



# BENCHMARKS

The University of North Texas Computing Center Newsletter

VOLUME 12 NUMBER 4

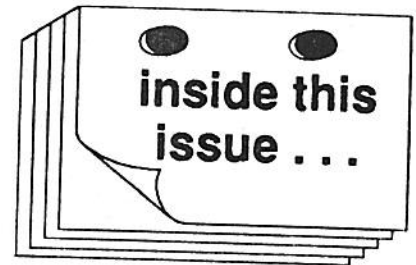
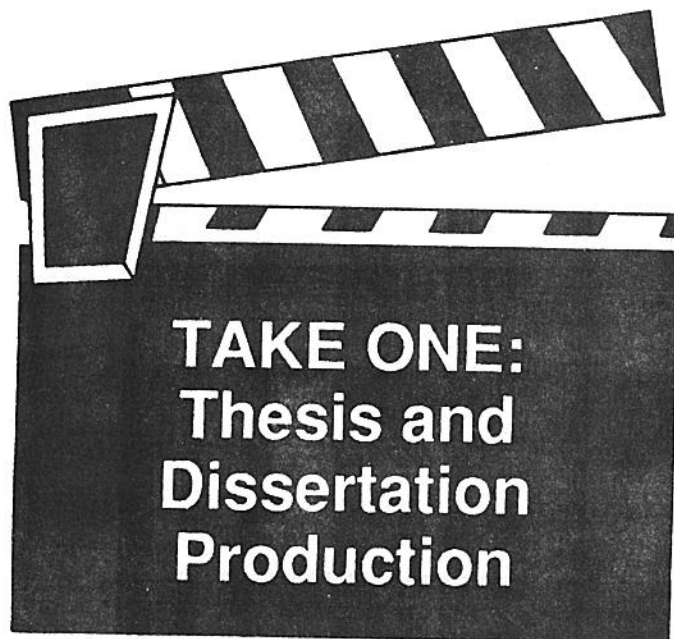
April 1991

## The Theme of This Issue

By Claudia Lynch, *Benchmarks* Editor (BITNET: AS04@UNTVM1)

As you have probably already surmised, the theme of the April 1991 issue of *Benchmarks* is "Thesis and Dissertation Production." This is a very broad topic, and it is by no means covered exhaustively within these pages. We have tried to present various aspects of the process of "doing" a thesis or dissertation. We hope that you will get some good ideas from this issue, whether you are actively engaged in producing a thesis or dissertation or just involved with ongoing scholarly research.

The first article, "Preparing Your Thesis or Dissertation With a Microcomputer," outlines the process one might go through in planning and producing such a document. There are four articles that discuss various sources for available data to aid in the research process. Two more articles discuss statistical tools to aid in data analysis. Rounding out the "Thesis and Dissertation Production" portion of the newsletter is an article from the University of Colorado at Boulder. This article, "The Digital Dissertation: A Better Answer for the Future," was included as a "thought piece." Perhaps you will read it and suddenly find yourself thinking about your research and the research and publication process in a whole new light. We hope so! As always, if you have any questions or comments about the topics presented here, contact Academic Computing Services, ISB 119 (565-2324). ■



<input type="checkbox"/>	<b>THESIS AND DISSERTATION PRODUCTION</b>	
	Preparing Your Thesis or Dissertation With a Microcomputer ...	3
	SPSS 4.1 Available for Testing .....	5
	Using Secondary Data Sources .....	5
	COMPUSTAT II .....	6
	Using IMSL for Research .....	9
	The Digital Dissertation .....	10
	How to Find Out if Your Data are Heteroscedastic .....	10
	1990 CRSP Data Information .....	11
<input type="checkbox"/>	<b>GENERAL INFORMATION</b>	
	ICPSR Summer Programs .....	17
	Report from SUGI 16 .....	71
	STATLIB .....	19
	More NETLIB Software .....	19
	NSF Grants Bulletin Board .....	20
	The BITNET Connection .....	20
	LIST of the Month .....	21
	BENCHMARKS FORUM .....	22
	IRCI News .....	23
	IEBCOPY .....	24
<input type="checkbox"/>	<b>MICROCOMPUTERS</b>	
	Some PC Viruses Described .....	25
	MICRO-TIPS .....	26
<input type="checkbox"/>	<b>VAX/UNIX SYSTEMS</b>	
	VAXCLUSTER Usage Statistics .....	32
	Best of the BBS .....	32
	Compressing VMS Mail Files .....	32
	The UNIX Shell .....	34
	UNIX News .....	34
<input type="checkbox"/>	<b>COMPUTING TECHNICAL SERVICES</b>	
	.....	25

## SERVICES AVAILABLE TO USERS OF THE UNT COMPUTING FACILITIES

The UNT Computing Center is located in the Information Sciences Building (ISB), Room 119. Phone Numbers:

**Computing Center:** (817) 565-2324  
**Help Desk:** (817) 565-4050  
**Micro Support:** (817) 565-2316, 565-2319  
**Graphics Lab:** (817) 565-3479  
**ISB I/O Area:** (817) 565-3890  
**BA I/O Area:** (817) 565-2350

All personnel listed below can be contacted either by calling the Computing Center or by sending them electronic mail on VM/CMS (ID-codes follow each name. All IDs are on BIT-NET node UNTVM1).

**Benchmarks** - Claudia Lynch (AS04)

**Information & ID-Codes; Disk Space Problems, Passwords** - Theresa Russell

**Statistical/Research Support** - George Morrow (AS01), Panu Sittiwong (AC09), Phanit Laosirirat (AC44)

**Academic ADABAS/COM-LETE** - Cathy Hardy (AC55)

**CRSP & COMPUSTAT Problems** - Panu Sittiwong (Panu), Phanit Laosirirat (AC44)

**Student Programming Problems** - CSCI Dept., GAB Room 550; BCIS Dept., BA Room 152

**Problems with JCL, Operating Systems, or Communication/Terminal Problems** - Help Desk

**Data Entry; Test Scoring & Analysis** - Betty Grise

**Administrative Applications** - Coy Hoggard

**Printout Retrieval** - ISB or BA I/O Operators

## DIALING-UP UNT COMPUTERS OVER THE TELEPHONE

Phone numbers for accessing UNT computing systems:

300 - 2400 BAUD: (817) 565-3300  
 300/1200 BAUD: (817) 565-3499  
 300 - 9600 BAUD: (817) 565-3461  
 300 - 2400 BAUD: D/FW METRO 792-4140

Area code 214 must dial 817 before the METRO #.

In your communications program, set Data Bits to 7, Parity to S, and Stop Bits to 1. The dial-up numbers have an autobaud feature that requires you to hit the <RETURN> key repeatedly once connection with the remote modem is made. This is so that the receiving modem can determine the appropriate baud rate. When you have established a communications link, a prompt (# for non-metro numbers, UNTModems> for the metro lines) will appear on your screen and you can enter one of following commands to connect with the system of your choice.

Metro Lines UNTModems>	Non-Metro Lines #	System
N/A	CALL 8040	MUSIC/SP (line editing and PCWS)
Connect VM3270	CALL 3270	Academic Mainframe Full Screen (MUSIC, CMS, Academic COM-LETE)
Connect DEC	CALL DEC	VAXcluster (VMS)
Connect Sol	CALL 900	Solbourne (Unix)
Connect Ponder	CALL 780	Sequent (Ponder)
Connect Library	CALL 3000	UNT Libraries' on-line card catalogue.

## HOURS FOR UNIVERSITY OF NORTH TEXAS COMPUTER ACCESS AREAS : Spring 1991\*

Location	Days	Times
Computing Center RJE	Sunday Monday-Saturday	Noon-Midnight 7 a.m. Mon.-Midnight Sat. (Open 24 hours/day)
ISB 110 Terminal Area	Sunday Monday-Thursday Friday Saturday	1 - 11:50 p.m. 8:00 a.m. - 11:50 8:00 a.m. - 8:50 p.m. 9 a.m. - 8:50 p.m.
College of Business	Sunday Monday-Thursday Friday, Saturday	Noon - Midnight 8:15 a.m. - 11:45 p.m. 8:15 a.m. - 7:45 p.m.
GAB 550	Sunday Monday-Thursday Friday Saturday	2 p.m. - Midnight 8 a.m. - Midnight 8 a.m. - 3 p.m. CLOSED
Graphics Lab	Sunday Monday-Thursday Friday Saturday	1 p.m. - Midnight 8 a.m. - Midnight 8 a.m. - 9 p.m. 9 a.m. - 9 p.m.
Willis Library	Sunday Monday-Thursday Friday Saturday	1 p.m. - Midnight 7:30 a.m. - Midnight 7:30 a.m. - 9 p.m. 9 a.m. - 9 p.m.

\*Hours may vary. Check MUSIC/SP, VM/CMS, VAX, or Solbourne NEWS and/or posted schedules for exceptions.

This issue of Benchmarks was produced by the Documentation Services section of Academic Computing using Xerox Ventura Publisher on a 386SX clone and printed on an Apple LaserWriter IINT. Unless otherwise noted, articles or information may be reproduced for nonprofit purposes provided the publication and issue are fully acknowledged.

# Preparing Your Thesis or Dissertation With a Microcomputer

By Panu Sittiwong, Academic Computing Consultant — "ABD" in Political Science (BITNET: PANU@UNTVM1)

The process of preparing your thesis or dissertation involves, in general, three activities:<sup>1</sup>

- 1 Preparing your thesis or dissertation prospectus.
- 2 Carrying out your research and data analysis.
- 3 Reporting the results and findings.

This article will attempt to show the advantages of using a microcomputer in those activities. "HOW TO" instructions are given when necessary.

## Preparing Your Prospectus

In this stage, unless you are planning a pilot study, you can use a microcomputer for word processing and a bibliographic and references database. Here WordPerfect version 5.1 is an excellent tool. Beyond the basic word processing functions and capabilities, WordPerfect provides a variety of special functions and capabilities. These include tables, graphics, equation composition, automatic table of contents and indices generation, spelling checker, on-line thesaurus, etc. In addition, third party software is available that can be used with WordPerfect to customize your document. One example of such software is called *Thesys*. It is a program written in WordPerfect macros. This program will help you format your document in various styles including APA, Turabien, etc. Another obvious advantage of a word processing program is its ability to reformat the document automatically as you write, rewrite, or rearrange your text.

You also can use WordPerfect as a reference and bibliography compilation tool. With WordPerfect's sort function, your references database can be arranged and organized in various formats as needed. For example, when sorted by author's last name, it then can be used as part of your references in your document.

Another product appropriate to use in organizing your reference materials is D-Base III. The example on page 4 shows a sample entry in a D-Base III file.

## Data Gathering and Analysis

The advantages a microcomputer can provide at this stage of your thesis or dissertation preparation are enormous. Most data that you want to use in your research can be arranged in a "Data Matrix" format where columns define data. For example, my dissertation requires biographical information of more than 100 Canadian Supreme Court variables or measuring items, and rows are subject. A spreadsheet or data-

base type program such as Lotus 1-2-3 or D-base can be used to collect and organize the court justices. I used D-Base III to compile this information. Since D-base III files can be directly used in SPSS/PC+ and SAS/PC, I can easily carry out the data analysis (see "Converting Foreign Files to SAS PC and SPSS PC Data Sets" in the September 1990 issue of *Benchmarks* for more information about converting D-Base and Lotus 1-2-3 to SAS PC or SPSS PC+ file).

