UCLA

2007 SOCR/CAUSEway/UCLA Continuing Statistics Education Workshop

Title

Handbook on Continuing Statistics Education with Technology Workshop

Permalink

https://escholarship.org/uc/item/7hz6t6hk

Authors

Dinov, Ivo D
Christou, Nicolas
Sanchez, Juana

Publication Date

2007-08-06

2007 Joint SOCR/CAUSE
Continuing Statistics Education Workshop

08/06/07 - 08/08/07

The SOCR Resource
Department of Statistics UCLA,
Box 155404
8125 Mathematical Sciences Bldg.
Los Angeles, CA 90095-1554, USA.
Tel. USA + 310-825-8430
Fax: USA + 310-206-5658

NSF DUE 0442992

Pedagogical Utilization of SOCR & CAUSEweb Resources in Probability and Statistics Courses
August 06-08, 2007, UCLA

SOCR/CAUSEway Continuing Statistics Education

Workshop Handbook

Table of Contents

SOCR	CAUSE	UCLA Statistics	UCLA OID	NSF

www.SOCR.ucla.edu
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

Welcome Letter

Dear Workshop Attendees:

On behalf of the Statistics Online Computational Resource (SOCR) and the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), we would like to welcome you to the 2007 SOCR/CAUSE Workshop. The theme of this year's workshop is *Pedagogical Utilization of Novel Technologies in Probability and Statistics Courses.*

We organized this workshop to achieve two major goals – provide *training for instructors* in using the latest SOCR and CAUSE educational resources and at the same time, provide an open learning forum for instructors to *communicate and exchange ideas* about existent educational materials, desired resources and useful instructional tools.

Under different tabs in this booklet, you will find the complete Workshop Program; an Outline of the Workshop Logistics & Goals; Workshop Evaluation Forms, which you should complete tear off and return to us as instructed for outside analysis; SOCR/CAUSE About/Background Section; the Complete Workshop Activities and Materials. The back cover also includes the Workshop Program-at-a-Glance and an area Map.

Over the past several years, SOCR and CAUSE have developed, catalogued, and annotated a large number of instructional materials, interactive applets, internet resources and collaborative resources. We will present many of these new developments and you can always find the complete collections of SOCR and CAUSE resources on the web at www.SOCR.ucla.edu and www.CAUSEweb.org.

We are looking forward to a productive and exciting session on statistics education in the next few days and hope that the exchange of ideas, instructional materials and Internet resources stimulates long-term collaborations and facilitates our teaching with technology in the classroom.

Ivo D. Dinov	Nicolas Christou	Juana Sanchez	Dennis Pearl
SOCR Resource	SOCR Resource	SOCR Resource	CAUSE

Workshop Program At-A-Glance

http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

Day 1 (Mon 08/06/07)

Morning Session - Distributions (8AM - 12PM)

- Registration and Coffee (8:00 - 9:00 AM)
- Welcome, Ivo Dinov, SOCR Director (9:00 - 9:10 AM)
- Participants Introductions (9:10 - 9:20 AM)
- Guest Accounts (9:20 - 9:30 AM)
- The State of the SOCR Resource, Ivo Dinov, (9:30 - 9:40 AM)
- SOCR Distributions - Overview and Class Utilization, Ivo Dinov (9:40 - 9:50 AM)
- Distribution Activities I, Juana Sanchez (9:50 - 10:30 AM)
  - Poisson Distribution
- Morning Break (10:30 - 10:45 AM)
- Distribution Activities II, Nicolas Christou (10:45 - 11:30 AM)
  - SOCR Discrete Distributions Activity
  - Explore relations between Distributions
  - Solving Practical Problems using SOCR Distributions Applets
- Interactive Discussion on Distributions - What works, what doesn't, how to extend the collection and how improve teaching of discrete and continuous distributions (11:30AM - 12:00 PM)

Lunch Break (12:00 - 1:00 PM), UCLA Faculty Center

Afternoon Session - Experiments & Games (1PM - 4PM)

- SOCR/VLPS Experiments/Games - Overview and Class Utilization, Ivo Dinov (1:00 - 1:15 PM)
- SOCR/VLPS Experiments/Games Activities I, Juana Sanchez (1:15 - 2:00 PM)
  - Bivariate Activity
  - Birthday Experiment
- SOCR/VLPS Experiments/Games Activities II, Nicolas Christou (2:00 - 2:45 PM)
  - Confidence Interval Experiment
  - Roulette Experiment
- Break (2:45 - 3:00 PM)
- Assessing Conceptual Understanding in a First Statistics Course, Rob Gould (3:00 - 3:50 PM)
- Interactive Group Discussion on Experiments - What works, what doesn't, how to extend the collection and enhance the experiences of others (4:00 - 4:30 PM)

Dinner, 6:00-8:30 PM at the Redwood Room in the UCLA Faculty Center

Day 2 (Tue 08/07/07)

Morning Session - SOCR Analyses & Modeler (9AM - 12PM)

- SOCR Analyses - Overview and Class Utilization, Ivo Dinov (9:00 - 9:15 AM)
- SOCR Analyses, Annie Chu & Ivo Dinov (9:15 - 10:15 AM)
  - ANOVA
- Break (10:15 - 10:30 AM)
- SOCR Modeler, Ivo Dinov (10:30 - 11:30 AM)
  - SOCR Normal & Beta Distribution Model Fitting
- Group Interactive Discussion on Analyses & Modeler - What works, what doesn't, how to extend the collection and how improve teaching of statistical analysis methodologies? (11:30AM - 12:00 PM)

Lunch Break (12:00 - 1:00 PM) , UCLA Faculty Center

Afternoon Session - CAUSE Resources and UCLA Materials (1PM - 4PM)

- SOCR/VLPS Experiments/Games Activities III, Ivo Dinov (1:00 - 1:50 PM)
  - General Central Limit Theorem (CLT)
  - Law of Large Numbers
- Break (1:50 - 2:00 PM)
- CAUSE Resources for Teaching Statistics, Brian Jersky, St. Mary's College of California (2:00 - 2:50 PM)
- Break (2:50 - 3:00 PM)
- Using technology to enhance the role of writing and assessment in statistics education, Mahtash Esfandiari, Hai Nguyen, and Chris Barr (3:00 - 3:50 PM)
- Break (3:50 - 4:00 PM)
- Group Interactive Discussion - What works, what doesn't, how to extend the collection and experiences of other participants (4:00 - 4:15 PM)

Social: Movie Night/LACMA/TBD (UCLA Transportation is arranged 6-9 PM).

Day 3 (Wed 08/08/07)

Morning Session - EDA & Development of new Instructional Materials (9AM - 12PM)

- SOCR Charts, Nicolas Christou, Ivo Dinov & Jenny Cui (9:00 - 10:00 AM)
    - Box-and-Whisker Plots
    - Histogram Generation Activity
    - Random Number Generation
    - Power Transformation Family
- Break (10:00 - 10:15 AM)
- Wiki Editing and Development of new Instructional Materials, Ivo Dinov & Nicolas Christou (10:15 - 11:15 AM)
- Group Interactive Discussion on EDA and Development of Wiki Resources - What works, what doesn't? (11:15AM - 11:30 AM)
- Collaborative AP Statistics E-Book (11:30 - 11:45 AM)
- Workshop Evaluation (11:45 - 12:00 Noon)

Lunch Break & Adjourn (12:00 - 1:00 PM)

Afternoon Session - Visit to J. Paul Getty Museum (1PM - 4PM)

Workshop Logistics

We anticipate having about 35 workshop participants for this 3-day event. All participants will be partially supported to attend the workshop (travel, accommodation), by the NSF-funded SOCR & CAUSEway resources.

i. **Dates**: Mon-Wed, August 06-08, 2007
ii. **Times**: AM & PM Sessions (9AM - 12PM & 1PM - 4 PM)
iii. **Venue/Place**: Powell Library (CLICC Classroom B, Powell 320B, see map on the back)
iv. **Accommodation:** UCLA Guest House & Hilgard House
v. **Local Information**: Maps & local visitor information (see back cover)
vi. **Funding Support Details**: Participants will be staying at UCLA Guest House or the Hilgard House for 3 nights (Aug. 05, 06, 07, 2007), these costs are covered by the conference organizers and will be paid directly. Only no-shows will be charged for accommodation. Most meals will be provided during the Workshop; however, participants will be responsible for some of their meals. There is no Workshop registration fee nor will there be any charges for the Workshop materials which will be distributed. Transportation costs for attendees traveling > 30 miles (to UCLA) will be reimbursed to the maximum level of $300, per participant.

Workshop Goals

The overarching goals of this workshop are to provide continuing education and training for instructors using the latest SOCR and CAUSE educational resources and at the same time, provide an open learning environment for attendees to communicate and exchange ideas about validated existent educational materials, desired new resources and useful instructional tools.

In particular, we will discuss the diverse SOCR and CAUSE Internet resources, their design, usage, evaluation and extensibility. Among these are the SOCR Java applets for distributions, experiments, analysis, modeling and data exploration, various activities for hands-on demonstrations, as well as, students' and instructors' Internet resources.

Presenters & Attendees

PRESENTERS	
PRESENTER	**AFFILIATION**
Chris Barr	UCLA Statistics
Nicolas Christou	UCLA Statistics
Annie Chu	UCLA Statistics
Ivo Dinov	UCLA Statistics & CCB
Mahtash Esfandiari	UCLA Statistics
Rob Gould	UCLA Statistics
Brian Jersky	St. Mary's College of California
Hai Nguyen	UCLA Statistics
Juana Sanchez	UCLA Statistics

PRESENTERS	
ATTENDEE	**AFFILIATION**
Aggi Dellevoet	Palos Verdes Peninsula HS
Alfred Akinsete	Marshall University
Andrea Yarbrough	Etiwanda HS
Art Martinez	El Camino College
Deepak Sanjel	Minnesota State University
Elaine Merrill	BYU Hawaii
Grace Thomson	Nevada State College
Huizhen (Jean) Guo	Xavier University
James Engles	Eastern Technical HS
Jamie Fife	Brigham Young University Hawaii
Karen Berkeley	Parkville High School
Karen Walker	Chaparral HS
Kaysa Laureano	El Camino College
Lingling Ma	University of Kentucky
Marjorie Bond	Monmouth College
Mitra Moassessi	Santa Monica College
Morton A. Frankson	Loma Linda University
Nasser Dastrange	Buena Vista University
Paul Hurst	BYU-Hawaii
Paul J. Fields	Brigham Young University
Paula Easton	Los Osos HS
Sam Forster	City University of New York (SUNY)
Scott Hyde	Brigham Young University
Shiva Metghalchi	Loma Linda University
Susan Barton	BYU-Hawaii
Tahereh Zamansani	UCI Medical Center
Valerie Pacheco	Schurr High School
Veda M. Abu-Bakare	Langara College
Wayne Smith	CSU Northridge
ZANGUIM Maurice Bertin	C.E.S. of ONGOT

www.SOCR.ucla.edu
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

Workshop Evaluation Forms

The next four surveys and questionnaires contain Workshop evaluation materials that should be returned as per the instructions on each form. Please complete, tear off and return these forms to us in due time.

SOCR/CAUSEway Workshop Evaluation

Workshop Participants Expectations and Understanding of Workshop Evaluation Expectations

The Statistics Online Computational Resource (www.SOCR.ucla.edu) and the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) have received a NSF funds to develop educational resources and organize workshops and professional development events for statistics educators across the United States. This requires that an external evaluation be conducted to determine the effectiveness of these events and to identify strengths and limitations. Science and Mathematics Program Improvement (SAMPI) at Western Michigan University is conducting the evaluation of this workshop. As a workshop participant, you are being asked to participate in the evaluation. All information you provide is strictly confidential. When reporting results, neither your name nor your school will be linked to the data. Only group data will be reported.

BENEFITS AND SERVICES

Participants will have access to all opportunities provided by the workshop, receive a variety of resources for supporting teaching and learning of statistics, and receive financial support as described in recruitment materials.

EXPECTATIONS. Beyond the project expectations for participation and receipt of benefits and services, workshop participants will be expected to participate in the following evaluation activities:

- Complete a pre-workshop survey (at beginning of workshop)
- Complete an end-of-workshop questionnaire (at end of workshop)
- Participate in a follow-up telephone interview during the school year following the workshop
- Complete an email survey at the end of the school year following the workshop
- Assist in arranging an evaluator site visit during the school year following the workshop (sample of participants only)
- Assist evaluators in administering pre/post statistics content assessments for college students in selected classes (sample of participants only)
- Provide copies of syllabi and other curriculum and instructional materials created as a result of participating in the SOCR/CAUSEway workshop

The evaluation team will coordinate these activities with project staff so as to minimize intrusions and inconveniences. Questions about SOCR, CAUSEway and SAMPI should be directed to the project PIs: Ivo Dinov (SOCR), Dennis Pearl (OSU) and Mark Jenness (SAMPI).

Pedagogical Utilization of SOCR & CAUSEweb Resources in Probability and Statistics Courses
August 06-08, 2007, UCLA
Please Complete This Form on the First Day of the Workshop (08/06/07)!

SOCR/CAUSEway Workshop Evaluation

Workshop Participants Expectations and Understanding of Workshop Evaluation Expectations

The Statistics Online Computational Resource (www.SOCR.ucla.edu) and the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) have received a NSF funds to develop educational resources and organize workshops and professional development events for statistics educators across the United States. This requires that an external evaluation be conducted to determine the effectiveness of these events and to identify strengths and limitations. Science and Mathematics Program Improvement (SAMPI) at Western Michigan University is conducting the evaluation of this workshop. As a workshop participant, you are being asked to participate in the evaluation. All information you provide is strictly confidential. When reporting results, neither your name nor your school will be linked to the data. Only group data will be reported.

BENEFITS AND SERVICES

Participants will have access to all opportunities provided by the workshop, receive a variety of resources for supporting teaching and learning of statistics, and receive financial support as described in recruitment materials.

EXPECTATIONS. Beyond the project expectations for participation and receipt of benefits and services, workshop participants will be expected to participate in the following evaluation activities:

- Complete a pre-workshop survey (at beginning of workshop)
- Complete an end-of-workshop questionnaire (at end of workshop)
- Participate in a follow-up telephone interview during the school year following the workshop
- Complete an email survey at the end of the school year following the workshop
- Assist in arranging an evaluator site visit during the school year following the workshop (sample of participants only)
- Assist evaluators in administering pre/post statistics content assessments for college students in selected classes (sample of participants only)
- Provide copies of syllabi and other curriculum and instructional materials created as a result of participating in the SOCR/CAUSEway workshop

The evaluation team will coordinate these activities with project staff so as to minimize intrusions and inconveniences. Questions about SOCR, CAUSEway and SAMPI should be directed to project PIs: Ivo Dinov (SOCR), Dennis Pearl (OSU) and Mark Jenness (SAMPI).

Signature of Workshop Participant: _____ Date: _____

Please print your name: _____

This memo of agreement is not legally binding, but represents a good faith commitment to full participation in the SOCR/CAUSEway evaluation by this participant for one year beginning with the workshop.

10

SOCR/CAUSEway Workshop Evaluation

Pre-Program Participant Survey

As part of the required evaluation of this workshop, participants are asked to complete the following *16-question pre-program survey*. The information will be used to help us learn about the effects the program. You will be asked to complete a similar survey at the end of next school year. The information is strictly confidential--no one except project evaluators will see individual survey results. Only group data will be reported. The code number is for follow-up purposes and to analyze pre- and post-program data.

When You Have Completed The Survey, Return It To The Workshop Organizers

Thanks for taking time to complete this survey! If you have questions, please ask the facilitator.

PART A: About you.

1. What courses do you teach that include instruction in statistics? _____

   _____

2. How many years have you been an instructor? _____

3. What is the subject-area of your highest college degree? _____

4. Are you a member of any national statistics education professional organizations? ____Yes ___ No

   If yes, which one(s)? _____

5. Why did you choose to participate in this SOCR/CAUSEway workshop?

6. What are your expectations for this workshop? What do you want to get out of it?

7. How did you learn about this workshop?

8. Were you familiar with SOCR (the Statistics Online Computational Resource) before enrolling in this workshop?

9. Were you familiar with CAUSE (the Consortium for the Advancement of Undergraduate Statistics Education) before enrolling in this workshop?

9. Are you familiar with the www.SOCR.ucla.edu and www.CAUSEweb.org?

10. How <u>well prepared</u> are you to teach the following statistics topics in the courses you teach? Rate each item on a 4 point scale, with 1 = not adequately prepared and 4 = very well prepared.

	Not Adequately Prepared			Very Well Prepared
a. Data collection (surveys and experiments)	1	2	3	4
b. Summary statistics and graphics (such as histograms and boxplots)	1	2	3	4
c. Probability	1	2	3	4
d. Sampling distributions	1	2	3	4
e. Confidence intervals	1	2	3	4
f. Hypothesis testing (one sample for means and proportions)	1	2	3	4
g. Simple linear regression and correlation	1	2	3	4
h. Using graphing calculators in statistics	1	2	3	4

11. How well prepared are you to facilitate the following in the classes you teach? Rate each item on a 4-point scale, with 1 = not adequately prepared and 4 = very well prepared.

	Not Adequately Prepared			Very Well Prepared
a. Emphasizing statistical literacy	1	2	3	4
b. Developing statistical thinking	1	2	3	4
c. Using real data	1	2	3	4
d. Focusing on conceptual understanding rather than mere knowledge of procedures	1	2	3	4
e. Fostering active learning among my students	1	2	3	4
f. Using technology for developing conceptual understanding	1	2	3	4
g. Using technology to analyze data	1	2	3	4
h. Use assessment to improve student learning	1	2	3	4
i. Using assessment to evaluate student learning	1	2	3	4

12. About <u>how often</u> do you do each of the following in a class? Rate each item on a 5-point scale:
1=never; 2=Rarely (1-2 times/ term); 3=Sometimes (3-4 times/ term); 4=Often; and 5=in almost all lessons

a. Introduce content through formal presentations	1	2	3	4	5
b. Use open-ended questioning strategies	1	2	3	4	5
c. Require students to explain their reasoning when giving an answer	1	2	3	4	5
d. Encourage students to explore alternative methods for solutions	1	2	3	4	5
e. Help students make connections between statistics and real-world situations	1	2	3	4	5

f. Read and comment on reflections students have written in their notebooks or journals	1	2	3	4	5

13. About how often do each of the following occur for your statistics classes? Rate each item on a 5-point scale, with 1 = never and 5 = during almost all lessons.

	Never	Rarely	Some times	Often	Almost always
a. I have access to a computer lab to use with my students during class.	1	2	3	4	5
b. When teaching, I have access to a computer projection system in my classroom.	1	2	3	4	5
c. When teaching, I have access to an Internet connection in my classroom.	1	2	3	4	5
d. My students have access to graphing calculators.	1	2	3	4	5

14. Below are several statements about statistics teaching and learning. Rate the degree to which you agree or disagree with each statement on a 5-point scale, with 1 = strongly disagree to 5 = strongly agree.

	Strongly Disagree				Strongly Agree
a. Every student should feel that statistics is something she/he can do.	1	2	3	4	5
b. It is sometimes productive for students to work together during statistics class to conduct investigations or solve statistics problems.	1	2	3	4	5
c. You have to study statistics for a long time before you see how useful it is.	1	2	3	4	5
d. Memorization plays an important role in learning statistics.	1	2	3	4	5
e. A lot of things in statistics must be simply accepted as true and remembered.	1	2	3	4	5
f. Students' achievement in statistics is directly related to their teacher's effectiveness in teaching these subjects.	1	2	3	4	5
g. When teaching statistics, I usually welcome student questions.	1	2	3	4	5

15. About how often do your <u>students</u> take part in each of the following types of activities as part of their statistics lessons? Rate each on a scale of 1 to 5:

<u>1=never; 2=Rarely (1-2 times/ term); 3=Sometimes (3-4 times/ term); 4=Often; and 5=in almost all lessons</u>

a. Participate in discussion with the teacher to further statistical understanding	1	2	3	4	5
b. Work independently	1	2	3	4	5
c. Make formal presentations to the class	1	2	3	4	5
d. Answer textbook/worksheet questions	1	2	3	4	5
e. Work on solving a real-world problem	1	2	3	4	5
f. Share ideas or solve problems in small groups	1	2	3	4	5
g. Engage in hands-on statistical activities (simulations, data collection, etc.)	1	2	3	4	5
h. Design or implement their own investigation	1	2	3	4	5
i. Work on extended investigations or projects (a week or more in duration)	1	2	3	4	5
j. Write a description of a plan, procedures, or problem-solving process	1	2	3	4	5
k. Write reflections in a notebook or journal	1	2	3	4	5
l. Use calculators for learning or practicing skills	1	2	3	4	5
m. Use calculators as a tool (e.g., spreadsheets, data analysis).	1	2	3	4	5
n. Use calculators to develop conceptual understanding	1	2	3	4	5
o. Use computers for learning or practicing skills	1	2	3	4	5
p. Use computers as a tool (e.g., spreadsheets, data analysis)	1	2	3	4	5
q. Use computers to develop conceptual understanding	1	2	3	4	5
r. Take short-answer tests (e.g., multiple choice, true/false, fill-in-the-blank)	1	2	3	4	5
s. Take tests requiring constructed responses (e.g. definition-type questions)	1	2	3	4	5
t. Engage in performance tasks (i.e. create a product, make a presentation) for assessment purposes	1	2	3	4	5

16. What are the major issues or concerns for you related to the teaching and learning of statistics at your grade level? Use back of this page if you need more room.

14

www.SOCR.ucla.edu
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

Pedagogical Utilization of SOCR & CAUSEweb Resources in Probability and Statistics Courses
August 06-08, 2007, UCLA
Please Complete this Form on the Last day of the Workshop (08/08/07)!

End-of-Workshop Evaluation Questionnaire

As part of the required evaluation of SOCR and CAUSEway, workshop participants are asked to respond to the following about the 2007 SOCR/CAUSEway workshop. Your comments are important in helping us improve our materials, extend our resources and fine-tune our pedagogical approaches. Your responses are anonymous & confidential. They will be compiled and reported only as group data. DO NOT WRITE YOUR NAME ON THIS FORM.

The evaluation of this Workshop is being conducted independently by SAMPI. If you have questions, please contact Mark Jenness (Email: mark.jenness@wmich.edu or http://www.wmich.edu/sampi). **We appreciate your comments!**

A. ABOUT YOU:

1. What **courses** do you teach (that the material of this workshop may be relevant to)? Please include: title, Upper/Lower/Graduate division, number of students and number of offerings per year.

_____

_____

2. How many years have you been an instructor in these courses?

3. What attracted you to apply and take part in this Workshop?

B. WORKSHOP OUTCOMES. The workshop sessions were designed to address various strategies for teaching introductory statistics. Please rate each one of the associated objectives according to:
1) Your perception of the VALUE (V) of the session objective and
2) Whether you think it was ACCOMPLISHED (A)
Note: "1" represents the lowest score; a "5" represents the highest score. Please make comments.

The workshop sessions helped me to do the following:

		1	2	3	4	5	
1. Learn how to emphasize statistical literacy	V						Comments:
	A						
2. Learn how to develop statistical thinking	V						Comments:
	A						
3. Learn to use real data	V						Comments:
	A						
4. Learn how to focus on conceptual understanding rather than only knowledge of procedures	V						Comments:
	A						

15

www.SOCR.ucla.edu
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

5. Foster active learning among your students	V						Comments:
	A						

6. Learn to use technology for developing conceptual understanding	V						Comments:
	A						

7. Learn to use technology to analyze data	V						Comments:
	A						

8. Use assessment to improve student learning	V						Comments:
	A						

9. Use assessment to evaluate student learning	V						Comments:
	A						

C. WORKSHOP SESSIONS. Rate the following on their usefulness in helping you with teaching and learning of mathematics on a scale of 1-5, with 1 = not useful and 5 = very useful. Please make comments.

	1	2	3	4	5	
10. Planning a conceptual course using common threads and big ideas						Comments:
11. Making the class interactive with activities						Comments:
12. Finding resources using CAUSEweb						Comments:
13. Finding and using real data						Comments:
14. Using technology in the classroom						Comments:
15. Assessment in the classroom						Comments:
16. Participant discussion and presentation						Comments:

D. WORKSHOP ARRANGEMENTS. Rate the following on a scale of 1-5, with 1 = Disagree and 5 = Agree.

(Disagree) 1   2   3   4   5 (Agree)

	1	2	3	4	5	
17. Workshop facilities were satisfactory.						Comments:
18. Workshop facilitators were effective in communicating ideas and issues						Comments:
19. Workshop facilitators were effective in organizing sessions so that I was actively involved.						Comments:
20. A collaborative and helpful tone was established during the session.						Comments:

21. What two or three BIG ideas about the teaching and learning of statistics did you learn during this workshop?

22. In what ways could this workshop be improved?

23. In what ways do you plan to use what you have learned in this workshop in your own teaching?

About the SOCR Resource & CAUSE Initiative

The SOCR Resource

SOCR (http://www.SOCR.ucla.edu) is an NSF-funded project (DUE 0442992) that designs, implements, validates and integrates various interactive tools for statistics and probability education and computing. Many of the projects and initiatives are directed at the basic introduction to probability and statistics courses. SOCR resource tools attempt to bridge between the introductory and the more advanced computational and applied probability and statistics courses. There are four major types of SOCR users: educators, students, researchers and tool developers. The proposed workshop is intended for educators. Course instructors and teachers find the SOCR class notes and interactive tools useful for student motivation, concept demonstrations and for enhancing their technology based pedagogical approaches to any study of variation and uncertainty. Students and trainees find the SOCR class notes, analyses, computational and graphing tools extremely useful in their learning/practicing pursuits. Model developers, software programmers and other engineering, biomedical and applied researchers may find the light-weight plug-in oriented SOCR computational libraries and infrastructure useful in their algorithm designs and research efforts.

The **main objective of SOCR** is to offer a homogeneous interface for online activities appropriate for the Introductory Statistics Course, Introductory Probability course and other advanced Statistics courses that rely on hands on demonstrations and simulation to illustrate Statistical concepts. A common portal for all SOCR activities is very important to minimize the amount of time that students have to spend learning the technology. SOCR materials and activities have received recognitions from several international, educational and technology-based initiatives (http://www.socr.ucla.edu/htmls/SOCR_Recognitions.html). SOCR has been, and continuous to be, tested in the classroom. Most recently, a large-scale experimental study we conducted led us to conclude that using SOCR for the teaching of Introductory Statistics and Probability was effective (Dinov and Sanchez 2006; Dinov, Sanchez et al. 2006). Thus, the SOCR/CAUSE workshop offers participants research-based knowledge on effective teaching with online learning resources, which is the best kind of teaching strategies (Richard, Shavelson et al. 2002).

Until now, many technology inclined educators have adopted in their course curriculum interactive aids (e.g., applets) from diverse and heterogeneous resources. Many instructors have also created their own IT-instruments to enhance their pedagogical approaches. The **implementation of the SOCR philosophy** for developing and utilizing new IT-resources (tools) and educational materials is depicted on Table 1. The subdivision of all SOCR resources into tools and materials is natural as our goals are similarly dichotomous; First,

SOCR PHILOSOPHY	
DEVELOPMENT OF TOOLS	**DEVELOPMENT OF EDUCATIONAL MATERIALS**
2.Tools must be freely available on the Web (SOCR Motto: *It's online, therefore it exists!*)	1. Newly developed materials must increase pedagogical content knowledge, be extensible, factually correct, validated and deployed on the web.
2. Tools must be well-designed, extensible, platform-independent, open-source and connected to some activity that increases pedagogical content knowledge.	2. Materials must explicitly utilize some SOCR interactive resources (e.g., activities with applet demos) and provide the means of cross reference by various SOCR tools (e.g., applet calculations citing formulas in instructional materials).

