

{Previous article in MSOR Connections Aug 2001 was here}

---

## Review of Cyberstats

---

**Keith Dear**  
**Australian National**  
**University**

keith.dear@anu.edu.au

***Review system***

Windows NT, Netscape  
Communicator 4.75, internet  
via LAN

Cyberstats consists of a large book published by Duxbury – without page numbers as reviewed, but I estimate perhaps 1200 pages with some sections still to come. The book really consists of an edited printout of the very polished website, which is hosted by CyberGnostics, a one-product company that sells online access to Cyberstats for US\$30 a user.

The website and its book provide a traditional general introductory statistics course, perhaps with a bias towards business rather than science or medicine. The editor in chief for both is Jessica Utts of UC-Davis, and there are 20 contributors from various academic institutions all in the United States.

There is a list of a similar number of institutions where the course is apparently already in use. Each new student signs up for their particular course when first accessing the site, and instructors can apparently customise the website by selecting units for the course content and by providing their own “syllabus”

page with links to other materials of their own choosing. I signed up as an “independent student”, and was not able to explore this side of the product. There is a message board and an integrated email utility, but I was not able to test these either.

I saw pleasantly fast response from the website, though I didn't test this in the small hours of the Australian night when server load in the USA would be heavier. The site has a rather short time-out requiring a fresh login, which I found particularly irritating because the system does not remember where you were but starts again at the homepage.

### Structure

The course consists of some 40 “units” plus several add-on utilities such as a glossary. A brave and I think generally worthwhile attempt has been made to impose a rather rigid editorial structure on the units. Each consists of:

- Introductory material consisting of the “Unit home”, with a one-paragraph summary outlining the concepts and the real-world significance of the topic of the Unit; and a page called “Think first”, about the purpose and usefulness of the Unit's subject;
- Pages on “The Three Keys”: “Basics”, which defines the terms and concepts needed to master the unit; “Uses”, on the techniques and methods pertinent to the unit's objective; and “Warnings”: “because statistics deals with complexity, it is easy to make false assumptions or draw erroneous conclusions, unless one is aware of a more sophisticated level of reasoning”. Each of the three key pages has a “Practice” page;
- Further practice and assessment pages: “Examples”, to provide the integration of the Three Keys, and practice for homework; “Self assessment”: with explanations of why answers are correct or wrong; and “Exercises”: for submission to the instructor.

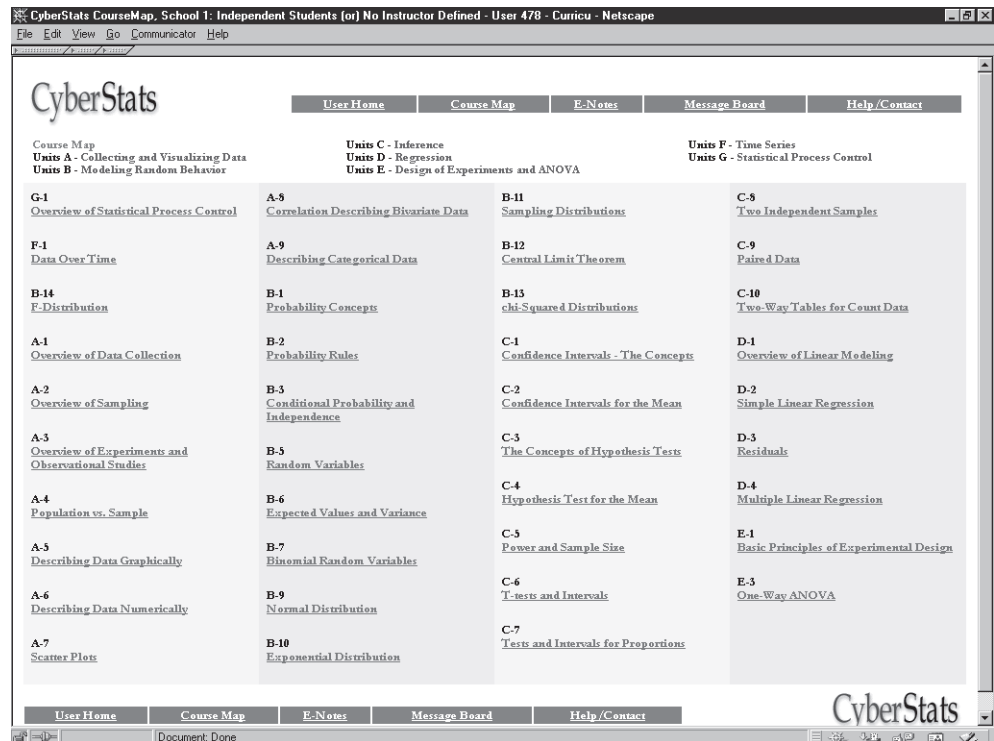


Fig 1 The course map

At the top of every web page is a menu that provides links to each of these elements of the current unit, and in addition links to the other facilities: a glossary, a “resources” page that turned out to contain only two indexes, a calculator with the usual scientific functions (but no statistical functions), and “Data tools”: Java applet implementing a rudimentary statistics package including a reasonable set of graphics quite nicely done. More about these later. Further self-study guidance is provided by pop-up windows listing “Essential terms for this unit” and “Points you need to understand”.

