Statistical User Interface for Spreadsheets

Kazunori YAMAGUCHI

Department of Industrial Relations, Rikkyo University, Tokyo 171-8501, Japan, kyamagu@rikkyo.ac.jp

Abstract

Increasingly spreadsheet packages are being used to perform statistical analysis in the workplace and at universities, for both research and education. However, spreadsheet packages have not necessarily been designed for statistical computing. For example, users have to specify range of cells that include data instead of variable names, and handling of data with missing values on the spreadsheets might cause some problems. In this paper, we develop several user interfaces for statistical computing on spreadsheet-type software.

1. Introduction

The recent rapid growth of computer technology has changed statistical education and statistical software. Statistical software might be expected to be used when undergraduate students study statistics. A general goal in the computer industry has been towards making software more user friendly. Recently each software system has a graphical user interface (GUI), but making a common GUI is a very difficult task. If we have to develop software for several platforms, the difficulty might be greater. Nakano (1997) tried to develop statistical software with a user interface which was as independent of the computational engine as possible (also see Liu et al., 1995). Nakano (1997) used a World Wide Web (WWW) interface for this purpose. Merits of the WWW interface have been discussed in Thioulouse and Chevent (1996), West (1997), etc. Yamaguchi et al. (1998) also developed the WWW interface for Microsoft Excel using the ActiveX technology (Farrar 1996), in order to develop a user friendly interface. Takeuchi et al. (2000b) gave some examples of user interface for Excel. Some of them are for experts and the others are for novices. The dialog box for novices is quite simple with a few options and that for experts has many options which provides many facilities. Both interfaces, however, used same statistical engine which is provided as a DLL (Dynamic Link Library) (see Takeuchi et al. (2000a) for statistical DLLs).

Increasingly spreadsheet packages are being used to perform statistical analysis in the workplace and at universities, for both research and education. However, spreadsheet packages have not necessarily been designed for statistical computing. For example, users have to specify range of cells which include data instead of variable names.

Furthermore, McCullough and Wilson (1999) pointed out several problems on the accuracy of statistical procedures in Microsoft Excel 97. Rybolt and McKenzie Jr. (2000) pointed out that Excel were found to have a number of weakness when performing standard statistical computing. McKenzie Jr. and Rybolt (2000) reviewed how the Microsoft handled missing data in Excel 2000 and made some recommendations so that the Microsoft could improve its next release of the Excel with respect to missing data.

In this paper, we introduce several user interfaces for statistical computing on the Excel and discuss what we do for more accurate statistical computing.

2. Statistical User Interface for Excel

Microsoft Excel has 81 statistical functions and the Analysis ToolPak. The Analysis ToolPak is an add-in that adds the statistical computing capabilities to Excel. When we use this add-in for statistical analysis, we have to select range of cells which include data instead of variable names (see Figure 1).

Yamaguchi *et al.* (1998) developed a WWW interface for statistical computing on Excel, on which we could specify variable names (Figure 2).

基本統計量	? ×	
入力元 入力範囲型 データ方向: ⑦ 列(①) ⑦ 行(限) 『 先頭行をラベルとして使用① 出力オジョン ○ 出力先 ②): ○ 新規又は次のワークシート(P) ○ 新規又は次のワークシート(P) ○ 新規ブック(型) 「統計情報(⑤) 平均の信頼区間の出力 (№) 「 K 番目に大きな値(Δ): 「 K 番目に小さな値(M):	OK キャンセル ヘルブ(H)	Principal Component Analysis Input Data Cell Range Variavle Names(in the First Row) Output Options Loadings Scree Plot Scores Principal Component Analysis

Figure 1 : Cell range based Interfaces (Analysis ToolPak on Excel and an example file of Takeuchi et al.(2000))

