4	51	Density
	C5	51	RLM
	C6	51	ULM
	C7	51	AVMTPC
	C8	51	Regist
	C9	51	FatCrash
	C10	51	TraffFat
	C11	51	BAC
	C12	51	DrivLic
	C13	51	PctLic
	C14	51	DefBridge
T	C15	51	AVMCat

Student Author.

Results of in-class student survey. [scales of measurement, variation, distributions, hypothesis development, independence, linear relationships]

	Column	Count	Missing	Name
T	C1	111	0	Gender
	C2	111	0	Ht
	C3	111	0	Idlht
	C4	111	1	Wt
	C5	111	1	Idlwt
T	C6	111	0	Birthday
T	C7	110	2	Hand
T	C8	111	0	Eyes
	C9	111	0	Exer
T	C10	111	0	Resident
	C11	111	10	WorkHrs
	C12	111	1	StudyHr
	C13	111	2	GPA
T	C14	111	0	UseJob
T	C15	111	0	Owncar
	C16	111	3	Miles
T	C17	111	0	HomeReg
T	C18	111	0	Driver
T	C19	111	0	Belts
	C20	111	1	Accident
	C21	111	0	Siblings
T	C22	111	0	Smoke
	C23	111	1	Haircut
T	C24	111	3	RegVote
T	C25	111	0	PresPref
T	C26	111	0	Dog
	C27	111	2	Beer
	C28	111	0	Female
	C29	111	1	WtLoss
T	C30	111	0	Work

**File Name Source and Description/ [applications]**

Tires NHTSA <http://www.nhtsa.dot.gov/cars/problems/Equipment/Tires/index.html>  
 Raw data from National Highway Transportation Safety Administration reporting vehicle accidents involving tire failures, July 2000. [data cleaning, operationalization, scales of measurement, assignable causes, independence]

	Column	Count	Missing	Name
T	C1	71	0	STATE
D	C2	71	0	COMPLAINT DATE
	C3	71	8	INJURIES
T	C4	71	0	FATALITIES
T	C5	71	0	TIRE_MODEL
T	C6	71	0	TIRE_SIZE
T	C7	71	0	OEM_TIRE
	C8	71	19	VEH_MILES
T	C9	71	0	VEH_MFR
T	C10	71	0	VEH_MODEL
	C11	71	6	VEH_MODEL_YEAR
D	C12	71	0	FAILURE_DATE
T	C13	71	0	BLOWOUT
T	C14	71	0	TREAD_SEPARATION
T	C15	71	0	ROLLOVER
T	C16	71	0	CRASH
	C17	71	10	DRIVFR
	C18	71	10	DRIVRR
	C19	71	10	PASSFR
	C20	71	10	PASSRR
	C21	71	7	NumFAIL

US Statistical Abstract of the United States, Economic Report of the President, various years.

Demographic and economic time series data. [Time series, distributions, forecasting, linear relationships, hypothesis development]

	Column	Count	Missing	Name
	C1	36	0	Year
	C2	36	1	Pop
	C3	36	1	Employ
	C4	36	1	Unemprt
	C5	36	1	GDP
	C6	36	1	PersCon
	C7	36	1	PersInc
	C8	36	1	PersSav
	C9	36	1	DefGDP
	C10	36	1	DefPC
	C11	36	1	CPIU
	C12	36	1	M1
	C13	36	2	Dow
	C14	36	1	Starts
	C15	36	1	SellPrc
	C16	36	1	ValNH
	C17	36	1	NHMort
	C18	36	1	Prime
	C19	36	1	PPIConst
	C20	36	5	Cars
	C21	36	2	MortDebt
	C22	36	1	Exports
	C23	36	1	Imports
	C24	36	0	FedRecpt
	C25	36	0	FedOut
	C26	36	0	FedInt

**File Name**    **Source and Description/ [applications]**

Utility        Author; Monthly Surface Data, October 8, 1999.

<http://cdo.ncdc.noaa.gov/cdo/3220doc.txt>

Monthly utility usage data from author's home for multiple years. [Seasonal variation, assignable causes, distribution, linear vs. non-linear relationships, hypothesis development, estimation.]

	Column	Count	Missing	Name
D	C1	138	0	Month
	C2	138	0	Days
	C3	138	0	MeanTemp
	C4	138	14	GaspDay
	C5	138	14	Therms
	C6	138	14	GasDays
	C7	138	0	KWH
	C8	138	0	KWHPDay
	C9	138	0	ElecDays
	C10	138	0	HDD
	C11	138	0	CDD
	C12	138	0	NewRoom
	C13	138	0	NewMeter
	C14	138	0	HVAC