Madron, Tate, and Brookshire (1985) point out several ways which microcomputers can be used at this stage of your research. These include using microcomputer software for note-taking and organizing, questionnaire construction, automated survey research, etc.

Microcomputer statistical packages provide varieties of statistical procedures ranging from simple procedures — frequency distributions, cross-tabulations, univariate descriptions, etc. — to advanced procedures — analysis of variance, multiple regression, factor analysis, multi-dimensional scaling, etc. In fact, some packages, SAS/PC and SPSS for the Macintosh and OS/2 provide procedures identical to their mainframe counterparts but with more user friendliness. SPSS/PC+ provides more than 90 percent of the procedures that are available on the mainframe version of SPSS.

There are several advantages in using statistical packages on the microcomputer rather than on the mainframe system.

- 1 First, microcomputer-based packages provide, in general, a more user-friendly environment than the mainframe based package. For example, SPSS provides a command generator which you can use to select and paste commands into your program. This helps reduce the time it takes to learn SPSS syntax,

<sup>1</sup> This article follows the same organization as *Using Microcomputers in Research* by Thomas WM. Madron, C.Neal Tate and Robert G. Brookshire (1985, Beverly Hills, CA: Sage Publications). It does not, however, try to do an extensive review of microcomputer software. The software covered is available to students in the Computing Center public access labs.

## Thesis and Dissertation Production

Author: Hanushek, Eric A. and John E. Jackson  
Title: Statistical Methods for Social Scientists  
Journal: None  
Date: 1977  
Publisher: Academic Press, New York  
Subject: Statistics  
Content1: Regression Analysis: Theory and Application  
Content2: Logit and Probit Analysis  
Content3: Matrix Algebra

### D-Base III Reference File

which in turn allows you to carry out your analyses faster.

- ② Second you don't have spend time learning the microcomputer's operating system, since microcomputer-based statistical packages require little knowledge about such things. It is likely that you will have adequate knowledge of the operating system if you use a microcomputer from the beginning of the project.
- ③ Third and most important is the sharability of data among microcomputer-based software. For example, the output from a statistical program, both text and graphic, can be directly incorporated into a word processing program. This will help reduce redundancy in preparing your document, and save you a lot of time and energy.

Statistical analysis based on microcomputers is not suitable for everyone, however. Microcomputers are limited as to the amount of disk storage and processing power that is available. If you are working with a large database, i.e., you have either a lot of observations or large number of variables, or both, you may not have enough hard disk space to store all your data.

Your analyses also may require a lot of processing power which is not avail-

able on your microcomputer or if available will take a long time processing. For example, a Monte Carlo study or complex simulation, which often needs hours of processing time on a mainframe or minicomputer, will probably run for days on microcomputer.

The Computing Center has SAS/PC, SPSS/PC+ and SPSS for the Macintosh available in most of the microcomputer labs on campus. All of these programs contain advanced statistical procedures. They also have the ability to produce high resolution presentation graphics.

### Reporting the Results and Findings

After data are collected and analysed, it's time to put every thing together. Your word processing program, again, will play a central role. Careful planning can help save your time in both compiling information and presenting it. As they say, word processing will not make bad writer into a good one, but it surely makes a document look good. A completed thesis or dissertation will go through several revisions before it is acceptable by your committee members. Revisions can be done easily with a word processing program. While the text is changed, some formatting such as page and chapter numbering, etc. will be updated. The ability to cut and

paste text allows you to move or re-arrange the document in various ways. It also allows you to re-use information from different sources. For example, a review of literature that is done for the proposal can be re-used again as part of the document.

It is likely that your document will include some graphical presentation of the data. Both SAS/PC GRAPH and SPSS/PC+, used in conjunction with either Harvard Graphics, Microsoft Chart, Chart-Master, Draw Applause, or Graftalk, are capable of generating graphic output files that are recognized by WordPerfect version 5.1. Those formats include HPGL, Postscript, CGM, etc. You can use the SAS commands below to create HPGL output in the file called `graf1.hpg` on the `c:\graf` directory.

```
FILENAME grafout 'c:\graf\graf1.hpg';  
GOPTIONS device=hp7550a noprompt  
rotate=landscape gsfmode=replace  
gsfname=grafout;  
Proc gchart;  
etc.;
```

When you request a graphics procedure in SPSS/PC+, the program will automatically launch the graphical program that is available to you. You will need to follow each packages' method of printing graphs to disk to save the needed output. You need to select the output medium that is capable of generating output in the format you want. For example, you can select any of the HP plotters if you want to produce the HPGL format file.

To include graphs in WordPerfect select the Filename option when invoking Graphics (Alt-F9). WordPerfect will prompt for the filename. After you enter the filename and press <ENTER>, the graph will be loaded and become part of your document.

A good document must be free of any spelling or grammatical errors. Most microcomputer-based word process-

ing programs provide an on-line spelling checker. In addition, they also have an on-line thesaurus which allows you to select an alternate words for your document.

GRAMMATIK is available in the public access labs maintained by the Computing Center. This software will check for grammatical errors and analyze your writing style. Although it is not a fool-proof document checker, you can use it as a primary checker for some obvious errors such as subject verb agreement, inconsistent capitalization, etc.

Microcomputers are powerful research tools. Proper use of the software and hardware can help you produce a good looking thesis or dissertation. If you have any questions about the topics discussed in this article, contact Academic Computing Services (565-2324) and we will be glad to help you. ■

### SPSS 4.1 Available for Testing on OS/MVS

The Computing Center has installed the new release of SPSS on the HDS-8083 Academic Mainframe running under OS/MVS. This release will solve some problems that occurred in SPSS4.0. If you want to execute this new version, you will need to change your EXEC card as follows: // EXEC SPSS41

Your SPSS program will run under the new release without any further modifications.

Please inform Academic Computing Services (565-2324) of any errors you find in this new release. If we do not encounter any difficulties, we will switch over from SPSS 4.0 to SPSS 4.1 during the semester break. ■

## Using Secondary Data Sources for Your Thesis or Dissertation

By Panu Sittiwong, Academic Computing Consultant — "ABD" in Political Science (BITNET: PANU@UNTVMI)

When preparing the prospectus for your thesis or dissertation, you may want to keep in mind some publicly available secondary data sources. You may be able to use these data as supplement to your primary data source or as the main data source for your study.<sup>1</sup>

UNT has vast quantities of machine readable data available to students and faculty members. Many of these data files are maintained by the Computing Center. These data are acquired from several sources including the Inter-university Consortium for Political and Social Research (ICPSR), the Department of Labor, Center for Research in Security Prices (CRSP), Standard and Poors (COMPUSTAT), etc. In addition, through the Internet connection via VAX/VMS, researchers can search and obtain the Louis Harris and Associates data holdings at the Institute of Research in Social Science (IRSS) at the University of North Carolina, Chapel Hill.

□ **CRSP Data:** Once a year UNT receives data from the Center for Research in Security Prices. The data include:

- Monthly NYSE and AMEX Returns and Master file;
- Daily NYSE and AMEX Returns file;
- Daily NASDAQ Returns and Master file;
- Daily, Monthly, Quarterly, and Annual Markets Indices file; and
- National Market System Securities file.

□ **COMPUSTAT II Data:** Similar to the CRSP data, UNT receives updates of COMPUSTAT II data from Standard and Poors once a year. The current holdings include:

- Annual Primary, Supplementary, and Tertiary (Industrial and Research) file;
- Annual Over the Counter file;
- Annual Bank file; and
- Annual Price, Dividends and Earnings file.

In addition, we are in the process of acquiring some quarterly data.

□ **ICPSR Data:** Data from ICPSR constitute the majority of the data archives maintained by the Computing Center at UNT. Currently, there are more than 200 data titles available locally at UNT. As a member of the Consortium, UNT students and faculty members can request any data from ICPSR. Data available from ICPSR cover a wide range of subjects and disciplines including:

- Public Opinion Surveys
- Election Studies of the U.S. and of foreign countries
- Congressional Roll Calls
- General Social Surveys

<sup>1</sup> Related articles on COMPUSTAT II and CRSP follow this article, on pages 6 & 7.

- Health Interview Survey
- Consumer Expenditure Survey
- Government Finance
- World Economic Indicators
- Population Surveys
- Censuses for the U.S. and foreign countries
- EURO-BAROMETER, etc.

You can search and locate ICPSR data holdings available at UNT by issuing the command **FINDICPSR** from either the MUSIC/SP or VM/CMS prompt.<sup>2</sup> If the data that you need are not available at UNT, you can request the data by contacting Panu Sittiwong at ext. 2324, or by sending electronic mail to either PANU on CMS or AC09 on MUSIC. It may take up to 4 weeks before the data can be accessed when it is ordered from ICPSR.

- **Louis Harris & Associates Data:** In addition to the above data, any researcher with access to the Internet can now search the Louis Harris and Associates data holdings at the Institute of Research in Social Science (IRSS) at the University of North Carolina, Chapel Hill. Keywords in combination can be used to locate items of interest. Soon the search will also display frequency distributions for each question retrieved.

IRSS Data Services can be reached through the internet from the UNT VAXcluster. The Internet address for UNCVMI is:

**UNCVMI.ACS.UNC.EDU or 128.109.157.5**

After logging on the VAX, to connect to UNCVMI, type:

**TN3270 UNCVMI.ACS.UNC.EDU or TN3270 128.109.157.5**

When you connect to UNCVMI, you'll see the standard VM logon banner. At the logon line, type:

**IRSS1**  
or  
**IRSS2**

The system will prompt for a password. The password is **IRSS**. You will now be logged on to the Computer at University of North Carolina, Chapel Hill. The procedures should be self-explanatory from this point on. Please note that this data archive is available on a remote computer system. Hence, the Computing Center staff can only assist you on a limited basis.

### Data from Other Sources

In addition to those public data sources mentioned above, you can acquire data from various public and private agencies yourself. The Computing Center can handle data on many mediums. If you plan to acquire a secondary data source yourself, please consult with an Academic Computing Services consultant concerning the method and medium that is most appropriate to use in storing the data when it arrives here at UNT. ■

<sup>2</sup> Currently, before using the **FINDICPSR** command on VM/CMS, you must first link to the ACAD 291 minidisk.

## COMPUSTAT II

By Phanit Laosirirat, Academic Computing  
Consultant — Political Science Doctoral  
Student (BITNET: PHANTIT@UNTVM1)

### Introduction

**T**he COMPUSTAT II data service consists of a computer-readable library of financial, statistical, and market information covering several thousand industrial and nonindustrial companies. There are more than three hundred variables available on the main industrial tape. The COMPUSTAT data service is used as a reference source for several publications. For example, it is the main source of *Business Week* magazine in ranking its most valuable companies each year. More specifically, COMPUSTAT II constantly updates its information on a yearly basis on the largest and most significant New York and American Stock Exchange companies, as well as firms trading over-the-counter. The coverage is extensive. A full range of fundamental data is provided, allowing for variety and flexibility of usage. Data offered includes key income statements, balance sheets, changes in financial position, and market items.

