EDC384 (Stroup) Spring 1999 --- Designing for Technological Innovation

FINAL PROJECT --- Hunter Ellinger (hunter.ellinger.org, ellinger@io.com) --- March 1999

CENTRAL LIMIT

A program to demonstrate & explain the normal convergence of distributions of sample means

Introduction:

The design described here is for the initial component of a program system planned to support statistics education. The primary target audience is people with a moderate algebra background (especially graphing of functions) but not necessarily any prior knowledge or training in sampling or estimation theory. Formal courses in this content are usually presented at the late secondary or early college level. The program will include connections to higher-level material, however, so that it will also be of some use to higher-level students and to those professionals who use sampling and estimation techniques.

While initial testing and distribution of the program is planned from the MAA/ASA statistics-education web sites, and certain parts of the program will make use (when available) of internet links to external sites, the program design given here would work about as well in separate-single-user form. For example, the optimal distribution of tasks between client and server machines, which can be expected to vary with hardware/network evolution, is not seen as having substantial impact on the design of this program set. It is therefore ignored here.

While I have made no significant revisions in my list of preferred design criteria since midterm, several details of the program design itself were changed or added in response to thoughts provoked by the second half of the course. Because I found that the interaction of the criteria with the specific design decisions was holographic rather than one-for-one, my discussion is centered on explaining the design rather than labeling the source of ideas.

Organization of this paper:

A. Outline of my preferred criteria for instructional-software design

B. Outline of the specific design for the Central Limit statistics-education program

C. Discussion of the design in light of the design criteria

A. DESIGN CRITERIA FOR EDUCATIONAL-SOFTWARE DEVELOPMENT

General instructional goals

- Growth in sophistication
- Awareness of interrelatedness
- Confrontation of uncertainties
- Knowledge of applicability limits
- Independent inquiry
- Validation of non-school experience
- Teamwork

Task-definition refinement

- Sources of external input
- Expectations
- Concerns
- Constraints
- Interferences
- Harmonies

Instructional strategies

- Declaration of intended use
- Communication of relevance
- Mutual introductions
- Adaptability to students
- User locus of control
- Encouragement of self-assessment
- Mutual assessment

Instructional tactics

- Interactivity
- Variety
- Multi-level glossary
- Modeling
- Scaffolding
- Progressive abstraction
- Dialectical understanding

General software-design criteria

- Match to task definition
- Matching user needs
- Focus on task
- Support requirements
- Software maintainability/extensibility
- Program assessment
- Documentation

General user-interface methods:

- Coherent model of operation
- Navigation
- Context/state cues
- Language/graphics integration
- Scripting/recording

B. DESIGN OUTLINE

CENTRAL LIMIT:

Overt instructional objective: *Demonstrate and explain the convergence of distributions of sample means (as sample size increases) to a normal function whose mean equals that of the population and whose variance is that of the population divided by the sample size.*

Educational agenda (beyond direct instruction):

-- develop understanding (both intuitive and articulate) of random choice

-- develop competence in the use of simulation to investigate and test expectations

-- illustrate cloud of related ideas around the central theme; show linkages

-- provide clear but smooth transitions to higher levels of abstraction

-- illustrate use of graphical modes well-adapted to particular mathematical ideas

-- elicit questions/investigations, rather than avoidance, when uncertainties are encountered

-- promote skill in, and inclination toward, self-assessment

-- point to relevant potential external experiences

-- enable customization and integration into classes by providing user-accessible scripting

-- support "what's under the hood?" inquiries about program mechanisms

-- encourage thoughtful instructional design by discussing decisions made for this program

-- raise user and programmer expectations from educational software

Follow-on instructional objectives:

-- derive theoretical distributions (of sample means as function of sample size) from population description [demonstrate efficiency of theory over exhaustive sampling]

-- investigate the question of "how different is the result from a 'normal' curve?" (How can difference best be measured in this context? When do differences matter?)

List of primary operating modes

Lecture: standard or teacher-supplied "instructional movies" based on user state and goals

Direct: user chooses parameters for specific simulations

Repeat: same parameters, new samples; "fade" previous runs

Comparison: compare sample-mean distributions of populations with same mean and variance

Theory-curve sequence -- illustrate computation of nested integrals, need for efficiency

Hit parade -- precomputed program-run results of particular interest, with discussions

Scientific mode (extensive runs, [potentially] full documentation of results, suppress running display, provide access to program text to clarify exact algorithms used)

<u>Outline of approach for "mutual introductions" between program and user:</u>

Communicate the basic question/problem: "If one takes averages from each of many sets of random samples of some population, how does *the way that the averages vary* depend on [a] the distribution of the population and [b] the number items in each sample?"

