

Chapter 1

Introduction

1.1 What Zelig and R Do

Zelig¹ is an easy-to-use program that can estimate and help interpret the results of an enormous and growing range of statistical models. It literally *is* “everyone’s statistical software” because Zelig’s unified framework incorporates everyone else’s (R) code. We also hope it will *become* “everyone’s statistical software” for applications, and we have designed Zelig so that anyone can use it or add their models to it.

When you are using Zelig, you are also using R, a powerful statistical software language. You do not need to learn R separately, however, since this manual introduces you to R through Zelig, which simplifies R and reduces the amount of programming knowledge you need to get started. Because so many individuals contribute different packages to R (each with their own syntax and documentation), estimating a statistical model can be a frustrating experience. Users need to know which package contains the model, find the modeling command within the package, and refer to the manual page for the model-specific arguments. In contrast, Zelig users can skip these start-up costs and move directly to data analyses. Using Zelig’s unified command syntax, you gain the convenience of a packaged program, without losing any of the power of R’s underlying statistical procedures.

In addition to generalizing R packages and making existing methods easier to use, Zelig includes infrastructure that can improve all existing methods and R programs. Even if you know R, using Zelig greatly simplifies your work. It mimics the popular Clarify program for Stata (and thus the suggestions of King, Tomz, and Wittenberg, 2000) by translating the raw output of existing statistical procedures into quantities that are of direct interest to researchers. Instead of trying to interpret coefficients parameterized for modeling convenience, Zelig makes it easy to compute quantities of real interest: probabilities, predicted values, expected values, first differences, and risk ratios, along with confidence intervals, standard errors, or full posterior (or sampling) densities for all quantities. Zelig extends Clarify by

¹Zelig is named after a Woody Allen movie about a man who had the strange ability to become the physical reflection of anyone he met — Scottish, African-American, Indian, Chinese, thin, obese, medical doctor, Hassidic rabbi, anything — and thus to fit well in any situation.

seamlessly integrating an option for bootstrapping into the simulation of quantities of interest. It also integrates a full suite of nonparametric matching methods as a preprocessing step to improve the performance of any parametric model for causal inference (see MatchIt). For missing data, Zelig accepts multiply imputed datasets created by Amelia (see King, Honaker, Joseph, and Scheve, 2001) and other programs, allowing users to analyze them as if they were a single, fully observed dataset. Zelig outputs replication data sets so that you (and if you wish, anyone else) will always be able to replicate the results of your analyses (see King, 1995). Several powerful Zelig commands also make running multiple analyses and recoding variables simple.

Using R in combination with Zelig has several advantages over commercial statistical software. R and Zelig are part of the open source movement, which is roughly based on the principles of science. That is, anyone who adds functionality to open source software or wishes to redistribute it (legally) must provide the software accompanied by its source free of charge.² If you find a bug in open source software and post a note to the appropriate mailing list, a solution you can use will likely be posted quickly by one of the thousands of people using the program all over the world. Since you can see the source code, you might even be able to fix it yourself. In contrast, if something goes wrong with commercial software, you have to wait for the programmers at the company to fix it (and speaking with them is probably out of the question), and wait for a new version to be released.

We find that Zelig makes students and colleagues more amenable to using R, since the startup costs are lower, and since the manual and software are relatively self-contained. This manual even includes an appendix devoted to the basics of advanced R programming, although you will not need it to run most procedures in Zelig. A large and growing fraction of the world's quantitative methodologists and statisticians are moving to R, and the base of programs available for R is quickly surpassing all alternatives. In addition to built-in functions, R is a complete programming language, which allows you to design new functions to suit your needs. R has the dual advantage that you do not need to understand how to program to use it, but if it turns out that you want to do something more complicated, you do not need to learn another program. In addition, methodologists all over the world add new functions all the time, so if the function you need wasn't there yesterday, it may be available today.

1.2 Getting Help

You may find documentation for Zelig on-line (and hence must be on-line to access it). If you are unable to connect to the Internet, we recommend that you print the pdf version of this document for your reference.

If you are on-line, you may access comprehensive help files for Zelig commands and for each of the models. For example, load the Zelig library and then type at the R prompt:

```
> help.zelig(command)           # For help with all zelig commands.
```

²As specified in the GNU General License v. 2 <http://www.gnu.org/copyleft>.

```
> help.zelig(logit)                                # For help with the logit model.
```

In addition, `help.zelig()` searches the manual pages for R in addition to the Zelig specific pages. On certain rare occasions, the name of the help topic in Zelig and in R are identical. In these cases, `help.zelig()` will return the Zelig help page by default. If you wish to access the R help page, you should use `help(topic)`.

In addition, built-in examples with sample data and plots are available for each model. For example, type `demo(logit)` to view the demo for the logit model. Commented code for each model is available under the examples section of each model reference page.

Please direct inquiries and problems about Zelig to our listserv at zelig@lists.gking.harvard.edu. We suggest you subscribe to this mailing list while learning and using Zelig: go to <http://lists.hmdc.harvard.edu/index.cgi?info=zelig>. (You can choose to receive email in digest form, so that you will never receive more than one message per day.) You can also browse or search our archive of previous messages before posting your query.

1.3 How to Cite Zelig

Please cite Zelig with reference to these two sources:

Kosuke Imai, Gary King, and Olivia Lau. 2007. “Zelig: Everyone’s Statistical Software,” <http://GKing.harvard.edu/zelig>.

Kosuke Imai, Gary King, and Olivia Lau. 2008. “Toward A Common Framework for Statistical Analysis and Development,” *Journal of Computational and Graphical Statistics*, forthcoming, <http://gking.harvard.edu/files/abs/z-abs.shtml>.

To refer to a particular Zelig model, please refer to the “how to cite” portion at the end of each model documentation section.