For convenience in accessing the COMPUSTAT data files, all the data have been converted to SAS datasets by Academic Computing Services. As a result, a SAS program is required for retrieving any of these data files (otherwise, FORTRAN or PL/1 programs are needed for the complex data array format). Sample SAS programs will be provided later on in this article.

It is extremely important that users of COMPUSTAT datasets consult with the COMPUSTAT manual for data items and their definitions. The manuals are available in the reference section in the main library. Call Numbers of the manuals as follows:

- COMPUSTAT II and RESEARCH and OTC COMPUSTAT (H 62 .C58483 1989 c.2)
- PDE COMPUSTAT (H62 .P35 1986)
- BANK COMPUSTAT (H62 .B32 1986)

## Data Availability

The COMPUSTAT II data library consists of several files; **INDUSTRIAL FILES**, a **BANK FILE**, and a **PDE FILE**. All files are in annual format with a 20 year span for each company covered.

### Industrial Files

The Industrial files can be subdivided into 5 categories:

- **The Primary Industrial file** (approximately 800 companies) specifically includes all companies in the S&P Industrial Index, some companies in the S&P Utilities Index, the Transportation Index, and the S&P Financial Index, plus companies of greatest interest - primarily companies on the New York Stock Exchange.
- **The Supplementary Industrial file** (approximately 800 companies) contains companies which on the major exchanges but which may have a lesser degree of investor interest.
- **The Tertiary file** (approximately 800 companies) completes the coverage of industrial companies with common stock listed on the New York and American Stock Exchanges. It also includes approximately 300 nonindustrial companies which have been modified for comparability to industrial companies. The nonindustrial companies are from the following areas: Banks, Utilities, Life Insurance, Railroads, Property and Liability, and Real Estate Investment Trusts (REIT). These nonindustrial com-

panies include some of the companies in the S&P Utilities Index, the S&P Transportation Index, and the S&P Financial index.

- **The Over-the-Counter file** (approximately 850 companies) contains those companies traded over-the-counter that command the greatest investor interest, i.e., number of institutional holders, volume traded, price movement, earnings growth, regional and/or economic importance.
- **The Industrial Research file** consists of companies which have been deleted from the COMPUSTAT II Primary, Supplementary, Tertiary, and Over-the-Counter files. Companies may be deleted from active files and placed on the research file for the following reasons:
  - Acquisition or Merger
  - Bankruptcy
  - Leveraged Buyout
  - Liquidation
  - Reverse acquisition (from 1983 forward)
  - No longer fits original file format (from 1978 forward)
  - Now a private company
  - Other (no longer files with SEC, etc.)

### Bank File

This file has approximately 150 companies and contains data on leading United States banking institutions. It is available in annual format with 20 years of data for each bank.

### PDE File

The Price-Dividends-Earnings (PDE) file contains key market information on approximately 7,000 companies and approximately 120 industry indexes and composites. Each data record con-

tains 10 data items. Each item contains 12 months of data.

## Effective COMPUSTAT Usage

Each company in the COMPUSTAT files has several index variables which can be used to group companies together for effective data retrieval. The names of the index variables can be found in the CompuSTAT II manual; Section 8-B, Appendix B and C in the library. The following list designates the various company groupings.

- **Group by File.** In this case the File Identification Code (FILE variable) is the variable index. For instance, one can extract all the Primary file companies from the Total industrial file by writing a SAS program to select the cases that have a FILE variable equal to 01 and 11. Likewise, Tertiary file companies can be extracted by checking for a 03.
- **Group by Trading Exchange.** The Exchange Listing Code (ZLIST variable) can be used to group companies by exchange. Thus all New York Exchange companies can be extracted from the Total Industrial file by checking the Exchange Listing Code for a 1, 2, 3, 4, and 10.
- **Group by Industry.** All companies in a specific SIC industry group can be identified on the tapes by checking the industry field (DNUM variable) for the desired industry group or groups. Thus, all companies in the Metal Mining industry group can be extracted by checking for an industry number of 1000.
- **Major S&P Index Companies.** All companies in the S&P 400 Index are carried on the Primary Industrial file, and there are two methods available for identifying these companies. Either the File Identification Code (FILE variable) can be checked for an 11, or the Exchange Listing Code (ZLIST variable) can be checked for a 1, 11, and 15.

## Thesis and Dissertation Production

After a decision about which group of companies is to be retrieved, the next step is to determine the year or years of data to be used. It is *highly recommended* that users retrieve a subset of companies either by major groupings or by year, instead of retrieving the entire data set because retrieval is more efficient and much faster that way.

### SAS Program Setup

The following statements are the Job Control Language for retrieving the most recent COMPUSTAT data library files.

#### Industrial Files

```
/OSJE
RETURN
//XXXXCOMP JOB (XXXX,;15,2),'YOUR—NAME',CLASS=A,PASSWORD=YYYY
// EXEC SAS
//COMPSTAT DD DSN=USER.A000.COMSTAT2.INDUS90,DISP=SHR
DATA NAME;
SET COMPSTAT.INDUSTRY;
```

#### Research File

```
/OSJE
RETURN
//XXXXCOMP JOB (XXXX,;15,2),'YOUR—NAME',CLASS=A,PASSWORD=YYYY
// EXEC SAS
//COMPSTAT DD DSN=USER.A000.COMSTAT2.RES90,DISP=SHR
DATA NAME;
SET COMPSTAT.RESEARCH;
```

#### OTC File

```
/OSJE
RETURN
//XXXXCOMP JOB (XXXX,;15,2),'YOUR—NAME',CLASS=B,PASSWORD=YYYY
// EXEC SAS
//COMPSTAT DD DSN=USER.AC44.COMSTAT2.OTC90,DISP=SHR
DATA NAME;
SET COMPSTAT.OTC;
```

#### Bank File

```
/OSJE
RETURN
//XXXXCOMP JOB (XXXX,;15,2),'YOUR—NAME',CLASS=B,PASSWORD=YYYY
// EXEC SAS
//COMPSTAT DD DSN=USER.AC44.COMSTAT2.BANK90,DISP=SHR
DATA NAME;
SET COMPSTAT.BANK;
```

#### PDE File

```
/OSJE
RETURN
//XXXXCOMP JOB (XXXX,;15,2),'YOUR—NAME',CLASS=B,PASSWORD=YYYY
// EXEC SAS
//COMPSTAT DD DSN=USER.AC44.COMSTAT2.PDE90,DISP=SHR
DATA NAME;
SET COMPSTAT.PDE;
```

The cut off date for these tapes is 07/31/90. The data are available from 1970 to 1990. The tapes from the previous years are also available. Please contact the Computing Center (565-2324, ISB 119) for more information.

### Examples

The SAS program set-up on the following page is for retrieving COMPUS-TAT data. In this example, it is assumed that you are interested in getting financial information about computer companies. Thus, you would look at the industrial code for these companies in Appendix C in the COMPUSTAT manual. You would find that the industrial codes for these companies span 5 numbers: 3570, 3571, 3572, 3575, and 3577. You also have decided that you want the information only in the year 1988. The financial information that you want consists of the following: Inventories (DATA003), Assets (DATA004), Liabilities (DATA005), Assets/Liabilities and Stockholders' Equity (DATA006), Sales (DATA007), and Fortune Rank (DATA279). The SAS program setup submitted from the MUSIC/SP operating system is shown on the following page.

### The BCIS Department is Really HOT!!

According to the 1991 Spring Edition of *Sourcebook Magazine*, the UNT BCIS Department has one of the five hottest undergraduate IS programs in America. Other schools sharing the honor are California Polytechnic State University at San Luis Obispo, Georgia State University at Atlanta, University of South Florida at Tampa and Bentley College, Waltham, Massachusetts. ■



## Sample Program Set-up for Retrieving COMPUSTAT Data

```

/OSJE
RETURN
//XXXXCOMP JOB (XXXX,15,2),'YOUR—NAME',CLASS=A,PASSWORD=YYYY
// EXEC SAS
//COMPSTAT DD DSN=USER.A000.COMSTAT2.INDUS90,DISP=SHR
/DATAOUT DD DSN=USER.XXXX.NAME,DISP=(NEW,KEEP),UNIT=SYSDA,
//          VOL=SER=ACADnn,SPACE=(TRK,(10,10),RLSE)
TITLE 'SAMPLE SAS PROGRAM TO READ COMPUSTAT DATA';
*
* Step One: Read COMPUSTAT DATA
*
DATA COMPUSTA;
SET COMSTAT.INDUSTRY;
KEEP CNUM CONAME YEAR DATA003 DATA004 DATA005 DATA006
    DATA012 DATA 279; /* Select data items */
IF YEAR = 88; /* Select data for year 1988 */
*
* Step Two: Read Companies' CUSIP number
*
DATA CUSIP;
INPUT @ 30 CNUM $CHAR8.;
CARDS;
/INC COMPANY.CUSIP
:
*
* Sort Both Data Sets Before Merging
*
PROC SORT DATA=COMPUSTA;
BY CNUM;
PROC SORT DATA=CUSIP;
BY CNUM;
*
* Final Step: Create a SAS DATA SET for the company required.
* The data set will be saved to disk for future use.
*
DATA DATAOUT.NAME;
MERGE COMPUSTA CUSIP (IN=A);
IF A; /* Keep only those companies that are needed */
RUN;

```

For the more information about the COMPUSTAT II data, please see the COMPUSTAT handout, "Introduction to COMPUSTAT II," which is available in the Computing Center reception area, ISB 119. ■

## Using IMSL for Research

By George Morrow, Academic Computing  
Consultant (BITNET: AS01@UNTMUSIC)

The IMSL Libraries now consist of three libraries containing computer routines for analysis in Math, Statistics, and Special Functions (SFUN). These routines are called from a program written in VS FORTRAN.

- ① The Math library contains routines for analysis in Linear Systems, Eigensystem Analysis, Interpolation and Approximation, Integration and Differentiation, Differential Equations, Transforms, Nonlinear Equations, Optimization, Basic Matrix/Vector Operations, and Utilities.
- ② The Stat library contains routines for Basic Statistics, Regression, Correlation, Analysis of Variance, Categorical and Discrete Data Analysis, Nonparametric Statistics, Tests of Goodness of Fit and Randomness, Time Series Analysis and Forecasting, Covariance Structures and Factor Analysis, and Discriminant Analysis, Cluster Analysis, Sampling, Survival Analysis, Life Testing and Reliability, Multidimensional Scaling, Density and Hazard Estimation, Line Printer Graphics, Probability Distribution Function and Inverses, Random Number Generation, and Utilities.
- ③ The SFUN Library contains routines on Elementary Functions, Trigonometric and Hyperbolic Functions, Exponential Integrals and Related Functions, Gamma Functions, Error and Related Functions, Bessel Functions, Kelvin Functions, Bessel Functions of Fractional Order, Elliptic Integrals, Weierstrass Elliptic and Related Functions, and Miscellaneous Functions.