State the significance of the question/problem: measurement, estimation, data-based decisions

Describe the program's function: simulate random sampling from distributions of different shapes for various sample sizes

Ask for information on the user's educational state and goals: statistical novice/user/expert? student/teacher/practitioner? estimation/decision-making/hypothesis-critique?

Explain types of user commands:

Action: ***Running the "sampling machine"***

Control Panel: select operating/display mode/parameters

Go button: run the machine and display the results

Scripts: use planned sequences (which external users [especially teachers] may supply)

Context: ***Where am I?***

Tasks: What questions are ready for investigation?

Why: What is the point of this section? (What does it do? Why does that matter?)

Before: What prior user knowledge is being assumed? (with links for review)

After: What further questions does the result raise? What statistical concepts and terms does it illustrate?

Information: ***Looking under the hood.***

Lookup: What does a particular term mean? Where is more information about it?

Pieces: What do the components of this visual element represent?

Machinery: How does this section of the program work?

Design: Why was this section of the program designed this way?

<u>Outline of design particulars:</u>

User inputs:

Population-distribution choices:

Continuous: rectangular, isosceles, right triangle, bimodal (rectangles), normal

Discrete: two values, multiple values

Number of samples in group (to form mean) -- defaults to 16 [logarithmic slider]

Number of groups in set (to show distribution of means) -- defaults to 256 [log slider]

Means update frequency (w/ mean displayed as X [darker for multiple hits]) -- defaults to all

Groups for which sample data is displayed -- defaults to 16 [log slider]

Display of distribution of means:

Compute expected mean and variance

Derive appropriate histogram bins and scale

Scale expanded-domain mean-distribution domain to +/- 4 sigma

Use "blowup" lines to connect population & expanded-domain (histogram/theory) graphs

Use scale markers on theory curves [nest sequence tags at 1:10 scale as curves expand]

Show "theoretical" distribution expected for mean; draw "target" normal; mark differences

Data display: mark 2-5 mm under population abscissa (short vertical lines, darker on overlap)

Means display: mark 8-12 mm under population abscissa (X marks, darker on overlap), also in blowup/histogram

Update histogram of means

Help system:

Discursive expansions (including FAQ lists) on particular items shown

Provide "simultaneous translation" between common words, technical terms, & symbols

Different presentations for different levels of user (different sets of defaults)

Mark pages with difficulty levels and list assumptions about what is already mastered

Links to other sites/topics

Problem-solution section:

Parameter-estimation problems (best estimate based on given data)

Confidence-estimation problems (how likely are various errors?)

Decision problems (what minimizes overall cost? when is more data needed?)

Reference problems (where should a solution be sought?)

Application analysis (how should questions about this problem be posed?)

Background topics:

