

Getting More SAS for Your Money

Thomas J. Scott, Ph.D.
Western Illinois University

I. INTRODUCTION

Statistical programs have been heavily used in many computing environments. This paper discusses the relative costs of running SAS[®] software in four different hardware environments. The SAS system was originally produced for IBM 360 and descendant machines. Before 1984, computing center personnel would normally install SAS on its mainframe environment. By the late 1980's, SAS was available for IBM mainframes, VAX/VMS systems, some versions of Unix, Stand-Alone PC systems, PC LANs, such as Novell, and Apple MacIntoshes. Most versions have been "improved" to utilize features of each host. SAS 1990 promotional literature states that SAS is used by over 2,000,000 users, and is installed on 32,704 mainframes, 14,693 minicomputers, 3,622 Unix workstations, and over 500,000 microcomputers. In 1990, SAS was installed on 83.9% of mainframe sites and 43.5% of Digital VAX sites, [SAS90]. The primary goal for this paper is to provide a set of parameters and stated assumptions to allow a computer center staff to decide how to provide the most cost effective SAS statistics capability.

Two workgroup sizes will be used: a small group of 35 simultaneous users, and a large group of 100 simultaneous users. For each of these two workgroups, we discuss four computing hardware environments, including mainframe, minicomputer, stand-alone PCs, and PC LANs.

These eight combinations of user group sizes and hardware environments represent what is typically found in a College or University setting. A matrix of eight rows, with each row representing one of the workgroup/hardware combinations, will be built. Each entry in the final matrix is a Five Year Average Cost, which is calculated by a formula similar to $(\text{InitialCost} + 4 * \text{UpgradeMaintenanceCost}) / 5$. Some factors used to determine the matrix entries are listed in Table 1.

Table 1: Matrix Entries

-
- A. The Cost of the Original Licenses and Annual Software Upgrades.
 - B. The Time and Costs Required to Install and Update Each Product.
 - C. The Pro-Rated User Workstation Costs for each Environment.
 - D. A Pro-Rated Cost of Each Hardware Host, called the Statistics Host Cost.

II. LICENSE & SOFTWARE UPGRADE COSTS

In the pre-PC days, most of the SAS capability was available in the standard distribution. Due to original PC hardware limitations, SAS Institute chose to bundle its packages differently, thus allowing users the choice of an initial base product, and then add-ons, such as graphics, tables, and other less frequently used capabilities. The bundled SAS programs allow the same modules to do the same statistical options for each of the four hardware systems combinations.

In 1991, a computer center manager must decide which capabilities his/her users need, and purchase the appropriate modules. Three assumptions will be used in this paper to determine licensing costs. These assumptions are listed in Table 2.

Table 2: Statistics Purchasing Assumptions

- A. Each Usergroup/machine Combination has the same Capability.
- B. SAS Users have Base SAS and SAS STAT.
- C. The Licenses Chosen for all Options Minimize the Total Five Year Costs.

The Five Year Average Per Year SAS Software License fees, as quoted to degree granting institutions, are listed in Table 3. These figures are derived by adding the initial license fee for Base SAS and SAS Stat, and four years of upgrade licenses, and then dividing the total by 5. The SAS license fees are listed in Table 3.

Table 3: SAS Educational Prices for Minimum 5 Year Average Cost

	Number of Users	
	Medium	Large
IBM Machine	4381	3090-200
5 Yr Avg Cost	\$3,100	\$3,600
VAX Machine	6310	8800-20
5 Yr Avg Cost	\$2,900	\$2,900
PC LAN and StandAlone PC		
Initial Cost	\$1,798	\$2,614

SAS Institute offers other licensing agreements. The listed Five Year Average Costs have been calculated by purchasing an initial license and then an annual renewal.

III. INSTALLATION COST ESTIMATION

The time required for initial installation and checkout of IBM Mainframe and VAX/VMS statistical software will be similar, as both procedures will be done on a

centralized system. The software maintenance upgrades and large data files, such as the US Census Data, will be installed similarly in each of these environments.