The following is an example of a FORTRAN run in which an IMSL routine is called (ZBRENT in this example).

```
//XXXXIMSL JOB (XXXX,1,3), 'Your—Name', PASSWORD=YYYYY
// EXEC FORVCLG
//FORT.SYSIN DD *
  INTEGER MAXFN, NSIG, IER
  DOUBLE PRECISION F, EPS, A, B
  EXTERNAL F
  A=-10.0
  B = 0.0
  EPS = 0.0
  NSIG = 3
  MAXFN = 100
  CALL ZBRENT(F, EPS, NSIG, A, B, MAXFN, IER)
  WRITE (6,99999) B, MAXFN
99999 FORMAT (' THE BEST APPROXIMATION TO THE ZERO OF F IS EQUAL
TO',
& F5.1, '.', '/', ' THE NUMBER OF FUNCTION EVALUATIONS',      NS',
& ' REQUIRED WAS ', I2, '.', '/')
END
  REAL FUNCTION F (X)
  REAL X
  F = X**2 + X - 2.0
  RETURN
END
```

If you have any problems using the IMSL program libraries, contact Academic Computing Services, ISB 119 (565-2324). ■

## The Digital Dissertation: A Better Answer for the Future

By Greg McArthur, Office of Vice Chancellor for Academic Affairs, University of Colorado at Boulder

*This article was reprinted from Buffer, the Newsjournal of Computing at the University of Denver, March 1991. It originally appeared in the May/June 1990 issue of the University of Colorado's Digit, vol. 25, no.3.*

**T**hirteen years ago I received my doctorate in Environmental Biology. My dissertation dealt with disturbed habitat revegetation and made use of computer-assisted methods to keep track of all the various data that impinge upon things environmental.

Unfortunately, that was about the only computer assistance I had to help me produce and then write my dissertation — wordprocessors (indeed, wordprocessing) was not yet widely available. This was 1977.

Continued on page 11.

## How to Find Out if Your Data are Heteroscedastic

By Phanit Laosirirat, Academic Computing Consultant — Political Science Doctoral Student (BITNET: PHANIT@UNTVM1)

**M**any situations arise during the process of data analysis that are less than desirable to the researcher. Heteroscedasticity is a condition encountered when performing regression analysis of data. It occurs when the variance of the residuals or the error terms is not constant. That means it changes from one setting of the independent variables to another.

In a general multiple regression model, each observation can be represented as:

$$Y = \beta_0 + \beta_1 + \beta_2 \bar{x}_2 + \dots + \beta_k \bar{x}_k + \epsilon$$

Where:

Y	is the dependent variable.
X <sub>1</sub> ...X <sub>k</sub>	are the independent variables.
β <sub>0</sub>	is the intercept.
β <sub>1</sub> ...β <sub>k</sub>	are the regression coefficient of the independent variables x <sub>1</sub> ...x <sub>k</sub> respectively.
ε	is the random error or residual.

In order to obtain an unbiased estimate of the *beta(s)* (β), the residuals must satisfy the following assumptions (other non residual-related assumptions are also needed to be met):

- ① Must have a normal distribution with mean equal to zero and variance equal to σ<sup>2</sup> for any given set of values of independent variables;
- ② The residuals are independent or have no evidence of serial correlation.

Continued on page 13.

## 1990 CRSP Data Information

By Panu Sittiwong, Academic Computing Consultant — "ABD" in Political Science (BITNET: PANU@UNTVM1)

The following information applies to CRSP tapes with data ending DEC. 1989. Since the Daily Return and OTC data are used frequently, they are now stored as cataloged data sets on one of the HDS-8083 academic mainframe system disks. The sample program and all sub-routines are also stored on a system disk. In addition, there are pre-compiled modules for all the sub-routines stored on disk. The pre-compiled modules will help speed-up your program if you haven't made any modifications to the sub-routines.

### CRSP Monthly Index and Return File: Data are Stored on Tape

TMS vol. 105410

File #	DSN	RECFM	LRECL	BLKSIZE
1	CRSP.MRET.COPYRIT	FB	80	80
2	CRSP.MRET.BPROGAM	FB	80	4000
3	CRSP.MRET.CPROGAM	FB	80	4000
4	CRSP.MRET.INCLUDE	FB	80	4000
5	CRSP.MRET.ACCESS	FB	80	4000
6	CRSP.MRET.UTILITY	FB	80	4000
7	CRSP.MRET.INC.PDS	FB	80	6160
8	CRSP.MRET.ACC.PDS	FB	80	6160
9	CRSP.MRET.UTILPDS	FB	80	6160
10	CRSP.MRET.OBJLIB	U	0	32760
11	CRSP.MRET.CALENDA	VB	48	32760
12	CRSP.MRET.DATA90	VB	26516	32760

### CRSP Daily Return File: Data are Stored on Disk

Contents	DSN	RECFM	LRECL	BLKSIZE
Sample Program	USER.CRSP.SAMPROG	FB	80	4000
Includes Sub-Routines	USER.CRSP.INCLUDE.PDS	FB	80	4000
Access Sub-Routines	USER.CRSP.ACCESS	FB	80	4000
Utility Sub-Routines	USER.CRSP.UTILITY	FB	80	4000
Sub-Routine Object Library	USER.CRSP.OBJLIB	U	0	32760
Calendar File	USER.CRSP.DRET.CALENDA	VB	48	32760
Data File	USER.CRSP.DRET.DATA90	VB	27680	32760

Continued on page 12

Dissertation continued from page 10.

Standalone wordprocessors were incredibly expensive and using them required a good deal of training. Graduate students, then as now, had very little in the way of money to buy cycles on a machine just to type their dissertation. In fact, most of the computer cognoscente sneered at the idea of using a computer's power to emulate the function of a lowly typewriter.

How things have changed. Wordprocessing is now a full blown and ubiquitous technology and it has made the production of dissertations a much less onerous task.

I realize, of course, that not everyone has access to personal computers or wordprocessing software. But most of us do and I suspect that nearly all graduate students have found a way to get their hands on both.

Now I note another kind of technology being developed that is as radical a change to 1990 dissertation writers as the use of a wordprocessor would have been to me in the late 1970s. It consists of a mixture of hypermedia elements written on CD-ROMs or high-density floppies.

The current representatives of the genre are but the earliest forms of "knowledge navigators" used primarily in the tutorials and instructional materials we have all seen whenever we unpack a new Macintosh or open up the latest version of PageMaker.

This nascent technology will definitely affect the way doctoral candidates assemble and defend their dissertations. I do not know what to call this technology; perhaps it does not yet have a name.

But for now I will refer to it as information processing as distinct from

Continued on page 12

CRSP continued from page 11

## NASDAQ File : Data are Stored on Disk

Contents	DSN	RECFM	LRECL	BLKSIZE
Sample Program	USER.CRSP.SAMPROG	FB	80	4000
Includes Sub-Routine	USER.CRSP.INCLUDE.PDS	FB	80	4000
Access Sub-Routines	USER.CRSP.ACCESS	FB	80	4000
Utility Sub-Routines	USER.CRSP.UTILITY	FB	80	4000
Sub-Routine Object Library	USER.CRSP.OBJLIB	U	0	32760
Calendar File	USER.CRSP.NASQ.CALENDAVB		48	32760
Data File (Part 1)	USER.CRSP.NASQ.DATA90	VB	17244	32760
Data File (Part 2)	USER.CRSP.NASQ.DATA90A	VB	17244	32760

You can use the **TMSINFO** program to find out more information about data sets stored on tape. Please note also that CRSP has changed the format of all tapes. Hence, you may need to get the new sample program sub-routine and source code in order to process the new data. They can be retrieved using the **IEBGENER** program.

### Sample TMSINFO Program

---

```

/INC OSJE
RETURN
//idnnTMS JOB (idnn.:10,1)'Your Name',PASSWORD=mvspw
// EXEC TMSINFO
//SYSIN DD *
VOL=105410
/*
    
```

---

### Sample IEBGENER Program

---

```

/INC OSJE
RETURN
//idnnENER JOB (idnn.:05,1,9999)'Your Name',PASSWORD=mvspw
// EXEC PGM=IEBGENER
//SYSPRINT DD SYSOUT=A
//SYSIN DD DUMMY
//SYSUT2DD SYSOUT=B
//SYSUT1 DD DSN=xxxxx,DISP=SHR
    
```

---

Where:                    **xxxxx** is a data set name.

---

Continued on page 13.

Dissertation continued from page 11.

wordprocessing. I believe it will lead to the creation of the digital dissertation. Such dissertations will be as different from the typewriter-produced versions as those produced on wordprocessors.

Indeed, digital dissertations will be orders of magnitude more sophisticated and have a richer information content than anyone can now imagine.

So, what is a digital dissertation? In its most basic form it is a multi-dimensional "document" linking discrete, yet highly associated pieces of information together such that the entire work is a definitive statement of newly-discovered knowledge.

Standard, written dissertations are essentially the same, with the major exception that they exist in linear form, line after line, page after page. They must be carefully constructed from beginning to end, to bring the reader to the conclusions arrived at in the candidate's original research.

Due to the limited nature of this highly linear process, large chunks of information that do not easily lend themselves to this form of presentation must be omitted. For example, color illustrations, sound, animation, and video segments that might have been part of the original research cannot be placed in current dissertations. Their omission is due to the limitations placed on the standard format required of all doctoral candidates — a written dissertation.

But why not make use of the entire range of information resources at your disposal, rather than limiting yourself to the written word? The technology is here, now, to incorporate all of the other, non-written elements of your research into a dissertation. All it takes is a bit of imagination on the part of the

Continued on page 13.

CRSP continued from page 12.

## Sample Program Set-up

CRSP requires that you write a FORTRAN program to access the data. The sample program and the sub-routines are required in order to access the data. The program, however, is for general purpose use and you must add your own codes in order to obtain particular data items. The following sample shows a job stream for CRSP data processing. If you have further questions about accessing CRSP data, please contact Academic Computing Services, ISB 119 (565-2324).

```

/INC OSJE
RETURN
//idnnENER JOB (idnn.:05,1,9999)'Your Name',PASSWORD=mvspw
// EXE FORVCLG
//FORT.SYSLIB DD DSN=USER.AC09.CRSP89.INCLUDE,DISP=SHR
//SYSIN DD *
//*****
//* FORTRAN Program
//*****
//LKED.SYSLIB DD DSN=USER.CRSP.OBJLIB,DISP=SHR
//          DD DSN=SYS1.VFORTLIB,DISP=SHR
//GO.SYSIN DD *
//GO.FT10F001 DD DSN=xxxxxxx,DISP=SHR
//GO.FT11F001 DD DSN=xxxxxxx,DISP=SHR

```

Where: FT10 and FT11 must be present in the program. FT10F001 refers to the DD for the Calendar file. The DSN is either USER.CRSP.DRET.CALENDA for the Daily Returns data or USER.CRSP.NASDAQ.CALENDA for the NASDAQ file.