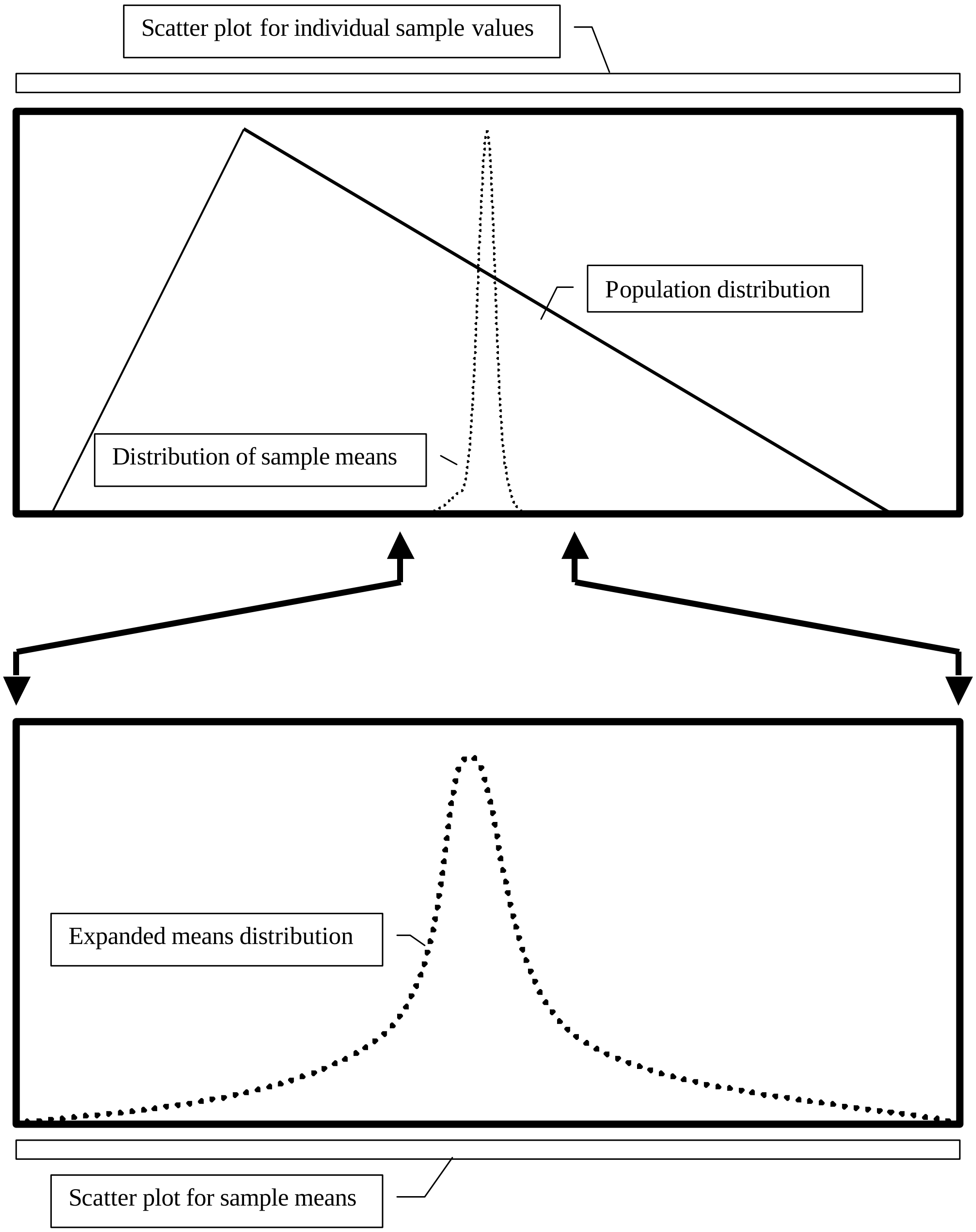
measurement (including non-statistical sources of error); precision/accuracy distinction

sampling; decision theory; graphs (1-D and 2-D); random-number generation

programming, scripting

Tracking methods to assess learning from sequence/timing of user responses

Diagram of basic display layout for Central Limit program



C. *Discussion* of the design in light of the design criteria

Overt instructional objective: *Illustrate the Central Limit Theorem by computer simulation*

This topic was chosen as the focal point for a program because it is one of the most accessible of useful deep mathematical results, providing connections into many areas and levels of mathematics and math-related activities but having only modest mathematical prerequisites. The topic has also been identified by practitioners and teachers of statistics as being an area in which instructional support is needed. It is well suited for the strengths of computer presentation and for use in illustrating the higher-order issues that arise in both mathematics application and instructional delivery.

Educational agenda (beyond direct instruction):

-- develop understandings (intuitive, articulate, and operational) of random choice
Randomness is a key explanatory concept. Sampling provides a non-trivial but clear context in which to acquaint students with random processes.

-- introduce the use of simulation to investigate and test expectations
The use of a method with relatively few knobs to turn (and no "topological" changes permitted) makes sampling an opportune place to introduce people to simulation. Even the simple "sampling engine" simulation used here can be a tool in a significant research program, and an interesting set of such projects will be suggested.

-- illustrate the cloud of related ideas around the central theme; show linkages
The central limit theorem has natural linkages in many directions: basic math, applications, calculus, probability, estimation, decision theory, error analysis, etc. Since these will be presented as optional digressions rather than as "on the test" essential elements, there is greater freedom both for students (who will look only where drawn by interest or need) and for the designer of the included elements (since a topic cannot be "too high" or "too low" if it is authentically connected through appropriate transitions).

-- provide clear but smooth transitions to higher levels of abstraction
There are several natural levels of abstraction for this topic: description of samples, statistical summary of samples (e.g., histograms), description of the distribution of a statistic, expected-value function for a statistic, and asymptotic functional limit. The display techniques and operating modes of this program are designed to both distinguish and link these levels.