Table 1: SOCR working philosophy in development of new tools and educational materials.

develop libraries and foundational instruments for demonstration, motivation, visualization and statistical computing, which are platform agnostic. And second, design instances of course-, topic- and student-specific educational materials (lecture notes, activities, assignments, etc.) which are agile and extensible wrappers around the available SOCR tools. Both of these categories are open-source and may be extended, revised and redistributed by others in the community. For example, technically savvy users may quickly implement a new SOCR Analysis object, add an additional SOCR Distribution, extend the functionality of a SOCR Experiment, etc. by simply implementing as a plug-in the corresponding SOCR Java object. Redistributing the new tool to the community only requires posting of the new tool on the SOCR web page, it does not require complete SOCR package rebuild or restructuring. Educators that are more interested in the application of the SOCR tools and their in-class utilization also may actively contribute to the SOCR efforts by developing new, improving existent and testing and validating the SOCR educational materials. In fact, just like Wikipedia, the entire SOCR effort is contingent upon the continued support and development efforts of the community (educators and researchers in the areas of probability, statistics, mathematics, data modeling, etc.). We simply provide the infrastructure for these developments; the user community is responsible for the rest.

The CAUSE Initiative

The Consortium for the Advancement of Undergraduate Statistics Education) is a product of a strategic initiative of the American Statistical Association, a national organization whose mission is to support and advance undergraduate statistics education, in four target areas: resources, professional development, outreach, and research. CAUSEweb (http://www.causeweb.org/) is an Internet-based collection of activities, lessons, literature and other resources put together by CAUSE. CAUSEway is an initiative that resulted from this effort and is intended to encourage the dissemination of new ideas in the teaching of Statistics, primarily the introductory Statistics courses. CAUSE brings to our Workshop experience in organizing Introductory Statistics education and research workshops and a whole set of resources compiled from all over the World. This integrated SOCR/CAUSEway workshop proposed is the first one of its kind, where a

unique and uniform interface is used and where the main focus is on the activities and how they can enhance pedagogical content knowledge. Other resources provided by CAUSE, such as literature on research on the effect of those online tools will help participants put the activities in context and judge when they have and where they have not worked.

A critical component of the SOCR resource development effort is that we build and disseminate IT-infrastructure, along with instances of use-cases demonstrating our classroom utilization of these resources. We do not design and distribute tools and materials that may be used only in the specific context for which they were developed. We strive to provide excellent IT-based instruments for STEM education for all undergraduate and AP students and to engage instructors in the process. The critical instructor involvement is in the form of development of custom tailored SOCR-based materials, exchange ideas and community interaction in the development and efficacy of available educational materials. Some other IT-based initiatives distribute excellent but static instructional materials. We have instead focused on development of a general framework that allows instructors to use the SOCR resources to design course-specific activities and tools that fit their syllabi, student populations and topics covered. Other statistics education workshops typically do not have this important feature, and we hope the SOCR/CAUSEway Workshop participants will remain engaged in the IT-resource development process long after the workshop.

Workshop Activities and Materials

Mon 08/06/07

Morning Session (9 AM – 12 Noon)

- o **9:50 - 10:30 AM** Distribution Activities I, Juana Sanchez

  i. *Poisson Distribution* – This is an activity to explore the Poisson Probability Distribution.

  - **Description**: You can access the applets for the above distributions at http://www.socr.ucla.edu/htmls/SOCR_Distributions.html .

  - **Exercise 1:** Use SOCR to graph and print the distribution of a Poisson random variable with $\lambda = 2$. What is the shape of this distribution?

  - **Exercise 2:** Use SOCR to graph and print the distribution of a Poisson random variable with $\lambda = 15$. What is the shape of this distribution? What happens when you keep increasing λ?

  - **Exercise 3:** Let $X \sim Poisson(5)$. Find $P(3 \leq X < 10)$, and $P(X > 10 | X \geq 4)$.

  - **Exercise 4:** Poisson approximation to binomial: Graph and print $X \sim b(60, 0.02)$. Approximate this probability distribution using Poisson. Choose three values of *X* and compute the probability for each one using Poisson and then using binomial. How good is the approximation?

    Below you can see the distribution of a Poisson random variable with $\lambda = 5$. In this graph you can also see the probability that between 2 and 5 events will occur.
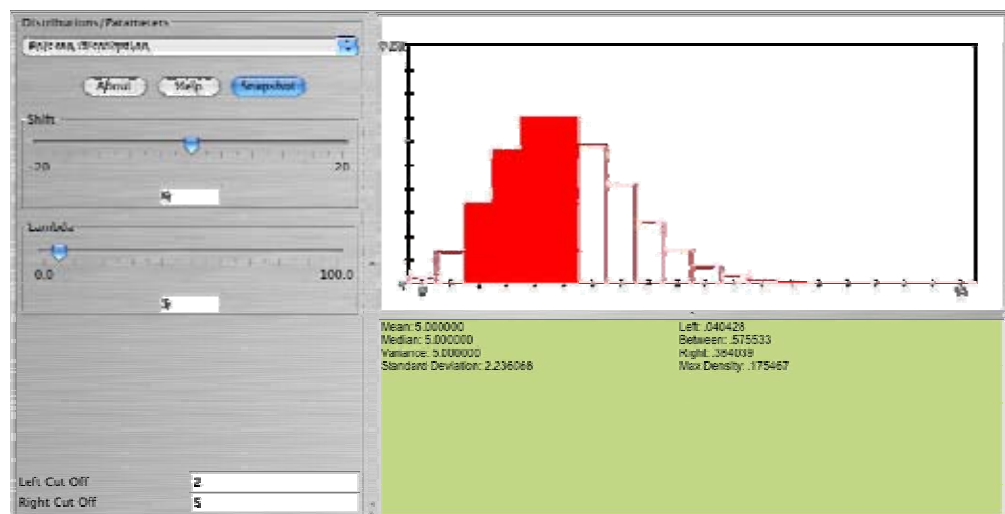
- o **10:45 - 11:30 AM** Distribution Activities II, Nicolas Christou
  - i. *SOCR Discrete Distributions Activity* – This is an activity to explore the Binomial, Geometric, and Hypergeometric Probability Distributions.

  - **Description**: You can access the applets for the above distributions at http://www.socr.ucla.edu/htmls/SOCR_Distributions.html .

  - **Exercise 1:** Use SOCR to graph and print the following distributions and answer the questions below. Also, comment on the shape of each one of these distributions:
    - o a. $X \sim b(10, 0.5)$, find $P(X = 3)$, $E(X)$, $sd(X)$, and verify them with the formulas discussed in class.
    - o b. $X \sim b(10, 0.1)$, find $P(1 \leq X \leq 3)$.
    - o c. $X \sim b(10, 0.9)$, find $P(5 < X < 8)$, $P(X < 8)$, $P(X \leq 7)$, $P(X \geq 9)$.
    - o d. $X \sim b(30, 0.1)$, find $P(X > 2)$.

    Below you can see a snapshot of the distribution of $X \sim b(20, 0.3)$
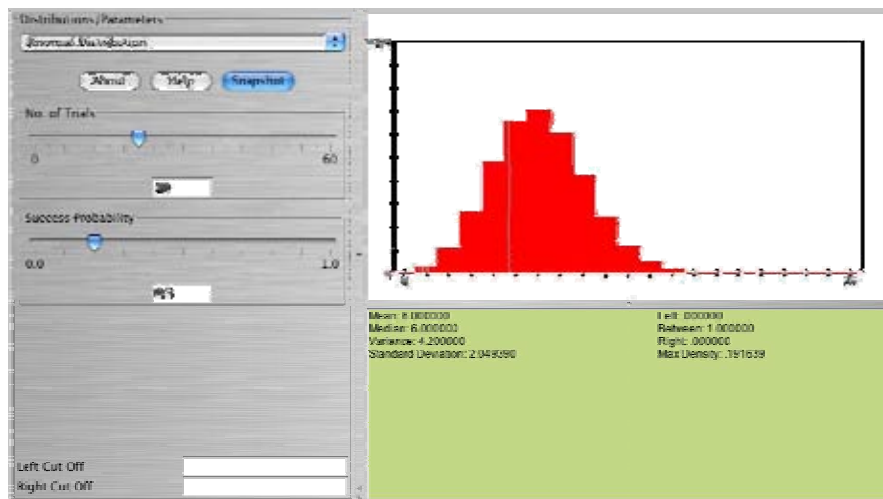
  - **Exercise 2:** Use SOCR to graph and print the distribution of a geometric random variable with $p = 0.2, p = 0.7$. What is the shape of these distributions? What happens when p is large? What happens when p is small?

    Below you can see a snapshot of the distribution of $X \sim geometric(0.4)$
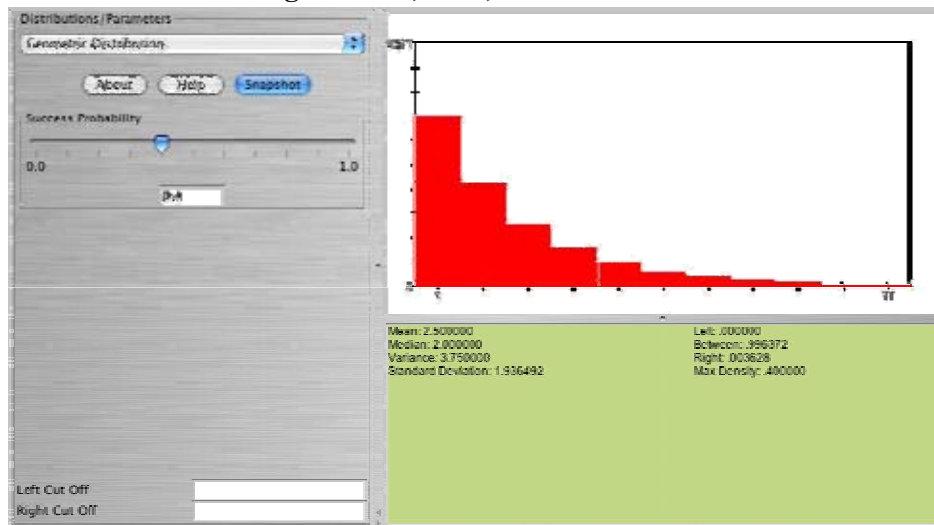
- **Exercise 3:** Select the geometric probability distribution with $p = 0.2$. Use SOCR to compute the following:

  a. $P(X = 5)$
  b. $P(X > 3)$
  c. $P(X \leq 5)$
  d. $P(X > 6)$
  e. $P(X \geq 8)$
  f. $P(4 \leq X \leq 9)$
  g. $P(4 < X < 9)$

- **Exercise 4:** Verify that your answers in exercise 3 agree with the formulas discussed in class, for example, $P(X = x) = (1 - p)^{x-1}p$, $P(X > k) = (1 - p)^k$, etc. Write all your answers in detail using those formulas.

- **Exercise 5:** Let X follow the hypergeometric probability distribution with $N = 52$, $n = 10$, and number of "hot" items 13. Use SOCR to graph and print this distribution.

  Below you can see a snapshot of the distribution of
  $$X \sim hypergeometric(N = 100, n = 15, r = 30)$$
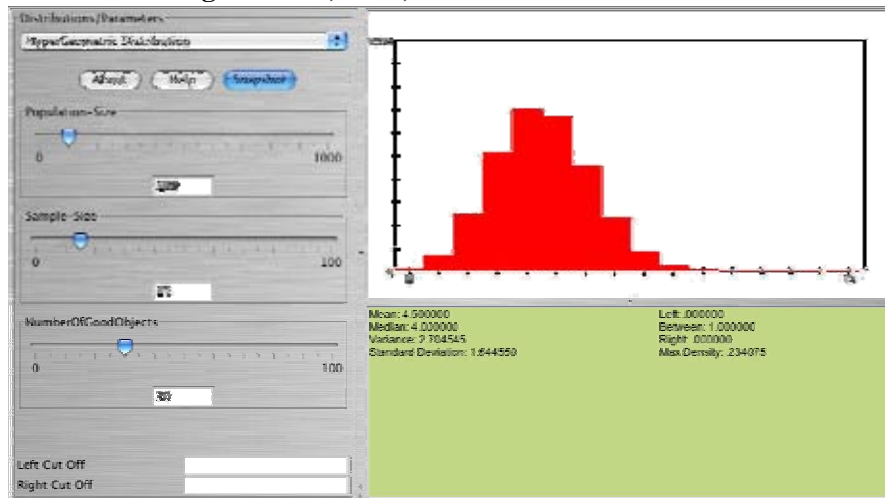
- **Exercise 6:** Refer to exercise 5. Use SOCR to compute $P(X = 5)$ and write down the formula that gives this answer.

- **Exercise 7:** Binomial approximation to hypergeometric: Let X follow the hypergeometric probability distribution with $N=1,000, n=10$ and number of "hot" items 50. Graph and print this distribution.

- **Exercise 8:** Refer to exercise 7. Use SOCR to compute the exact probability: $P(X = 2)$. Approximate $P(X = 2)$ using the binomial distribution. Is the approximation good? Why?

- **Exercise 9:** Do you think you can approximate well the hypergeometric probability distribution with $N=50, n=10$, and number of "hot" items 40 using the binomial probability distribution? Explain.

ii. *Explore relations between Distributions* – This is a SOCR activity to explore the relationships among some of the commonly used probability distributions. This is a complementary SOCR activity to the SOCR Distributions Activity.

Set-up: go to the SOCR Distribution applets for these activities by going to this URL page http://www.socr.ucla.edu/htmls/SOCR_Distributions.html and use a Java-enabled browser,

Geometric probability distribution

Suppose we roll two dice until a sum of 10 is obtained. What is the probability that the first sum of 10 will occur after the 5th trial? The answer to this question is

$$P(X > 5) = (1 - \frac{3}{36})^5 = 0.6472.$$ This is equivalent to the event that no sum of 10 is observed on the first 5 trials (5 failures). Now, using SOCR we can

24

obtain this probability easily by entering in the SOCR geometric distribution applet $p = \dfrac{3}{36} = 0.0833$ and in the Right Cut-Off box 5. We can find the desire probability on the right corner of the applet. The figure below clearly displays this probability.
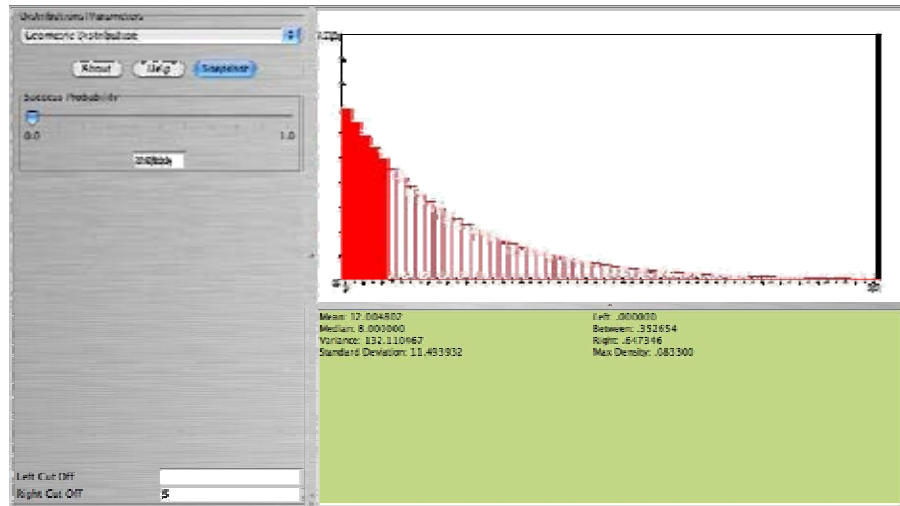
Binomial approximation to hypergeometric

An urn contains 50 marbles (35 green and 15 white). Fifteen marbles are selected without replacement. Find the probability that exactly 10 out of the 15 selected are green marbles. The answer to this question can be found using the formula:

$$P(X = 10) = \frac{\binom{35}{10}\binom{15}{5}}{\binom{50}{15}} = 0.2449.$$

Using SOCR simply enter population size 50, sample size 15, and number of good objects 35, to get the figure below.
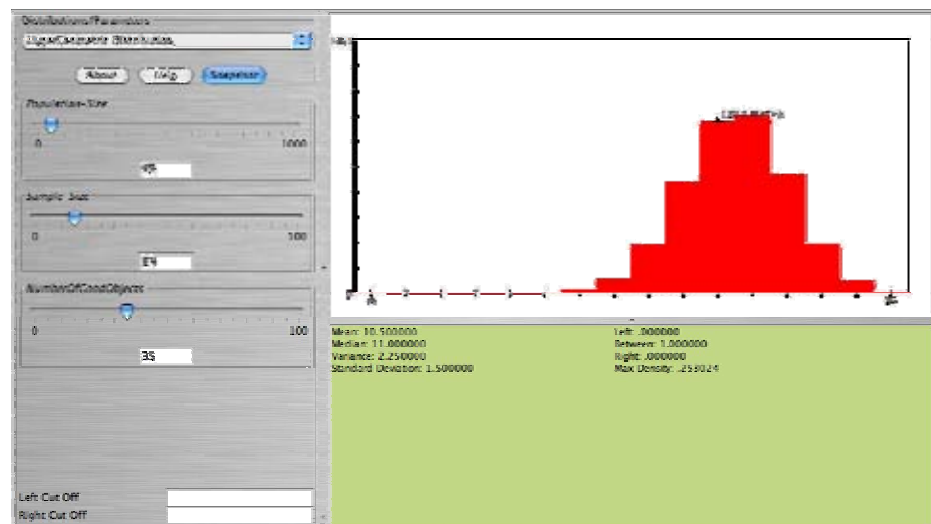
Now, select without replacement only 2 marbles. Compute the exact probability

$$P(X = 1) = \frac{\binom{35}{1}\binom{15}{1}}{\binom{50}{2}} = 0.4286.$$

that 1 green marble is obtained. This is equal to
This is also shown on the figure below.
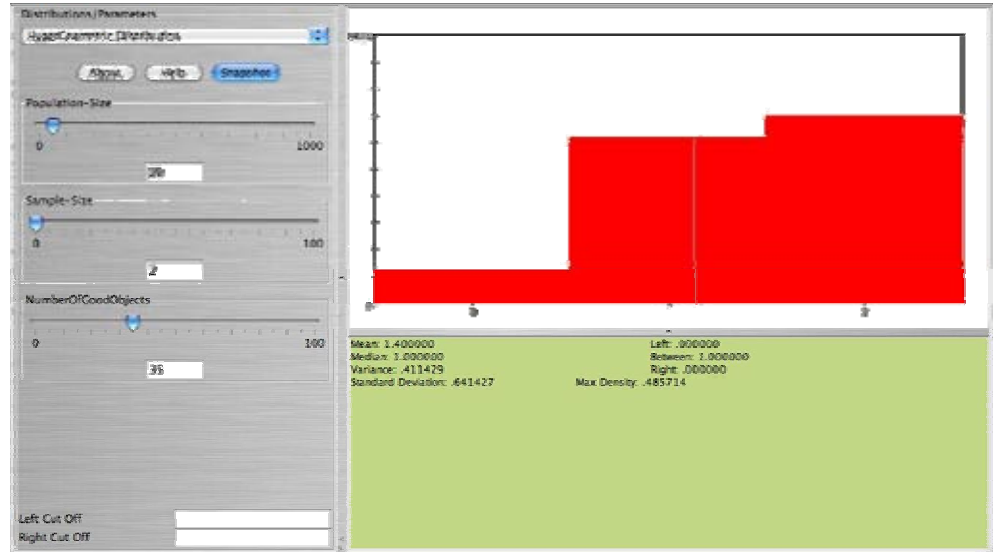
We will approximate the probability of obtaining 1 green marble using Binomial as follows. Select the SOCR Binomial distribution and choose number of trials 2 and probability of success $p = \frac{35}{50} = 0.7$. Compare the figure below with the figure above. They are almost the same! Why? Using the Binomial formula, we can compute the approximate probability of observing 1 green marble as

$$P(X = 1) = \binom{2}{1}0.70^1 0.30^1 = 0.42$$

(very close to the exact probability, **0.4286**).
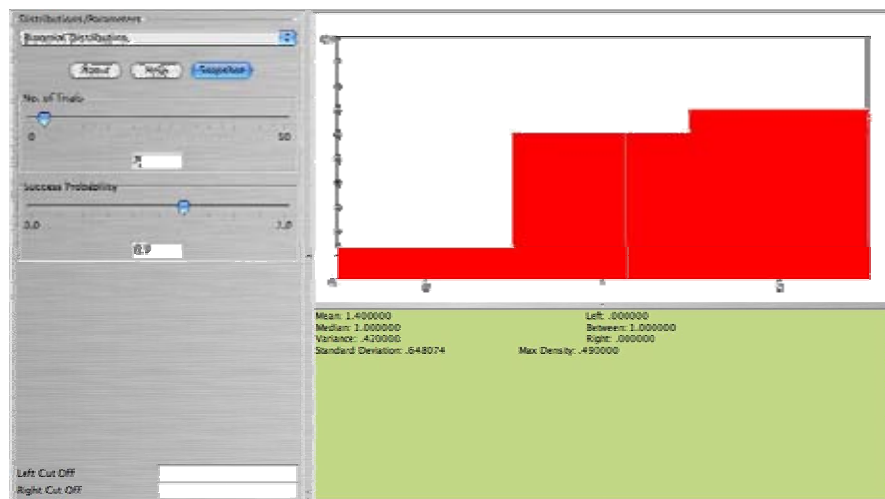
Normal approximation to Binomial

Graph and comment on the shape of binomial with $n = 20, p = 0.1$ and $n = 20, p = 0.9$. Now, keep $n = 20$ but change $p = 0.45$. What do you observe now? How about when $n = 80, p = 0.1$. See the four figures below.
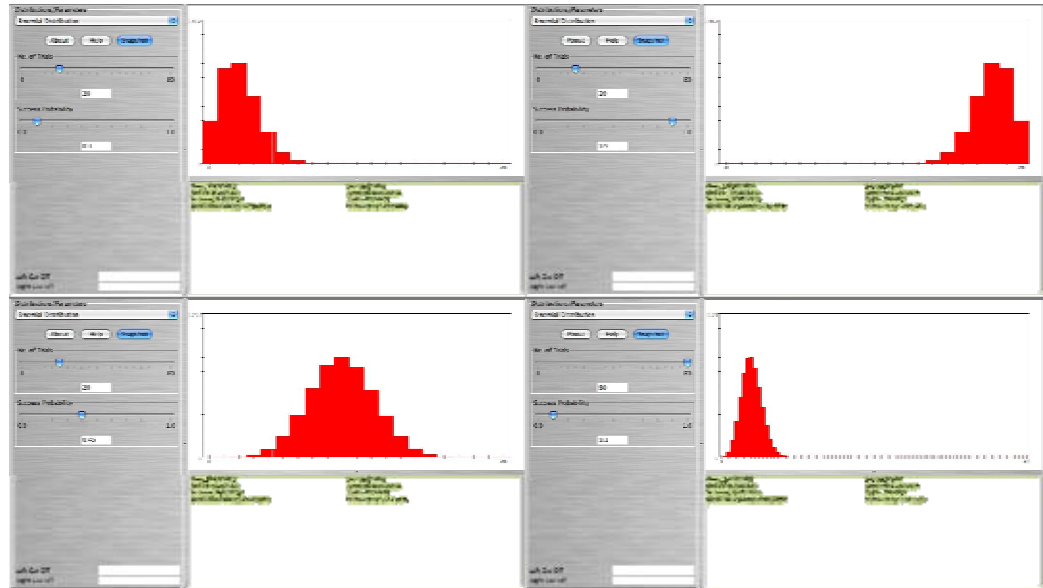
What is your conclusion on the shape of the Binomial distribution in relation to its parameters n, p? Clearly when n is large and p small or large the result is a bell-shaped distribution. When n is small (10-20) we still get approximately a bell-shaped distribution as long as $p \approx 0.5$. Because of this feature of the Binomial distribution we can approximate Binomial distributions using the normal distribution when the above requirements hold. Here is one example: Eighty cards are drawn with replacement from the standard 52-card deck. Find the exact probability that at least 8 aces are obtained. This can be computed using the formula

$$P(X \geq 8) = \sum_{x=8}^{80} (\frac{4}{52})^x (\frac{48}{52})^{80-x} = 0.2725 \quad .$$

Much easier, we can use SOCR to compute this probability (see figure below).
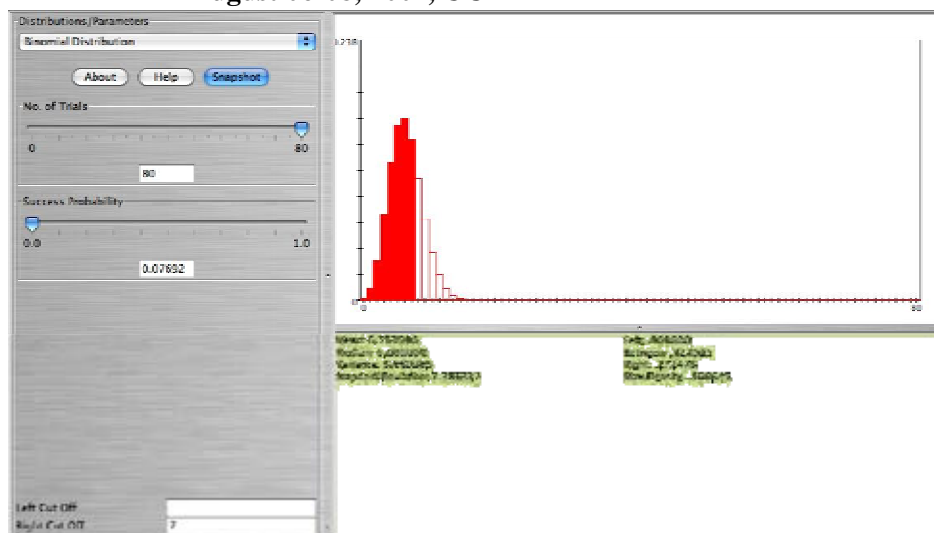
But we can also *approximate* this probability using the normal distribution. We will need the mean and the standard deviation of this normal distribution. These are

$$\mu = np = 80\frac{4}{52} = 6.154 \qquad \sigma = \sqrt{80\frac{4}{52}\frac{48}{52}} = 2.383.$$