FT11F001 refers to the DD for the Data file. The DSN is USER.CRSP.DRET.DATA90 for the Daily Returns data. The NASDAQ data are contained in two separate files. Hence the DD is as follows:

```

//GO.FT11F001 DD DSN=USER.CRSP.NASQ.DATA90,DISP=SHR
//          DD DSN=USER.CRSP.NASQ.DATA90A,DISP=SHR

```

### Heteroscedasticity continued from page 10.

Heteroscedasticity happens when the first assumption is violated. The variance of the residuals varies across observations. For example, the variance of the residuals for observations 1 to 20 is different from variance of the residuals calculated from observations 15 to 40. If the residuals are randomly or normally distributed, their variance will be the same or constant for any given set of observations.

There are many possibilities that render the heteroscedastic contamination. Two are mentioned in this article. For example, the first possibility is when you include travel expenditures of a cross section of households in your model. Heteroscedasticity is a reasonable problem in this situation, since low income families will almost

Continued on page 14.

Dissertation continued from page 12.

doctoral candidate, some software like HyperCard and wordprocessing programs, and access to a video camera, a tape recorder, and a Macintosh.

Let's consider a hypothetical doctoral candidate to show how a digital dissertation would run rings around your standard dissertation to represent the same fundamental research.

Suppose our candidate is interested in the reproductive biology of a species of killer Californian butterflies (*Obstreperous californicus var smogensis*) that are spreading into Southern Colorado.

Our researcher has concluded, after four years of research, that the butterfly is entering Colorado along old Spanish trading routes, but that its expansion throughout the entire state is being limited by a particular bird that happens to enjoy eating this species of butterfly.

Our researcher has discovered that not only is the butterfly's expansion from California to Southern Colorado influenced by activities (e.g., the Spanish trade routes), but a nice natural check has kept the population in balance and prevented a devastating overpopulation of these nasty butterflies in the southwestern portion of Colorado.

Using a video camera, our intrepid candidate has videotaped the breeding behavior of the butterflies, made a detailed visual study of its morphology (particularly the female/male differences), and has good footage showing a mass ascension of over two million individuals in their preferred habitat.

The research also uncovered a cache of old Spanish mission documents describing the movement of the but -

Continued on page 14.

## Thesis and Dissertation Production

Heteroscedasticity continued from page 13.

certainly have low average expenditures on travel and also a low variance of actual travel expenditures. High income families, on the another hand, will spend more on travel which will reflect in both high mean of travel expenditures and greater variance.

Another possibility of heteroscedasticity is when your observations are *Pooled Cross Sectional Time Series or Pooled Time Series*. Social scientists always encounter the problem of too many variables and too small a number of observations. Many researchers have used pooled time series in order to increase the number of observations in their study. As its name implies, the pooled time series is the combination of cross sections and time series. The variables used in the pooled time series are observed from a number of different cross-sections over a time span. For example, cities observed over a 5-year period will look like:

OBS	GROUP	CITY	YEAR	REVENUE PER CAPITA	POPULAR CHANGE	INCOME CHANGE
1	1	ABILENE	1976	7.4106	0.96121	21.5010
2	1	ABILENE	1977	7.3416	0.96121	21.5010
3	1	ABILENE	1978	7.2240	0.96121	24.3554
4	1	ABILENE	1979	7.7365	0.96121	24.3554
5	1	ABILENE	1980	8.1783	0.96121	25.4624
6	2	DALLAS	1976	23.4137	6.5088	19.7725
7	2	DALLAS	1977	19.9819	6.5088	19.7725
8	2	DALLAS	1978	18.7005	6.5088	25.7852
9	2	DALLAS	1979	17.8606	6.5088	25.7852
10	2	DALLAS	1980	16.7092	6.5088	19.9652
11	3	FT.WORTH	1976	13.8924	4.6900	16.3844
12	3	FT.WORTH	1977	13.4445	4.6900	16.3844
13	3	FT.WORTH	1978	13.4426	4.6900	24.1286
14	3	FT.WORTH	1979	12.9661	4.6900	24.1286
15	3	FT.WORTH	1980	11.8687	4.6900	21.3059
16	4	WICHITA	1976	9.5915	-2.7823	24.6832
17	4	WICHITA	1977	9.4000	-2.7823	24.6832
18	4	WICHITA	1978	9.0924	-2.7823	18.4469
19	4	WICHITA	1979	9.4597	-2.7823	18.4469
20	4	WICHITA	1980	9.3594	-2.7823	23.325.

When these kind of data are used in an ordinary regression analysis, the model is likely to be contaminated by heteroscedasticity. Applying multiple regression to the above data, one is sure to get unequal variances of the residuals across the observations. It is obvious for anyone to expect that the error or residual variance for Dallas is larger than the error variance for Abilene. Although, this article is not dealing with alternative data analysis models, it would be clear that ordinary regression analysis is not appropriate for this kind of data set. More complex models are more suitable for these types of data. Those complex models are Pooled Cross Sectional Time Series Regression and Seemingly Unrelated Regression.

Continued on page 15.

Dissertation continued from page 12.

terflies along the trade routes, and even some ancient photos showing the butterflies flitting around Pancho Villa's head outside of Tuscon in the early 1890's.

Finally, our researcher has recorded the sounds of the butterflies' mating call, recorded anecdotal stories from ranchers along the butterflies' migration routes, and captured on tape the rapacious sounds of the principal bird predator dining on thousands of hapless butterflies.

Our doctoral candidate is now ready to commit the results of years of research to close scrutiny in the form of a dissertation. But this time, rather than writing it all down, our candidate chooses to make use of the technology available and place the entire research undertaking into a digital dissertation (with supporting narrative, of course).

Instead of presenting the usual "book" to the committee, they are astounded to receive a 3.5- inch diskette, with the simple instructions, "Please insert this diskette into any Macintosh with at least one megabyte of memory."

What the committee members see, hear, read, and experience as they wander through this dissertation we must leave to conjecture, for each digital dissertation will be unique and put together in a distinctly individual way.

At a minimum, we can guess that the committee members will "read" about the research project, click on various parts of the screen to see the video clips of the butterflies ascending into a clear, blue Arizona sky, look at highly detailed images of the anatomy and reproductive morphology of the butterflies, hear their mating call, and read from excerpted portions of old Spanish mission documents about the

Continued on page 15.

## Heteroscedasticity continued from page 14.

In the example above, the travel expenditures, it is easier to correct for heteroscedasticity than the second example. Ordinary regression can still be used to analyze the data, but only after some of the variables are transformed by using one of the arithmetic functions, for example log, sine, or square root. Depending on the characteristic of the distribution of the residuals, these transformations will stabilize the variance of the residuals. However, remedies for heteroscedasticity are beyond the scope of this article. The focus of this article is to detect heteroscedasticity if it is present.

### Test Statistics

There is a univariate test of homogeneity of variance in SPSS under MANOVA. This test is called the Bartlett test. It is a univariate test because the only variable that we are interested in is the residual. In SAS, there is a short routine that gives the Bartlett test. The only requirement of the Bartlett test in both statistical packages is the index variable. The index variable must be a series of integers. In the sample data set above, the index variable is GROUP. The index number is assigned such that members of the same cross section belong to the same group. The following steps are necessary to obtain the Bartlett test for the residuals.

- 1 Apply OLS regression to your model.
- 2 Extract the residuals from the regression model.
- 3 Apply MANOVA, if you are using SPSS or apply the Bartlett routine in case of SAS to the residuals.

### Examples

Two examples are illustrated in this article, one using SAS and another using SPSS. The samples were processed on SAS PC and SPSS PC+, however these two programs will work fine on both CMS and MVS on the mainframe.

In the first example, on the following page, the residuals from the regression model are saved in a new SAS data set named RCITIES. The name of the residual variable is RESID. PROC SUMMARY uses the RCITIES data set to calculate the variances of the residuals for each city and save them in the sas data set named BART. Two new variables are created in the BART data set; VARIANCE and NUM. VARIANCE is the name of the variable which contains the variance of the residuals for each city. NUM is the number of observations for each city. In the next step, SAS uses the BART data set to calculate the CHI-SQUARE for the BARTLETT test statistics. Output from this program is displayed in the next table.

In this example, the CHI-SQUARE statistic is very significant. This indicates that we can reject the null hypothesis that the variance in all cities are the same. It means that heteroscedasticity is present in the model. A normal regression model is not suitable for this data set. A new model should be used for the data analysis.

The second example shows how to use SPSS PC + to analyze the same data set. Although, this program is processed on SPSS PC +, the same program can be used on the mainframe version.

## Dissertation continued from page 12.

spread of the butterflies during the spring of 1876.

As they continue to explore the rich amount of material supporting the dissertation, the committee will learn as much about the subject as the candidate — an in-depth analysis of this particular research problem seen from a totally new perspective. Understanding of the subject is enhanced; new knowledge is generated and transmitted.

Defending this type of dissertation would be no less difficult than defending a traditional, typed version. The candidate would still be expected to show a mastery of the discipline, complete command of particulars encountered in the research, and have the full support of all of the committee members to be granted a doctoral degree.

But think how much more interesting and rich the entire experience would be for all concerned. The years of effort that went into obtaining all of the pieces of the research puzzle, the video segments, the original documents, etc., would all be part of the finished product.

The spurious example I chose here was from my own field. But the concept works equally well for any discipline, for any research undertaking. There is no limit to what a digital dissertation could cover. As with the research itself, it is limited only by the imagination and creativity of the author.

I look forward to a graduate student creating the first such dissertation. ■

## Example 1: Using SAS to test for Heteroscedasticity

```

0001 LIBNAME SASIN 'C:\';
0002 PROC REG DATA=SASIN.CITIES;
0003 MODEL PCRY = TRUTHYR POCHANGE INCHANGE PCSALES LPCDEBT;
0004 OUTPUT OUT=RCITIES R=RESID;
0005 RUN;
0006 /******
0007 /* BARTLETT ROUTINE STARTS HERE */
0008 /******
0009 PROC SUMMARY NWAY; /* COMPUTE & STORE THE VARIANCE & */
0010 CLASS GROUP; /* NUMBER OF OBSERVATIONS FOR EACH LEVEL. */
0011 VAR RESID;
0012 OUTPUT OUT=BART VAR=VARIANCE N=NUM;
0013 RUN;
0014 DATA _NULL_;
0015 SET BART END=EOF;
0016 LOGVARI=LOG(VARIANCE);
0017 N=NUM-1; /* DEGREES OF FREEDOM FOR THE CURRENT LEVEL */
0018 SLOGVAR+LOGVARI*N;
0019 TOTN+N;
0020 NVAR=N*VARIANCE;
0021 SNVAR+NVAR;
0022 A+1; /* NUMBER OF LEVELS */
0023 SFRACT+1/N;
0024 IF EOF THEN DO;
0025 M=TOTN*LOG(SNVAR/TOTN)-SLOGVAR;
0026 C=1+(1/(3*(A-1)))*(SFRACT-1/TOTN);
0027 CHISQ=M/C;
0028 PROBCHI=PROBCHI(CHISQ,(A-1));
0029 ALPHA=1-PROBCHI;
0030 FILE PRINT;
0031 PUT 'BARTLETT'S TEST: CHI-SQUARE=' CHISQ ' ALPHA=' ALPHA ' ';
0032 END;
0033 RUN;
    
```

### Output From Example 1

```

SAS 10:59 Friday, March 22, 1991 3
BARTLETT'S TEST: CHI-SQUARE=23.652458265 ALPHA=0.0000937633.
    
```

## Example 2: Using SPSS to test for Heteroscedasticity

```

GET FILE='C:\CITIES.SYS'.
REGRESSION /VARIABLES ALL
/DEPENDENT PCRY
/METHOD ENTER TRUTHYR POCHANGE INCHANGE PCSALES LPCDEBT
/SAVE RESID (RESIDUAL).
MANOVA RESIDUAL BY GROUP (1,5)
/PRINT HOMOGENEITY (BARTLETT)
/DESIGN.
    
```

### Output From Example 2

Univariate Homogeneity of Variance Tests			
Variable ..	RESIDUAL	Residual	
	Bartlett-Box	F (4,3750) =	5.94439, P = .000

In Example 2, SPSS adds a new variable to the active file after REGRESSION. The new variable is RESIDUAL. RESIDUAL contains the residuals calculated in the REGRESSION step. MANOVA will use this new variable in the next step to calculate BARTLETT test statistics based on 5 cities. The output from this program is shown in the next table.

