Programming Statistical Applications in R

Some Project Suggestions

The Rcmdr Package

You should feel free to explore whatever additions or modifications to the **Rcmdr** package you wish to pursue, but here are a few suggestions (in no particular order).

1. There are several places in which the **Rcmdr** accepts a single variable, but where the user might prefer to get parallel output for several variables. For example, consider the dialog box for *Statistics → Summaries → Numerical summaries*. Identify places where it would make sense to allow more than one variable to be selected, but where currently only a single variable is accepted, and make the necessary changes.

2. Is there a general, preferably automatic, way to report the number of observations (and number of missing observations) used to produce **Rcmdr** output? If so, implement it.

3. Explore the possibility of integrating (parts of) the **survival** package, including Cox regression models, with the **Rcmdr**.

4. Long variable lists in the **Rcmdr** are difficult to navigate. It would be nice to be able to move to the next name in the list that begins with a particular letter simply by pressing the corresponding letter-key. Add this capability to the `variableListBox` function (which constructs **Rcmdr** variable lists). While you are working on this function, modify its behaviour so that when a variable list contains only one item, that item is initially selected. (The current default behaviour of `variableListBox()` is for *no* item to be selected initially even when there is just one item in the list.)

5. Add the capability to import data from the clipboard, via *Data → Import Data → from clipboard*.

6. Make provision for the generation of random data, including (multiple?) random samples drawn from the current data set, by adding appropriate items to the *Distributions* and *Data → Active data set* menus.

7. Add a *Demos* menu, and integrate (some of) the demonstrations in the **TeachingDemos** package into the **Rcmdr**. Consider adding other demonstrations that would be useful in an introductory statistics class.

8. Some plot windows require user interaction to identify points. Not exiting from `identify()` causes the **Rcmdr** to appear to freeze. Figure out a better way to notify the user that the **Rcmdr** is waiting for input. Alternatively (or additionally) is there a way to allow `identify()` to "time out" if the wait is too long?

9. Add more of the **Rcmdr** options to the *Tools → Options* dialog.

10. Add dialogs that provide interfaces to `merge()` and `stack()` under the *Data → Active data set* menu. Investigate the possibility of incorporating `unstack()` as well. Consider implementing other enhancements to the **Rcmdr**'s data-management facilities.

The sem Package

You should, likewise feel free to make whatever additions or modifications you wish to the **sem** package. I expect, however, that this would be a less natural project than working with the **Rcmdr** package for two reasons: (1) The statistical programming likely would be much more technical. (2) There are plans to build a new structural-equation modeling package with a front-end in **R** and computations to be handled by an open-source version of **Mx**. If this project goes ahead, it will supersede the current **sem**. Nevertheless, here are a few suggestions:

1. Provide a facility for fitting multiple-group SEMs, with provision for cross-group equality constraints.

2. Provide correct modification indices for `sem` objects. (I have some work on this subject that you might want to make use of.)

3. Create a graphical interface for specifying structural-equation models, perhaps using the **dynamicGraph** package as a starting point.

4. Remove model-fit indices from the `summary` method for `sem` objects, and instead write a `fit.indices` function that provides an expanded range of fit measures.