Of course, this can be obtained directly from the SOCR Binomial applet. Now, all you need to do is to select the SOCR normal distribution applet and enter for the mean 6.154, and for the standard deviation 2.383. To obtain the desire probability in the right cut-off box enter 7.5 (using the continuity correction for better approximation). The approximate probability is $P(X \geq 8) \approx 0.2861$ (see figure below).
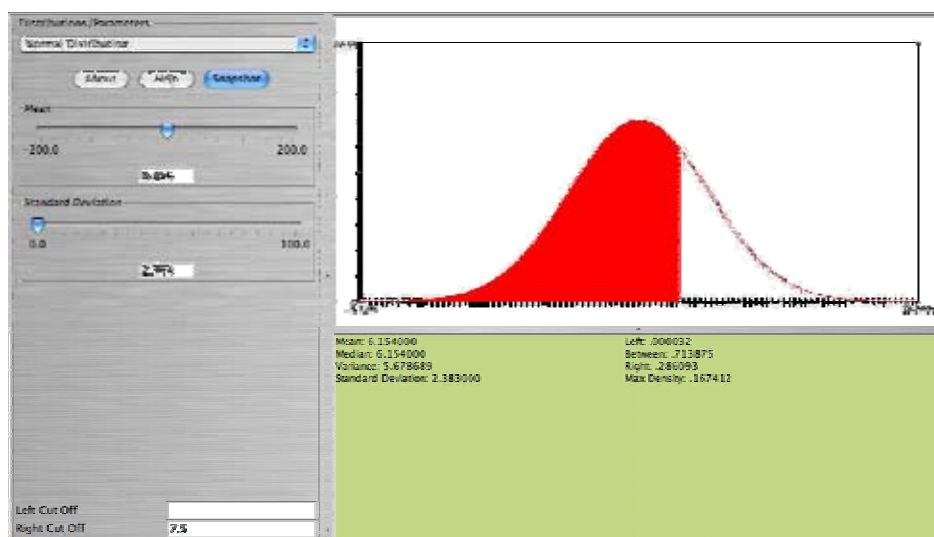
Normal approximation to Poisson

The [Poisson distribution](#) with parameter λ can be approximated with normal when λ is large. Here is one example. Suppose cars arrive at a parking lot at a rate of 50 per hour. Let's assume that the process is a Poisson random variable with λ = 50. Compute the probability that in the next hour the number of cars that arrive at this parking lot will be between 54 and 62. We can compute this as follows:

$$P(54 < X < 62) = \sum_{x=54}^{62} \frac{50^x e^{-50}}{x!} = 0.2617.$$

The figure below from SOCR shows this probability.
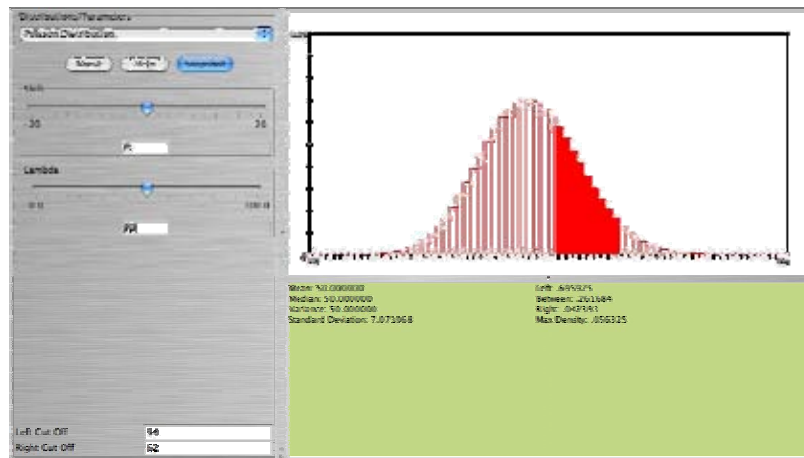
- **Note**: We observe that this distribution is bell-shaped. We can use the normal distribution to approximate this probability. Using $N(\mu = 50, \sigma = \sqrt{50} = 7.071)$, together with the continuity correction for better approximation we obtain $P(54 \leq X \leq 62) = 0.2718$, which is close to the exact that was found earlier. The figure below shows this probability.
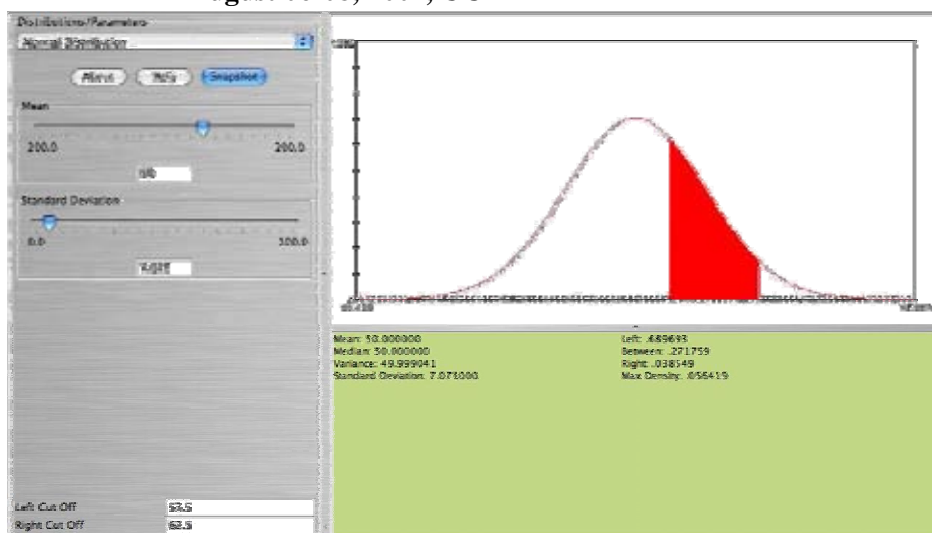
Poisson approximation to Binomial

The [Binomial distribution](#) can be approximated well by Poisson when *n* is large and *p* is small with $np < 7$. This is true because $\lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!}$, where $\lambda = np$. Here is an example. Suppose 2% of a certain population have Type AB blood. Suppose 60 people from this population are randomly selected. The number of people *X* among the 60 that have Type AB blood follows the Binomial distribution with $n = 60, p = 0.02$. The figure below represents the distribution of *X*. This figure also shows $P(X = 0)$.
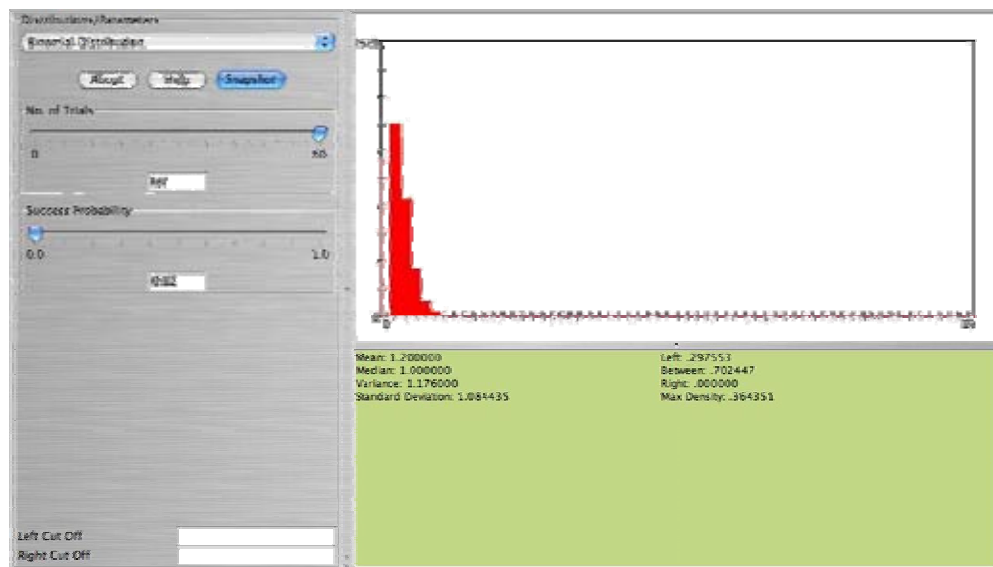
- **Note**: This distribution can be approximated well with Poisson with $\lambda = np = 60(0.02) = 1.2$. The figure below is approximately the same as the figure above (the width of the bars is not important here. The height of each bar represents the probability for each value of *X,* which is about the same for both distributions).

30
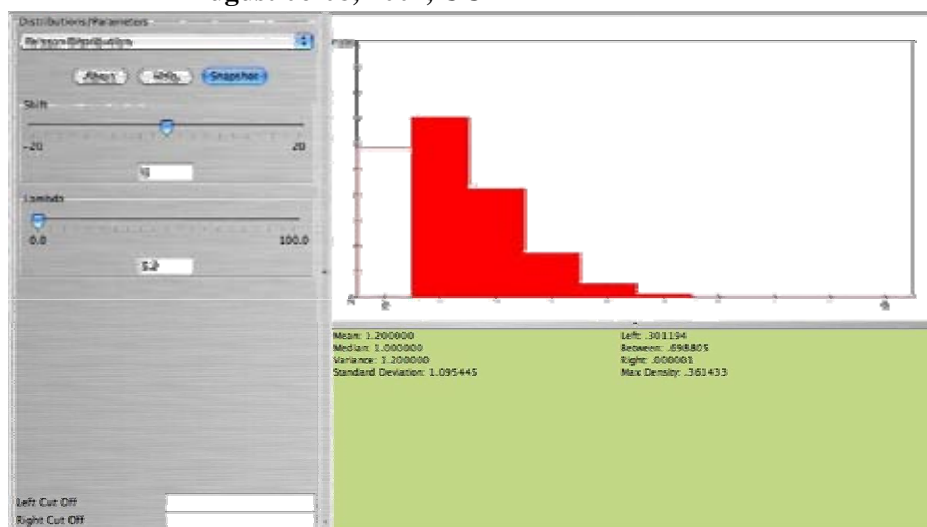
iii. *Solving Practical Problems using SOCR Distributions Applets*

This activity demonstrates solving practical problems using the SOCR Distribution Applets.

Example 1: From a large shipment of peaches, 12 are selected for quality control. Suppose that in this particular shipment only *65%* of the peaches are unbruised. If among the 12 peaches 9 or more are unbruised the shipment is classified A. If between 5 and 8 are unbruised the shipment is classified B. If fewer than 5 are unbruised the shipment is classified C. Compute the probability that the shipment will be classified A, B, C. Of course, computing the probabilities of two of these events yields the third probability, as

$$1 = P(A \cup B \cup C) = P(A) + P(B) + P(C).$$

We can use the Binomial distribution formula to compute

$$P(A) = P(X \geq 9) = \sum_{x=9}^{12} \binom{12}{x} 0.65^{x} 0.35^{12-x} = \cdots$$

$$P(B) = P(5 \leq X \leq 8) = \sum_{x=5}^{8} \binom{12}{x} 0.65^{x} 0.35^{12-x} = \cdots$$

$$P(C) = P(X < 5) = \sum_{x=0}^{4} \binom{12}{x} 0.65^{x} 0.35^{12-x} = \cdots$$

Alternatively, we may solve the problem much easier using SOCR ...

31

Here is the distribution of the number of unbruised peaches among the 12 selected. After we enter $n = 12$ and $p = 0.65$ we get the distribution below:
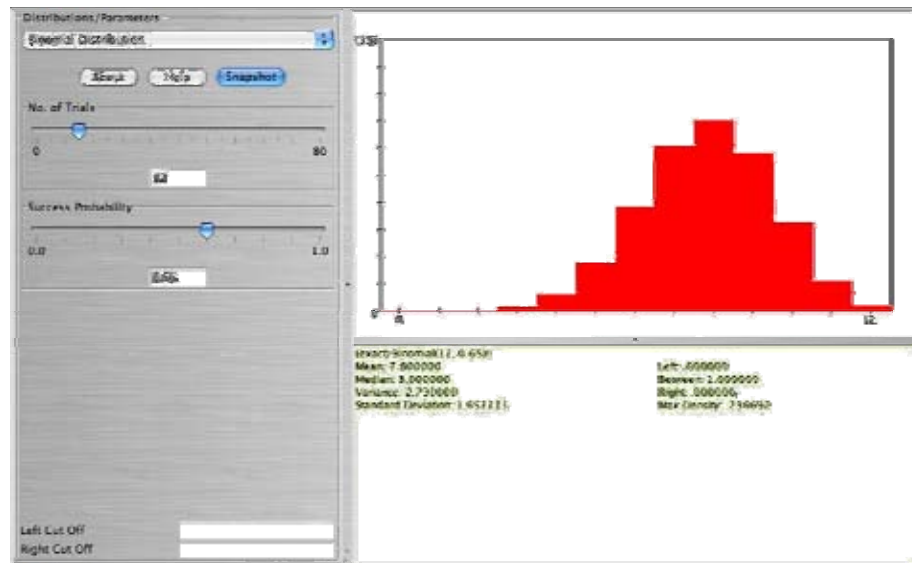
Now, in the **Left Cut Off** and **Right Cut Off** boxes (bottom left corner of the applet) enter the numbers 5 and 8 respectively. What do you observe?
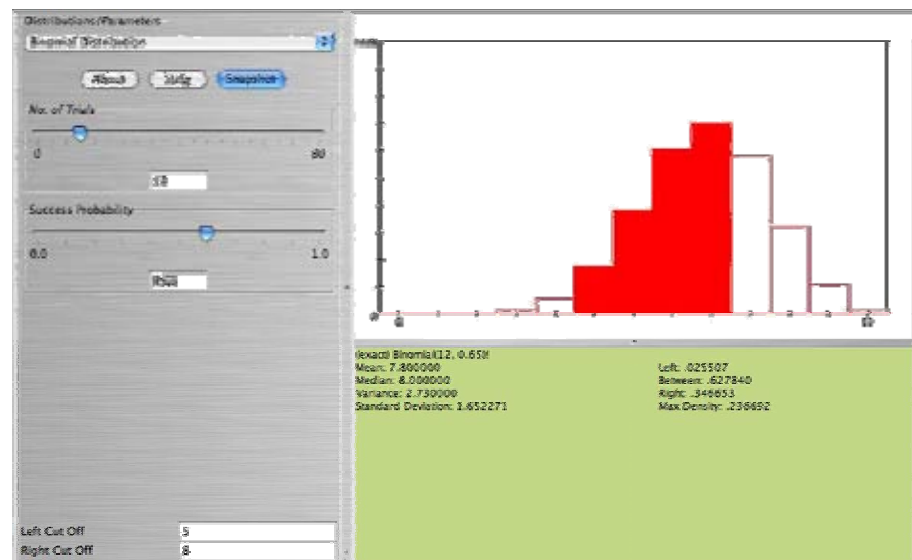
The distribution is divided into three parts. The left part (less than 5), the right part (above 8), and the between part (between 5 and 8 included). All the SOCR distributions applets are designed in the same way. From the applet the probabilities are $P(A) = 0.346653, P(B) = 0.627840, P(C) = 0.025507$.

Example 2: Suppose a lot of size N is accepted if it contains no more than c defective components. A production manager selects at random a sample of n components from this lot and determines the number of defective components. If

32

he finds more than c defective components then the lot is rejected, otherwise it is accepted. Answer the following questions:

a. Suppose the manager wants to choose between two lot sizes: $N = 500$ or $N = 1,000$. Both lots will contain *1%* defective components and he will sample in both cases *n=5%* of the lot. Which sampling scheme will have a higher probability of rejecting the lot if $c = 0$? Use SOCR and print the two snapshots.

b. Repeat (a) with $c = 1$. Answer the question using SOCR.

c. Repeat (a) with $c = 1$ and defective rate *10%*. Use SOCR.

Example 3: The admissions office of a small, selective liberal-arts college will only offer admission to applicants who have a certain mix of accomplishments, including a high SAT score. Based on past records, the head of the admissions feels that the probability is 0.58 that an admitted applicant will come to the college. Based on financial considerations, the college would like a class of size 310 or more. Find the smallest *n*, number of people to admit, for which the probability of getting 310 or more to come to the college is at least 0.95. Use SOCR to find the answer and print the appropriate snapshots.

Example 4: The diameters of apples from the Milo Farm have diameters that follow the normal distribution with mean 3 inches and standard deviation 0.3 inch. Apples can be size-sorted by being made to roll over a mesh screens. First, the apples are rolled over a screen with mesh size 2.5 inches. This separates out all the apples with diameters < 2.5 inches. Second, the remaining apples are rolled over a screen with mash size 3.2 inches. Find the proportion of apples with diameters < 2.5 inches, between 2.5 and 3.2 inches, and greater than 3.2 inches. Use SOCR to find the answers and print the appropriate snapshots.

o **11:30AM - 12:00 PM** Interactive Discussion on Distributions

o **12:00 - 1:00 PM** Lunch Break - UCLA Faculty Center

Afternoon Session (1 PM – 4 PM)

- o **1:00 - 1:15 PM** SOCR/VLPS Experiments/Games - Overview and Class Utilization, Ivo Dinov

- o **1:15 - 2:00 PM** SOCR Experiments Activities I, Juana Sanchez
  ### i. *Bivariate Activity*

  The Bivariate Normal Experiment: The Java applet needed for the following two activities can be found in the SOCR site:
  http://socr.stat.ucla.edu/htmls/SOCR_experiments.html
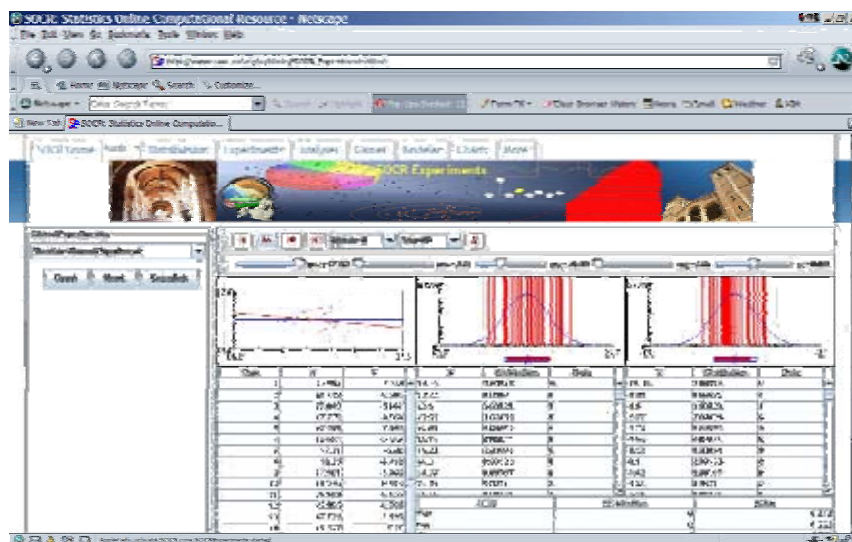
  - **Setting**: This experiment consists of selecting values for the random variables X and Y which are jointly normally distributed as a bivariate normal f(X,Y) with parameters $\mu_x = 0$, $\mu_y = 0$, σ_x="a value of your choice" , σ_y="a value of your choice" , and ρ="a value of your choice". The **first objective** of our activity is to see how the location of the base of the distribution and its spread changes as the parameters change. The **second objective** is to see how no matter what the values of the parameters are, the marginal distribution of X and the marginal distribution of Y are both normal, with more or less spread depending on the values you assign to the parameters. The points you select on the left hand side diagram, which shows the area above which the normal density lies (or area of integration), can be chosen by setting stop=10,000 update=100 and then clicking on the <RUN> button.

    What happens when ρ and the standard deviation of Y, σ_y, are constant, and the standard deviations of X, σ_x, increases?

- **Exercise 1**: Next, set $\rho = 0.6, \sigma_x = 1.3$ and $\sigma_y = 1.1$ and select the 10,000 points as indicated above. Write down the mean and standard deviation of the numbers you generated and the correlation. Save the picture.

- **Exercise 2**: Now, change only the standard deviation of X from 1.3 to 2. Write down the means, standard deviations and correlations of numbers you generated. Save the picture.

- **Exercise 3**: Compare the pictures generated in **Exercise 1** and **2** above. What would you say is the effect on the joint distribution $f(X,Y)$ of increasing the standard deviation of X, other things held constant. Write your comments here.

- **Exercise 4**: Compare the marginal densities for X and for Y in **Exercise 1** and **2**. Which density is more spread out?

- **Exercise 5**: Compare the regression lines in **Exercise 1** and **2**. Make your comparison in terms of the slope.

See what happens when the standard deviations of X and Y are fixed and the correlation increases

- **Exercise 6**: Fix now the standard deviation of X to 1.3 and the standard deviation of Y to 1.1 and the correlation coefficient to 0.9. Print a screenshot of the pictures, means, standard deviations and correlations in your data. Compare your pictures to those in **Exercise 1**.

- **Exercise 7**: According to your results in **Exercise 1**, what has happened to the joint density function of X and Y as the correlation coefficient has increased? What has happened to the regression line? What has happened to the marginal densities of X and of Y?

Write here the joint density function of X and Y with parameter values as in Exercise 1

- **Exercise 8**: Write the formula for the marginal density of X and the marginal density of Y with parameter values as given in Exercise 1. Write the formulas for the conditional densities of X given Y and Y given X, with parameter values as given in Exercise 1. Write then the regression lines that follow from these densities.

ii. *Birthday Experiment*

Experiment Goals: To understand the computation of the probability that at least two people in a group of m people share the same birthday and the probability that nobody in a group of m people share the same birthday.

Setting: To understand the idea of those probabilities try the Birthday Experiment applet at the following website:
http://socr.stat.ucla.edu/htmls/SOCR_Experiments.html. Select the *Birthday Experiment* from the drop-down list on the top-left.
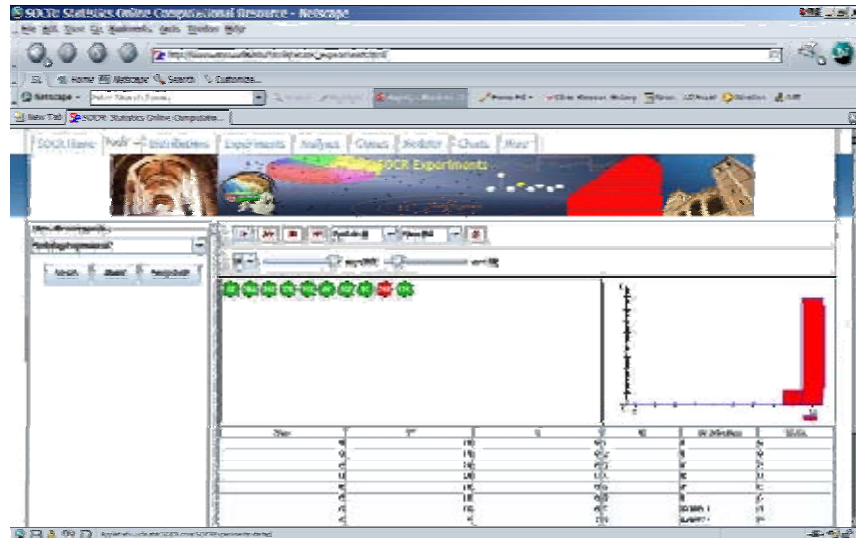
- **Exercise 1**: Draw 10 people at random from the population and observe their birthdays. This is like drawing 10 balls (**n=10**) at random, with replacement, from an urn containing 365 balls (**m=365**). To do that, set **n=1**, **m=365** and set **update=1** and **stop=1**. Click on the <RUN> button or the one-step button <STEP>.
    - How many distinct balls did you get? _____
    - Is that number the same as the value of V?
    - What is the value of V ? _____
    - Were there any two balls that repeated? (i.e., any red ball in your sample set?) _____. *I* should be 1, if there are two balls with the same number, (the number in the green ball). If not, *I* should be 0. What is your I? _____

- **Exercise 2**: Above we did only one run of the experiment. That won't take us too far. We need to repeat the same experiment many times to see how often we get 10 distinct (green) balls and how often we get at least one red ball. Keep **n=10** and **m=365** and set **update=1** and **stop=10**. You will only see that last run of the experiment in the coins drawn, but you will see all 10 runs and the **V**'s and **I**'s and the distribution of the **I**'s.
    - Was there a red ball in the last run? (The last one is the one shown on the bottom of the screen)____
    - In how many runs was the number of distinct balls equal to 10? _____
    - In how many runs was V equal to 10? _____
    - In how many runs was I equal to 0? _____
    - In how many runs was I equal to 1? _____

36

- o What is the distribution of I on the top-right hand side giving you then? _____
- o According to your 10 runs, the probability that no two people in a group of 10 share the same birthday is? _____
- o The probability that at least two people in a group of 10 share the same birthday is, according to your 10 runs _____

- **Exercise 3**: With 10 runs, you have not got too close to the true probability. So, what about trying **10,000** runs. Don't even dream of seeing every single one of these. So, <RESET> the screen, and set the stop at 10,000 and the update at 1,000. This means you will only see the outcome of one in 1,000 runs, even though the computer is generating all 10,000 experiments. But, the final distribution with all 10,000 runs will appear on the right. In blue you'll se the theoretical distribution and in red you'll see the sampling distribution. Are these similar?
  - o Look at the distribution of I on the right. What proportion of times did the run have at least a red ball? _____. This is the probability that at least two people in a group of 10 share the same birthday. So what is that probability?
  - o What proportion of the 10,000 experiments had all 10 balls be distinct? _____This is the probability that nobody in a group of 10 share the same birthday.
  - o What is then that probability?_____
  - o How is the first probability, which you calculated using only 10 experiments above, related to the second one you just did?

- **Exercise 4**: Use what you have learned above to determine empirically the probability that at least 2 people in a group of 5 share the same birthmonth. Determine also the probability that nobody in a group of 5 share the same month. Write your answer here and attach a printout/screenshot of the applet with the final runs.

- **Exercise 5**: Use what you have learned above to determine empirically how large should the group of people observed be, for the probabilities of at least two same birthdays and the probability of nobody same birthday to be 50%-50%. This may take some trial and error. Turn in your final answer written here and attach your applet. Use the sampling distribution (in red) and the theoretical exact distribution (in blue).

o **2:00 - 2:45 PM** SOCR Experiments Activities II, Nicolas Christou
   i. *Confidence Interval Experiment*

This experiment randomly samples from a N(μ=0, σ^2=1) distribution and constructs a confidence interval for the mean (μ). Note how confidence interval widens as the confidence level increases. This is counterbalanced by the sample-size increase. Also, see the effect of the confidence level, $100(1-\alpha)\%$, on the frequency of intervals missing the estimated parameter (μ=0, in this case).

Goal: To provide a simple simulation demonstrating the effects of confidence level and sample-size on the size of the constructed confidence interval, and utilizing this experiment when sampling from almost all symmetric unimodal distributions.

Experiments

Go to the SOCR Experiments and select the Confidence Interval Experiment from the drop-down list of experiments on the top left. The image below shows the initial view of this experiment:
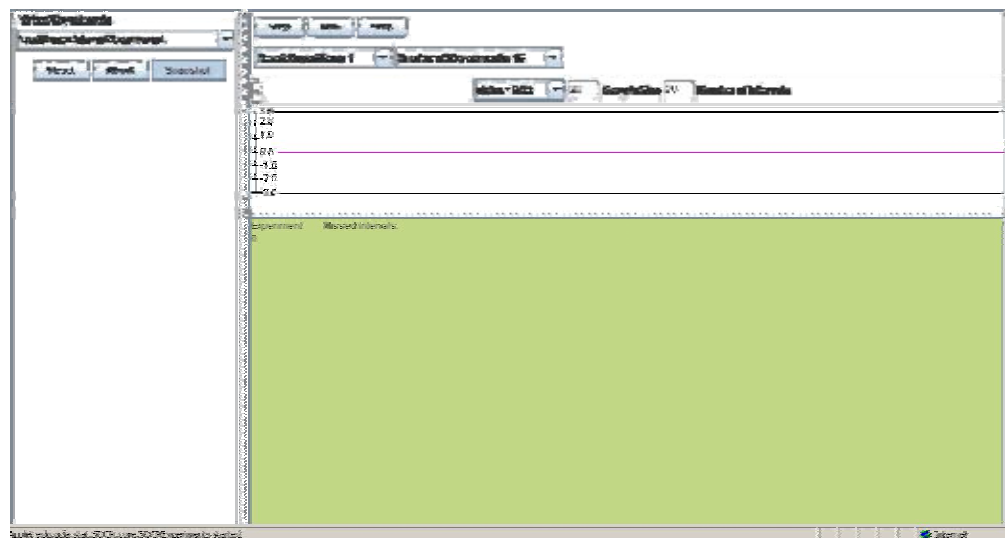
The control buttons on the top allow the user when to begin and end the experiment. "Step" runs one trial, "Run" executes a number of trials depending on the settings of the list boxes below, and "Stop" ends the trials.