The significant level indicates that there is no reason to accept the null hypothesis that the variances in all cities are equal. The conclusion is there is heteroscedastic contamination in the model.

More information about heteroscedasticity can be found in econometric textbooks. Some are mentioned in the references below.

### References

- Fomby, Thomas B., R. Carter Hill, and Stanley R. Johnson. *Advanced Econometric Methods*. New York: Springer-Verlag Inc., 1984.
- Johnston, J. *Econometric Method*. McGraw-Hill, Inc., 1984.
- Maddala, G. S. *Econometrics*. New York: McGraw-Hill, Inc., 1977. ■





## ICPSR Summer Programs

By Panu Sittiwong, Academic Computing Consultant (BITNET: PANU@UNTVM1)

The Inter-university Consortium for Political and Social Research (ICPSR) is sponsoring a series of quantitative methods training workshops and seminars during July 1, 1991 to August 23, 1991. This summer program is divided into two sessions. The first session starts on July 1 and ends on July 26. The second session starts on July 29 and ends on August 23. Each session will have a series of lectures, and workshops. The first session schedule includes:

### Lectures:

- Quantitative Methods of Program Evaluation.
- Basic Mathematics.
- Intro. to Computing.
- The Logic of Data Analysis: Measurement and Design.
- Mathematics for Social Scientists.
- Advanced topics in Social Research.

### Workshops:

- Mathematical Models: Game Theory.
- Likelihood Models and Statistical Inference.
- Latino Research Issues.
- Intro. to Regression Analysis.
- Multivariate Statistical Methods.
- Quantitative Historical Analysis.
- Quantitative Analysis of Crime and Criminal Justice
- Intro. to Statistics and Data Analysis I.
- Regression Analysis.
- Scaling and Dimensional Analysis.

### Special Workshops:

- Survey of Income and Program Participation (SIPP).
- Logit and Log-linear Models.
- Network Analysis.
- Management of Machine-Readable Social Science Information.
- Regression Diagnostics.

The second session schedules include:

### Lectures:

- Intro. to Computing.
- Matrix Algebra.
- Dynamic and Longitudinal Analysis.
- Formal Models of Social Systems.
- Advanced topics in Social Research.

### Workshops:

- Regression Analysis
- Structural Equation (Causal) Models
- Mathematical Models: Rational choice.
- Categorical Data Analysis.
- LISREL Models.
- Time Series Analysis.
- Intro. to Statistics and Data Analysis II.
- Advanced Analysis of Variance.

### Special Workshops:

- LISREL Models: Introduction.
- Utilization of Data Resources from the 1990 Census
- LISREL Models: Intermediate.
- Intergenerational Relations: Research Issues, Methodology and Data.

For more information contact either Dr. Valerie Martinez — Political Science — (817) 565-2276 or Panu Sittiwong — Computing Center — (817) 565-2324. ■

## Report from SUGI 16

By Dr. Philip Baczewski, Academic Mainframe User Services Manager (BITNET: AC12@UNTVM1)

Recently, Panu Sittiwong and I attended the sixteenth annual SAS User's Group International (SUGI) conference, held February 17-20, in New Orleans, Louisiana. With over 4000 people in attendance, there was a lot of opportunity to exchange information with people using SAS for a variety of applications. Attendees included university and corporate users of SAS as well as a very large contingent of staff members from SAS Institute.

One highlight of the conference was the opening session in the New Orleans convention center. With almost all conference attendees present, all of one hall was used for a multi-media program emceed by conference chairman Martin J. Rosenberg of MAJARO Info-Systems, Inc. This included presentations from several key managers from SAS Institute as well as from Dr. James Goodnight, creator of SAS and President and founder of SAS Institute.

### SAS 6.07

Much of Goodnight's talk centered around the next version of SAS, version 6.07, and its implementation of SAS's Multi-Vendor Architecture, in which the greatest portion of the SAS system can run on a number of differing platforms with only a small portion needed to interface to that particular system. Also discussed was the availability of this new version in four stages over the next year.

Version 6.07 is scheduled for first quarter 1991 release for the IBM RS/6000, the DECstation 5000 series, the HP 9000/300 series, MIPS, Apollo, and Data General AViiON workstations. The end of the second quarter should yield this release for Sun-3 and

## General Information

Sun-4 operating systems, Silicon Graphics, HP 9000/800 series, and the IBM AIX/370 operating system. Version 6.07 for IBM MVS and CMS, as well as VAX VMS is scheduled to ship by the end of third quarter, 1991, and the fourth family of version 6.07, due by end of fourth quarter, includes IBM VSE, OS/2, Windows 3.0, AIX PS/2, SCO/UNIX for 386 and 486 machines, and the Sequent and NeXT workstations.

One exciting component of version 6.07 is the utilization of the native Graphical User Interfaces (GUI) of the Unix, OS/2, and Windows 3.0 operating systems to enhance the usability of the SAS system. Goodnight mentioned that many of the future developments in the SAS system will be geared to users of GUIs. Along these lines, he announced that version 6.04 will be the last version of SAS for the PC/MS-DOS platform, with the OS/2 and Windows versions updated as new releases are developed. He stated that version 6.04 will be maintained for several more years, but recommended the move to a GUI.

Goodnight went on to announce two new products entering the SAS line. SAS/CALC, described by Goodnight as "the ultimate spreadsheet," is scheduled to ship with release 6.07 for MVS, CMS, and VMS. Also announced was SAS/EIS, "an object-oriented, applications development interface to the SAS Applications System." This product was designed for development of specially tailored interfaces to the SAS system, especially Executive Information Systems.

Goodnight's speech included many exciting developments for SAS, and the end of the opening session of SUGI was quite exciting as well. Even though the conference was held a week after Mardi Gras, SAS arranged to stage a mini-Mardi Gras parade right there in the convention center. This included motorcycle policemen, several bands, floats, and of course, a lot of beads

being thrown from those floats. SAS even provided Mardi Gras masks to each attendee to help get into the spirit of the evening.

In the three days following the opening, there was the opportunity to attend a variety of paper sessions covering many aspects of the SAS system. There was also an extensive demo area which allowed hands-on experience with the version 6.07 on a number of different platforms. I even had the chance to try out SAS on the NeXT workstation. (My impression was that it was a bit slow on the 68030-based "cube," and will probably be much more responsive on a 68040 NeXTStation or on an upgraded cube.) Also in the demo area were problem desks manned by SAS Institute staffers. I had the opportunity to talk at length

they're know in SUGI lingo), as well as in several papers. Several points quickly became evident. Version 6.06 was a complete re-write of SAS for MVS and CMS, switching from PL/I to C as the development language. As a result, some aspects of this version yield poorer performance than the previous version 5.18, and in some cases cause compatibility problems with version 5.18. Many of these problems are alleviated in version 6.07, but for the mean time, the BOFs and paper sessions yielded some valuable tips for making SAS perform better on both MVS and CMS, some of which we can apply here at UNT to the test versions of SAS 6.06 on both CMS and MVS.

Panu attended a BOF on the Windows 3.0 version of SAS. He also had the opportunity to attend the SAS-L BIT-

---

*If you are a Windows 3.0 user, you can participate in evaluating the [SAS] product. Call Panu Sittiwong at 565-2324 and ask to have your name put on the list.*

---

with Kieth Collins, the manager of product development for the CMS platform, and express to him our concerns here at UNT related to SAS on CMS.

Another result of the demo area was that we were able to become a Beta-Test site for SAS 6.07 under Windows 3.0., and expect to receive the software some time during the summer semester. If you are a Windows 3.0 user, you can participate in evaluating the product. Call Panu Sittiwong at 565-2324 and ask to have your name put on the list. You will need a 386SX or more powerful PC with a minimum of 6M of memory in order to effectively run the SAS software under Windows.

One topic of interest throughout the conference was the performance of SAS version 6.06 on the MVS and CMS platforms. This was discussed in several "Birds-of-a-Feather" special-interest-group sessions (or BOFs, as

NET list BOF, and came away with a hot-pink official SAS-L SUGI 16 T-shirt. This was emblazoned with the unofficial "slogan" of SAS-L, RTFM, interpreted by one SAS staffer in a later session as "Read The Friendly Manual." The SAS-L BOF was a chance to associate faces with the names of those posting messages to that mailing list.

All-in-all, it was a memorable conference that yielded much valuable information. SAS is aggressively marketing its system for solutions in the corporate world, while at the same time continuing to provide its powerful statistical analysis capabilities. Of course, another reason this conference will linger in my memory may have something to do with its location: the Crescent City; ships steaming up and down the Mississippi; the St. Charles street car; the French Quarter; beignets and coffee at *Cafe Du Monde*. ■

## STATLIB: Public Domain Statistical Software

*This article was adapted from an article by the same name in the February 1991 issue of the University of Minnesota Academic Computing Services Newsletter, acs.*

**S**TATLIB is an electronic mail system to distribute public-domain statistical software and data. STATLIB is a minor modification of the NETLIB software [see related article on this page and/or consult *Benchmarks*, February 1991 (page 9) and November/December 1990 (pages 16-19) for information on NETLIB]. All of the STATLIB routines come in source code form and require the appropriate compiler (usually FORTRAN). Keep in mind that though the routines are free of charge, they come with no guarantee. See Table 1 for a list of STATLIB libraries.

Some types of problems can occur when using the routines. We recommend that you carefully test and check each routine. In particular, look for machine constants in the codes. Machine constants for a CRAY might produce incorrect results when used on an IBM-PC.