The PC LAN environment is somewhat different. Nowadays, most academic institutions have more than one server interconnected. Installing the statistical software and frequently used data sets on several servers is normally recommended to provide both a redundancy and to balance server loads. SAS license agreements allow the installation on as many servers as there are licenses, and network operating systems facilitate this process by giving the systems personnel the option of quickly installing an already installed product on another server. The installation of statistics software is clearly more time consuming in the Stand-Alone PC environment. The 1986 version of three SAS PC products was contained on 24 360K floppy disks, and, when correctly installed, required over 8 megabytes of hard disk space. Good PC systems personnel will correctly install the products and standard datafiles on a test PC, and then create a backup saveset for this installation. If large capacity disks and a good, commercial-grade backup program are both used, the entire SAS distribution could be kept on 3-5 floppy disks. The backup saveset could be installed quickly at each PC. Using the restore option of a standard backup program helps alleviate time consuming disk related problems of an installation on another stand-alone PC or LAN. Besides being faster, the recommended installation procedure also promotes a similar disk structure for all stand-alone PCs.

Including the necessary reading, a good PC systems installer would probably require about two hours for the initial installation and checkout of the two SAS PC modules. It is hard to estimate how

long a subsequent installation from the backup saveset would take, but 30 minutes would not be uncommon, especially if the PC is menu driven, etc. Installation of large Data-Sets, such as US Census Data, would be similar, and would need to be done regularly to keep the data current. The data set installation would probably take an average of 30 minutes per installation.

Conservative estimates, based on actual experience, of the times spent for original installations, maintenance upgrades, and data-set installations are listed in Table 4. These time estimates will be used to determine yearly support costs for statistics programs.

Table 4: Installation and Maintenance Times

Product Type	Install Time	Maintain Time	Data-Set Install Time
IBM Mainframe	7	2.5	.5
VAX/VMS	6	2.0	.5
PC LAN Medium	4	2	1.0
PC LAN Large	5	3	1.5
Stand-Alone PC			
Medium	19	18	17.5
Large	52	51	50.0

The PC LAN times are based on 2 hours for the initial installation and checkout for one server, and then one additional hour for each server. The stand-alone PC times are based on 2 hours initial installation to build the backup saveset, and then 30 minutes for the installation of the backup saveset on each stand-alone PC.

These times can be converted into annual statistics support staff costs, by assuming: 1) The hourly wage for a technician who installs and update system software and data is \$30/hour, and 2) Each five year cycle consists of the initial installation, the original data installation, and then four upgrades and data installations (one each per year). Thus, the Total Five-year Installation Cost is

$[(\text{Year-1-Cost}) + 4 * (\text{Year-2-Cost})] * \30 , and the Average Five-year Installation Cost is the Total Five-year Cost Installation Cost divided by 5. For the Medium PC LAN, these costs are $[(4 + 1) + (4 * 3)] * 30 = \510 and $\$510 / 5 = \102 respectively. Using the data from Table 4, we then have the following table of upgrade costs:

Table 5: Statistics Support Labor Costs

	Total	Average
IBM Mainframe	\$585	\$117
VAX Systems	\$495	\$99
PC Lan Medium	\$510	\$102
PC Lan Large	\$735	\$147
Stand Alone PC		
Medium	\$5,355	\$1,071
Large	\$15,180	\$3,036

IV. USER WORKSTATION COSTS

Workstation costs can vary tremendously. One common workstation setup gives each user a stand-alone PC as a workstation. Because of the many options for stand-alone PC workstations, especially those suited for heavy statistical analysis, workstation costs can be quite variable [SCC91]. If the statistics software host is an IBM mainframe or a VAX/VMS system, some users would have "dumb" terminals, while some users would still have PC's as their terminals. Today's data demands more processor independence, and raw data for statistical analyses come from a variety of sources, some of which are originally hosted on PCs. When the statistics software is hosted on a PC LAN or a stand-alone PC system, serious statistics users would have high-performance, co-processor equipped PC workstations. In University environments, "Stat Labs" with such PC systems are becoming more common.

To make uniform comparisons, assume the following costs for workstations, which are listed in Table 6.