The sample size and number of intervals may be modified by the user's discretion. Notice that when alpha has a very small value, the results are also small in value and deviation. The image below shows this information:
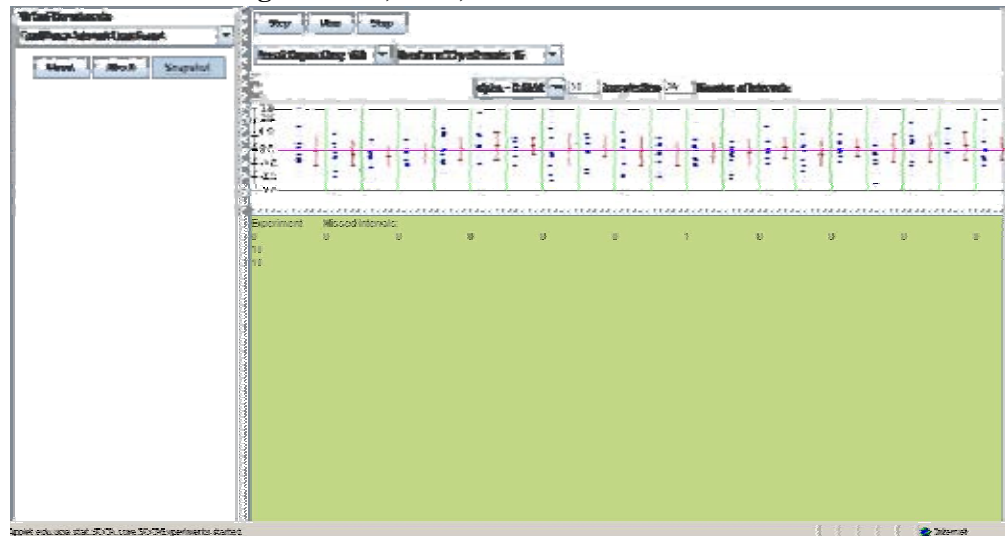
Applications

The Confidence Interval Experiment may be applied to any situation regarding a population parameter associated with a probability that is generated from a random sample.

For example, adjusting a machine that is to fill cans with soda so that the mean content of the cups is approximately 8 ounces. Since it is not possible for every cup to be filled at exactly 8 ounces, setting the random variable X as the weight of the filling can be applied to this applet. The distribution of X is assumed to be of a normal distribution with an expected mean value and standard deviation.

Suppose engineers want to check for the amount of pollution in the waters of Los Angeles and discover the average amount with standard deviation from the past couple of years. By using this applet, they draw generalizations of the average amount of pollution in Los Angeles with a 95% confidence level.

ii. *Roulette Experiment*

Description: The American Roulette wheel has 38 slots numbered 00, 0, and 1-36. Slots 00 and 0 are green. Half of the slots numbered 1-36 are red and half are black. The experiment consists of rolling a ball in a groove in the wheel; the ball eventually falls randomly into one of the 38 slots. The roulette wheel is shown in the left graph panel; the ball is shown on each update.

One of seven different bets can be selected from the list box:

o   Bet on 1: this is an example of a straight bet, and the bet pays 35:1
o   Bet on 1, 2: this is an example of a split bet, and pays 17:1
o   Bet on 1, 2, 3: this is an example of a row bet, and bet pays 11:1
o   Bet on 1, 2, 4, 5: this is an example of 4-number bet, and pays 8:1

39

- o Bet on 1-6: this is an example of a 2-row bet, and pays 5:1
- o Bet on 1-12: this is an example of a 12-number bet, and pays 2:1
- o Bet on 1-18: this is an example of an 18-number bet, and pays 1:1

On each update, the outcome X is shown graphically in the first panel and recorded numerically in the first table. Random variable W gives the net winnings for the chosen bet; this variable is recorded in the first table on each update. The density function and moments of W are shown in blue in the distribution graph and are recorded in the distribution table. On each update, the empirical density and moments of W are shown in red in the distribution graph and are recorded in the distribution table.

Goal: The Roulette Experiments provides a real-life simulation of the American Roulette game. It demonstrates the probabilities and outcomes for every trial under different conditions. By using this applet, users should be able to have a better understanding of the game, Roulette.

Experiment: Go to the SOCR Experiment [SOCR Experiment] and select the Roulette Experiment from the drop-down list of experiments on the top left. The image below shows the initial view of this experiment:
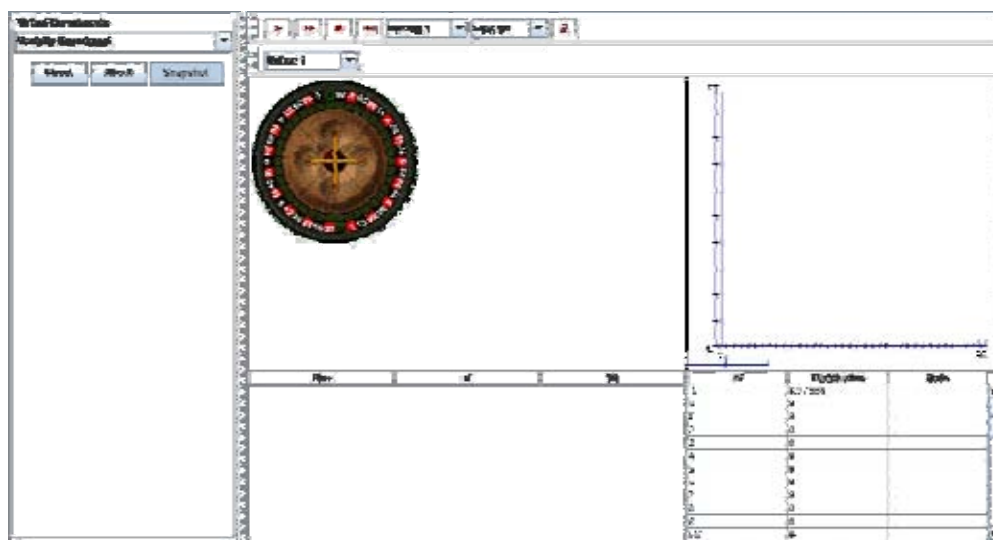
When pressing the play button, one trial will be executed and recorded in the distribution table below. The fast forward button symbolizes the nth number of trials to be executed each time. The stop button ceases any activity and is helpful when the experimenter chooses "continuous," indicating an infinite number of events. The fourth button will reset the entire experiment, deleting all previous information and data collected. The "update" scroll indicates nth number of trials (1, 10, 100, or 1000) performed when selecting the fast forward button and the "stop" scroll indicates the maximum number of trials in the experiment.

40

Bet on 1

The distribution graph illustrates a high probability for values of 1 and a low value for 35. As the number of trials increase, the empirical density graph will converge to the distribution graph as demonstrated below:
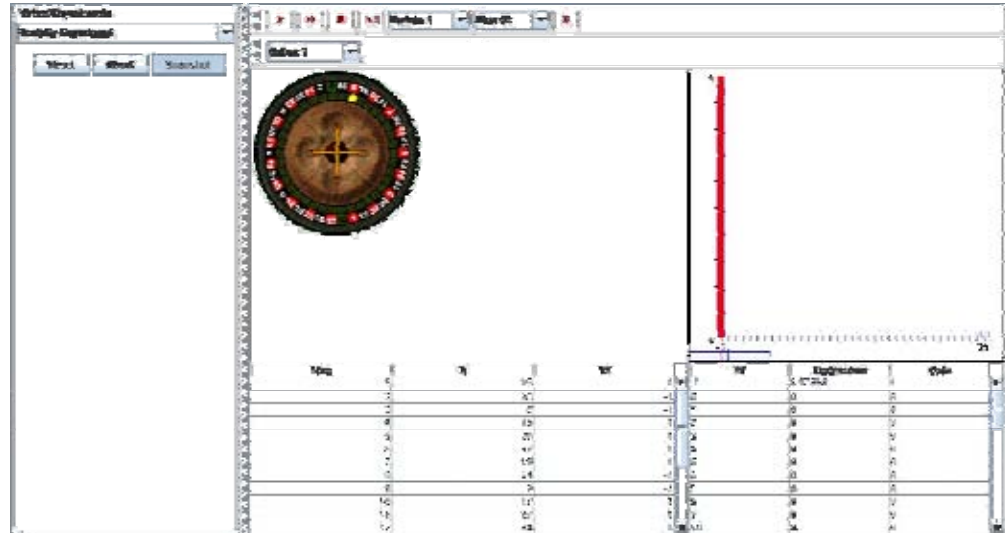
Bet on 1, 2

Similarly, the distribution graph is most saturated on the left side with a very small proportion for the highest value in the x axis. As the number of trials increase, the empirical density graph will converge to the distribution graph as demonstrated below:
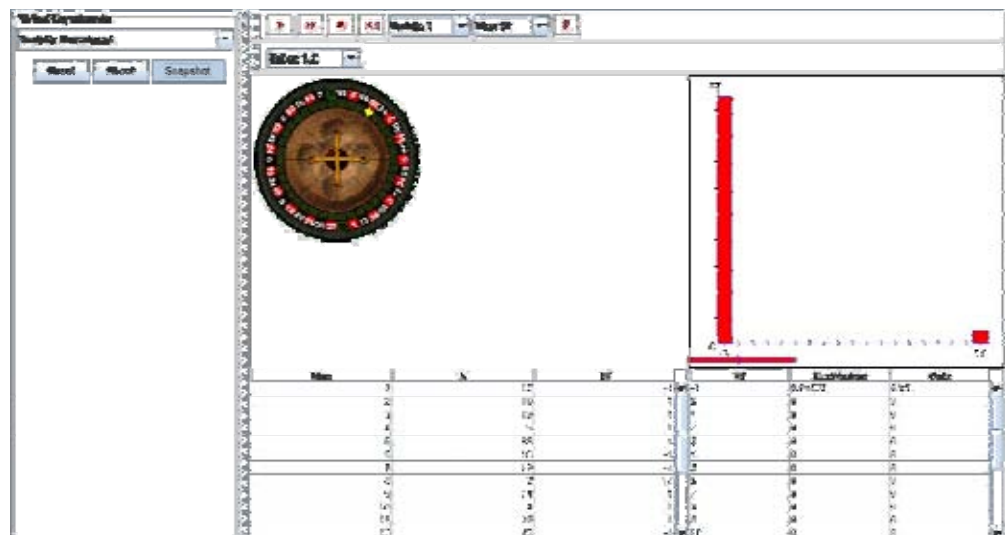
While going down the list box and selecting the specific bet, the first value on the x axis obtains a much higher proportion than the last value on the x axis for every distribution graph. But as the ratio of pay increases, the distribution graph begins

41

to shift until Bet 1-18, in which the first and last values are equal. This is shown in the image below:
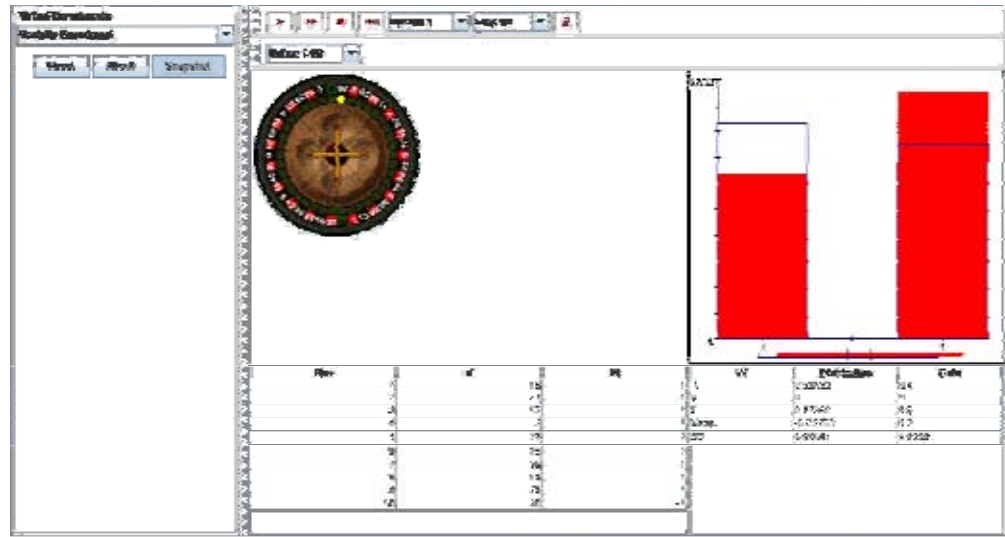
Since the distribution graph represents the chances of winning, users should understand that the lower the proportion value is on the right side of the graph, the lower their chances are of winning. That is why selecting Bet 1-18 has an equal chance of winning and losing with a graph illustrating equal values for -1, 1.

Applications

o The Roulette Experiment is best applicable in gambling situations such as:
o A gambler would like to illustrate the chances of winning in the American Roulette game.
o A player wants to determine which of the Bets will give him the highest chance of winning after ten trials.

o **3:00 - 3:50 PM** Assessing Conceptual Understanding in a First Statistics Course, Rob Gould

Statistical Thinking and Reasoning Test (START)

Developed by the Web ARTIST Project
https://app.gen.umn.edu/artist/

Funded by a grant from the National Science Foundation
NSF CCLI ASA- 0206571
Principal Investigators:
Joan Garfield and Bob delMas, University of Minnesota
Beth Chance, Cal Poly – San Luis Obispo
Post-doctoral Research Assistant:
Ann Ooms, University of Minnesota
August 30, 2006

Tue 08/07/09

Morning Session (9 AM – 12 Noon)

- **9:00 - 9:15 AM** SOCR Analyses - Overview, Ivo Dinov

- **9:15 - 10:15 AM** SOCR Analyses, Annie Chu & Ivo Dinov

i. ANOVA

This SOCR Activity demonstrates the utilization of the SOCR Analyses package for statistical Computing. In particular, it shows how to use Analysis of Variance (ANOVA) and how to interpret the results.

- **ANOVA Background**: Analysis of variance (ANOVA) is a class of statistical analysis models and procedures, which compare means by splitting the overall observed variance into different parts. The initial techniques of the analysis of variance were pioneered by the statistician and geneticist R. A. Fisher in the 1920s and 1930s, and is sometimes known as Fisher's ANOVA or Fisher's analysis of variance, due to the use of Fisher's F-distribution as part of the test of statistical significance. Read more about ANOVA.

- **SOCR ANOVA**: Go to SOCR Analyses and select **One-way ANOVA** from the drop-down list of SOCR analyses, in the left panel. There are three ways to enter data in the SOCR ANOVA applet
    - Click on the **Example** button on the top of the right panel.
    - **Generate random data by clicking on the** Random Example **button**
    - Pasting your own data from a spreadsheet into SOCR ANOVA data table.
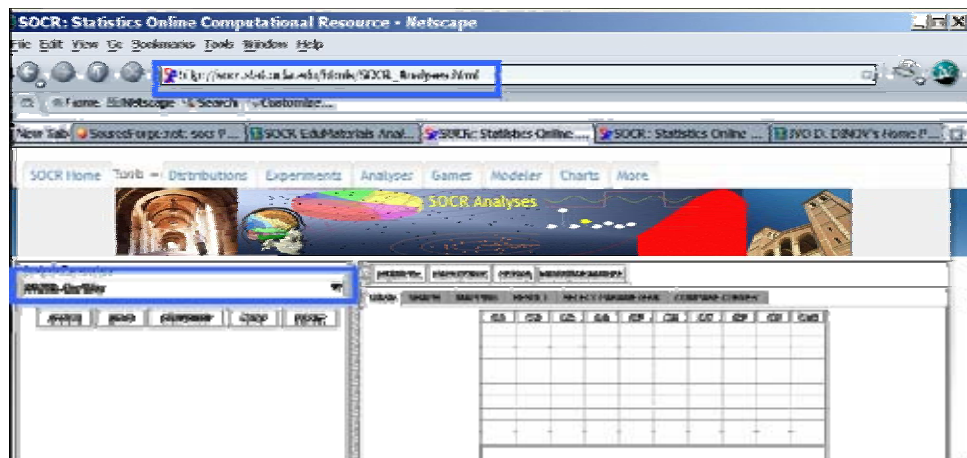
- Now, map the dependent and independent variables, by going to the **Mapping** tab, selecting columns from the available list and sending them to the

50

corresponding bins on the right (see figure). Then press Compute button to carry the ANOVA analysis.
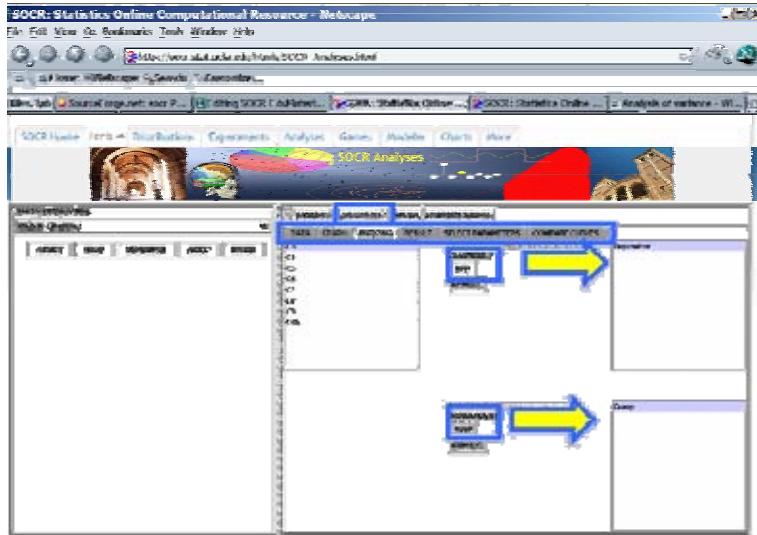
- The quantitative results will be in the tab labeled **Results**. The **Graphs** tab contains the QQ Normal plot for the residuals. In this case, we have a very significant grouping effect, indicated by the p-value $< 10^{-4}$.
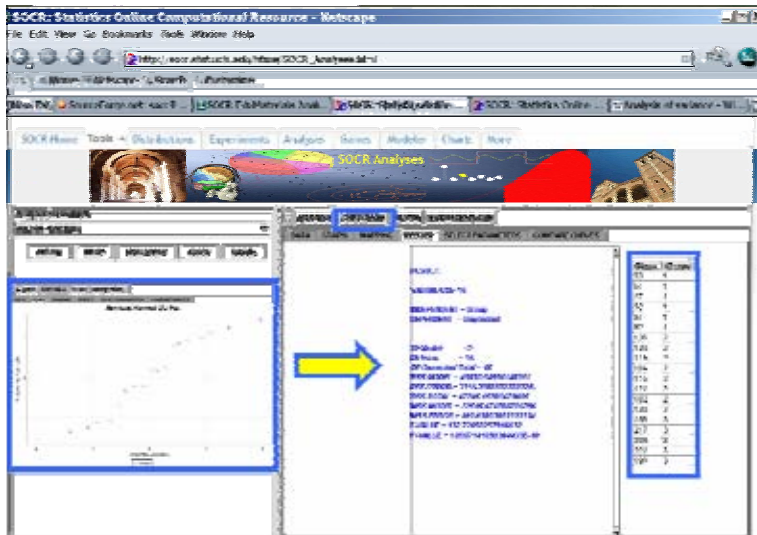
o **10:15 - 10:30 AM** Break

o **10:30 - 11:30 AM** SOCR Modeler, Ivo Dinov

i. *SOCR Normal & Beta Distribution Model Fitting*

Summary: This activity describes the process of **SOCR** model fitting in the case of using Normal or Beta distribution models. *Model fitting* is the process of determining the parameters for an analytical model in such a way that we obtain optimal parameter estimates according to some criterion. There are many strategies for **parameter estimation**. The differences between most of these are the underlying cost-functions and the optimization strategies applied to maximize/minimize the cost-function.

Goals

- motivate the need for (analytical) modeling of natural processes
- illustrate how to use the SOCR Modeler to fit models to real data
- present applications of model fitting

Background & Motivation

Suppose we are given the sequence of numbers {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} and asked to find the best (Continuous) Uniform Distribution that fits that data. In this case, there are two parameters that need to be estimated - the minimum (*m*) and the maximum (*M*) of the data. These parameters determine exactly the support (domain) of the continuous distribution and we can explicitly write the density for the (best fit) continuous uniform distribution as:

$$f(x) = \frac{1}{M - m}, \text{ for } m \leq x \leq M \text{ and } f(x) = 0, \text{ for } x \notin [m : M].$$

Having this model distribution, we can use it's analytical form, *f(x)*, to compute probabilities of events, critical functional values and, in general, do inference on the native process without acquiring additional data. Hence, a good strategy for model fitting is extremely useful in data analysis and statistical inference. Of course, any inference based on models is only going to be as good as the data and the optimization strategy used to generate the model.

Let's look at another motivational example. This time, suppose we have recorded the following (sample) measurements from some process {1.2, 1.7, 3.4, 1.5, 1.1, 1.7, 3.5, 2.5}. Taking bin-size of 1, we can easily calculate the frequency histogram for this sample, {6, 1, 2}, as there are 6 observations in the interval [1:2), 1 measurement in the interval [2:3) and 2 measurements in the interval [3:4).

52
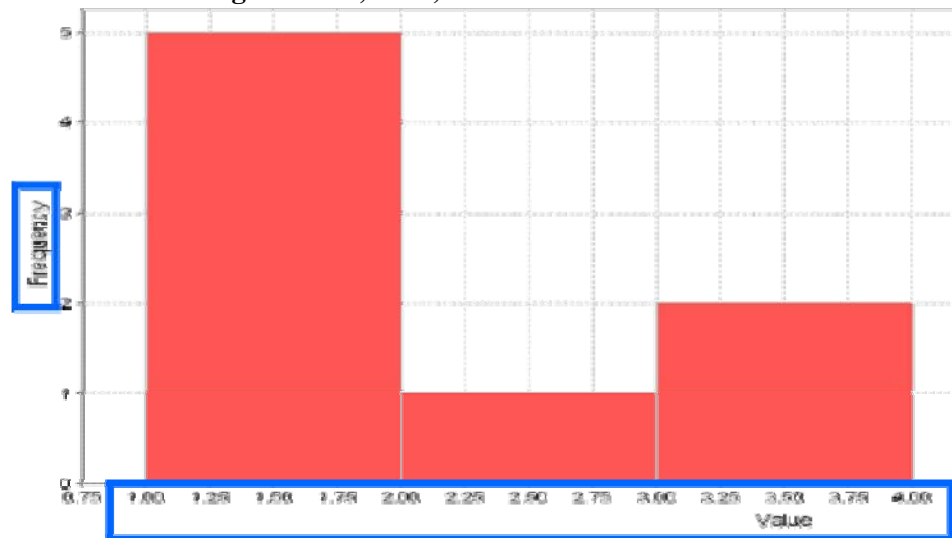
We can now ask about the *best Beta distribution model fit to the histogram of the data*!

Most of the time when we study natural processes using probability distributions, it makes sense to fit distribution models to the frequency histogram of a sample, not the actual sample. This is because our general goals are to model the behavior of the native process, understand its distribution and quantify likelihoods of various events of interest (e.g., in terms of the example above, we may be interested in the probability of observing an outcome in the interval [1.50:2.15) or the chance that an observation exceeds 2.8).

Exercise 1

Let's first solve the challenge we presented in the background section, where we calculated the frequency histogram for a sample to be {6, 1, 2}. Go to the SOCR Modeler and click on the **Data** tab. Paste in the two columns of data. Column 1 {1, 2, 3} - these are the ranges of the sample values and correspond to measurements in the intervals [1:2), [2:3) and [3:4). The second column represents the actual frequency counts of measurements within each of these 3 histogram bins - these are the values {6, 1, 2}. Now press the **Graphs** tab. You should see an image like the one below. Then choose **Beta_Fit_Modeler** from the drop-down list of models in the top-left and click the estimate parameters check-box, also on the top-left. The graph now shows you the best Beta distribution model fit to the frequency histogram {6, 1, 2}. Click the **Results** tab to see the actual estimates of the two parameters of the corresponding Beta distribution (*Left Parameter = 0.0446428571428572; Right Parameter = 0.11607142857142871; Left Limit = 1.0; Right Limit = 3.0*).
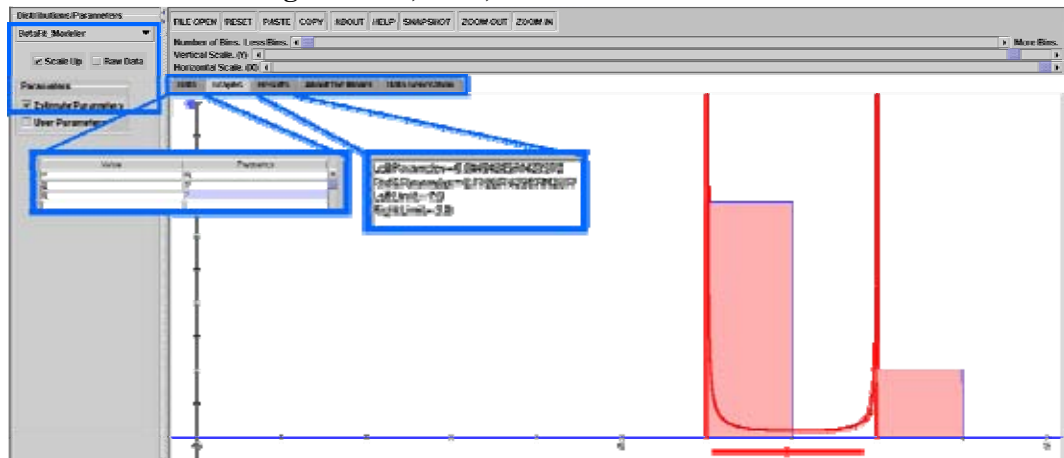
You can also see how the (general) Beta distribution degenerates to this shape by going to SOCR Distributions, selecting the **(Generalized) Beta Distribution** from the top-left and setting the 4 parameters to the 4 values we computed above. Notice how the shape of the Beta distribution changes with each change of the parameters. This is also a good demonstration of why we did the distribution model fitting to the frequency histogram in the first place - precisely to obtain an analytic model for studying the general process without acquiring mode data. Notice how we can compute the odds (probability) of any event of interest, once we have an analytical model for the distribution of the process. For example, this figure depicts the probabilities that a random observation from this process exceeds 2.8 (the right limit). This probability is computed to be 0.756
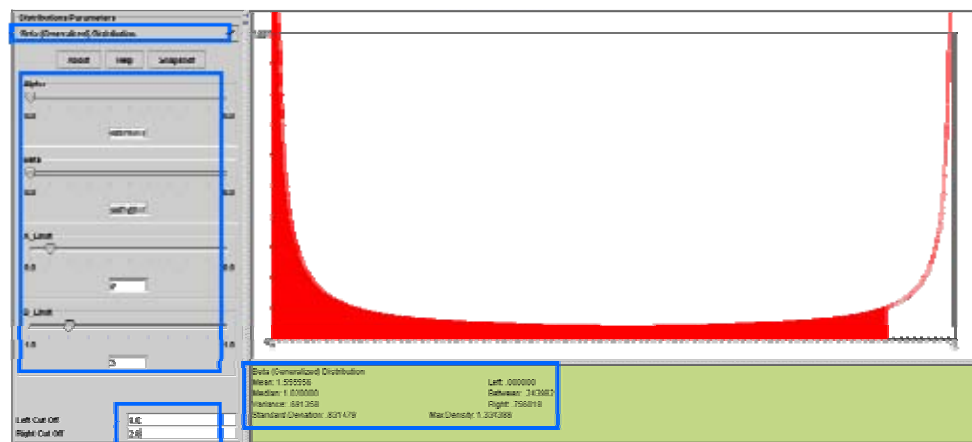
Exercise 2

Go to the SOCR Modeler and select the **Graphs** tab and click the "Scale Up" check-box. Then select **Normal_Model_Fit** from the drop-down list of models and begin clicking on the graph panel. The latter allows you to construct manually a histogram of interest. Notice that these are not random measurements, but rather frequency counts that you are manually constructing the histogram of. Try to make the histogram bins form a unimodal, bell-shaped and symmetric graph.

