

**ZMATH 2011f.00649**

**Sanchez, Juana; He, Yan**

**Internet data analysis for the undergraduate statistics curriculum.**

J. Stat. Educ. 13, No. 3 (2005).

Summary: Statistics textbooks for undergraduates have not caught up with the enormous amount of analysis of Internet data that is taking place these days. Case studies that use Web server log data or Internet network traffic data are rare in undergraduate Statistics education. And yet these data provide numerous examples of skewed and bimodal distributions, of distributions with thick tails that do not follow the usual models studied in class, and many other interesting statistical curiosities. This paper summarizes the results of research in two areas of Internet data analysis: users' web browsing behavior and network performance. We present some of the main questions analyzed in the literature, some unsolved problems, and some typical data analysis methods used. We illustrate the questions and the methods with large data sets. The data sets were obtained from the publicly available pool of data and had to be processed and transformed to make them available for classroom exercises. Students in Introductory Statistics classes as well as Probability and Mathematical Statistics courses have responded to the stories behind these data sets and their analysis very well. The message in the stories can be conveyed at a descriptive or a more advanced level.

*Classification:* K65 K85

*Keywords:* undergraduate study; internet; data analysis; statistics; textbooks; course content; case studies; probability; online searching; search strategies; computer networks; probability distributions