### Course content

The 40 or so units are grouped into seven chapters (though the word is not used), and for some reason they are listed in the online Course Map in a non-obvious order (Fig 1). The course map is the heart of the site. Navigation tools on the page while within a unit are limited to the various unit components and the permanent utilities such as the message board, so the only way to move from one unit to another is via the Map page. This avoids imposing any particular sequence through the course in a way that is not possible in a printed book.

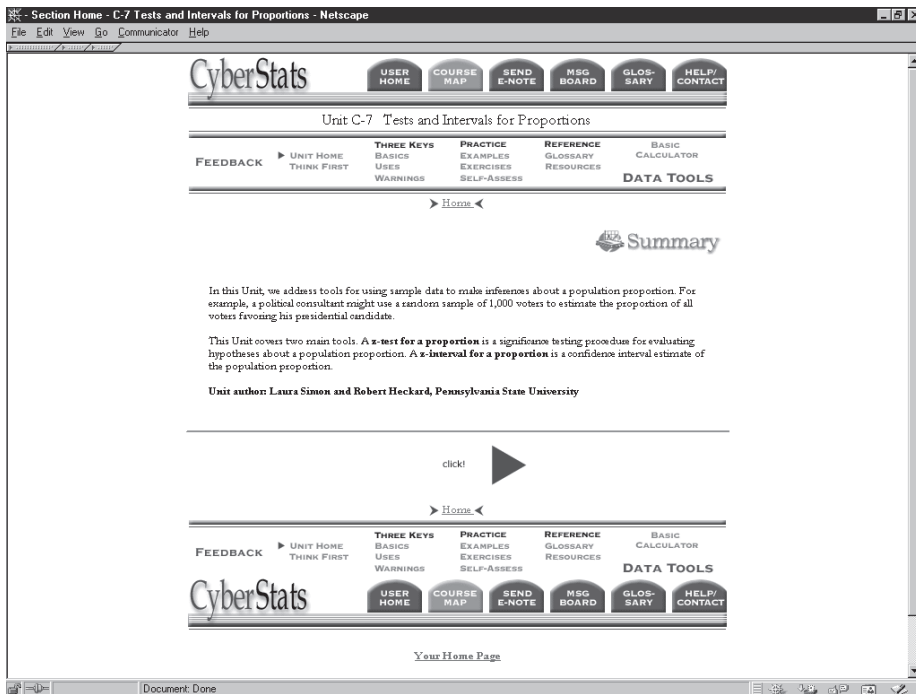


Fig 2 A unit's home page

- Units A - Collecting and Visualizing Data (10 units)
- Units B - Modelling Random Behavior (14 units)
- Units C – Inference (10 units)
- Units D – Regression (4 units)
- Units E - Design of Experiments and ANOVA (3 units)
- Units F - Time Series (1 unit)
- Units G - Statistical Process Control (1 unit)

Not all units exist at present in the book or on the website, and there is some mismatching: for example the website is missing B8: Poisson Distribution but does have a unit B14: F-Distribution that is not mentioned in the book. Some units are foreshadowed in the book but with no material as yet: G1: Statistical Process Control is one such, though there is material on the web. The same goes for F1: Data Over Time. Another is unit E2: Anova, which is represented online, perhaps as an interim measure only, by E3: One-Way Anova. Generally the website is slightly ahead of the book as I received it, although the book lists some web components that are yet to appear.

### Other Facilities

- The glossary is very well done. A unit-specific glossary can be opened in a separate window from within any unit, and provides an index of terms used in that unit. Each term defined is also linked back to where it is first used in any unit. In addition, within the unit text, occasional hyperlinked terms provide pop-up definitions: I should think this might save students a lot of page-thumbing time relative to using a printed text, though it needs to be implemented more completely. Some of the definitions are a bit over-casual: “mean : for a set of numbers, it is the sum of the numbers divided by  $n$ ; also called the ‘average’.” – but what is ‘ $n$ ’?