54

Observe that as you click, new histogram bins will appear and the model fit will update. Now click the Estimate Parameters check-box on the top-left and see the best-fit Normal curve appear superimposed on the manually constructed histogram. Under the **Results** tab you can find the maximum likelihood estimates for the mean and the standard deviation for the best Normal distribution fit to this specific frequency histogram.
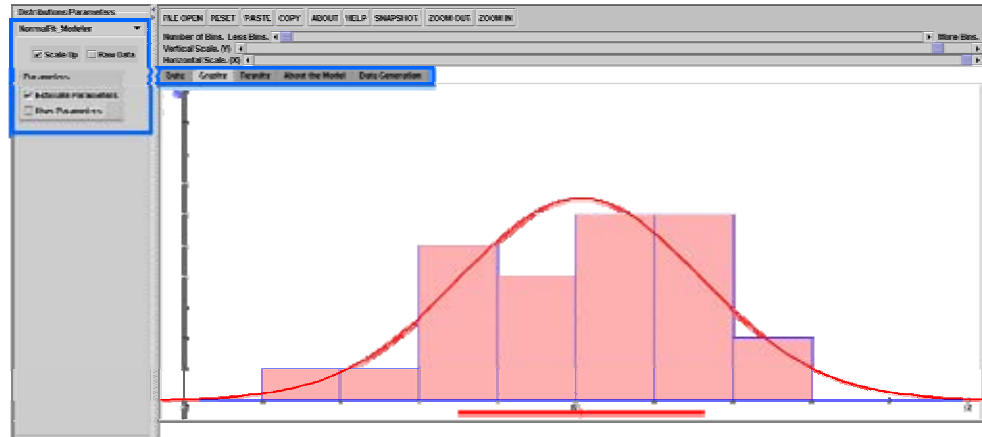
Applications

- On the SOCR Wiki, you can see more instances of using the SOCR Modeler to fit distribution models to real data.
- SOCR Modeler allows one to fit in distribution, polynomial or spectral models to real data - more information about these is available at the SOCR Modeler Activities.

o **11:30AM - 12:00 PM** Interactive Discussion on Analyses & Modeler

o **12:00 - 1:00 PM** Lunch Break UCLA Faculty Center

55

Afternoon Session (1 PM – 4 PM)

o **1:00 - 1:50 PM** SOCR Experiments/Games Activities III, Ivo Dinov

i. *General Central Limit Theorem (CLT)*

This activity represents a very general demonstration of the the Central Limit Theorem (CLT). The activity is based on the SOCR Sampling Distribution CLT Experiment. This experiment builds upon a RVLS CLT applet by extending the applet functionality and providing the capability of sampling from any SOCR Distribution.

Goals

- Provide intuitive notion of sampling from any process with a well-defined distribution.
- Motivate and facilitate learning of the central limit theorem.
- Empirically validate that sample-averages of random observations (most processes) follow approximately normal distribution.
- Empirically demonstrate that the *sample-average* is special and other sample statistics (e.g., median, variance, range, etc.) generally do not have distributions that are normal.
- Illustrate that the expectation of the sample-average equals the population mean (and the sample-average is typically a good measure of centrality for a population/process).
- Show that the variation of the sampling distribution of the mean rapidly decreases as the sample size increases ($\frac{1}{\sqrt{n}}$).
- Reinforce the concepts of a native distribution, sample, sample distribution, sampling distribution, parameter estimator and data-driven numerical parameter estimate.

The SOCR CLT Experiment

To start the this Experiment, go to SOCR Experiments and select the SOCR Sampling Distribution CLT Experiment from the drop-down list of experiments in the left panel. The image below shows the interface to this experiment. Notice the main control widgets on this image (boxed in blue and pointed to by arrows). The generic control buttons on the top allow you to do one or multiple steps/runs, stop and reset this experiment. The two tabs in the main frame provide graphical access to the results of the experiment (Histograms and Summaries) or the Distribution selection panel (Distributions). Remember that choosing sample-sizes <= 16 will animate the samples (second graphing row), whereas larger sample-sizes (N>20) will only show the updates of the sampling distributions (bottom two graphing rows).

56
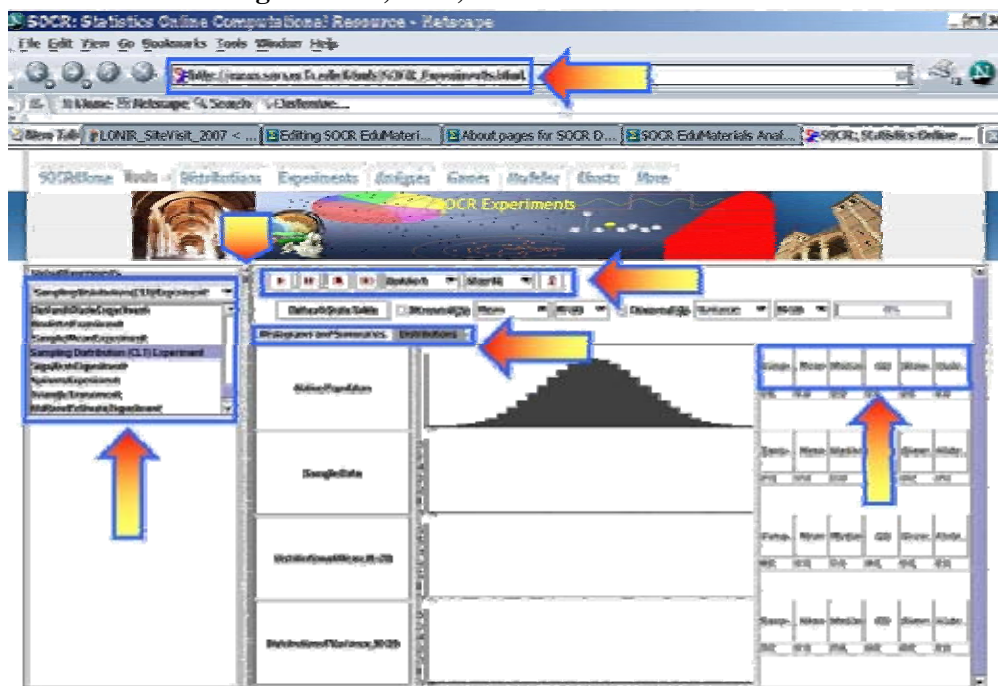
Experiment 1

Expand your Experiment panel (right panel) by clicking/dragging the vertical split-pane bar. Choose the two sample sizes for the two statistics to be 10. Press the **step**-button a few of times (2-5) to see the experiment run several times. Notice how data is being sampled from the native population (the distribution of the process on the top). For each step, the process of sampling 2 samples of 10 observations will generate 2 sample statistics of the 2 parameters of interest (these are defaulted to *mean* and *variance*). At each step, you can see the plots of all sample values, as well as the computed sample statistics for each parameter. The sample values are shown on the second row graph, below the distribution of the process, and the two sample statistics are plotted on the bottom two rows. If we run this experiment many times, the bottom two graphs/histograms become good approximations to the corresponding sampling distributions. If we did this infinitely many times these two graphs become the sampling distributions of the chosen sample statistics (as the observations/measurements are independent within each sample and between samples). Finally, press the **Refresh Stats Table** button on the top to see the sample summary statistics for the native population distribution (row 1), last sample (row 2) and the two sampling distributions, in this case *mean* and *variance* (rows 3 and 4).
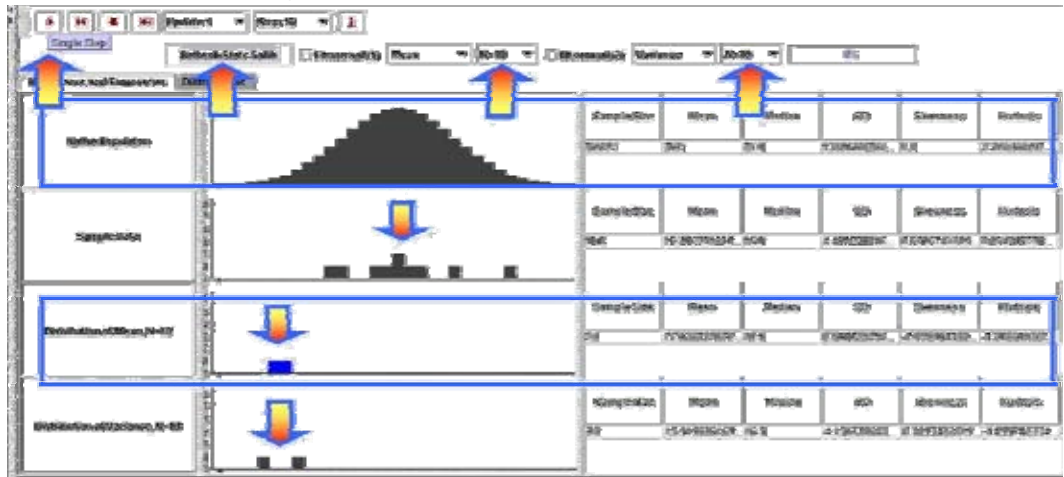
Experiment 2

For this experiment, we'll look at the mean, standard deviation, skewness and kurtosis of the sample-average and the sample-variance (these two sample data-driven statistical estimates). Choose sample-sizes of 50, for both estimates (mean and variance). Select the **Fit Normal Curve** check-boxes for both sample distributions. **Step** through the experiment a few times (by clicking the Run button) and then click **Refresh Stats Table** button on the top to see the sample summary statistics. Try to understand and relate these sample-distribution statistics to their analogues from the native population (on the top row). For example, the mean of the multiple sample-averages is about the same as the mean of the native population, but the standard deviation of the sampling distribution of the average is about $\dfrac{\sigma}{\sqrt{n}}$, where σ is the standard deviation of the original native process/distribution.
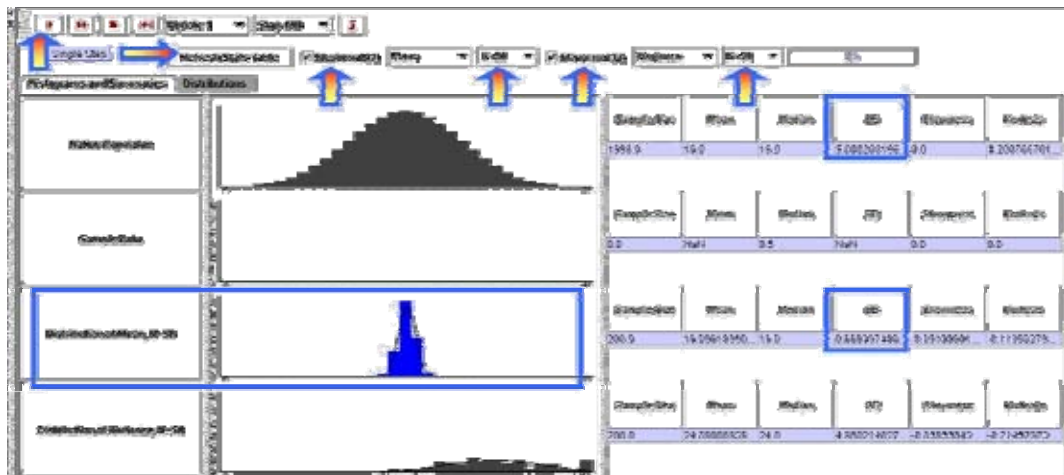
Experiment 3

Now let's select any of the SOCR Distributions, sample from it repeatedly and see if the central limit theorem is valid for the process we have selected. Try Normal, Poisson, Beta, Gamma, Cauchy and other continuous or discrete distributions. Are our empirical results in agreement with the CLT? Go to the **Distributions** tab on the top of the graphing panel. Reset the experiments panel (button on the top). Select a distribution from the drop-down list of distributions in this list. Choose appropriate parameters for your distribution, if any, and click the **Sample from this Current Distribution** button to send this distribution to the graphing panel in the **Histograms and Summaries** tab. Go to this panel and again run the experiment several times. Notice how we now sample from a Non-Normal Distribution for the first time. In this case we had chosen the Beta distribution ($\alpha = 6.7, \beta = 0.5$).

Experiment 4

Suppose the distribution we want to sample from is not included in the list of SOCR Distributions, under the **Distributions** tab. We can then draw a shape for a hypothetical distribution by clicking and dragging the mouse in the top graphing canvas (Histograms and Summaries tab panel). This away you can construct contiguous and discontinuous, symmetric and asymmetric, unimodal and multi-modal, leptokurtic and mesokurtic and other types of distributions. In the figure below, we had demonstrated this functionality to study differences between two data-driven estimates for the population center - sample mean and sample median. Look how the sampling distribution of the sample-average is very close to Normal, where as the sampling distribution of the sample median is not.

59

Questions

- What effects will asymmetry, gaps and continuity of the native distribution have on the applicability of the CLT, or on the asymptotic distribution of various sample statistics?
- When can we reasonably expect statistics, other than the sample mean, to have CLT properties?
- If a native process has $\sigma_X = 10$ and we take a sample of size 10, what will be $\sigma_{\overline{X}}$? Does it depend on the shape of the original process? How large should the sample-size be so that $\sigma_{\overline{X}} = \frac{2}{3}\sigma_X$?

Applications

The second part of this SOCR Activity demonstrates the applications of the Central Limit Theorem.

ii. Law of Large Numbers

This is part I of a heterogeneous activity that demonstrates the theory and applications of the Law of Large Numbers (LLN). **Part II of this activity** contains more examples and diverse experiments.

Example

The average weight of 10 students from a class of 100 students is most likely closer to the *real average* weight of all 100 students, compared to the average weight of 3 randomly chosen students from that same class. This is because the sample of 10 is a *larger number* than the sample of only 3 and better represents the entire class. At the extreme, a sample of 99 of the 100 students will produce a sample average almost exactly the same as the average for all 100 students. On the other extreme, sampling a single student will be an extremely variant estimate of the overall class average weight.

Statement of the Law of Large Numbers

If an event of probability p is observed repeatedly during **independent repetitions**, the ratio of the observed frequency of that event to the total number of repetitions converges towards p as the number of repetitions becomes arbitrarily large.

Complete details about the LLN can be found on the SOCR Wiki

Exercise 1

This exercise illustrates the statement and validity of the LLN in the situation of tossing (biased or fair) coins repeatedly. Suppose we let H and T denote Heads and Tails, the probabilities of observing a Head or a Tail at each trial are $0 < p < 1$ and $0 < 1 - p < 1$, respectfully. The sample space of this experiment consist of sequences of H's and Ts. For example, an outcome may be $\{H,H,T,H,H,T,T,T,....\}$. If we toss a coin n times, the size of the sample-space is 2^n, as the coin tosses are independent. [Binomial Distribution](#) governs the probability of observing $0 \leq k \leq n$ Heads in *n* experiments, which is evaluated by the binomial density at *k*.

In this case we will be interested in two random variables associated with this process. The first variable will be the *proportion of Heads* and the second will be the *differences of the number of Heads and Tails*. This will empirically demonstrate the LLN and it's most common misconceptions (presented below). Point your browser to the [SOCR Experiments](#) and select the **Coin Toss LLN Experiment** from the drop-down list of experiments in the top-left panel. This applet consists of a control toolbar on the top followed by a graph panel in the middle and a results table at the bottom. Use the toolbar to flip coins one at a time, 10, 100, 1,000 at a time or continuously! The toolbar also allows you to stop or reset an experiment and select the probability of Heads (**p**) using the slider. The graph panel in the middle will dynamically plot the values of the two variables of interest (*proportion of heads* and *difference of Heads and Tails*). The outcome table at the bottom presents the summaries of all trials of this experiment. From this table, you can copy and paste the summary for further processing using other computational resources (e.g., [SOCR Modeler](#) or [MS Excel](#)).

- **Note:** We report the normalized differences of the number of Heads minus the number of Tails in the graph and result table. Let H and T are the number of Heads and Tails, up to the current trial (k), respectively. Then we define the normalized difference $|H - T| = p + ((1-p)H - pT)/(2/3 \times Max_k)$, where $Max_k = \max_{i \leq i \leq k} \|H - T\|_i$ and $||H - T||_i$ is the maximum difference of Heads and Tails up to the ith trial. Observe that the expectation of the normalized difference $E(|H - T|) = p$, since $E((1-p)H - pT) = 0$. This ensures that the normalized differences oscillate around the chosen *p* (the LLN limit of the proportion of Heads) and they are visible within the graph window.
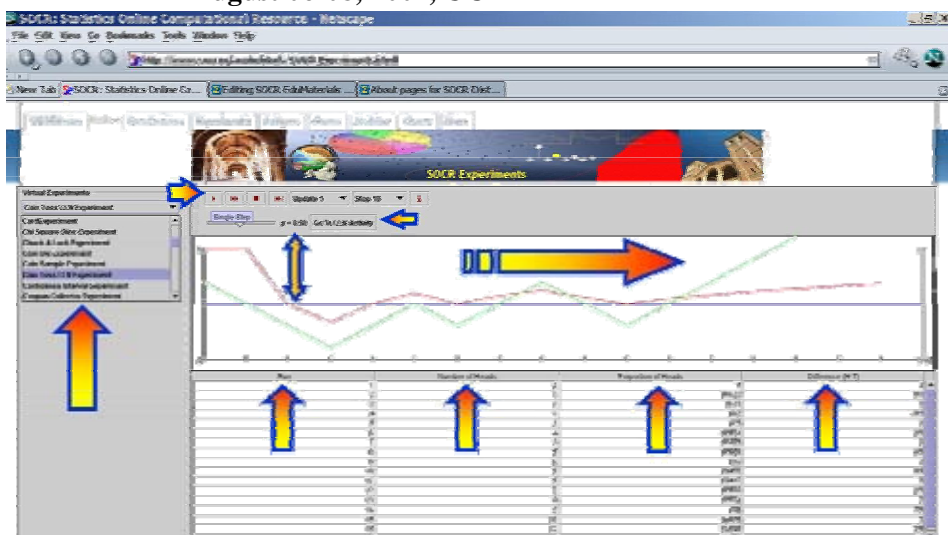
Now, select **n=100** and **p=0.5**. The figure below shows a snapshot of the applet. Remember that each time you run the applet the random samples will be different and the figures and results will generally vary. Click on the **Run** or **Step** buttons to perform the experiment and observe the *proportion of heads* and *differences* evolve over time. Choosing **Continuous** from the number of experiments drop-down list in the tool bar will run the experiment in a continuous mode (use the **Stop** button to terminate the experiment in this case). The statement of the LLN in this experiment is simply that **as the number of experiments increases the sample proportion of Heads (red curve) will approach the theoretical (user preset) value of p (in this case *p=0.5*)**. Try to change the value of **p** and run the experiment interactively several times. Notice the behavior of the graphs of the two variables we study. Try to pose and answer questions like these:

- If we set **p=0.4**, how large of a sample-size is needed to ensure that the sample-proportion stays within [0.4; 0.6]?
- What is the behavior of the curve representing the differences of Heads and Tails (red curve)?
- Is the convergence of the sample-proportion to the theoretical proportion (that we preset) dependent on p?
- Remember that the more experiments you run the closer the theoretical and sample proportions will be (by LLN). Go in **Continuous run mode** and watch the convergence of the sample proportion to *p*. Can you explain in words, why can't we expect the second variable of interest (the differences of Heads and Tails) to converge?
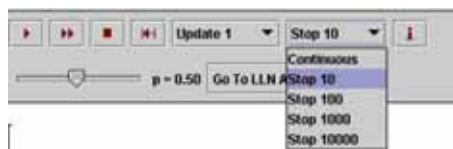
Common Misconceptions regarding the LLN

- **Misconception 1**: If we observe a streak of 10 consecutive heads (when p=0.5, say) the odds of the 11th trial being a Head is > p! This is of course, incorrect, as the coin tosses are independent trials (an example of a *memoryless* process).
- **Misconception 2**: If run large number of coin tosses, the **number of heads** and **number of tails** become more and more equal. This is incorrect, as the LLN only guarantees that the sample proportion of heads will converge to the true population proportion (the p parameter that we selected). In fact, the difference |Heads - Tails| diverges!

o **1:50 - 2:00 PM** Break

o **2:00 - 2:50 PM** CAUSE Resources for Teaching Statistics, Brian Jersky,
http://www.causeweb.org/resources/

www.SOCR.ucla.edu
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

o **2:50 - 3:00 PM** Break

o **3:00 - 3:50 PM** Using technology to enhance the role of writing and assessment in statistics education, Mahtash Esfandiari, Hai Nguyen, & Chris Barr

Formative Evaluations as a Means of Improving Learning and Teaching

• A continuous and ongoing process
• Pace student learning
• Enhance scaffolding
• Help students monitor their own progress

Introductory Statistics at UCLA: 1997 – 2004

• Three hours of lecture per week (with the Professor)
• One hour of section per week (with the T.A.)
• Homework, midterm and exam schedules are typical

Moodle
• Open source course management system
• Piloted by the statistics department since 2004
• Adopted campus-wide in 2007
• Chosen from among 30 possible candidates
• Key features include
  – Password protection
  – Online discussion forums
  – A central venue for posting teaching materials, data, etc.
  – Timed quizzes which are automatically graded
  – A flexible question bank in which quiz materials can be developed, refined and stored

An example of a Course Homepage

An Online Quiz

Real Time Quiz Results

www.SOCR.ucla.edu
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007

Q#	Question text	Answer's text	partial credit	R. Counts	R.%
(2217)	Q4A1 addition rule for disjoint events : In a company with 10,000 employees, there are 3000 engineers, 2000 scientists, 2000 technicians, and the rest are staff. If you pick a person at random, what is the probability that s(he) will be an engineer or staff?	0.30	(0.00)	2/141	(1%)
		0.60	(1.00)	129/141	(91%)
		0.50	(0.00)	3/141	(2%)
		0.06	(0.00)	3/141	(2%)
(2218)	Q4A1 addition rule for overlapping events : In a certain country, 30% of the people have blonde hair, 20% of the people have blue eyes, and 5% have both. If you pick a person at random, what is the probability that they will have blonde hair or blue eyes?	0.50	(0.00)	24/141	(17%)
		0.45	(1.00)	80/141	(57%)
		0.40	(0.00)	29/141	(21%)
		0.60	(0.00)	4/141	(3%)

Question Bank

Introductory Statistics at UCLA: 2005 - present

- Two hours of lecture per week (with the Professor)
- Two hours of section per week (with the T.A.)
- Homework, midterm and exams are typical
- Five or six quizzes are proctored via Moodle

- Students are allowed two attempts, with one lecture and one section meeting in between the attempts

Sat	Sun	Mon	Tues	Wed	Thurs	Fri
		Lecture	Section		Section	Lecture
First Attempt of the Quiz			Second Attempt of the Quiz			

Evaluating the New Approach

- Multiple quiz attempts and group discussion help students link new concepts to old, and help students generate their own knowledge
- Moodle's resources allow the instructor and teaching assistants' to operate more effectively as facilitators
- Immediate feedback allows students to identify the concepts which they find most challenging, review them and enhance scaffolding

Data Collection

- Control group (Without Moodle): Spring 2004
- Experimental groups (With Moodle): Fall 2005, Summer 2006, Fall 2006
- Class demographics and time of day were similar
- All four groups had the same instructor (Esfandiari), textbook, lecture notes, grading and assignment structure
- The control group received mock quizzes
- Teaching assistants varied across classes
  (Results gathered the day of the final)

Multiple Choice Questions

Difference (between experimental and control group) in percent of students answering question correctly.