### Accessing STATLIB

The index files and the software itself are available from STATLIB by sending requests in the form of electronic mail to [STATLIB@LIB.STAT.CMU.EDU](mailto:STATLIB@LIB.STAT.CMU.EDU). The requests are one-line commands that can be the subject part of a mail message or can be the mail message itself. STATLIB interprets the command, executes it, and sends you a mail message response. ■

apstat	Select algorithms transcribed from the journal of Applied Statistics
crab	Kodiak Island king crab survey data.
datasets	Various data sets.
directory	Lists of addresses and E-mail addresses of statisticians.
Disease	National Notifiable Diseases Data.
general	Software of general statistical interest.
griffiths-hill	Applied Statistics algorithms contained in the book of that name by Griffiths and Hill
multi	Multivariate Analysis and Clustering. An annotated directory and selected algorithms.
S	S functions, device drivers and related software.
s-news	Archives of the S-news mail, in digest format.
xlispstat	Luke Tierney's XlispStat system for UNIX systems. (Luke Tierney is a professor in the University of Minnesota's Statistics department.)

Table 1: List of STATLIB Libraries

Command	Requests
SEND INDEX	Index listing of all available software.
SEND INDEX FROM <i>collection</i>	Index of routines within a specific collection; for example: SEND INDEX FROM S
SEND <i>routine</i> FROM <i>collection</i>	Routine contained in a particular collection; for example: SEND SPLINES FROM S

Table 2: Some STATLIB Commands

## More NETLIB Public Domain Mathematical Software

*This article was adapted from an article by the same name in the February 1991 issue of the University of Minnesota Academic Computing Services Newsletter, acs.*

**T**he following has been recently added to [NETLIB@ORNL.GOV](mailto:NETLIB@ORNL.GOV):

CRPC                      Software available from the NSF Science and Technology Center for Research in Parallel Computation.

The following have recently have added at [NETLIB@RESERCH.ATT.COM](mailto:NETLIB@RESERCH.ATT.COM).

Amos                      Special functions by D. AMos (=TOMS/644).  
BLAS3                    Basic Linear Algebra Subroutines, level 3, for matrix\*matrix operations.

C                            Another miscellaneous library for software written in C.  
cascade                  Analysis and design of linear control systems.  
Diercks                   Spline fitting routines for various kinds of data and geometries.  
f2c                         FORTRAN to C converter. ■

## NSF Grants Bulletin Board Established

From an article by Geraldine Sonnesso (ACSGERI@UBVM) that appeared in the State University of New York at Buffalo Computer Center newsletter, Interface (February/March 1991).

The National Science Foundation (NSF) Grants electronic bulletin board was established by the NSF Division of Grants and Contracts. The purpose of the bulletin board is to provide a forum for discussion and communication among grant administrators, NSF grant officials, and other interested parties on grant administration and policy matters.

In addition, NSF posts official notices of interest to the grant administration community.

Users who would like to be added to the electronic mailing list should address requests to [grants-request@NSF.GOV](mailto:grants-request@NSF.GOV). Notices may be posted to the bulletin board at [grants@NSF.GOV](mailto:grants@NSF.GOV).

Please include with your request, your name, title, institution, U.S. Mail ad-

### Baczewski Compositions Performed

Two compositions by Dr. Philip Baczewski, Academic Mainframe User Services Manager, were performed at the *Society of Composers Inc. Region VI* conference, March 7-9 at UT Arlington. *I=3=I*, for computer-generated tape was presented on March 7, and *Seven Preludes for Dinner*, for clarinet and guitar, was performed on March 9. *I=3=I* was realized during the summer of 1990 while Dr. Baczewski was attending a seminar on programming using NeXT workstations at Stanford University's Center for Computer Research in Music and Acoustics. ■

## THE BITNET CONNECTION



By Dr. Philip Baczewski, BITNET INFOREP (BITNET: AC12@UNTVM1)

*This Column is a continuing feature of **Benchmarks** intended to present news and information on various aspects of the BITNET wide area network.*

### Guidelines for Electronic Mail on BITNET (or Anywhere Else)

Back in the "old" days people were forced to perform their communication through writing, by hand, on blank sheets of paper and then conveying those blank sheets to others. It is even reported that the U.S. Postal Service, today known primarily for its ability to deliver multitudes of "junk mail" and bills to your door, conveyed large numbers of these hand-written communiques, known as "letters." The exchange of letters tended to foster a certain style of communication: letters were received and thoughtfully read; letters were generally more formal than spoken communication; letters had permanence and could be saved for later reference; in responding to letters, people would often think and then write one sentence, think and write another sentence, etc. Now days, electronic mail has revolutionized communication. There is no need for paper or messy hand writing utensils. You no longer have to use that formal writing style. Mail is received and read with heretofore unknown speed. It's now possible to dash off a reply to an electronic mail message without even thinking.

Well, maybe it's not quite that cut and dried, but electronic mail does seem to have the ability to evoke what are known in e-mail circles as "flames:" emotional responses to messages which don't necessarily do much for fostering effective communication. Norman Z. Shapiro and Robert H. Anderson, in a report prepared for the National Science Foundation and published by the Rand Corporation, list several possible causes for the flame phenomenon:

- Difficulty in determining the formality of a message from its appearance.
- Attempts at humor, irony, sarcasm, and wit are often misinterpreted.
- Cues such as body language [or voice inflection] are lacking in electronic mail.
- The ease of an immediate "reply" encourages "off the top of the head" responses.
- Electronic messages containing hasty or ill-chosen words can stay in electronic in boxes or can be printed in a way that gives them importance never intended.
- Although anonymity is often mentioned as a factor, we have observed no significant difference in "flaming" between remote correspondents who don't know each other personally, compared with communication among people who know each other.

Shapiro and Anderson go on to give several suggestions for minimizing the possible problems of "escalating emotions."

- Carefully label message that have a deliberate emotional content. Sometimes just the annotation "Flame! Flame!" alerts the reader to the fact that the writer knows he or she is being emotional.
- Resist the temptation to fire off a response. Write the response, file it away, and wait 24 hours. Reconsider the response later, in the light of a new day (and perhaps a rereading and reinterpretation of the original message).
- Use alternative media to break the cycle of message-and-response. A telephone call or personal conversation can do wonders, when we can use body language, eye contact, and the other cues we've developed.

Just as in other human situations, the development of an etiquette can help solve some of the problems which potentially arise with electronic communication. John Quarterman, in his book entitled *The Matrix*, offers a number of suggestions concerning e-mail etiquette when sending messages to others or posting messages to BITNET LISTSERV or USENET mailing lists:<sup>1</sup>

*Electronic mail is not like other media.* Treating e-mail just like the telephone, paper mail, or any other medium can lead to misunderstandings and mistakes.

*Emulate experienced users.* See how those already posting to mailing lists make the most effective use of those forums.

*Be brief.* Often a few well-chosen words are better than long-winded elaborations.

*Label your message.* Choose a title that fits the subject and stick to it.

<sup>1</sup> Quarterman, John S., *The Matrix* (Digital Press, 1990) 34.

### LIST of the Month

*Each month we will highlight one of the BITNET LISTSERV Special Interest Group (SIG) mailing lists. This month's list...*

#### HUMANIST@UTORONTO

Coordinator: Willard McCarty (MCCARTY@UTOREPAS)

HUMANIST is a LISTSERV discussion group for people who support computing in the humanities. Those who teach, review software, answer questions, give advice, program, write documentation, or otherwise support research and teaching in this area are included. The list is primarily intended for interaction rather than publication or advertisement.

*BITNET is not just for those "techie" types, you know, and it's not just limited to the sciences. It's equally useful for exchange of information for a variety of areas. This list provides a good example in its forum for those using computers in support of a "non-technical" field of study, the humanities. If you are interested in subscribing to this mailing list, send the following command, either via an interactive message or as the first line of a mail message, to LISTSERV@UTORONTO:*

**SUBSCRIBE HUMANIST <your name>**

*where <your name> is your first and last name.*

*Remember your audience.* Use language, references, and subjects that will be comprehensible and not objectionable.

*Choose an appropriate medium and forum.* Use a conference or mailing list on a topic related to that of your message.

*Identify yourself.* Sign your message with some appropriate information such as your name and affiliation.

*Post new ideas.* Try not to repeat what has already been said except in brief confirmation.

*Respond to the topic and not the person.* Try to understand what the person is saying. If you can't understand what the person is saying, ask. If you

must criticize someone, give them a chance to respond. If you comment on the style of a message, respond to the content as well.

*Read other messages before responding.* Others may have already made the same obvious response.

*Don't respond in anger.* Wait a few minutes or hours, or even until the next day. If you are still angry when you respond, say so.

*Give the benefit of the doubt.* Mistakes, misunderstandings, and ignorance are far more common than maliciousness.

*Be careful with humor and sarcasm.* Many people have trouble recognizing these things even in person. Some networks have developed typographic conventions to get around the difficulties of expressing subtleties of expression through ASCII characters. One of the more universal is that UPPER CASE means shouting. Another is the use of the sideways "smiley face," :) or :-), to indicate lack of serious intent.

*Do be encouraging and polite.* The most effective encouragement is often a simple response acknowledging a posting.

*Discourage when necessary.* But do it privately and politely when possible. Don't discourage at all unless you're sure it's needed and that you are an appropriate one to do it.

*Assume Permanence and ubiquity.* Mail posted to discussion lists and sometimes even mail to individuals may be saved permanently, with or without your knowledge, and may be read by anyone, at any time, anywhere. Remember that even if a mail message has been delete, it may exist somewhere on a backup tape.

It's not enough just to observe etiquette. Quarterman also provides some valuable guidelines for e-mail *ethics*:

*Observe copyrights.*

*Cite sources.*

*Be careful with private correspondence.* Do not redistribute private correspondence without permission. Don't read other people's mail without permission. If you receive a message by accident, return it to the sender or forward it to the intended recipient.

*Be honest.* Don't distribute false information, and don't pretend to be someone you aren't in order to take unfair advantage of someone else.

*Someone is paying the bills.* Remember that what you post may cost others time and money. Try to stick to useful information distributed to appropriate people.

*Don't post harmful instructions or information.*

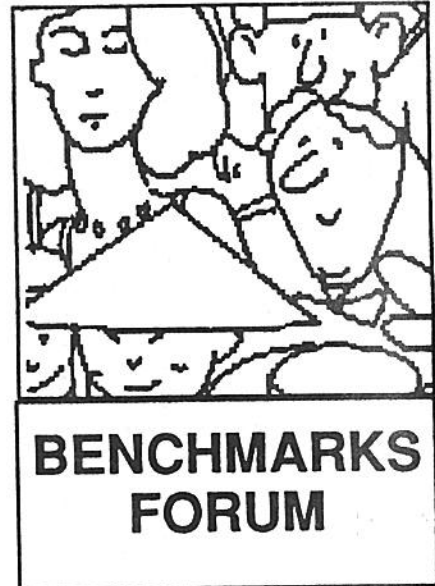
*Resource sharing systems are not like anything else.* A computer network is neither like a home computer system nor like any other single computer system. The damage that can be caused by mistakes or malevolence increases with the power and extent of the system.

*People depend on networks and conferencing systems.*

*Don't leave a security hole unfixed.* This applies to system administrators, system vendors, and users who choose obvious passwords.