🔀 Microsoft Excel - bat96.xls					_ 8 ×
2] ファイル(E) 編集(E) 表示(V) 挿入(D) 書式(D) (ソール① グラフ◎ ウィンドウѠ ヘルプ(H)				_8×
🗅 😅 🖶 🔒 🎒 🖪 🖏 🖤 🗈 💼 🗠 • 🤗	。Σ 🏂 🛃 🛍 🕄 💙 MS Ρゴシック	• 11 • B	• <i>I</i> <u>U</u> = = =	🖻 , % 🔅 🗉	- 🔌 - , 🄁 自動
グラフェリア 💌 😑	· · ·				
A B C	DEFG	Н	I J	K L	M
1					
2 40	StatPage : Histgram - Microsoft Internet	Explorer		- O ×	
3 35		入り(A) ツール(T) /	ヘルプ(日)	1	
4 30 -		- <u>-</u>		674 »	
<u> </u>	戻る 進む 中止 :	● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	検索お気に入り	履歴	
7 20	アドレス(D) 🧧 http://www.ir.rikkyo.ac.jp/*	kazunori/stat/Vbscript	nt/histgram.htm 📃	·	
8					
9,15					
10, 10		変数表	示ボタンをクリックし	して、まず	
11 5 -		データ ます。	·ファイル内の変数名? 次に変数を選択して、	を読み込み 、下の作成	
	ファイル内の変数の数	18 ボタン	をクリックします		
13					
15			ヒストグラム	「成	
16	区間の設定方法		3		
17		ポノート リスニューク			
18	O HICTORE C	9(1)21-×	ແລະທະທາບ)		
19	○ 最小値と区	間幅を指定			
20					
21					
22	◎ 区間数を指		J		
23			J		
24				-	
26				🥑 インターネット 👘	
27					
28					
Sheet1 /Sheet2 /Sheet3 Sheet4 /9	6年打撃成績/				
🋐 7 5 – F 🛛 🍪 🗊 💋 🤷 👋 🔤 html	🔁 rikkyo_KK 🚽 🖻 kazunori.doc 🎼	StatPage : 🔣	Microsoft Exc	2 😓 🗖 🌾 🛃 🤣 🕅 🏦	8 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Figure 2 : Variable oriented WWW-Interface for the Excel (Yamaguchi et al. 1998)

Create Boxplots
Input options O Use column of category levels O Values in separate columns
Data Values 打傘
Categories リーヴ
Boxplot appearance
Chart Options
Display sample Connect medians mean lines between boxes
Output Chart Sheet: box2
OK Cancel Help

Figure 3 : Interface of StatPlus 2.0 (Bark & Carey 2000)

Excel has a facility to define the range name. Range names are name given to specific cells or cells ranges. An add-in package StatPlus provides a "Use Range Names" option. As seen in Figure 3, we can use the range names as like as we use variable names.

We are developing variable name oriented interface for Excel now. As we avoid Excel's weakness mentioned in Section 1, we will use external statistical numerical library, for example, DLLs of DLLSA (Takeuchi *et al.* 2000a).

References

Berk, K.N. and Carey, P. (2000). Data Analysis with Microsoft Excel. Duxbury, Pacific Grove, CA.

- Farrar, B. (1996). Special Edition Using ActiveX Indianapolis: QUE Corp.
- Liu, L.M., Chan, K.K., Montgomery, A.L. and Muller, M.E. (1995). A system- independent graphical user interface for statistical software. *Computational Statistics and Data Analysis*, **19**, 23-44.
- McKenzie Jr., J.D. and Rybolt, W. H.(2000). Missing data in Microsoft Excel 2000. *Proc. Comp. Statist.* 2000, *Short Communications and Posters*, 63-64.
- McCullough, B. D. and Wilson, B. (1999) On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics and Data Analysis*, **31**, 27-37.
- Nakano, J. (1997). An internet user interface for statistical software. {¥it Bulletin of the International Statistical Institute, 51st Session Contributed Papers Book 2, 35-36.
- Rybolt, W. H. and Mckenzie Jr., J. D. (2000). The role of data filtering in solving problems in Excel 2000's statistical algorithms, *Proc. Comp. Statist.* 2000, *Short Communications and Posters*, 89-90.
- Takeuchi, A. *et al.*(2000a). Dynamic link library for statistical analysis, *Proc. Comp. Statist.* 2000, *Short Communications and Posters*, 267-268.
- Takeuchi, A. *et al* (2000b). Application of DLLSA for practical and educational use, *Proc the 7th Japan-China Symposium on Statistics* 211-214.
- Thioulouse, J. and Chevent, F. (1996). NetMul, a world-wide web user interface for multivariate analysis software. *Computational Statistics and Data Analysis*, 21, 369-372.
- West, R.W.(1997). Statistical applications for the world wide web. *Bulletin of the International Statistical Institute, 51st Session Proceedings Book* 2, 7-10.
- Yamaguchi, K, Inoue, T., Yadohisa, H and Watanabe, M. (1998). Educational statistical software on the world wide web, *Proc. Comp. Statist.* 98, *Short Communications and Posters*, 127-128.