For complete results, visit http://www.stat.ucla.edu/~chrisbarr/research1/EBS.pdf.
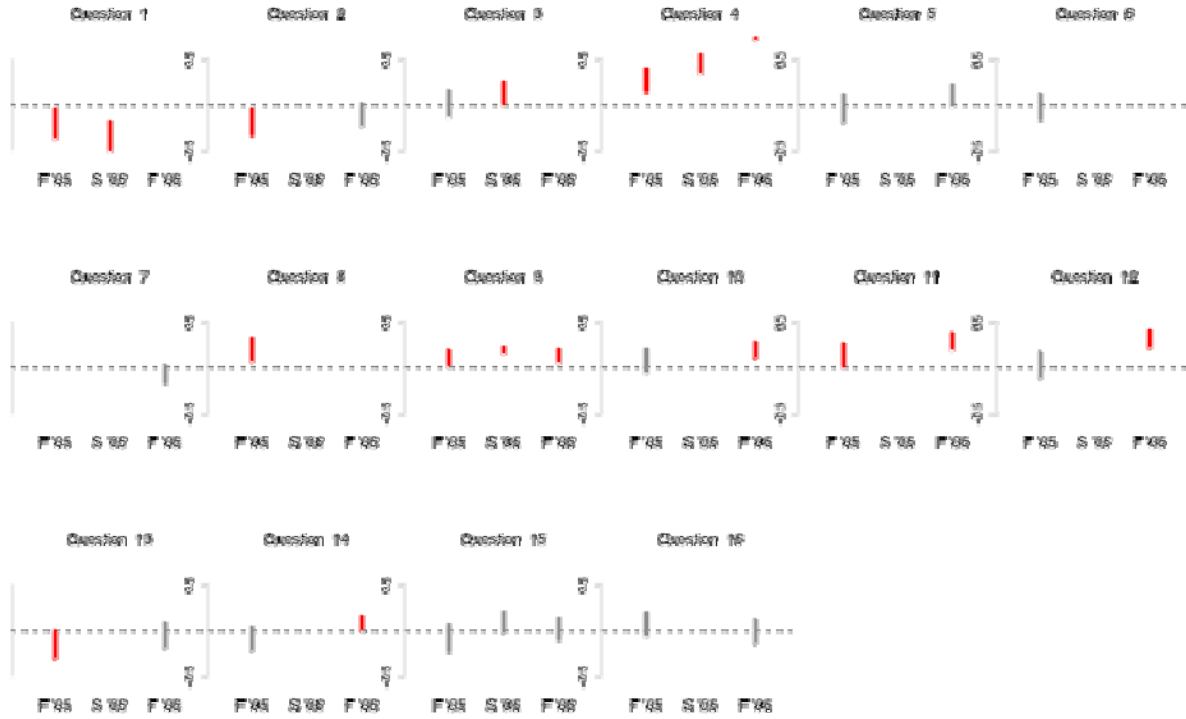
Free Response Questions
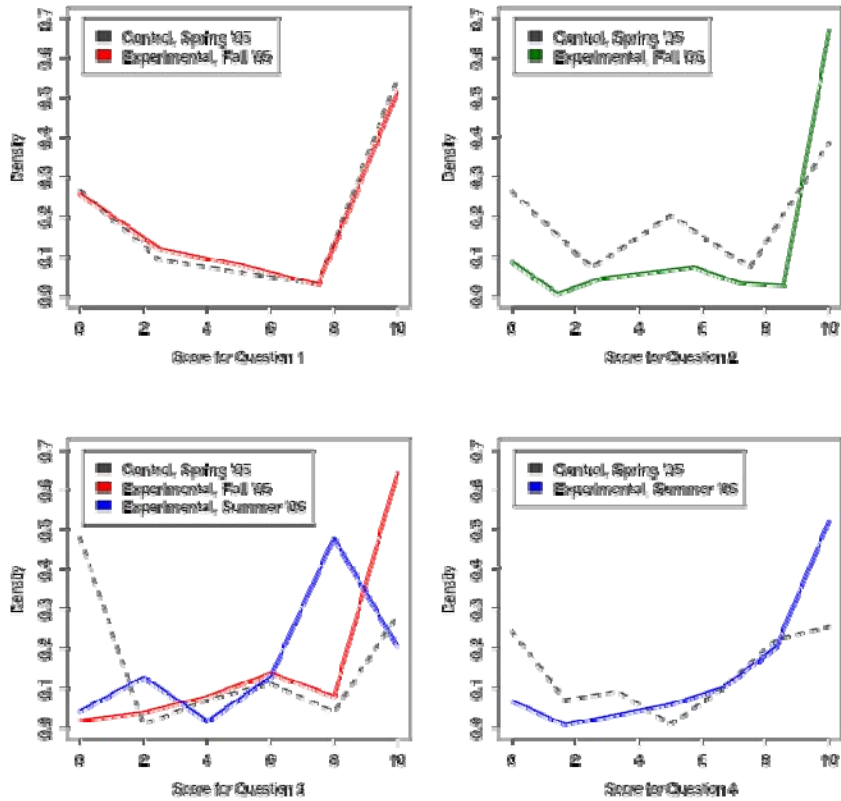
Summary of Survey Results
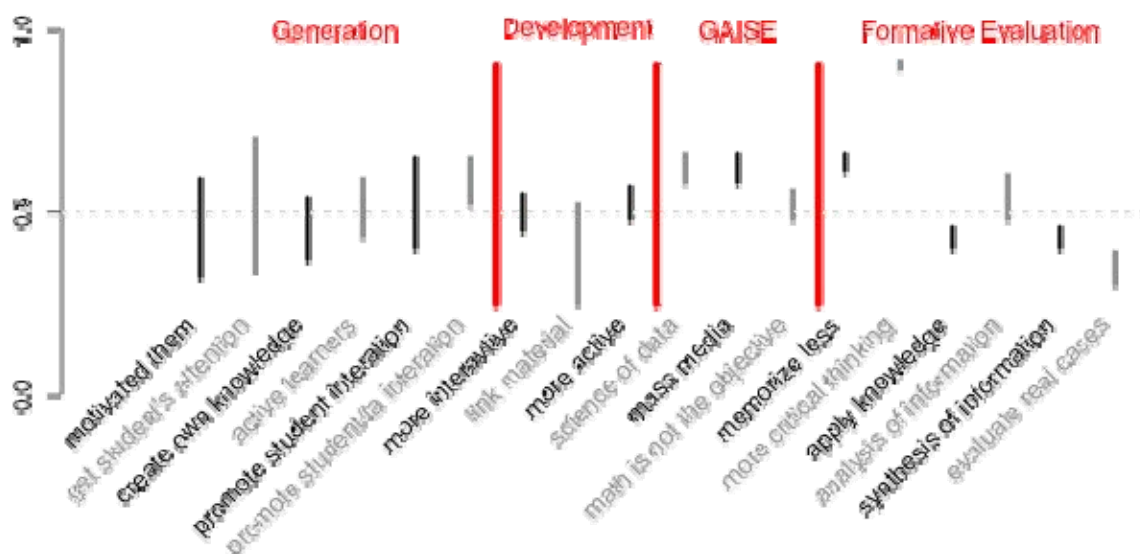
Moodle Application at the UCLA Department of Statistics

- Funding of "blended instruction case study" by the College of Letters and Science and Office of Instructional Development in 2004
- Research on possible course management systems and choosing the Moodle
- Why Moodle was chosen (free as opposed to Blackboard, customizable, based on testing and measurement theory)
- Moodle was used to start developing a test bank of multiple-choice questions and on line testing of students in Winter 2005 and Spring 2006 to pilot the BICS case study or restructured Statistics 10 on small groups of students,
- Fall 2006 Moodle was implemented in teaching large groups of Statistics 10
- Winter 2007 Moodle was implemented as a course management system for upper division and graduate classes (a total of 22 courses)

Using Moodle to develop the test bank: Part 1

- One of the major features of the restructured Statistics 10 is an automated test bank. We have taken the following steps toward developing this bank since Winter 2005.
- Used Moodle to develop a series of multiple-choice questions based on the major concepts and strategies covered in the restructured Statistics 10 so that by now we have close to 1000 multiple-choice questions most of which are written at the higher thinking level including application, analysis, synthesis, and evaluation.
- Developing a system that was effective for categorization and identification of the questions by content.

Using Moodle to develop the test bank: Part 2

- Checking the questions in the test bank for statistical soundness and accuracy,
- Editing the questions in the test bank,
- Classifying the questions in the test bank to a common bank to which all of the instructors can have access as well as banks that are specific to each instructor,
- Using Moodle to includes plots and graphics in the quizzes,
- Calculating the item difficulty (the percentage of students that got the correct answer) for each question since Winter 2005.

Item Difficulty

Distribution of Item Difficulty
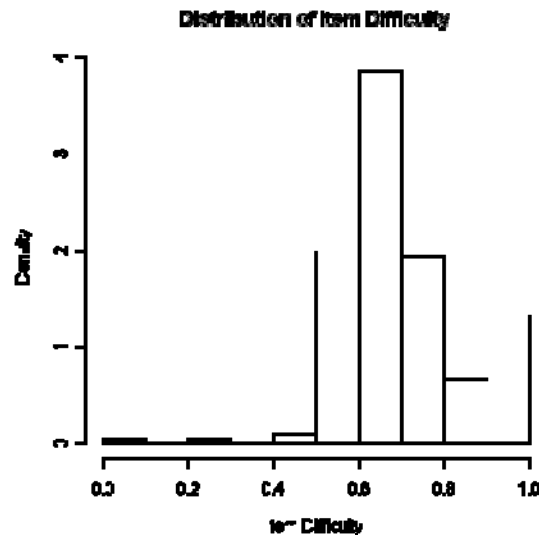
Quiz Grades Distribution
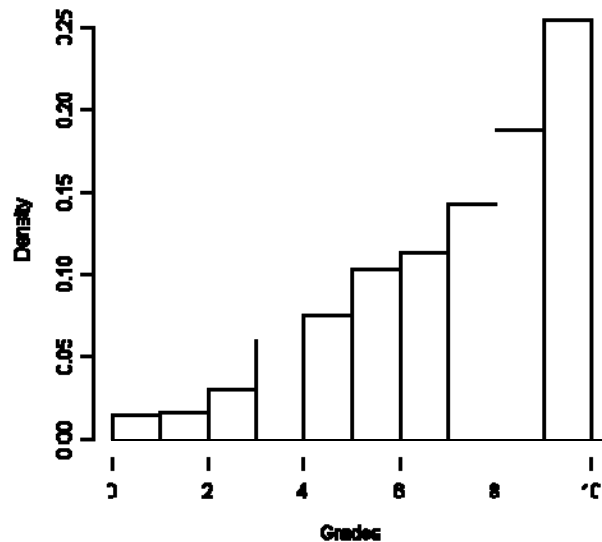
Sample Essay Prompt

1	You are working with a researcher who has collected the following data on 500 people
Marks: 1	• gender (male and female) • political affiliation (republican, democrat, and independent) • annuanl income in dollars • age in years He wants you to plot two graphs that would display: 1) the relationship between gender and political affiliation, and 2) the relationship between political affiliation and income. Q1) What plot would you draw for case 1 and why? Q2) what plot would you draw for case 2 and whY?

Sample Student Responses

- "I would draw three pie charts, one each for republican, democrat, and independent, with each chart having the percentage of males and females that are of that political affiliation, since it's easy to see how many males and females have a certain political affiliation with this plot"
- "I would draw three bar charts for this case since annual income has to be categorized by ranges of income and that's more fluently done using bar charts. For each chart, I would have the horizontal axis having annual income broken down by ranges (e.g. $50K-$100K being one range) and each of the horizontal axes of each chart being the political affiliation (republican, democrat, or independent)"

Why use Open-Ended Questions

- So far the instructors have used Moodle to pinpoint the percentage of correct answers and thus discuss what the students need to relearn or what they need to re-teach. The next step that we would like to take toward improvement of instruction and student learning is the identification of students' misconceptions and specially those misconceptions that are very common among students.
- Limitations
    - Not enough resources to read the answers.
    - Turn-around time is slow. Students get feedback about a week afterwards.

Leading Automatic Grading Software Tools

Name	Cost
IntelliMetric	$30/Student/Year
Criterion	$12/Student/Year
IEA	More than the department's Budget
BETSY	Free

First Try with BETSY

True vs. Computed Scores
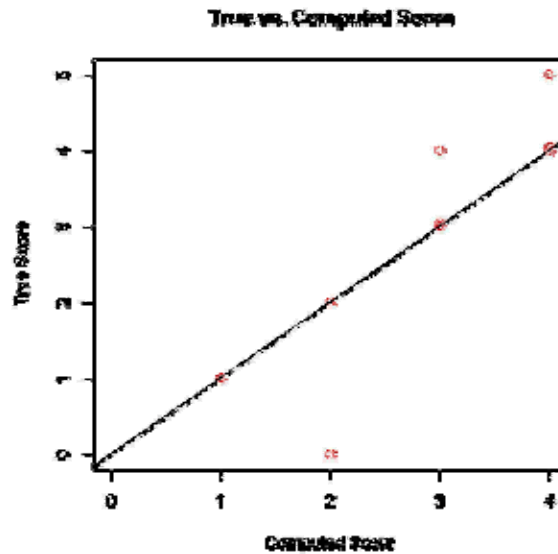
Future Plans

- Build a database of essay prompts for the Automated Grading Software Tools.
- Gather training essay responses.
- Implement an online access site where students can submit responses and have immediate feedback.

- o **3:50 - 4:00 PM** Break

- o **4:00 - 4:15 PM** Group Interactive Discussion

- o **6:00 – 9:00 PM** Social, LACMA (UCLA Transportation is arranged)

Wed 08/08/09

Morning Session (9 AM – 12 Noon)

o **9:00 - 10:00 AM** SOCR Charts, Nicolas Christou, Ivo Dinov & Jenny Cui

i. *Box-and-Whisker Plots*

Summary

This activity describes the construction of the box-and-whisker plot (or simply box plot) in SOCR. The applets can be accessed at SOCR Charts under the Miscellaneous folder.

Goals

- show the importance of the box plot in explanatory data analysis (EDA)
- illustrate how to use SOCR to construct a box plot
- present some unusual pathologies of a box plot

Background & Motivation

The boxplot (or box-and-whisker-plot), invented by John Tukey in 1977, is an efficient way for presenting data, especially for comparing multiple groups of data. In the box plot we can mark-off the five-number summary of a data set (minimum, 25th percentile, median, 75th percentile, maximum). The box contains the 50% of the data. The upper edge of the box represents the 75th percentile, while the lower edge the 35th percentile. The median is represented by a line drawn in the middle of the box. If the median is not in the middle of the box then the data are skewed. The ends of the lines (called whiskers) represent the minimum and maximum values of the data set, unless there are outliers. Outliers are observations below $Q_1 - 1.5(IQR)$ or above $Q_3 + 1.5(IQR)$, where Q_1 is the 25th percentile, Q_3 is the 75th percentile, and $IQR = Q_3 - Q_1$ (called the interquartile range). The advantage of a box plot is that it provides graphically the location and the spread of the data set, it provides an idea about the skewness of the data set, and can provide a comparison between variables by constructing a side-by-side box plots.

Examples & Exercises

- **Example 1**: Go to the SOCR Charts and first, click on the **Miscellaneous** folder and then on **BoxAndWhiskerChartDemo1**. In the Demo1 boxplot we

can see side-by-side box plots of two categories for each of three series. These demonstration data can be viewed by clicking on DATA. Clicking on MAPPING you can choose the variables. Clicking on SHOW ALL the applet will present the graph, the data, and the mapping environment. Let's clear this data set (click on CLEAR) so that we can enter our own data. After you click on CLEAR button, click on the DATA tab to enter data into the spreadsheet. The following data will be entered (don't forget to separate the data by commas!):

| C1	C2	C3
Series 1	1,2,3,4,5,6	2,4,6,8,10,12
Series 2	3,4,5,6,7,8	6,8,10,12,14,16,18
Series 3	5,6,7,8,9	10,16,18,20,22

When you finish entering your data, click on MAPPING to select the series and categories, and finally click on UPDATE_CHART to view the box plots. The following snapshot shows how the above data entered into SOCR:
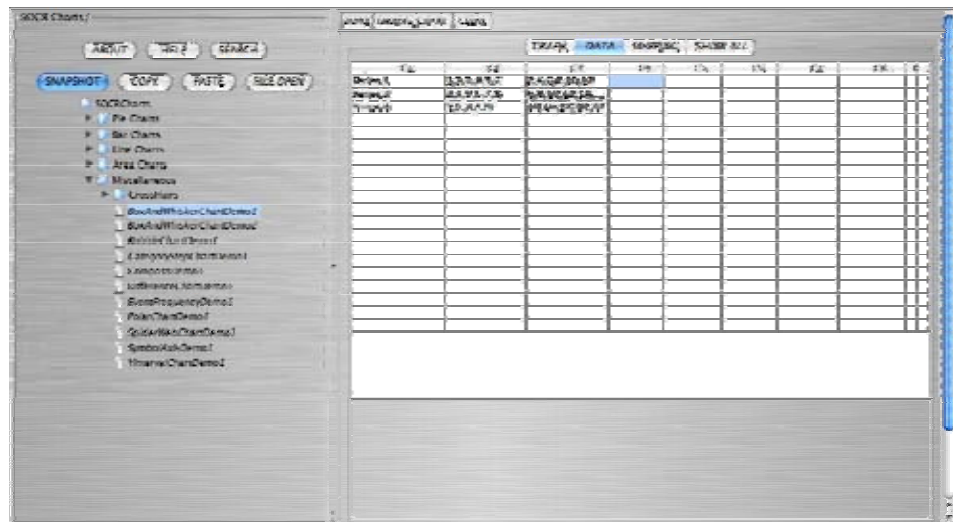
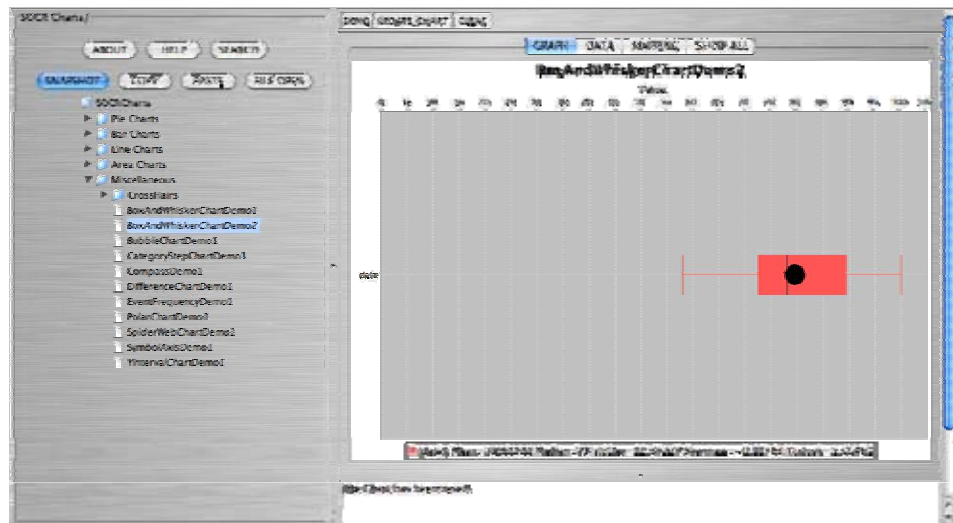The following snapshot shows the mapping of the data:
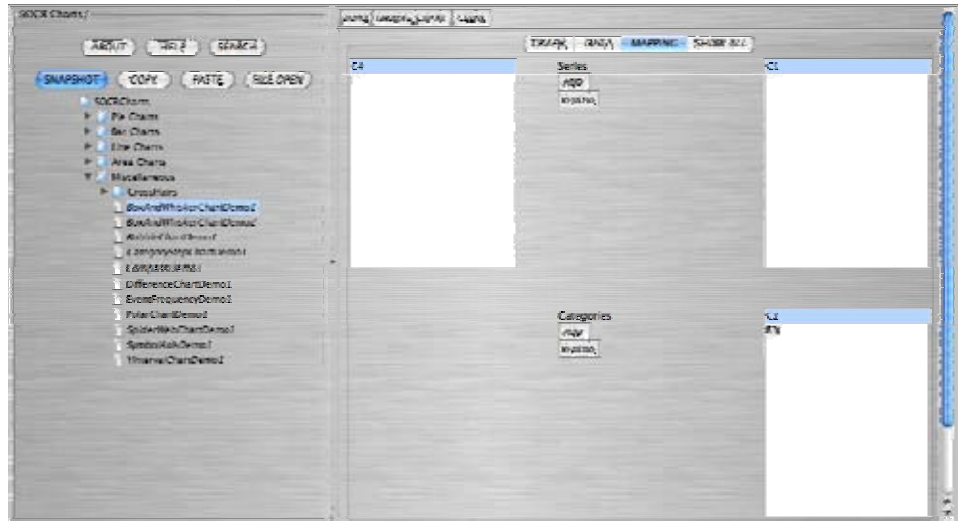
The following snapshot shows the side-by-side box plots:
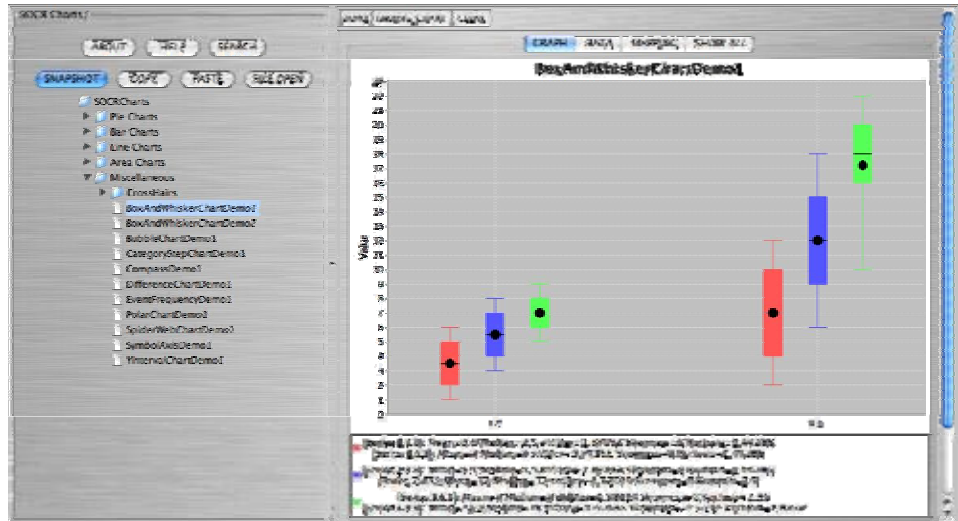
The following snapshot shows the data, the mapping, and the box plots in one screen:
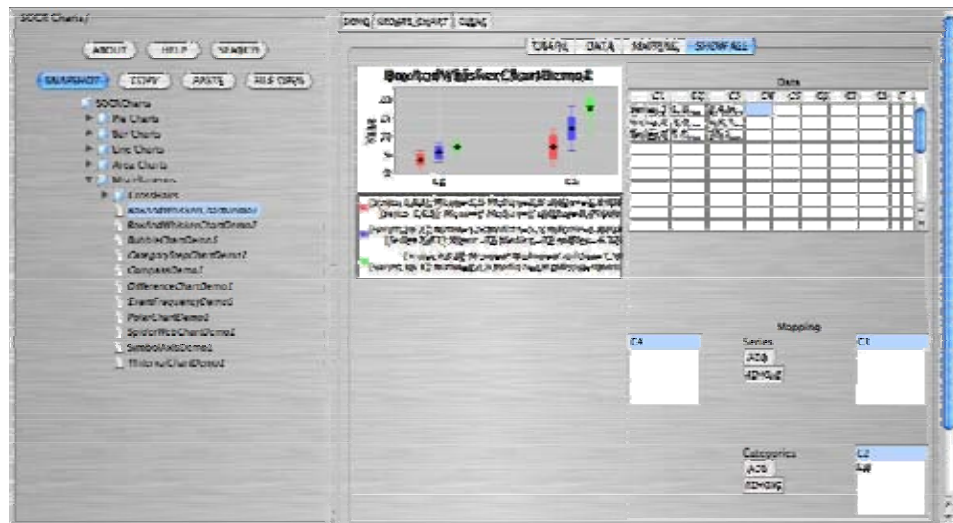
Example 2:

If we are working with a single variable, we can use the **BoxAndWhiskerChartDemo2**. Double click this link to see the demonstration of the construction of the box plot with one variable. As we did in example 1, we will enter our own data. Click on CLEAR to enter your data in the spreadsheet. The data we want to enter are the following: 60, 95, 72, 87, 88, 75, 76, 91, 100, 58, 78, 81, 73, 94, 65.

When you finish entering your data, click on MAPPING to select the category (here only C1), and finally click on UPDATE_CHART to view the box plot.

The following snapshot shows how the above data entered into SOCR:
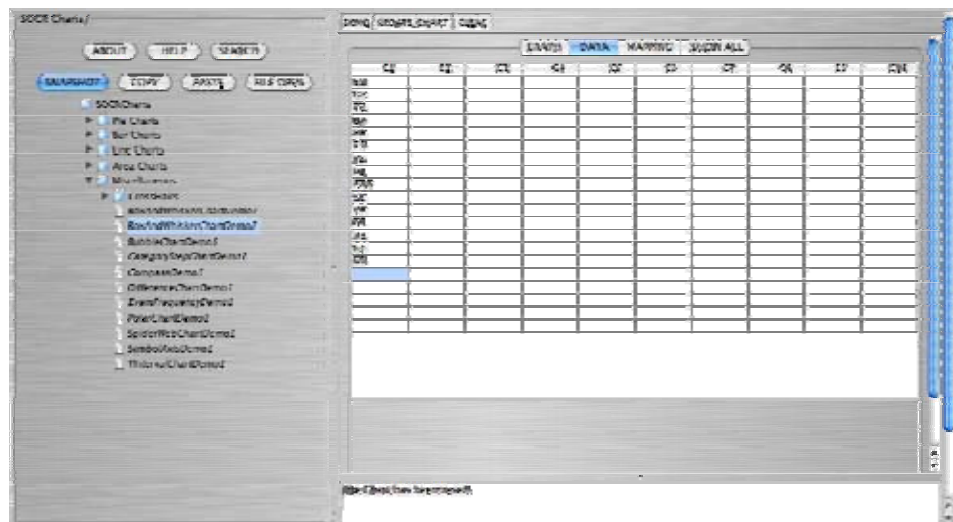
The following snapshot shows the mapping of the data:
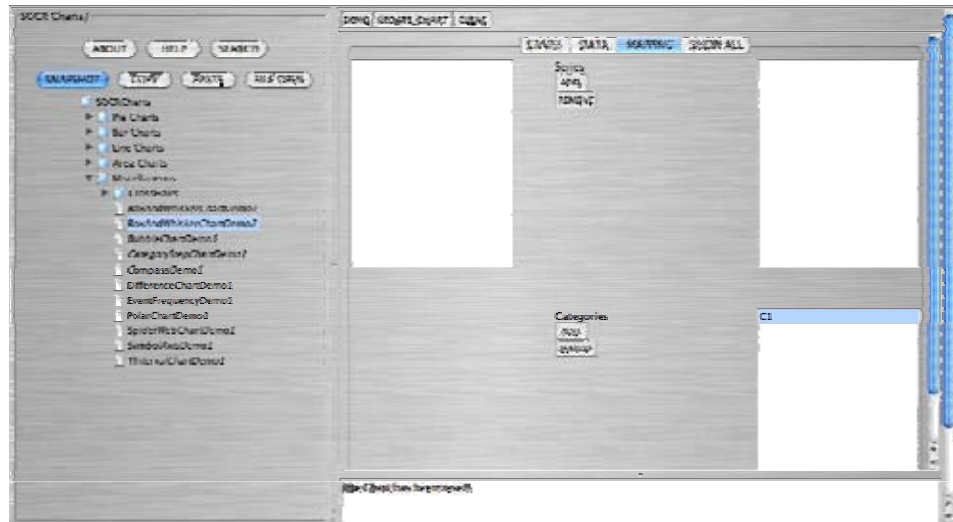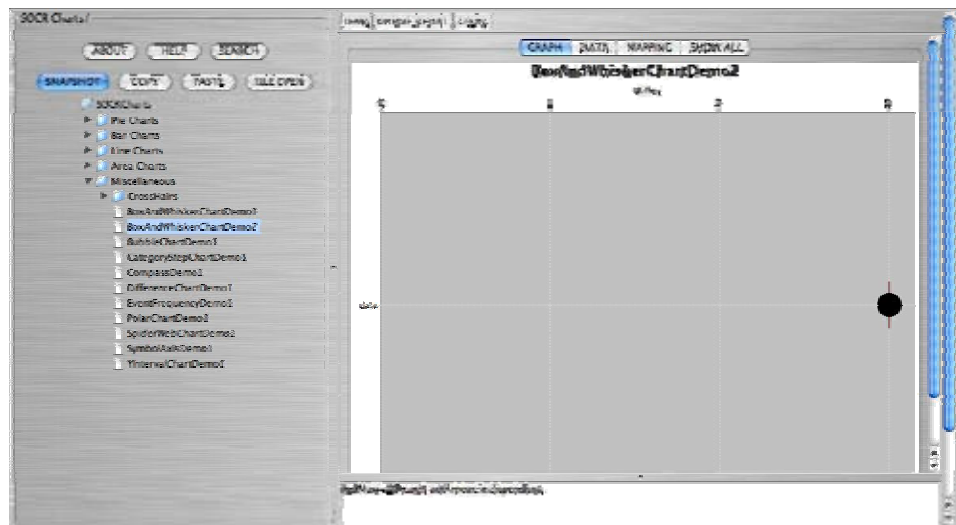

The following snapshot shows the box plot:

