IR-STAT-PAK

James Blustein ⟨jamie@uwo.ca⟩
Jean Tague-Sutcliffe ⟨tague@uwo.ca⟩

IR-STAT-PAK is a menu-driven suite of statistical tests for analysing and comparing the performance of a variety of queries over a variety of IR systems. To make a selection from a menu: type the corresponding number then press the Enter key. Figure 1 shows the opening menu.

Using data you choose from the entire data set the program can be used to:

- create Recall-Precision plots and histograms of precision data;
- perform two- and three-way ANOVAs;
- create tables.

```
IR-STAT-PAK -- version 1.01

Top level menu
 Statistical analyses
   1.  Descriptive statistics, tables and plots of TREC results
   2.  Comparison statistics with selected data

 Select Data to analyse or display
   3.  Select (or display) participants and queries
   4.  Select measures for analysis or display

 Miscellaneous options
   5.  Data files: reading them and information about them
   6.  Change program options
   7.  About the program

   8.  Quit

Type the number of your selection and press Return
```

Figure 1: The IR-STAT-PAK main menu

The program is designed to work with TREC result files but can be used with any dataset generated by standard IR tests as in TREC. TREC data consists of results of trials of many systems (called participants) on the same 50 queries.

1 Detailed Instructions

The program can only work with data that it reads in from a file in its own format. That format is explained in the file menu. The `extract.sh` and `extract.awk` files can be used together to create files in IR-STAT-PAK format from TREC mailfiles. An example of their use is in the file menu. Once you have data files in the correct format, follow these instructions:

1. Read a datafile into the program. A file can be read from the command line when the program begins and at anytime from the file menu.

   For example, to start the program and read in the file `all-input`, type `irsp all-input`

2. Select the participants and queries to include in the analysis or display you intend to create. Initially, all participants and all queries are selected. The selections you make in menu 3 move participants and queries in an out of the list of selected data. Only the data in the selected list will be used for the analysis or display. The last item on the menu will allow you to select from the main menu.

3. Select the type of data to analyse or display. You must choose either single level or multilevel measure. You may select one or both of the two-way measures: Average precision or R-precision[1]. You may choose only one multilevel measure: precision at fixed recall level $(0, 0.1, \ldots, 0.9, 1)$ or precision at fixed number of documents retrieved $(5, 10, 15, 20, 30, 100, 200, 500, 1000)$. To select the value of the measure to include, select its corresponding menu item. Selected single level measures appear in the menu preceded by an X. The selected levels of multilevel measures appear on the menu immediately beneath the measure's name. The last item on every menu returns you to the previous menu.

4. If you want to create plots or tables, select descriptive statistics from the main menu. You can display tables of precision data or create Postscript figures for display, on an X Window system, or print.

   If you have selected multilevel measures (see 3, above) then Recall-Precision curves will be plotted for all selected participants at all levels of interpolated recall. If you have selected single level measures, then the graph option will produce histograms.

   If you elect to have tables or histograms produced, you should first select the level of averaging to use.

[1] R-precision is the precision after the number of documents relevant to the query have been retrieved.

5. When you are finished using the program, select the last option from the main menu to quit.

2 How it works

There are four main categories of menus (see Figure 1):

1. statistical analysis (menus 1 and 2);
2. data selection (menus 3 and 4);
3. data files (menu 5);
4. further options and documentation (menus 6 and 7).

Before any analysis or printing can take place, data must be read into the program. You can specify a filename to read on the command line or from the file menu. When the program begins, it reads the data file specified on the command line. If no filename is specified then data must be read from the file menu before any processing can occur. Once data has been read in, it can be analysed or used to print tables and graphs.

Before you select the analysis to perform (or tables and graphs to create) you will need to select a subset of the data with which to work. By default: the average precision for all participants and all queries is selected for all tasks. Changes can be made in the data selection menus. If you do not make any selections then the entire data set will be selected. Changes to the participant and query selections are made from menu 3. Changes to the type of data (average precision, R-precision, interpolated precision at fixed recall level, precision at number of retrieved documents) can be made from menu 4.

Once the data have been read in and selected, the program will be ready to perform analysis or create tables and graphs. The choices are available from the first two menus.

3 Some Details

3.1 The ANOVAs

If the data selected for analysis is multi-point (e.g., precision measured at fixed recall levels) then you will have the choice of analysing it as by a series of two-way ANOVAs or one three-way ANOVA. Otherwise, a two-way ANOVA is computed. If the variances differ and the data has not been previously transformed, then the selected transform is applied. Currently, only the arcsine and logit transforms are available. Other such transforms can easily be included in the program code. The ANOVAs produce tables showing the critical values and a display of means in groups that do not differ by more than Tukey's honestly significant difference.

3.2 File format

To use the program you must supply a file of data in the correct format. You can display a specification of that format from the data files menu. You can also use the accompanying `extract.sh` and `extract.awk` scripts to convert TREC data files into the IR-STAT-PAK format. Once you have a file in the correct format, it can be read into the program via the command line or file menu.

3.3 What options are available

The options menu allows you to change the way the program works. You can use it to:

- change the transform to be applied;
- set the condition for the transform to be applied;
- select verbose output;
- change the scrolling behaviour of the output;
- change the maximum width of tables.

By default, the selected transform is applied to change the data if, during an ANOVA calculation, the variances of the participant data are found to differ significantly. You may choose not to have a transform applied or to have it applied before any analysis occurs. With the last option, you can print tables and create figures using transformed data. The program will never transform data that has already been transformed.

By default, only the most important results of the program are displayed. From the options menus you can select to have much of the intermediate results displayed during computation. If you select to have intermediate results displayed then you can see what tests and critical values the program is using.

The program pauses after printing a screen full of text (except in menus). If your screen is larger or smaller than the default 24 lines then you should enter the correct figure from the options menu. If you are capturing output to a file you may want to turn off this behaviour – set the screen height to 0 lines.

The program tries to display tables of descriptive statistics within the width of your display. By default, tables are wrapped at column 80. You may set a different size from the options menu. To see the default sizes of hard coded constants, such as the maximum table width, select 'display program limits' from the 'About the program' menu.

3.4 The Program Code

The program is written in ANSI C. The ABOUT and HOW-TO files in the distribution package document describe how various parts of the program work. ABOUT describes the data structures in the program. HOW-TO provides suggestions for how to make a number of changes to the program, e.g. to include a new transform.