

INFORMATION FOR PDA APPLICATION DESIGN: CALENDAR ENTRY AND NAME LENGTH STATISTICS

James R. Lewis
International Business Machines Corp.
West Palm Beach, Florida

Personal Digital Assistants (PDAs) are small computing devices that support a user's personal organization activities. Because they are small, PDAs have limited memory and display area. The first study determined the values of calendar entry statistics for computer calendar users. These statistics can help architects of PDA calendar programs design their programs to simultaneously (1) satisfy their users' requirements and (2) take up the least possible amount of memory in the PDA. The second study estimated the distribution of U.S. name lengths. This information is useful for designers of PDA software who need to present names to users in touch-screen buttons. The mean length of names was 13.8 characters with a standard deviation of 2.543 (presented in a last name, first name format). Thus, a touch-screen button that can show 20 characters should show a resident of the United State's complete name approximately 99% of the time.

INTRODUCTION

Computer software designers must meet multiple objectives when they design a program. For example, they must both satisfy their users' requirements and write a program that will work within the constraints of the computing device. The more constrained the device, the greater the challenge. This paper describes two studies designed to obtain information useful to designers of two types of PDA applications: calendar programs and address books.

CALENDAR ENTRY STATISTICS

Personal Digital Assistants (PDAs) are small computing devices that support a user's personal organization activities. One important software application for such activities is a calendar program. The purpose of this study was to measure the characteristics of meeting notices in a computer calendar program. These statistics (number of entries per day, characters per line, lines per entry, and characters per entry) can help the architects of PDA calendar programs design their programs to simultaneously (1) satisfy their users' requirements and (2) take up the least possible amount of memory in the PDA.

Method

Virtually all IBM employees have electronic mail addresses on a mainframe computer. All IBM employees who have an electronic mail address also have an electronic calendar, but they do not necessarily use their electronic calendar unless they are managers. I collected an initial random sample of 40 IBM Boca Raton employees' computer calendar entries for the week ending April 30, 1993. The process of viewing an employee's calendar also indicates whether the employee is a manager or a non-manager. The initial sample contained only four managers, so I collected another eight random samples from IBM managers for the same week. The key word "Requester:" in a calendar entry indicated that the user moved the data in the entry automatically from an electronic meeting notice rather than typing the contents of the entry. The sample of 48 users produced 333 calendar entries. The structure of the resulting database allowed analysis, as a function of manager/non-manager and manual/automatic calendar entry, of:

- Number of calendar entries per day,
- Number of lines per entry,
- Number of characters per line, and
- Number of characters per entry.

Results

Proportion of calendar users who used automatic calendar entry. The electronic mail system allows users to distribute meeting notices that the recipients can automatically send to their calendars. These meeting notices always contain the key word "Requester:". Of the 31 calendar users, 74% had at least one calendar entry that they had automatically placed in their calendar from their electronic mail meeting notice. A 90% binomial confidence interval on this percentage ranged from 58 to 87%. The distribution of this proportion as a function of employment type (manager or non-manager) was not significantly different (Fisher test, $p=.64$).

Number of calendar entries. The average number of calendar entries per day differed as a function of employee type (manager or non-manager) ($F(1,145)=48.1, p<.001$). Table 1 shows the average, the standard deviation, and estimated 95th and 99th percentiles for the number of calendar entries per day as a function of employee type.

Table 1. Number of Daily Calendar Entries as a Function of Employee Type

Employee Type	Standard		Percentile	
	Mean	Deviation	95th	99th
Manager	3.6	2.75	9.1	11.9
Non-Manager	1.2	1.44	4.1	5.5

Entry analyses. The database contained a total of 333 entries, with information about average line length, the number of lines per entry, and the number of characters per entry. Two variables that could influence these entry characteristics were employee type (manager or non-manager) and entry method (manual or automatic). Table 2 shows the results of the entry analyses, divided into two sections (employee type and entry method). If, for a given measurement, a t -test for a variable was significant, then the table includes statistics as a

function of that variable. Otherwise, the table provides overall statistics.

Table 2. Entry Statistics as a Function of Employee Type and Entry Method

<u>Employee Type (Manager or Non-Manager)</u>						
Variable	t -test	Based on	Mean	Standard Deviation	Percentile	
					95th	99th
Line length	$t=2.97$ $p=.003$	Manager	21.0	12.0	45.0	57.0
		Non-mgr.	25.0	11.0	47.0	58.0
Number of Entry Lines	$t=0.62$ $p=.54$	Overall	3.0	2.0	7.0	9.0
Characters Per Entry	$t=1.24$ $p=.22$	Overall	76.0	88.5	253.0	341.5
<u>Entry Method (Automatic or Manual)</u>						
Variable	t -test	Based on	Mean	Standard Deviation	Percentile	
					95th	99th
Line length	$t=8.88$ $p<.0001$	Automatic	29.8	8.2	46.2	54.4
		Manual	18.7	11.4	41.5	52.9
Number of Entry Lines	$t=16.3$ $p<.0001$	Automatic	4.9	2.0	8.9	10.9
		Manual	2.1	1.2	4.5	5.7
Characters Per Entry	$t=15.4$ $p<.0001$	Automatic	154.0	101.0	356.0	457.0
		Manual	39.0	34.0	107.0	141.0

Note: All t -tests had 331 df

Discussion

This information should be useful to programmers who design electronic calendar applications. The results show that employee type (manager or non-manager) and entry method (manual or automatic) affect the characteristics of computer calendar entries. Managers had more calendar entries per day and slightly more characters per line than non-managers. Automatic calendar entries had slightly longer lines and substantially more lines and characters per entry than manual entries.