The following snapshot shows the data, the mapping, and the box plots in one screen:
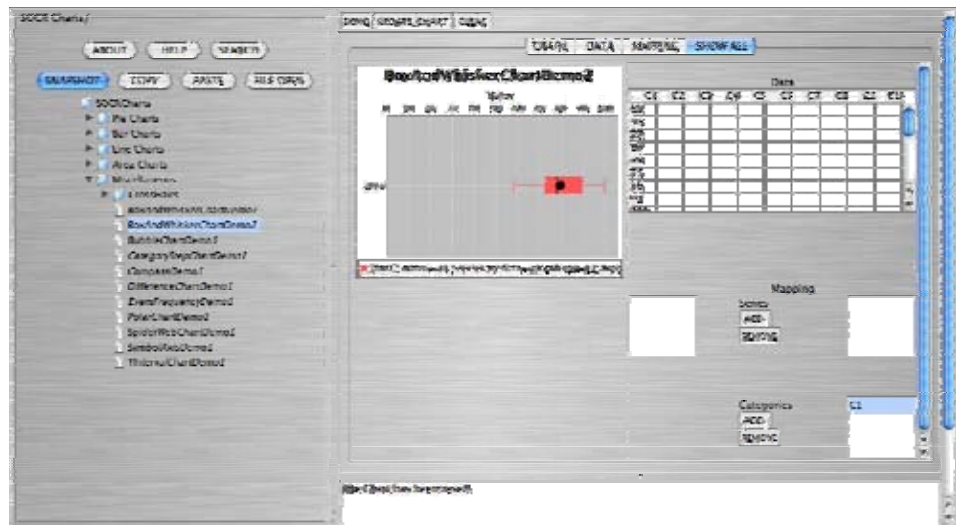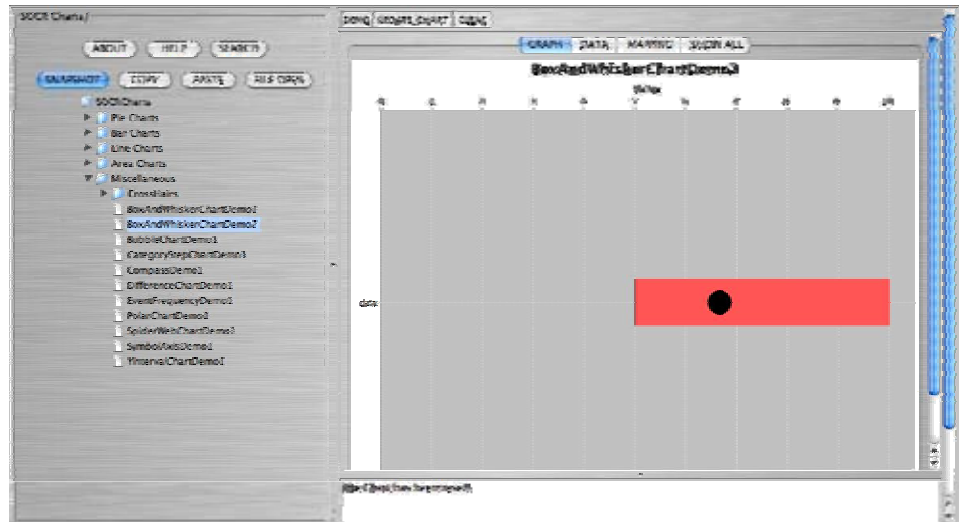

Box Plot Pathologies

Box plots can show unusual pathologies. For the following box plots enter the data in the SOCR Charts spreadsheet that created them.
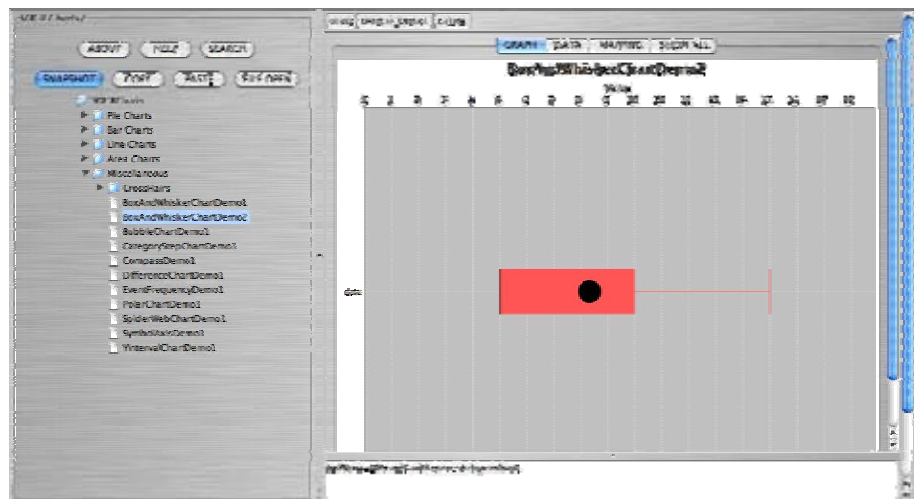
Example 1:

- **Example 2**:
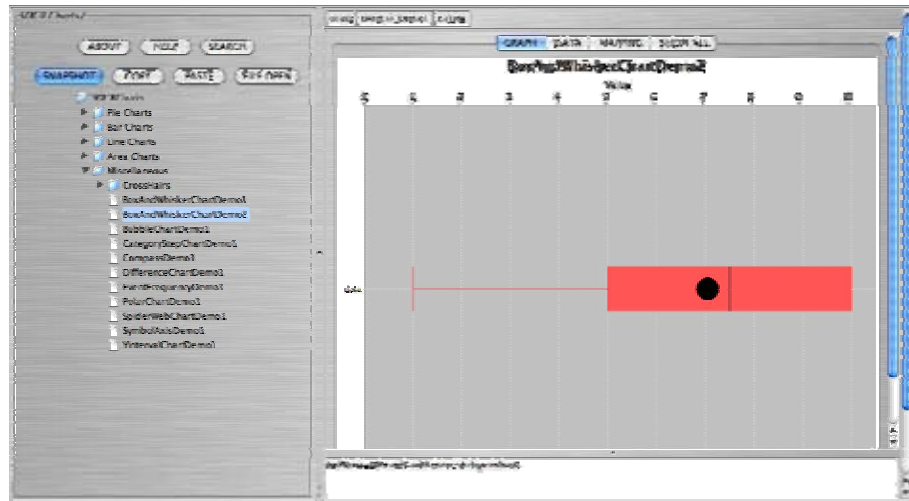
- **Example 3**:

- **Example 4**:

- **Example 5**:
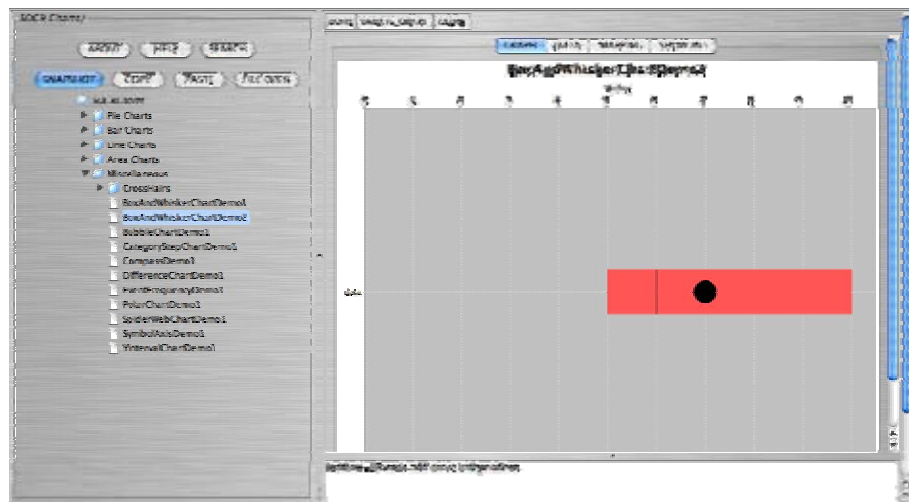
Other Forms of Data

Alternatively, the user can import data by clicking on FILE OPEN. Note here that the data must be saved previously as a comma delimited (CSV) in order to be accessed in SOCR.

ii. *Histogram Generation Activity*

Summary

This is an exploratory data analysis SOCR activity that illustrates the generation and interpretation of the histogram of quantitative data. The complete details about histograms can be found here. In a nutshell, a histogram of a dataset is a

82

graphical visualization of tabulated frequencies or counts of data within equispaced partition of the range of the data. A histogram shows what proportion of measurements that fall into each of the categories defined by the partition of the data range space.

Exercise 1: Simple Histogram from Raw Data

- This exercise demonstrates the construction of a histogram plot from raw quantitative data.
- First, point your browser to **SOCR Charts** and select the **HistogramChartDemo** (under BarCharts --> XYChart). There are three different ways to select data for this histogram chart:
  - o Use the default data provided for this chart (**DEMO** button);
  - o Enter your own data. This can be done by copying to the mouse buffer data from external spreadsheet/table, clicking on the top-left cell in the SOCR Histogram Data table and pasting (**Paste** button) the data into the histogram data table. Remember to **MAP** the data - this indicates what columns rows, parts of the data need to be used in the histogram calculations. Then you click **UPDATE** chart to have the new graph drawn in the **Graph** tab-pane;
  - o Obtain SOCR simulated data from the **Data-Generation** tab of the SOCR Modeler (an example is shown in exercise 3, as well as in the SOCR Power Transform Activity).
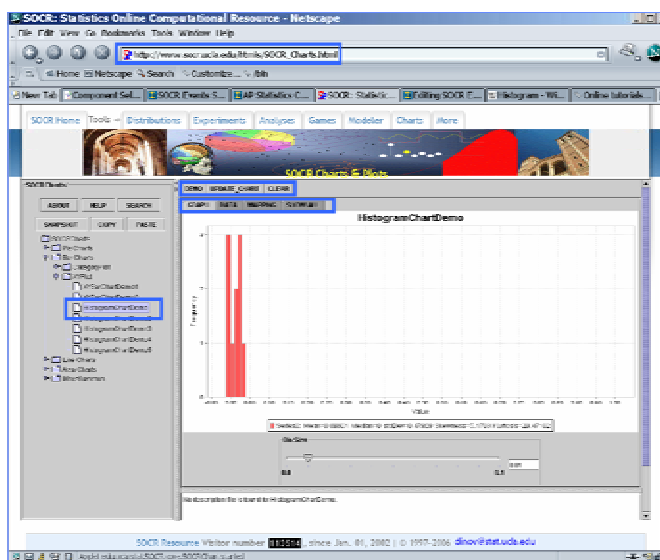
Exercise 2: Histogram from Categories and Frequencies

- Again, point your browser to **SOCR Charts**. This time select the **HistogramChartDemo3** chart (under BarCharts --> XYChart). Use the default data provided for this chart (**DEMO** button).

83

- Notice that this time, the chart requires the user to enter the counts/frequencies of observations within each of the range categories (in this default data case, *year*).
- Try revising some of the numbers in the second (frequency) column and click **UPDATE** button to see the effect of these changes on the histogram.
- Remember that if you enter your own data you need to go to the **MAP** tab-pane and select the columns that contain your histogram bin and frequency columns.
- Using the **SHOW_ALL** tab-pane you can see all three (graph, data and mapping) in the same view.

Exercise 3: Histogram from Simulated Data

- Let's first get some data: Go to SOCR Modeler and generate 100 Cauchy Distributed variables. Copy these data in your mouse buffer (CNT-C). Of course, you may use your own data throughout this exercise.
- Next, paste (CNT-V) these 100 observations in SOCR Charts **HistogramChartDemo** (BarCharts -> XYChart). Go to the **MAP** tab-pane and select the first column (where you pasted your data) in the **XValue** bin. Click **Update Chart** to see the histogram plot of these 100 Cauchy observations in RED!
- Note that the shape of this data histogram resembles the shape of the Cauchy distribution that we sampled this data from.

84

Questions

- What is the effect of the width/size of the histogram bin on the shape of the resulting histogram? Would the shape of the histogram change significantly, if we alter the bin-size? Does the sample-size play role in this?
- Would you expect the shape of the sample histogram to *look like* the shape of the population distribution the data sample came from?

iii. Random Number Generation

Summary

This activity describes the need, general methods and SOCR utilities for random number generation and simulation. SOCR Modeler allows interactive sampling from any SOCR Distribution. The simulated data may then be easily copied and pasted in different SOCR Analyses or Graphing tools for further interrogation.

Goals

- motivate the need for robust random number generators
- illustrate how to use the SOCR random number generators
- present applications of random number generation

Background & Motivation

How many natural processes or phenomena in real life can you describe that have an exact mathematical close-form description and are completely deterministic? Arrival times to school each day? Motion of the Moon around the Earth? The computer CPU? The atomic clock? It is an unsettling paradox that all natural phenomena we observe are stochastic in nature. Yet, we do not know how to replicate any of them exactly. There are good computational strategies to approximate natural processes using analytical mathematical models; however, upon careful review one always finds out a deterministic pattern in all purely computationally generated processes.

There are two strategies to generate random numbers. The first one relies on a *physical process* which is expected to be random. The other uses *computational algorithms* that produce long sequences of apparently random results, which are in fact determined by a shorter initial seed. Random number generators based on *physical processes* may be based on random particles' momentum or position or any of the three fundamental physical forces. Examples of such processes are the Atari gaming console (noise from an analog circuits to generate true random numbers), radioactive decay, thermal noise, shot noise and clock drift. A random number generator (RNG) based solely on deterministic computation is referred to

85

pseudo-random number generator. There are <u>various techniques for obtaining</u> <u>computational (pseudo)random numbers</u>. Virtually all RNG's used in practice are pseudo-RNGs. To distinguish real random numbers from the pseudo-random numbers is a very difficult problem.

If all natural processes are inherently random and at the same time we can not generate ourselves good (non-deterministic) RNG processes why are we even attempting to do that? Wouldn't it be much easier to just use measurements of the natural physical processes? The answer is simple: We typically need to sample/simulate data from a *specific process* and it is not easy to show that a physical phenomena we observe has the same distribution as the process of interest! So, our need of sampling from a specific distribution demands that we ensure the proper characteristics of the sample.

Where does this sampling-need come from? Random number generators have several important applications in statistical modeling, computer simulation, cryptography, etc. For example, data collection is often very expensive. Hence, to do appropriate inference on datasets of smaller sizes, we may consider simulating repeatedly from appropriate distributions, instead of using real observations. Another example of why are random number generators so important comes from cryptography. It is a commonly held misconception that every encryption method can be broken. Claude Shannon, Bell Labs, 1948, proved that the <u>one-time pad</u> <u>cipher</u> is unbreakable, provided the secret key is truly random and of length equal or greater than the length of the encoded message. <u>Monte Carlo simulations</u> are also based on RNGs and are used for finding numerical solutions to (multi-dimensional) mathematical problems that cannot easily be solved exactly. For example, integration, differentiation, root-finding, etc.

- **Exercise 1**: Go to the <u>SOCR Modeler</u> and click on the **Data Generation** tab. Select 200 observations from the <u>Generalized Beta Distribution</u>, as shown on the image below. Choose this four-tuple for the parameters $\alpha = 1.5; \beta = 3; A = 0; B = 7$. Copy these 200 values in your mouse buffer (CNT-C) and paste them in the **Data** tab of the **LineCharts --> PowerTransformHistogramChart** under <u>SOCR</u> <u>Charts</u>. Then *Map* this column to *XYValue* (under the **MAP** tab) and click **Update_Chart**. This will generate the histogram of the 200 observations. Indeed, this graph should look like a discrete analog of the <u>Generalized Beta</u> density curve. You can see exactly what the <u>Generalized Beta Distribution</u> looks like by going to <u>SOCR Distributions</u> and selecting *Beta($\alpha = 1.5; \beta = 3; A = 0; B = 7$)*.

- **Exercise 2**: Let's get some more simulated data: Go to SOCR Modeler and generate 100 Cauchy Distributed variables. Copy these data in your mouse buffer (CNT-C). Of course, you may choose any other distribution. Next, paste (CNT-V) these 100 observations in SOCR Charts (**Line-Charts -> Power Transform Chart**). *Map* this column to *XYValue* (under the **MAP** tab) and click **Update_Chart** to see the index plot of this data in RED!
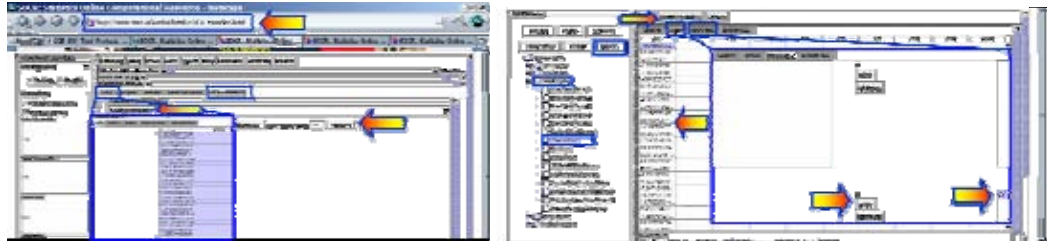
Applications

The RGN background and motivation section clearly described some of the critical scientific and technological challenges that rely upon the existence of quality RNGs. Here we present the applications of the SOCR RNG's for various interactive activities and demonstrations.

- Power-Transform Family Graphs
- Mixture Model Experiment

iv. *Power Transformation Family*

Summary

This is activity demonstrates the usage, effects and properties of the modified power transformation family applied to real or simulated data to reduce variation and enhance Normality. There are 4 exercises each demonstrating the properties of the power transform in different settings for observed or simulated data: X-Y scatter plot, QQ-Normal plot, Histogram plot and Time/Index plot.

Background

The **power transformation family** is often used for transforming data for the purpose of making it more Normal-like. The power transformation is continuously varying with respect to the power parameter λ and defined, as continuous piece-wise function, for all $y > 0$ by

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda}, \ if \ \lambda \neq 0 \\ log(\ y\ ), \ if \ \lambda = 0 \end{cases}$$

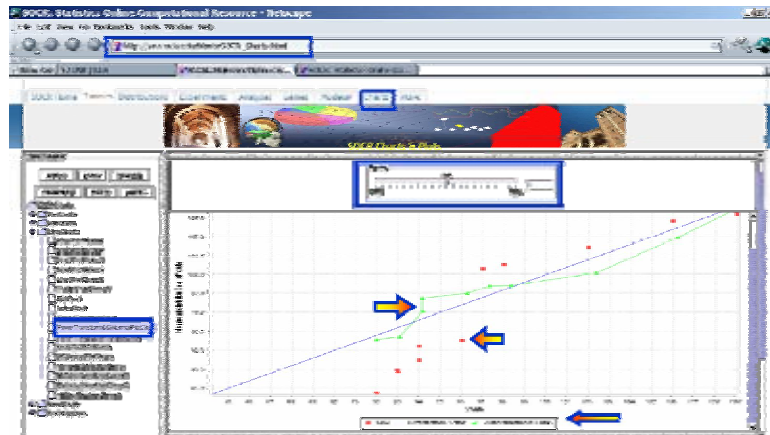## Exercise 1: Power Transformation Family in a X-Y scatter Plot Setting

- This exercise demonstrates the characteristics of the power-transform when applied independently to the two processes in an X-Y scatter plot setting. In this situation, one observed paired (X,Y) observations which are typically plotted X vs. Y in the 2D plane. We are interested in studying the effects of independently applying the power transforms to the X and Y processes. How and why would the corresponding scatter plot change as we vary the power parameters for X and Y?

- First, point your browser to **SOCR Charts** and select the **PowerTransformXYStatterChart** (Line-Charts -> PowerTransformXYStatterChart). Then either use the default data provided for this chart, enter your own data (remember to **MAP** the data before your **UPDATE** the chart) or obtain SOCR simulated data from the **Data-Generation** tab of the **SOCR Modeler** (an example is shown later in Exercise 4). As shown on the image below, try changing the power parameters for the X and Y power-transforms and observe the graphical behavior of the transformed scatter-plot (blue points connected by a thin line) versus the native (original) data (red color points). We have applied a linear rescaling to the power-transform data to map it in the same space as the original data. This is done purely for visualization purposes, as without this rescaling it will be difficult to see the correspondence of the transformed and original data. Also note the changes of the numerical summaries for the transformed data (bottom text area) as you update the power parameters. What power parameters would you suggest that make the X-Y relation most linear?

Exercise 2: Power Transformation Family in a QQ-Normal Plot Setting

- The second exercise demonstrates the effects of the power-transform applied to data in a QQ-Normal plot setting. We are interested in studying the effects of power transforming the native (original) data on the quantiles, relative the Normal quantiles (i.e., QQ-Normal plot effects). How and why do you expect the QQ-Normal plot to change as we vary the power parameter?

- Again go to SOCR Charts and select the **PowerTransformQQNormalPlotChart** (Line-Charts -> PowerTransformQQNormalPlotChart). You can use different data for this experiment - either use the default data provided with the QQ-Normal chart, enter your own data (remember to **MAP** the data before your **UPDATE** the chart) or obtain SOCR simulated data from the **Data-Generation** tab of the SOCR Modeler (an example is shown later in Exercise 4). Change the power-transform parameter (using the slider or the by typing int he text area) and observe the graphical behavior of the transformed data in the QQ-Normal plot (green points connected by a thin line) versus the plot of the native data (red color points). What power parameter would you suggest that make the (transformed) data quantiles similar to those of the Normal distribution? Why?


Exercise 3: Power Transformation Family in a Histogram Plot Setting

- This exercise demonstrates the effects on the histogram distribution after applying the power-transform to the (observed or simulated) data. In this experiment, we want to see whether we can reduce the variance of a dataset and make its histogram more symmetric, unimodal and bell-shaped.

- Again go to SOCR Charts and select the **PowerTransformHistogramChart** (Bar-Charts -> XYPlot -> PowerTransformHistogramChart). We will use SOCR simulated data from the **Data-Generation** tab of the SOCR Modeler, however you may chose to use the default data for this chart or enter your own data. The image below shows you the Generalized Beta Distribution using SOCR Distributions.
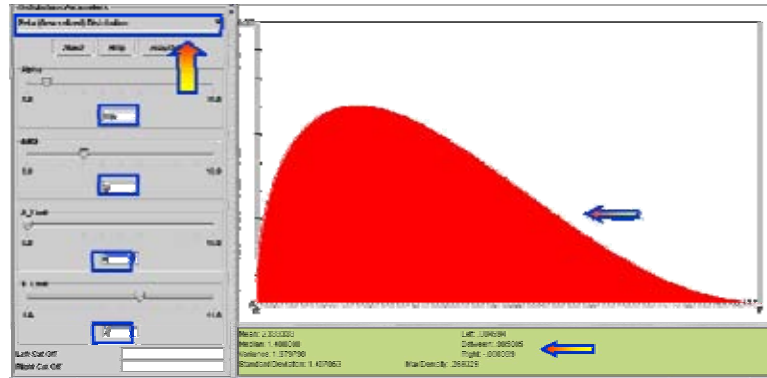
- Go to the SOCR Modeler and select 200 observations from the Generalized Beta Distribution ($\alpha = 1.5; \beta = 3; A = 0; B = 7$), as shown on the image below. Copy these 200 values in your mouse buffer (CNT-C) and paste them in the **Data** tab of the **PowerTransformHistogramChart**. Then *map* this column to *XYValue* (under the **MAP** tab) and click **Update_Chart**. This will generate the histogram of the 200 observations. Indeed, this graph should look like a discrete analogue of the Generalized Beta density curve above.
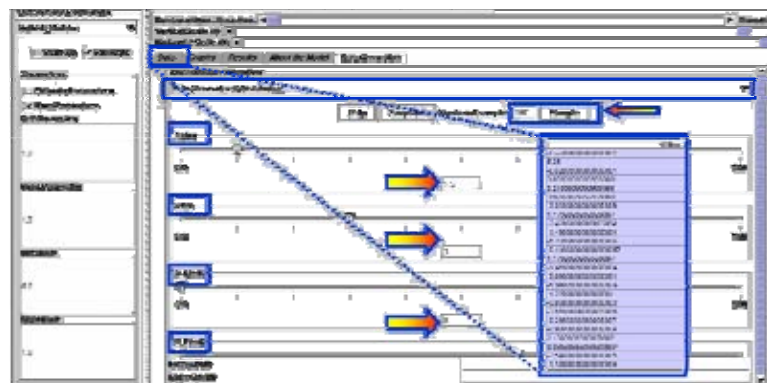

- In the **Graph** tab of the **PowerTransformHistogramChart**, change the power-transform parameter (using the slider on the top). All SOCR Histogram charts allow you to choose the width of the histogram bins, using the second slider on the top. Observe the graphical behavior of the **histogram** of the transformed data (blue bins) and compare it to the **histogram** of the native data (red bins). What power parameter would you suggest that make the **histogram** of the power-transformed data better? Why?