-- illustrate use of graphical modes well-adapted to particular mathematical ideas
The primary unusual graphical elements used are one-dimensional gray-scale scatter plots, scale markers, and linkage to a primary graph of an expanded-domain secondary graph which (with appropriate visual cues) varies its horizontal scale factor so as to maintain a constant expected standard deviation. The particular motivations and mechanisms are described individually below.

-- elicit questions/investigations, rather than avoidance, when uncertainties are encountered
The main mechanisms to be used here for this purpose are "sympathetic" questions ("Why might one think that . . . ?") which present uncertainty as evidence of insight and flexibility rather than error and ignorance. Declaration of limits to techniques also contributes to reaching this goal.

-- promote skill in, and inclination toward, self-assessment

One tool used in support of this goal is the initial profile solicited from the user, which is designed to provide a nonjudgmental outline of a variety of possible user capabilities, interests, and purposes -- the idea is to encourage self-knowledge rather than self-judgment (sympathetic-question lists should also be helpful here). The specific self-assessment routines (primarily in "predict what will happen in this case" form) will be nested: an initial form will ask high-level questions about both the prediction and the confidence with which it is made, while secondary forms will either ask for more precise predictions if the initial prediction is correct or suggest a list of appropriate review links otherwise.

-- point to relevant potential external experiences

This will be implemented through a set of selected links to both "world at work" examples and instructional-project lists. This is an opportunity for supporting cooperative extensibility by providing a framework of application areas for which users can suggest good-example links.

-- enable customization and integration into classes by providing user-accessible scripting

The program will support a simple command language that will permit exercise of its functions via scripts, as well as inclusion of teacher-supplied screens and "narration" commentary. A session-trace mode producing a script reflecting the session choices will also be provided.

-- support "what's under the hood?" inquiries about program mechanisms

Users will be able to access nested explanations of the exact mechanisms used to produce program results. These will include verbal, math-symbolic, and programming-language descriptions.

-- encourage thoughtful instructional design by discussing decisions made for this program

This section will share the design discussions (such as this paper) with interested programmers, and will also be "general education" for nonprogrammers as a window into programmer discourse. Feedback from it will also aid in the cumulative improvement of this program.

-- raise user and programmer expectations for educational software

A fully successful program resets the standards of excellence for its sector. This topic is not at all the only (or most important) area that needs instructional software meeting these criteria, and one goal is to promote collateral development on the same lines. Of course, the indication of ultimate success in this task is extensive plagiarism, which the open format encourages.

Follow-on instructional objectives:

-- derive theoretical distributions (of sample means as function of sample size) from population descriptions [show efficiency of theory approach]

This question is approached only after the student is already familiar with the actual pattern of dependence. The focus here is to illustrate the <u>efficiency</u> of theory (compared to nested numerical integration), as well as to show how a theoretical result is derived at several levels of rigor.

-- investigate the question of "how different is the result from a 'normal' curve?" (How can difference best be measured in this context? When do differences matter?)

This seemingly-simple question is the launching point for several meta-mathematical excursions. The higher-level educational goal is to show how open-ended and context-dependent answers to such questions can be.

Operating modes

Lecture: standard or teacher-supplied "instructional movies" based on user state and goals
This will be implemented via the scripting capability. Planned scripts include:

- *[1] "tours" of the site tailored for novices, statistics-course students, and practitioners*
- *[2] "lectures" focused on:*
  - *[a] populations and distributions of numerical population values*
  - *[b] random sampling from populations*
  - *[c] computation of statistics from samples*
    - *[i] measures of typicality: mean, median, midrange, mode*
    - *[ii] measures of dispersion: range, peak-width, std. dev., percentile-based*
    - *[iii] advanced measures: symmetry, normality, moments*
  - *[d] dependence on sample size of distribution of sample means from uniform distribution*
    - *[i] characteristic difference from population mean*
    - *[ii] convergence on normal distribution*
  - *[e] dependence of the means-distribution convergence on population distribution*
    - *[i] illustration of convergence for different populations as sample size increases*
    - *[ii] use of convergence to determine the best dispersion measure for general use*

Direct: user chooses parameters for specific simulations
This mode gives direct access to the simulation parameters. While of some use for exploratory investigation, the main point of this mode is to familiarize the user with the machinery of the sampling simulation.

Repeat: same parameters, new samples; "fade" previous runs
The primary instructional purpose of this mode is to permit users to develop their perception of random processes by seeing which things change and which persist in multiple runs using the same control parameters. New results are overlaid on the previous results, which are displayed in muted colors.

Comparison: compare sample-mean distributions of populations with same mean and variance
This mode is the centerpiece of the program. The distributions of sample means from two differing but comparably-scaled distributions (selected from a menu or entered by the user) are shown to converge as sample size increases. The expanded-domain display is crucial for this demonstration.