Table 6: Statistics Workstation Assumptions

- A. A "dumb" IBM mainframe terminal costs \$750.
- B. A "dumb" VAX/VMS terminal costs \$500.
- C. A "standard PC workstation" costs \$2,500,
- D. A "LAN PC workstation" costs \$2,800,
- E. A "high-performance PC workstation" costs \$4,000.
- F. A "high-performance PC LAN workstation" costs \$4,000.
- G. Because of other uses for each type workstation, the "statistics" cost of each workstation will be assumed to be 5% of the actual cost.
- H. The annual maintenance for each workstation will be 10% of its "statistics" cost, which is $0.10 * 0.05 * \text{Initial_Workstation_Cost}$.
- I. For the IBM Mainframe and VAX environments, the numbers of "dumb terminals" and "LAN PC Workstations" are assumed to be 17/18, and 50/50 in the two environments.
- J. For the PC and PC LAN environments, the numbers of "Standard PC Workstations" and "High performance PC Workstations" are similarly assumed to be 17/18, and 50/50.

A standard PC workstation consists of an Intel 386 SX-20 processor with a 40 megabyte disk drive, 2 megabytes of RAM memory, VGA video, and a math co-processor. A high-performance PC workstation consists of an Intel 486-25 CPU, 4 megabytes of RAM, an 80 megabyte disk drive, VGA, and a math co-processor. A Standard PC LAN Workstation is a Standard PC Workstation with a LAN card. If a high-performance PC is attached to a LAN, it has a 40 megabyte disk drive, and a LAN adaptor; and thus, for our purposes, costs the same as a stand-alone high performance PC workstation.

The Statistics Workstations Assumptions and cost figures given in Table 6 can be used to calculate workstation costs. First, let ITC denote the Initial Terminal Cost for a given hardware environment with a particular number of users. Then the Total 5 Year Statistics

Workstation Cost is $(ITC + .1 * ITC * 4) * 0.05 = 0.05 * 1.4 * ITC = 0.07 * ITC$. The Average 5 Year Statistics Workstation Cost is this result divided by 5, which is $0.014 * ITC$. Thus, for the Medium VAX Environment, the Average 5 Year Statistics Workstation Cost is $0.014 * (17 * 500 + 18 * 2,800) = 0.014 * (8,500 + 50,400) = \825 . Table 7 lists the Average 5 Year Stat Workstation Costs.

Table 7: Average 5 Year Statistics Workstation Costs

Machine	Medium	Large
Mainframe	\$884	\$2,485
VAX/VMS	\$825	\$2,310
PC LAN	\$1,675	\$4,760
PC	\$1,603	\$4,550

V. STATISTICS HOST COSTS

Because part of the justification of each hardware environment includes the jobs to be run, the initial and Average 5 Year Statistics Hardware Cost for each of the four hardware environments will now be calculated.

Since finding exact prices for different hardware configurations is not an easy task, especially for larger systems, it was decided to use prices for used, but often found computing systems. There is a large Used IBM Mainframe market, as well as a large Used VAX Hardware market. Since LANs are more recent, and LAN component prices have dropped in the last five years, the LAN prices given are for new equipment.

The Medium IBM machine is a 4381-P14 equipped with 32 megabytes of RAM, 10 gigabytes of DASD, and terminal support for 50-100 users. This machine is running VM/CMS and needs VM Sort and VM Secure from VM Center, as fundamental systems software. The software has been purchased through the IBM Computer Consortium at very reduced prices. The Large IBM

Machine is a 3090, Model 200, equipped with 128 megabytes of RAM, 10 gigabytes of DASD, and terminal support for well over 200 users. It is running MVS/TSO, RCAF, and other systems software. This large system initially costs over \$1,750,000 used, and the required annual MVS licenses exceed \$180,000.

The Medium Digital VAX machine is the 6310, and the Large VAX is the 8800-Model 20. Both these machines run the standard VMS operating system, which contains all the necessary systems software. The VAX 6310 has 32 megabytes of RAM and 10 gigabytes of DASD and terminal support for 96 users. The larger VAX 8800 machine has 128 megabytes of RAM, 10 gigabytes of DASD, and terminal support for 200 users.

The PC LAN Servers use Intel 80486 25 Mhz processors, and come with 8 megabytes of RAM, Novell Netware 386, and dual mirrored 320 megabyte DASD. The medium LAN consists of two servers and the large LAN has three interconnected servers. The workstations for all systems have already been described.

To calculate the Statistics Host Cost, a number of assumptions must be made, which are listed in Table 8.