90
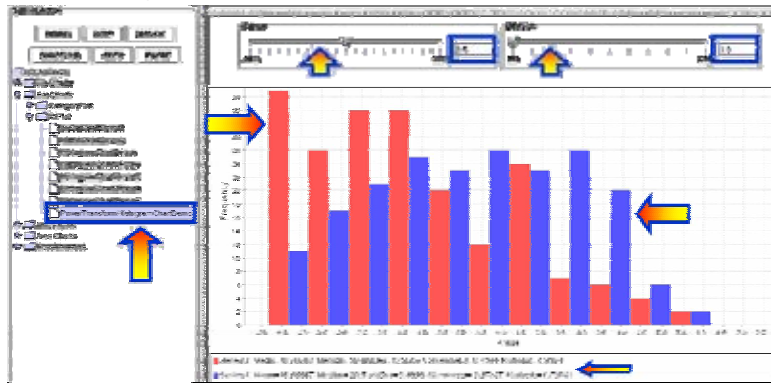
Exercise 4: Power Transformation Family in a Time/Index Plot Setting

- Let's first get some data: Go to [SOCR Modeler](#) and generate 100 Cauchy Distributed variables. Copy these data in your mouse buffer (CNT-C). Of course, you may use your own data throughout. We choose Cauchy data to demonstrate how the Power Transform Family allows us to normalize data that is far from being Normal-like.
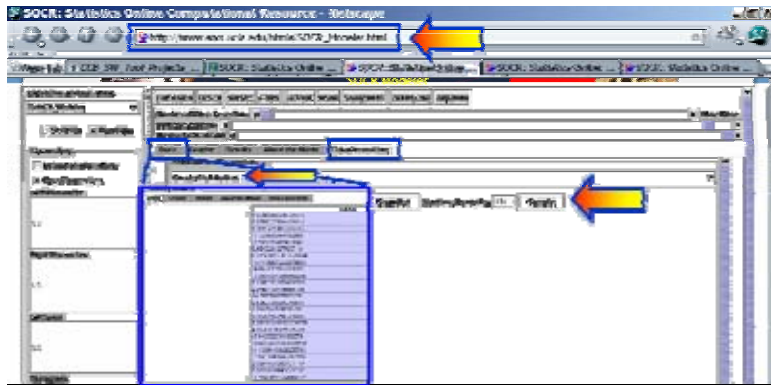

- Next, paste (CNT-V) these 100 observations in [SOCR Charts](#) (Line-Charts -> Power Transform Chart). Click **Update Chart** to see the index plot of this data in RED!
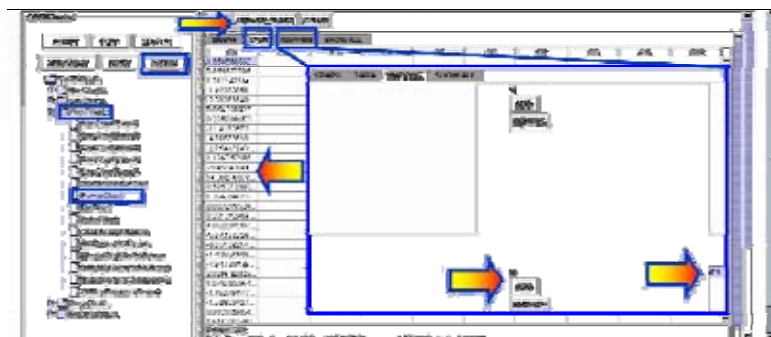
- Now go to the **Graph Tab-Pane** and choose λ = 0 (the power parameter). Why is λ = 0 the best choice for this data? Try experimenting with different values of λ. Observe the variability in the Graph of the transformed data in Blue (relative to the variability of the native data in Red).
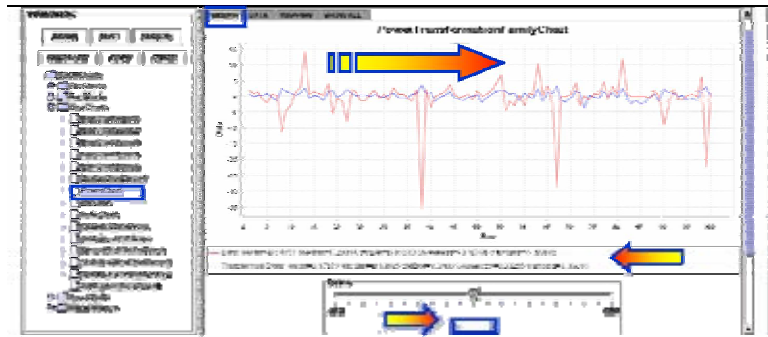

- Then go back to the **Data Tab-Pane** and copy in your mouse buffer the transformed data. We will compare how well does Normal distribution fit the histograms of the raw data (Cauchy distribution) and the transformed data. One can experiment with other powers of λ, as well! In the case of λ = 0, the power transform reduces to a **log transform**, which is generally a good way to make the histogram of a data set well approximated by a Normal Distribution. In our case, the histogram of the original data is close to Cauchy distribution, which is heavy tailed and far from Normal (Recall that the *T(df)* distribution provides a 1-parameter homotopy between Cauchy and Normal).
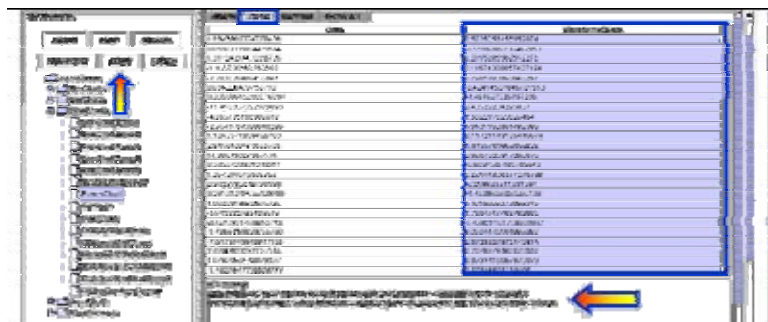

- Now copy in your mouse buffer the transformed data and paste it in the SOCR Modeler. Check the **Estimate Parameters** check-box on the top-left. This will allow you to fit a Normal curve to the histogram of the (log) Power Family Transformed Data. You see that Normal Distribution is a great fit to the histogram of the transformed Data. Be sure to check the parameters of the Normal Distribution (these are estimated using least squares and reported in the **Results** Tab-Pane). In this case, these parameters are: *Mean = 0.177, Variance = 1.77*, however, these will vary, in general.
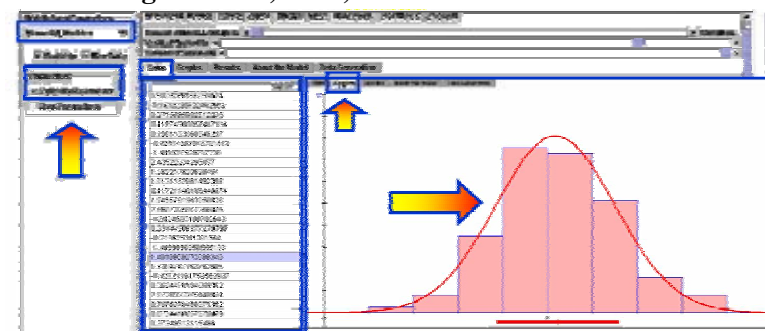
- Let's try to fit a Normal model to the histogram of the native data (recall that this histogram should be shaped as Cauchy, as we sampled from Cauchy distribution – therefore, we would not expect a Normal Distribution to be a good fit for these data. This fact, by itself, demonstrates the importance of the Power Transformation Family. Basically we were able to *Normalize* a significantly Non-Normal data set. Go back to the original SOCR Modeler, where you sampled the 100 Cauchy observations. Select **NormalFit_Modeler** from the drop-down list of models in the top-left and click on the **Graphs** and **Results** Tab-Panes to see the graphical results of the histogram of the native (heavy-tailed) data and the parameters of its best Normal Fit. Clearly, as expected, we do not have a got match.
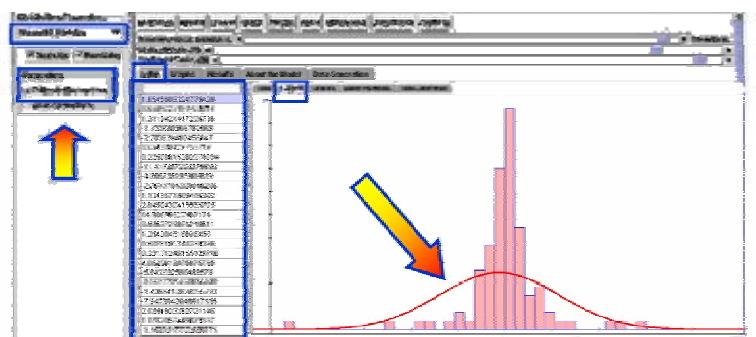
Questions

o Try experimenting with other (real or simulated) data sets and different Power parameters (λ). What are the general effects of increasing/decreasing λ in any of these domains [-10;0], [0;1] and [1;10]?

o For each of the exercises (X-Y scatter-plot, QQ-Normal plot, Histogram plot and Time/Index plot) empirically study the effects of the power transform as a tool for normalizing the data. You can take samples of size 100 from Student's T-distribution (low df) and determine appropriate levels of λ for which the transformed data is (visually) well approximated by a Normal Distribution.
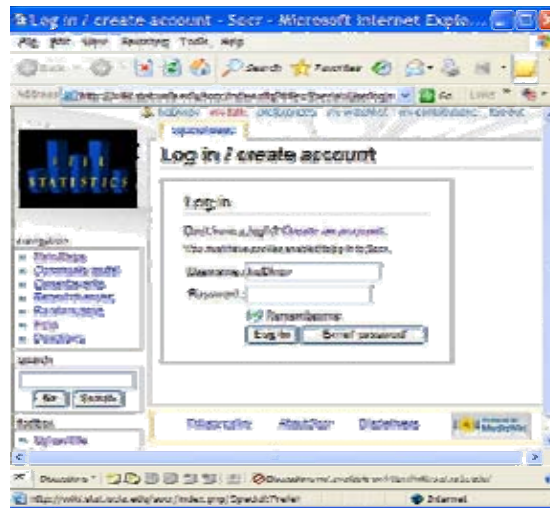
- o **<u>10:00 - 10:15 AM</u>** Break

- o **<u>10:15 - 11:15 AM</u>** Wiki Editing and Development of new Instructional Materials, Ivo Dinov & Nicolas Christou

This is a brief outline of the most common SOCR Wiki Resource Editing Tasks and should be used in addition to the **MediaWiki Editing Tutorials**

Get a SOCR Wiki Account

- Go to User Login/Create an Account first.


Choose a (root) page to link off your new document

- Go to the Main SOCR Wiki (Main Page), follow the hierarchy and start at the page that is the most natural parent of the material you are going to propose. You can also generate a stand-alone Wiki document, but you should avoid that. Always try to find where your future document belongs first. Then go to this page and edit this page (**edit** tab on top of page) to include a link to your new document.
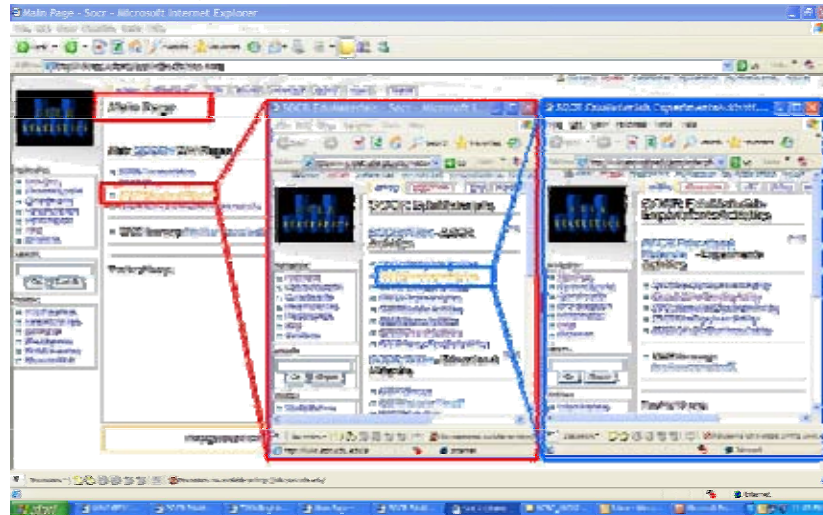
- In the root page you may include a code like this to generate a link to your future SOCR Wiki Page:
  [[SOCR_MainTopic_SubTopic_SubSubTopic_NewTopic_Author_Date]]. This link should be properly placed in the parent (root) page (location, context, etc.)
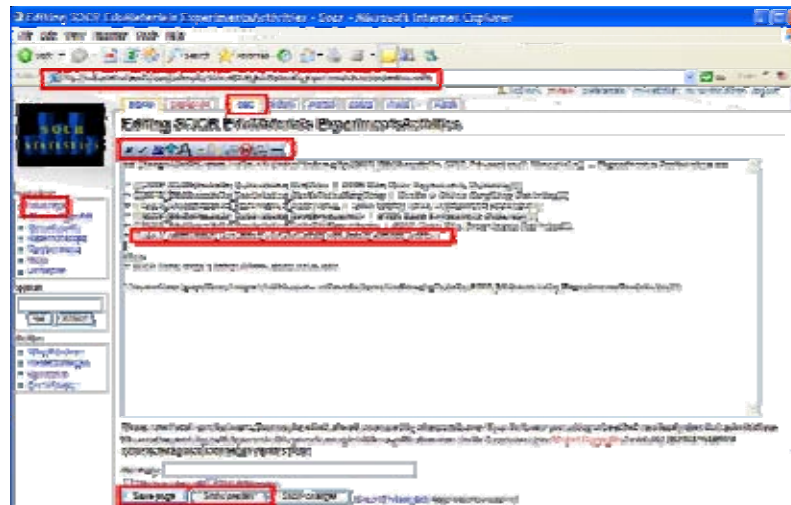
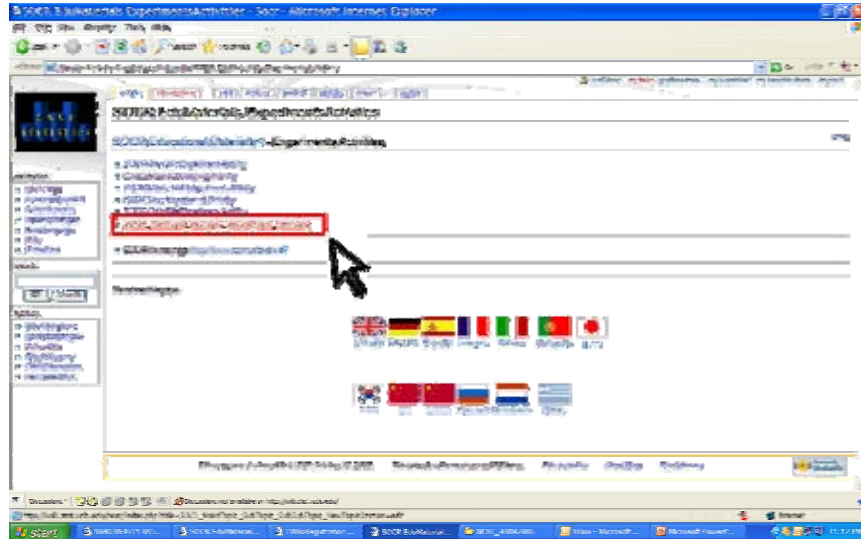## Naming of your new SOCR Wiki documents

- We recommend the following naming-convention to avoid potential document confusions
  o For more stable links:
    [[SOCR_MainTopic_SubTopic_SubSubTopic_NewTopic]], e.g.,
    SOCR_EduMaterials_Activities_Standard52CardDeck
  o For more dynamic and frequently revised documents by many people:
    [[SOCR_MainTopic_SubTopic_SubSubTopic_NewTopic_Author_Date_Ve
   rsion]]

95

Saving New-Page-Link in Root Page

- Now save your link in the parent (root) page: click the **save page** button on the bottom of the editing-mode Wiki pages. This will show you the newly redesigned root page that contains an explicit link to your new page, e.g., [[SOCR_MainTopic_SubTopic_SubSubTopic_NewTopic_Author_Date_Version]]


Edit the new SOCR Wiki document

- Click on the new link that you created in the root-page, [[SOCR_MainTopic_SubTopic_SubSubTopic_NewTopic_Author_Date_Version]]. This should put you directly in editing-mode of your new SOCR Wiki document (unless such a page already exists, then you will see this page, and you still can get in its editing mode by clicking the **edit** tab-pane on the top, as before).

96
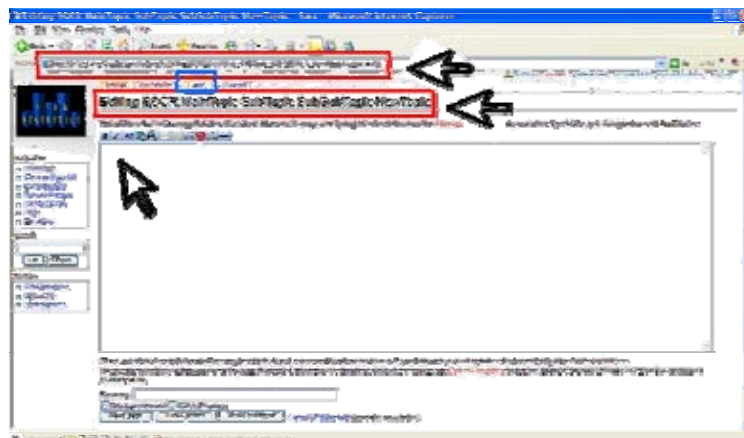
Images in SOCR Wiki documents

- Before you include (JPG) images in your SOCR Wiki pages you must save the images (use proper names, see below) on your client computer.
  o Naming your SOCR Wiki Images:
    ▪ Example: SOCR_Activities_CardCoinSampling_Dinov_092206_Fig1.jpg
    ▪ General Image Naming Convention: SOCR_MainTopic_SubTopic_SubSubTopic_Title_Author_Date_FigNo.jpg
  o Uploading Images: Click on the bottom-left side Toolbox link: Upload file and follow the instructions to upload your < 1MB JPG image(s).
  o Include your Image in your new Wiki document: Go back in the editing mode of your new SOCR Wiki page. A variant of this command [[Image:http://wiki.stat.ucla.edu/socr/uploads/e/e1/SOCR_Activities_CI_Dinov_092206_Fig1.jpg|300px]] will include your image as follows:
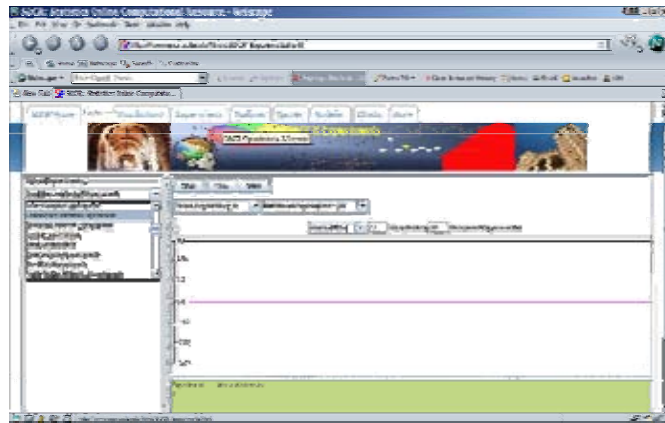
Tables

- This code generates the table below (**entered one-cell per line - e.g., | cell(2,2)**): {| align="center" border="1" | colspan="5" align="center"| Common Top Row |- | rowspan="3"| Common First Column | cell(2,2) | cell(2,3) | cell(2,4) | cell(2,5) |- | cell(3,2) | cell(3,3) | cell(3,4) | cell(3,5) |- | cell(4,2) | cell(4,3) | cell(4,4) | cell(4,5) |}. You can copy, paste and revise this code according to your table needs in the editing mode of your new Wiki page (remember that each cell value must the entered on a new line!)

Common Top Row				
Common First Column	cell(2,2)	cell(2,3)	cell(2,4)	cell(2,5)
	cell(3,2)	cell(3,3)	cell(3,4)	cell(3,5)
	cell(4,2)	cell(4,3)	cell(4,4)	cell(4,5)

External/Internal Links

- You can either use the 3rd and 4th tools on the top of each editing-mode Wiki page or you can use these to directly to do Internal SOCR Wiki Link or External SOCR Wiki Page.

Math and LaTeX expressions

- You can use either
  o The complete Math/LaTeX MediaWiki doc
  o In editing-mode, enter your TeX/LaTeX math expression and bound it by: $\sqrt{n}$
  o Or highlight any TeX/LaTeX expression and use the $\sqrt{n}$ tool on the top of any editing-mode Wiki page to mark it specifically as LaTeX-interpreted. Once you save your new Wiki page the math expression should appear correctly rendered on the new Wiki page.
  o See examples here: http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_CLT

Wiki Page Internationalization

- Remember that all SOCR Wiki pages should be made available in other languages using machine translations. This is accomplished via placing the following two lines at the bottom of your new document (you need to enter the new name of your SOCR Wiki page, as indicated). See the translation buttons below for examples of other languages that SOCR materials are automatically translated to: <hr>* SOCR Home page: http://www.socr.ucla.edu
 {{translate|pageName=http://wiki.stat.ucla.edu/socr/index.php?title=NEW_SOCR_Page_MainTopic_SubTopic_SubSubTopic_NewTopic_Author_Date_Version}}. See the bottom of this page for example of how your language translation will look like.

Hints

- Remember that the easiest way to construct a valid and well-designed SOCR Wiki page is the generate the materials, first, outside the SOCR Wiki. Make sure you are satisfied with the look-feel-content of your new materials, spell-ckeck the entire document (grammer, style, etc.) Finally, when you are happy with the content, you can quickly paste your text, figure, tables, links, expressions, diagrams, etc. in the SOCR Wiki page and make them available online to the community!

- o **11:15 - 11:30 AM** Group Interactive Discussion on EDA and Development of Wiki Resources

- o **11:30 - 11:45 AM** Collaborative AP Statistics E-Book
  http://wiki.stat.ucla.edu/socr/index.php/EBook
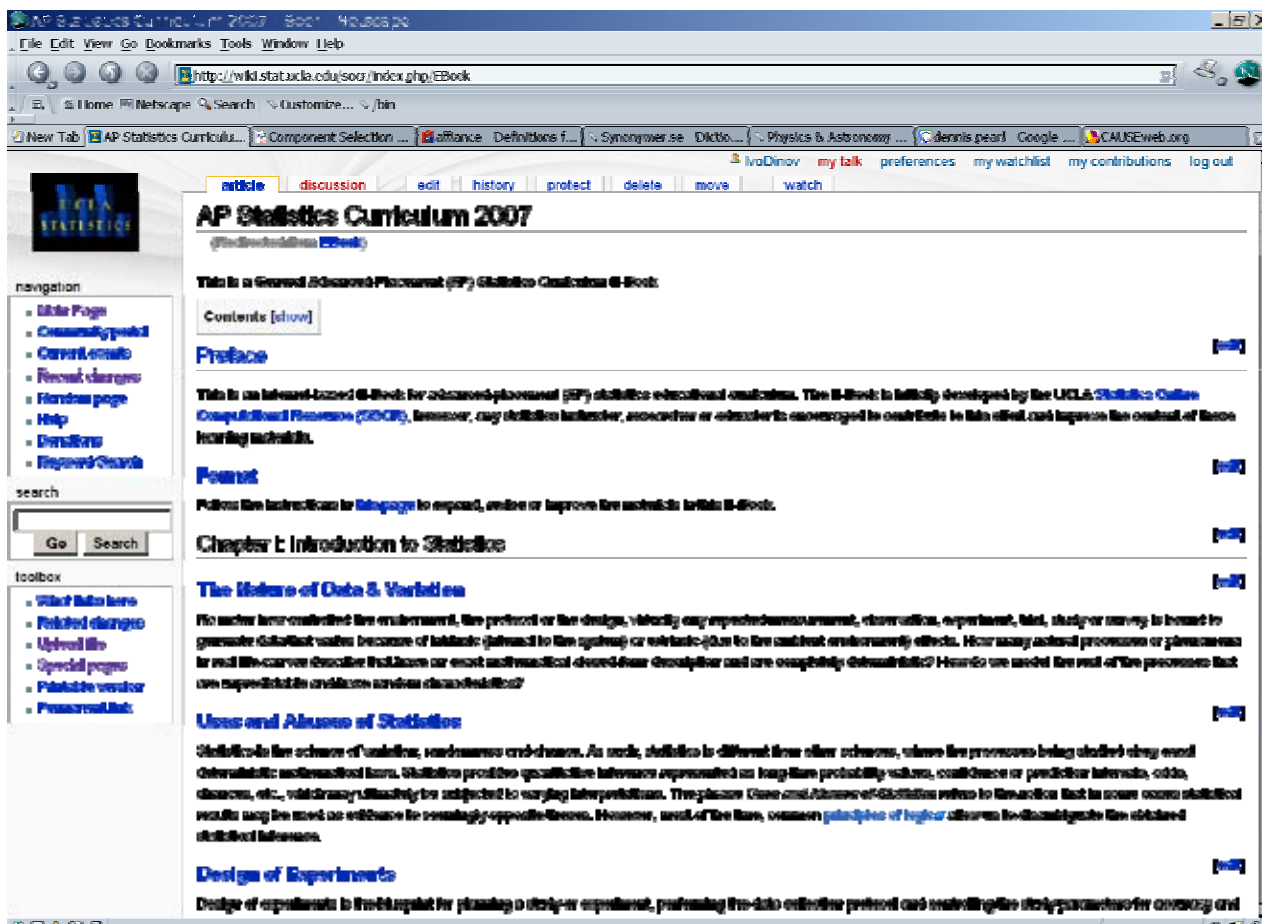
- o **11:45 - 12:00 Noon** Workshop Evaluation (see forms in this booklet)

- o **12:00 - 1:00 PM** Lunch Break & Adjourn

Afternoon Session - Visit to J. Paul Getty Museum - 1:00 – 4:00 PM. Bus pick up at Ackerman Union Turnaround. Getty Tour starts at 2 PM.

Program

DAY 1 - 08/06/07

8:00AM-12:00PM
DISTRIBUTIONS
Participants Introductions,
SOCR Distributions,
Distribution Activities I &
II, Discrete Distributions
Activities, Explore Relations
Between Distributions,
Interactive Discussion

12:00PM-1:00PM
LUNCH BREAK

1:00PM-4:00PM
EXPERIMENTS & GAMES
SOCR/VLPS Experiments,
SOCR/VLPS Experiments/Games
Activities I & II, Datasets,
Lab Manuals, UCLA Stat
Education Resources,
Interactive Discussions

DAY 2 - 08/07/07

9:00AM-12:00PM
ANALYSES & MODELER
SOCR ANALYSES
ANOVA, SOCR Modeler,
Interactive Discussion

12:00PM-1:00PM
LUNCH BREAK

1:00PM-4:00PM
CAUSE & UCLA RESOURCES
Experiments/Games
Activities III, CAUSE
Resources, Use of
Technology, Group
Discussion

DAY 3 - 08/08/07

9:00AM-12:00PM
EDA, CHARTS, &
INSTRUCTIONAL MATERIALS
SOCR Charts, Wiki Editing,
Development of New
Instructional Materials,
Interactive Discussion,
Collaborative AP Statistics
E-Book, Workshop Evaluation

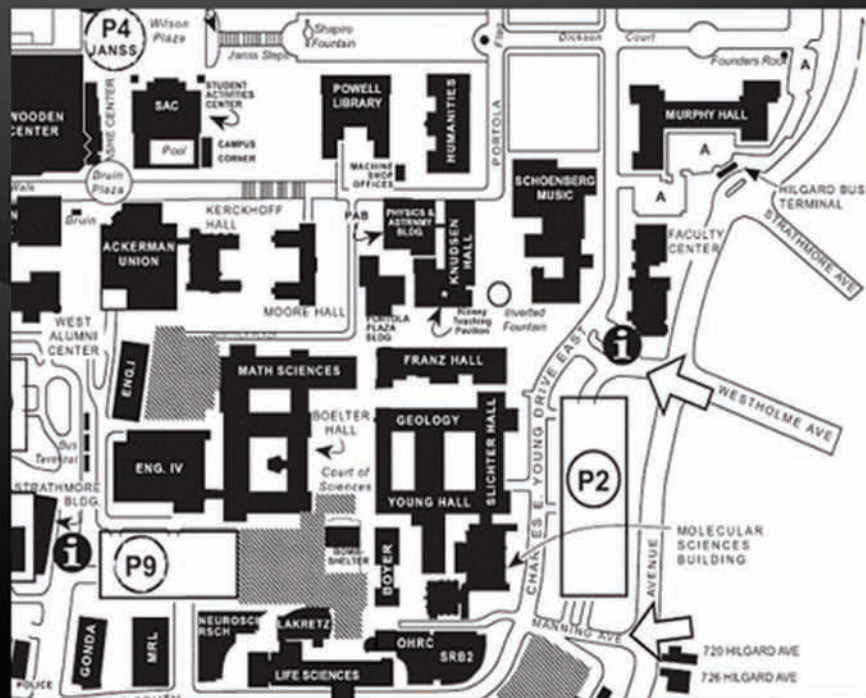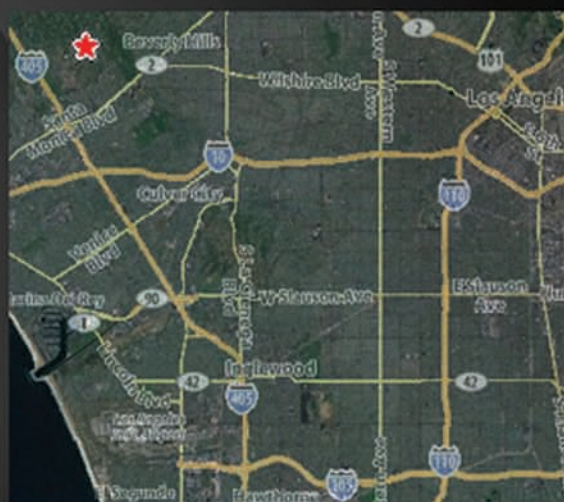**12:00PM-1:00PM**
LUNCH BREAK & ADJOURN

1:00PM-4:00PM
Visit to J. Getty Museum

Directions from LAX to UCLA

Start EAST on WORLD WAY/CENTER WAY > turn LEFT > take
CENTURY BLVD/CA-1N/SEPULVEDA BLVD RAMP > CENTURY BLVD
RAMP > merge W CENTURY BLVD > take I-405 toward
SACRAMENTO > merge I-405 N toward SANTA MONICA > merge
WILSHIRE BLVD toward WESTWOOD > LEFT onto WESTWOOD BLVD >
RIGHT onto LINDBROOK > STRAIGHT onto HILGARD > END at
ucla

Map of UCLA South Campus