Theory-curve sequence -- illustrate computation of nested integrals, need for efficiency
This mode has two purposes. It shows how the distribution of sample means can be directly computed from the population description by numerical integration, and it also illustrates the difficulties of such an approach for large sample sizes. This both connects simulation to direct computation of the "expected value" function and lays the groundwork for demonstrating the advantages of analytic solutions over direct computation.

Hit parade -- precomputed program-run results of particular interest, with discussions
This mode is of use when the emphasis is on results, rather than process, especially for cases where computational time is significant.

Scientific mode (extensive runs, [potentially] full documentation of results, suppress running display, provide access to program text to clarify exact algorithms used)
This mode permits the use of the program to provide input to other analysis tools.

Approach for "mutual introductions" between program and user:

Communicate the basic question/problem:

" If one takes the averages from each of many sets of random samples of a population, how does ***the way that the averages vary*** depend on [a] the distribution of the population and [b] the number of items in each sample? "

This statement of the theme will be elaborated by a list of plausible (but mutually-exclusive) possible answers. While most of these will of course be incorrect, they will be linked to descriptions of situations in which an analogous answer is correct. These descriptions will also include "reasons" (from historical arguments when possible) for each answer, helping communicate that mathematics is subject to argument and development. The correct answer (that the population distribution doesn't matter except for its mean and variance) will be included but not identified, although the convenience that would result if it turned out to be true will be pointed out.

State the significance of the question/problem: measurement, estimation, data-based decisions

These connections provide motivation and context for later examples. One mechanism used will be lists of specific real-world problems for which this question is relevant.

Describe the program's function: simulate random sampling from distributions of different shapes for various sample sizes

This screen also gives an outline of the various modes and subsystems of the program.

Ask for information on the user's educational state and goals: statistical novice/user/expert? student/teacher/practitioner? estimation/decision-making/hypothesis-critique?

This information will be used in setting various program defaults. The form will communicate a diversity of adult states and purposes, so students don't feel "pipelined" into purely academic (as distinct from educational) goals. This is also the point at which the prerequisite math skills are described, with links to review/development.

Types of user commands

The command menu toolbar outlined below is intended to present the capabilities described elsewhere to the user in a suggestive and accessible form. The "context" menu is used to give instructional coherence while avoiding both learning-machine rigidity and free-discovery inefficiency. The absence of a separate "Help" menu is deliberate, since the relegation of explanation and assistance to a separated secondary mechanism is seen as inappropriate for an instructional program.

Action: *Running the "sampling machine"*

Control Panel: select operating/display mode/parameters

Go button: run the machine and display the results

Scripts: use planned sequences (which external users [especially teachers] may supply)

Context: *Where am I?*

Tasks: What questions are ready for investigation?

Why: What is the point of this section? (What does it do? Why does that matter?)

Before: What prior user knowledge is being assumed? (with links for review)

After: What further questions does the result raise? What statistical concepts and terms does it illustrate?

Information: ***Looking under the hood.***

Lookup: What does a particular term mean? Where is more information about it?

Pieces: What do the components of this visual element represent?

Machinery: How does this section of the program work?

Design: Why was this section of the program designed this way?

<u>Design particulars:</u>

User inputs:

These inputs to the simulation engine would be made via standard GUI components, except that the slider positions of number-of-samples designations would be used as exponents rather than coefficients (better reflecting the relative significance of small and large samples).

Population-distribution choices:

Continuous: rectangular, isosceles, right triangle, bimodal (rectangles), normal
Discrete: two values, multiple values (implied linear connection)

Number of samples in group (to form mean) -- defaults to 16 [logarithmic slider]

Number of groups in set (to show distribution of means) -- defaults to 256 [log slider]

Means update frequency (w/ mean displayed as X [darker for multiple hits]) -- defaults to all

Groups for which sample data is displayed -- defaults to 16 [log slider]

Display of distribution of means:

The program will use a primary display region in which graphs of the population(s) and histograms of the samples will be displayed. Graphs and histograms of the distribution of means (and the corresponding limiting normal) will also be shown here, although their main display will be in the domain-expanded display (see description below and the attached diagram). The use of two display regions is essential to enable students to separately master the two effects of increasing sample size: [1] smaller variance in the sample mean, and [2] convergence of the sample means to a normal distribution regardless of the nature of the population distribution.