Table 8: Statistics Host Cost Assumptions

1. No additional Stand-Alone PC costs are needed, as these have already been included in the Workstation Costs.
2. The Initial Hardware Host Costs have been determined by using actual budgeted costs for some systems. If no such system was found "in-house", an approximate price was obtained from a reputable used equipment dealer.
3. If exact maintenance costs were known, they were used; otherwise a yearly maintenance cost of 10% of Original Purchase Price was used.
4. All prices were gathered in December, 1990.

5. For all systems, it was assumed that Statistics amounted to only 5% of the system load.
6. The Average Five Year Statistics Host Costs are determined by a calculation similar to ones already used, which is $0.05 * (\text{Initial-Cost} + 4 * \text{Yearly-MaintenanceCost}) / 5$.

Using the assumptions from Table 8, the Average 5 Year Statistics Host Costs can now be calculated. These costs are listed in Table 9.

Table 9: 5 Year Average Statistics Host Costs

Machine	Medium	Large
IBM Mainframe	\$3,237	\$28,343
VAX/VMS	\$3,040	\$5,284
PC LAN	\$714	\$1,071

VI. TOTAL STATISTICS COSTS

Using the data from Tables 3, 5, 7, and 9, the actual costs of supporting SAS in each of the eight usergroup/hardware environments can be calculated. These items will be divided by the number of users, and a resulting Average Five Year Per User Statistics Cost can then be calculated. The results of these calculations are listed in Table 10.

Table 10: Average 5 Year Per User Statistics Cost

Machine Type	Total Cost	Group Size	Average Cost/User
IBM 4381	\$8,570	35	\$246
IBM 3090	\$35,893	100	\$359
VAX 6310	\$8,096	35	\$231
VAX 8800	\$12,441	100	\$124
PC LAN			
MEDIUM	\$6,178	35	\$176
LARGE	\$10,776	100	\$107
STAND-ALONE PC			
MEDIUM	\$6,361	35	\$182
LARGE	\$12,384	100	\$124

VII. EVALUATIONS AND CONCLUSIONS

The numbers in Table 10 allow some fairly easy conclusions to the original question, "How do we get more SAS statistics performance for less money?" For example, the IBM mainframe costs far exceed the other system costs, and the Digital VAX systems perform the same tasks cheaper than IBM systems. Given the assumptions used to build the different tables, the PC LAN is the most cost effective system. The Stand-Alone PC environment allows great freedom, but the costs to manage as many as 100 isolated, independent workstations are greater.

It is true that an IBM 3090-200 is a much more capable machine than a PC. It is hard to compare performance 200 users will experience on the 30 mip 3090 with what one user will feel on a 3 mip 486-25 workstation. Also the larger systems have 10 gigabytes of DASD, while the LAN has 320-640 megabytes. Thus, the 3090 has support for much larger data sets. In truth, most statistics jobs run in University environments can be nicely handled on a PC; but there are always a few really large jobs that need a much larger machine.

Another item is the comparison of the Digital machines with the IBM 3090. The Digital VAX machine performance ratings are 7 Vax Units of Performance (VUPs) for the 6310 and 10 VUPs for the 8800-20. One common view is that one VUP and 8-10 megabytes of RAM provide adequate support for 20 highly interactive users. Thus the 6310 with 64 megabytes of RAM and the 8800-20 can support 80 and 200 users, respectively.

Finally, the PC systems chosen are all quite powerful. The worst performing PC has a 386 SX CPU running at 20 Megahertz with a math co-processor. This machine will yield good performance for over 85% of the statistics tasks it would be given. The faster PC worksta-

tions will adequately handle 98% of the jobs they would normally receive. Larger DASD and even faster Intel based workstations are available to handle larger jobs.

As we move toward the 21st century, mainframes, VAXes, PC LANs, and Stand-Alone PCs will all be faster, cheaper, and have more memory and DASD capability. If the 90's unfold in a manner similar to the 80's, the performance gap that separates PC's and mainframes today will be even narrower. Thus, from currently available information, the PC LAN appears to be the hardware environment that gives the most statistics for the least dollars, and this solution will be even more cost effective in the year 2000 than it is today.

VIII. TRADEMARK NOTICE

*SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA.

IX. BIBLIOGRAPHY

- [SCC91] Calhoun, J. & Scott, Thomas *Estimating the Marginal Costs of PC LAN Workstations*, Proceedings of the 1991 Symposium on Applied Computing, Kansas City, Mo.
- [SAS90] SAS, Institute Inc. *Promotional Literature*.