Both of these results seem reasonable. Due to the demands of their jobs, managers commit more of their time to meetings than non-managers. When

people write meeting reminders for themselves, a brief note is all they need to remember the purpose and other details about the meeting. When writing a meeting notice to send to others, however, it is important to provide all of the necessary information in sufficient quantity to convey the meeting's purpose and details.

This computer calendar had no constraints on entry length. Also, users had full-size displays and keyboards for viewing and entering calendar data. Therefore, the estimates of computer calendar statistics reflect this type of computer use. Personal digital assistants (PDAs) have reduced-size displays and keyboards relative to computer systems, so users who make entries manually should tend to have smaller entries. On the other hand, if the PDA allowed users to update the PDA calendar from a personal computer, the PDA calendar application would need to be able to handle entries similar to manual entries for a computer calendar. Furthermore, if the PDA allowed users to place meeting notices into the calendar automatically, the PDA calendar application would need to be able to handle entries similar to automatic entries for a computer calendar.

Note that the differences between managers' and non-managers' calendar entry statistics do not imply that one should design different PDAs for these different groups. To support both groups, however, the PDA must be able to handle the demands placed upon it by managers. If it can handle this demand, then it can handle the demands of non-managers.

NAME LENGTH STATISTICS

A common type of data to present to a PDA user is a person's name in applications such as address books and telephone quick-dialers. The purpose of this study was to estimate the distribution of lengths of names in the United States. This information is useful for designers of PDA software who need to present names to users in touch-screen buttons, but must work within the size constraints imposed by PDAs.

Method

I collected a random sample of 100 names from the Pompano, FL white pages telephone book. I

organized the names in a database using a "Last name, First Name" format, then calculated the length of each name, including the space and comma between the last and first names.

Results

The mean name length was 13.8 characters, with a standard deviation of 2.543. The standard error of the mean was 0.025. The minimum length was 8 characters, and the maximum length was 18. While not perfectly normal, the distribution was approximately normal (see Figure 1, Kolmogorov-Smirnov goodness of fit: $J_3(12,12)=3, p>.25$). Given a normal distribution, Table 3 shows the percentage of names that would fit in touch-screen buttons that vary from 10 to 22 characters wide. For example, the data show that a button 20 characters wide should accommodate approximately 99% of names in the United States. (In the sample actually taken, all names would fit in an 18-character button.)

Figure 1. Plot of Observed and Expected Normal Values for Percentage of Names Fitting as a Function of Touch-Screen Button Width

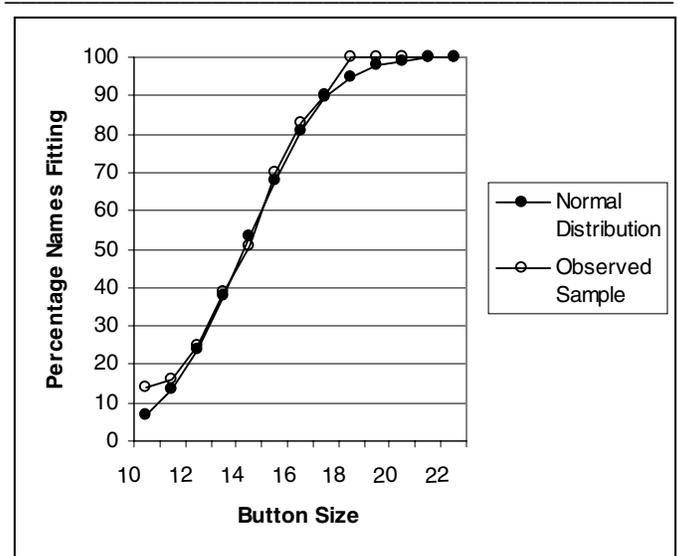


Table 3. Percentage of Names Fitting as a Function of Touch-Screen Button Width

Button Width (in characters)	Z-score	Fitting	Percent Improvement
10	-1.49	6.8	
11	-1.10	13.6	6.8
12	-0.71	23.9	10.3
13	-0.31	37.8	13.9
14	0.08	53.2	15.4
15	0.47	68.1	14.9
16	0.87	80.8	12.7
17	1.26	89.6	8.8
18	1.65	95.0	5.4
19	2.04	97.9	2.9
20	2.43	99.2	1.3
21	2.83	99.8	0.6
22	3.22	99.9	0.1

Discussion

This information is useful for designers of PDA software who need to present names to users in touch-screen buttons, but must work within the size constraints imposed by PDAs. Most PDAs should be able to support button widths of 20 characters, which will show a person's complete name

approximately 99% of the time. If a designer must use buttons smaller than 20 characters, Table 3 will help the designer make the trade-off between button size and the percentage of full names the button will show.

Because the sample consisted of names from only one city in South Florida, it may not be perfectly representative of the United States at large. However, increasing the scope of sampling within the United States would probably not change the estimate of the distribution substantially. The names in some countries, such as Greece, Germany, and Thailand, are usually longer than names in the United States. However, an informal examination of names from a database of IBM employees in Thailand (where some last names are very long relative to typical last names in the United States) showed that a 20-character button would display the entire last name and first initial of the longest name found. In most cases, this should be an adequate amount of information to show to a PDA user in a touch-screen button.

OVERALL DISCUSSION

Two of the most common PDA applications are calendar and address book. The data presented in this paper should help guide the development of PDA applications that can meet the needs of the PDA users and do so within the physical constraints of PDA design.