- A “resources” page, that turned out to contain only two indexes in the online text. It is odd that the index of “interactivities” does not link to them, and the General Index does not link to the referenced pages. The book includes lists of data sites including Statlib and DASL and to a short list of online “discussions of statistical concepts and methods” developed by others. No doubt these will be continually expanded and added to the online version.
- An off-the-shelf calculator with the usual basic scientific functions. I don’t mind that there are no statistical functions – the datatools applet provides a better place for calculations involving data – but the lack of any probability functions is a pity. They would surely be more useful than the trigonometric functions that are provided.
- “Data tools”, a Java applet implementing a rudimentary statistics package including a reasonable set of graphics. DataTools is a version of WebStat software: [www.stat.sc.edu/~west/webstat/](http://www.stat.sc.edu/~west/webstat/). Although I can see uses for such a tool, one can only hope that every instructor would introduce the class to a real statistics package rather early on. Datatools includes a large number of sample datasets, but there is no information about them and they do not seem to be the datasets used in the course.

## Data

There is a serious missed opportunity in that although there is much use of data in the text, there is no systematic attempt to provide the datasets for students to play with. Unit D2: Simple Linear Regression uses as an example Pierce's 1948 data on cricket chirps against temperature, but as far as I could see the only way I could get the data into a statistics package, even the "Datatools" applet provided with the course, would be to retype them from the table on the page. An index of datasets with downloadable text files of the data would be extremely valuable.

Some of the units refer to data held on independent websites – a dangerous practice as the target site can change without warning and without notifying anyone linking to it. For example, unit A5 begins (in its ThinkFirst section) with a discussion of presenting US Census data on state poverty rates – but the link provided is to a Government website that seems to hold quite different data. Fortunately the particular values discussed are reproduced on the page, but the otherwise commendable attempt to validate the data by giving the real source and context has backfired.

## Overall assessment

Overall, I am impressed. Although a fair bit of work remains to be done including adding several entire missing units, this is already a fairly comprehensive online text that makes sense and I think successful use of the technology.

The printed book seems to serve little additional purpose, except to pre-empt the urge to print everything that one sees on a computer screen and may want again. I should think that once over the initial feelings of unfamiliarity and insecurity with online study, most students would have little use for the printed version.

I wonder whether all the authors were entirely comfortable with the unit structure to which they had to conform. The "ThinkFirst" component in particular is novel and sometimes awkward. In each unit, it is placed after a short summary and before the text proper begins. According to the printed book, this component "will cause you to think actively about the purpose and usefulness of the subject of the Unit". This seems to be a lot to ask of the student BEFORE the material has been studied, and some authors seem to have struggled with how to approach it.

One unit where I think it **does** work is A5: Describing Data Graphically. The Summary begins "Graphical

methods for displaying a set of data and comparing two or more sets of data help reveal major features of the data. One can compare... Graphs reveal features such as..." which is spot on. The ThinkFirst component then continues: "When someone dumps a load of numbers on your desk, how can you understand them?" and presents a small dataset (the Census example mentioned above) with questions such as "How would you summarize this set of 51 numbers?". I can see that this might indeed prompt a sufficiently motivated student to start appreciating the need for statistical graphics before plunging into the technicalities of how to construct a stem and leaf plot. But the ThinkFirst is less successful in unit B13: Chi-squared Distributions. Here we find ourselves tossing a coin 100 times and recording the discrepancy from 50 Heads (thus approximating a normal distribution with mean zero, though we are not told this), and then paying a dollar amount equal to the square of this discrepancy. The questions are along the lines of "How much would you expect to pay"? There is no apparent relevance to the unit Summary ("The chi-squared distribution is used in many statistical areas, including hypothesis testing..."), and I should think this beginning would leave a student entirely bewildered as to what the unit was going to be about.

Including a Warnings component in each unit is commendable, and a clever way, I think, in which to add spice to the study of introductory statistics. I picked at random on unit B2: Overview of Sampling. The warnings here are "Be aware that erroneous conclusions can occur when:

1. The sample is drawn from a list that isn't representative of the population.
2. Many of the individuals chosen for the sample cannot be contacted.
3. Many of the individuals contacted do not respond.
4. Volunteers are asked to respond, rather than participants being chosen through a random sample.
5. A convenience or haphazard sample is used rather than a random sample."

Each of these is then explored through an example. This is I think a rather thorough treatment for this level of text, and the whole idea of emphasizing the pitfalls of statistical practice adds a distinctive tone to the book that I liked. Stressing the importance and the difficulty of getting it right is I think a good way of conveying to students that there is substance as well as cookery in Statistics.