Compute expected mean and variance

In order to most effectively display the means-distribution graphs, the program will unobtrusively "look in the back of the book" by computing the mean and variance of any populations to be used for sampling. This will be combined with sample size to set appropriate ranges for the histograms and theory graphs, and to permit meaningful comparisons between means from different sample sizes.

Derive appropriate histogram bins and scale

Again, theory will be used to set reasonable display parameters (expected value plus a few standard deviations), including histogram bin sizes such that fewer than half the bins are expected to be empty.

Scale expanded-domain mean-distribution domain to +/- 4 sigma

A secondary display region is used to provide horizontal magnification of the central region of the means-related information from the primary display. The horizontal scale will be set to +/- 4 standard deviations of the expected means distribution, thus automatically adjusting to changes in the distributions and/or the sample size.

Use "blowup" lines to connect population & expanded-domain (histogram/theory) graphs

Visual cues showing the amount of horizontal-scale expansion (and the source region in the primary graph of the secondary one) will be provided to help students maintain the connection.

Use scale markers on theory curves [nest sequence tags at 1:10 scale as curves expand]

The scale of the expansion will be indicated over an arbitrarily-large range by the use of scale markers (usually placed on the theory curves). These small squares will be spaced at tenfold fractions of the population range, with the order of expansion indicated by the digit inside the square (e.g., order-3 markers are spaced at 1/1000 of the population range). Intermediate points are intensified (but not labeled) as needed to have subdivisions at least every 5% of display width. This mechanism permits visual comparisons of scaled normal-approximating curves of very different widths while retaining the scale information.

Show "theoretical" distribution expected for mean; draw "target" normal; mark differences

Much of the interest in the investigations is the contrast of the expected-value graph for the distribution of sample means with the actual result on the one hand and with the large-sample-size convergence limit on the other.

Data display: mark 2-5 mm under population abscissa (short vertical lines, darker on overlap)

Means display: mark 8-12 mm under population abscissa (X marks, darker on overlap), also in blowup/histogram

One-dimension variable-intensity scatter plots:

The two-dimensional graphs typically used for functional representation are not optimal for visualization of sampling processes because their conventions and implicit continuity are too abstract. Even histograms are distractingly complex (e.g., in their sensitivity to bin size and registration), although they are an invaluable bridge between direct and abstract representations.

Scatter plots are better for direct illustration of samples. One or more one-dimensional scatter plots can be integrated with appropriate histograms and theory functions (especially since the ones of interest are all nonnegative) by placing them parallel to and registered with the abscissa. This is planned for both the sample and means data displays in this program.

Because of limits to display and visual spatial resolution, such scatter plots should provide a "darkening" effect when multiple values map to the same display slot. Such effects mean that this same display element naturally serves as a scatter plot at low densities and a color-scale histogram at high densities.

Update histogram of sample means

Help system:

As explained above, the help system is scattered among several menus rather than concentrated. The point is that help information needs to be linked at least as strongly to the instructional routines as to the other help information. A portion of the screen will be reserved as a "comment bar" in which helpful information can be displayed; pop-up notes are another standard GUI mechanism that will be used. Most help information will be supplied only on request, however, to avoid distracting students from working many things out for themselves.

Discursive expansions (including FAQ lists) on particular items shown

Provide "simultaneous translation" between common words, technical terms, & symbols

Different presentations for different levels of user (different sets of defaults)

Mark pages with difficulty levels and list assumptions about what is already mastered

Problem-solution section:

The opportunity to apply skills related to the topic to specific problems will be scattered throughout the program. While these can be thought of as assessment instruments (especially self-assessments), they are actually intended more for motivation and to model application of the techniques and concepts taught.

Parameter-estimation problems (best estimate based on given data)

Confidence-estimation problems (how likely are various errors?)

Decision problems (what minimizes overall cost? when is more data needed?)

Reference problems (where should a solution be sought?)

Application analysis (how should questions about this problem be posed?)

Background topics:

The material on the topics in this section will be straightforward commentary, with little attempt at sophisticated instructional techniques. Links will be provided to more instruction-oriented sites or programs if they become available.

measurement (including non-statistical sources of error); precision/accuracy distinction

sampling; decision theory; graphs (1-D and 2-D); random-number generation

programming, scripting

<u>Feedback support:</u>

Tracking methods to assess learning from sequence/timing of user responses:

At least in the initial stages of program testing (and in expansion to new audiences), user response data will be captured and archived to assist in the refinement of the program.

User commentary:

A user-comment system will send comments to a collection web page, annotated with the user profile and other information.