*Don't use security holes to cause damage.*

Some of these points of etiquette or ethics are obvious; others perhaps wouldn't occur to you. By following these guidelines we can make electronic mail on BITNET or any other network a very effective and efficient means of communication. Or we can wax nostalgic for the good "old" days of paper, fountain pens, envelopes, stamps, waiting five days to get a letter.... ■



*BENCHMARKS FORUM is intended to serve as a vehicle for answering questions that may be of general interest to the user community. If you have a question, please send electronic mail to the Benchmarks editor (BITNET: AS04@UNTVM1) or write it down and drop it by the Computing Center. We will try to answer it in the next issue.*

**Question:** *I am fairly new to computing and have a lot of questions about various topics. Where is the best place for me to look for answers to my questions?*

**Answer:** One of the things computers do for us is give us access to all kinds of information. Since many people have similar questions about various topics, "Frequently Asked Questions" (FAQs) have become a hot item on the Internet. The following is a posting taken from comp.archives (on ANU on the VAX).

Frequently Asked Questions re-update TSFAQ17C.ARC

Sun 3-Mar-91: I have updated my Frequently Asked Questions collection almost immediately after the last update to be /pc/ts/tsfaq17c.arc. The collection includes among other things the prerecorded instructions for garbo.uwasa.fi and I noticed that they were somewhat out of date.

# General Information

Benchmarks Reader/User feedback is encouraged. Send all letters, suggestions, etc to (AS04@UNTVM1), FAX 817-565-4060 or to the Benchmarks Editor at:

Academic Computing Services  
University of N. Texas Computing Center  
P.O. Box 13495, Denton, Texas 76203



## TSFAQ17C.ARC - Frequently asked questions & ans

Filename	Comment	Date	Time
FAQ.INF	Document (a readme)	03-03-91	19:18:10
FAQ.NWS	News announcement about tsfaq	03-03-91	19:11:36
FAQNEWS.TXT	Q&As relating to UseNet news	03-03-91	19:13:10
FAQPAS.TXT	Turbo Pascal questions & answers	03-02-91	19:05:16
FAQPROGS.TXT	Questions answers about programs	03-03-91	19:15:56
FAQQUOTE.TXT	Postings and email of interest	01-25-91	16:08:18
PD2ANS.TXT	Q&As on uwasa.fi archives & more	03-03-91	18:57:30
TSPROG.INF	List of PD programs from T.Salmi	01-25-91	16:05:54
VAASA.INF	Info: Finland, Vaasa, U of Vaasa	02-02-90	11:52:54

The wares are available by anonymous ftp from garbo.uwasa.fi, Vaasa, Finland, 128.214.12.37, or by using our mail server (use the latter if, and only if you don't have anonymous ftp). If you are not familiar with anonymous ftp or mail servers, I am prepared to send prerecorded instructions on request. (If you don't get the instructions from me within a few days, it will mean that your email address cannot be reached by a simple email reply. Contact your system manager for devising a proper mail path for you, because unless you do, you wouldn't be able to utilize the mail server anyway). (North American users should first consider a near site because of the overseas load.)

Prof. Timo Salmi Moderating at garbo.uwasa.fi anonymous ftp archives 128.214.12.37  
School of Business Studies, University of Vaasa, SF-65101, Finland Internet:  
ts@chyde.uwasa.fi Funet: gado::salmi Bitnet: salmi@finfun

Besides these FAQ archives, FAQs are published monthly on some Internet lists. Peruse the topics you are interested in on ANU and watch for a FAQ posting. If you have questions about ANU, the Internet, or anything else, contact Academic Computing Services (ISB 119, 565-2324) and someone will help you. ■



## Information Resources Council News

Minutes provided by Sue Harrison, IRC Recording Secretary

**T**he Information Resources Council met on Tuesday, January 15, 1991 and conducted items of business which are briefly summarized here:

Chairman Vondran reported that the General Access Computer Lab Report was on the agenda of the Information Resources Steering Committee meeting on Jan. 3. After discussion on the use of Student fees to fund General Access Labs, a study committee was named to assist in developing an acceptable formula for the allocation of any academic microcomputer monies that would be collected by means of the general computer fee. Vondran reiterated that the General Access Computer Lab Report with its recommendations, prepared by Cengiz Capan's Committee, was officially approved by the Information Resources Steering Committee with the one exception of method of funding, which this new committee will now address.

Chairman Vondran reported that he and Coy Hoggard met with the Information Resources Steering Committee on Jan. 3, at which time the Committee approved the recommendation of the IRC and the specific recommendation of Hoggard's sub-committee to upgrade the administrative mainframe computer. President Hurley also approved the recommendation and the subject will be brought before the Board of Regents at their February meeting. Chairman Vondran commended Coy Hoggard and his committee for their representation of the IRC and the Computing Center.

The recommended usage policy for the new Academic Unix System which was prepared by Dave Molta's subcommittee, was briefly discussed by the IRC. Cengiz Capan moved that the policy be approved; the motion was seconded by Don Grose, and unanimously passed. It was agreed that the usage policy should be publicized to the university community through *Benchmarks*, through distribution of the IRC meeting minutes, as well as by direct contact with major users such as Computer Science and the College of Business.

Dave Molta reported that his subcommittee on Microcomputer Data Integrity was not prepared to present recommended solutions at this time. He stated that the committee will move forward on the issue and welcomes any input from the Council members. Cengiz Capan suggested that Molta's committee consider a broader aspect of the subject and look at overall disaster recovery needs. He further suggested that the committee deal with the issue of providing support centers for networks that would be consistent across campus. It was pointed out that planning is necessary to guard against having computer networks distributed across campus without having enough appropriate technical support available for them. Molta stated that there are two functioning groups of network managers who are dealing with the tactical end of the policy; one is a Network

Managers Group, made up of users, that provides daily support; and another is a Micro Network Advisory Group made up of people from across campus. Both of these groups include administrative as well as academic people.

Richard Harris announced that he will be requesting a new Computing Center position, which will possibly be a Director of Networking and Distributive Computing Support. This position would combine microcomputer and data communications support within the Center.

Chairman Vondran asked that the Strategic Planning Subcommittee meet to establish the strategy for planning for the next five years. Richard Harris agreed to continue chairing that subcommittee and to meet prior to the next IRC meeting on Feb. 19.

Coy Hoggard reported that the Policy for Adding Terminals and Workstation Access to the Central Administrative Computer is still in use. Though this was only implemented as a temporary solution to the mainframe overload, it is a workable policy until the mainframe can be upgraded. The consensus was that there was no need to change the policy at this time.

Chairman Vondran asked for discussion on expanding the IRC membership. It was agreed that the IRC Charter would be consulted for direction on how to expand the Council; and Vondran stated that he would ask the Vice Presidents for their suggestions on additional representation from the administration. Vondran also asked IRC members to think about this subject so that it can be discussed further at the February 19 meeting.

Richard Harris reported that the University's Initial Operating Plan has been approved by the DIR. ■

## OS/MVS JCL Utilities: IEBCOPY

By Cathy Hardy, Academic Database Consultant (BITNET: AC55@UNTVM1)

*This is the fifth in a series of articles dealing with JCL (job control language). This series is aimed at the current JCL user who would like to have a better understanding of statement use, utilities, and coding options. If you are not currently a JCL user, but would like to begin learning about JCL, Academic Computing has a free handout available in ISB 119. Stop by and ask for "IBM Job Control Language," or contact an Academic Mainframe Users Support consultant for further information.*

**I**EBCOPY is an IBM data set utility used to copy or merge partitioned data sets. This utility can be used to:

- create a backup of a partitioned data set.
- copy data sets which were unloaded to a direct access device (load).
- rename members.
- compress data sets in place.
- recreate a data set that has used 16 extents or is out of directory space.
- copy a partitioned data set to a sequential data set (unload).
- replace data set members.
- exclude members from a data set that is being copied, loaded, or unloaded.
- merge data sets.

### IEBCOPY Example

In this example, members are to be selected, excluded, unloaded, and copied. Processing will occur, as follows:

- ① unload, excluding members
- ② unload, selecting members
- ③ load and copy to merge members

```
//JOB CARD
//STEP1 EXEC PGM=IEBCOPY
//SYSPRINT DD SYSOUT=A
//PDS1 DD DSN=USER.XXNN.CLASS.NAME.A,UNIT=SYSDA,
// VOL=SER=ACAD03,DISP=OLD
//PDS2 DD DSN=USER.XXNN.CLASS.NAME.B,UNIT=SYSDA,
// VOL=SER=ACAD03,DISP=OLD
//SEQ1 DD DSN=USER.XXNN.CLASS.NAME.NEWA,UNIT=SYSDA,
// VOL=SER=ACAD03,DISP=(NEW,KEEP),
// SPACE=(TRK,(5,1),RLSE)
//SEQ2 DD DSN=USER.XXNN.CLASS.NAME.TAPB,UNIT=TAPE9,
// VOL=nnnnnn,DISP=(NEW,KEEP),
// LABEL=(,SL)
//UPDATE DD DSN=USER.XXNN.CLASS.NEW,UNIT=TAPE9,
// VOL=nnnnnn,DISP=OLD,LABEL=(,SL)
//MERGE DD DSN=USER.XXNN.CLASS.MERGE,UNIT=SYSDA,
// VOL=SER=ACAD03,DISP=OLD
//SYSUT3 DD DSN=&&TEMP1,UNIT=SYSDA,DISP=(NEW,DELETE),
// SPACE=(TRK,(5,1),RLSE)
//SYSUT4 DD DSN=&&TEMP2,UNIT=SYSDA,DISP=(NEW,DELETE),
// SPACE=(TRK,(5,1),RLSE)
//SYSIN DD *
COPY OUTDD=SEQ1,INDD=PDS1
EXCLUDE MEMBER=(D,C)
COPY OUTDD=SEQ2,INDD=PDS2
SELECT MEMBER=(A,K)
COPY OUTDD=MERGE,INDD=((NEWUP,R),PDS1,PDS2)
EXCLUDE MEMBER=A
```

/\*



## General Information

In the previous example:

- PDS1 DD defines a partitioned data set that contains six members (A, B, C, D, E, F) and resides on volume ACAD03.
- PDS2 DD defines a partitioned data set that contains three members (A, K, L) and resides on volume ACAD03.
- SEQ1 DD defines a new sequential data set on volume ACAD03.
- SEQ2 DD defines a new sequential data set on tape.
- UPDATE DD defines an old sequential data set that is the unloaded form of a partitioned data set that contains eight members (A, B, C, D, M, N, O, P) on tape.
- MERGE DD defines a partitioned data set that contains six members (A, B, C, D, Q, R) and resides on volume ACAD03.
- The first COPY statement indicates the start of the first unload operation. (The input data set is partitioned, the output data set is sequential.)
- The first EXCLUDE statement specifies the members D and C are to be excluded from the unload operation specified by the preceding COPY statement.

The excluded members are searched for on every input data set and are always omitted.

- The second COPY statement indicates the start of the second unload operation. (Again, the input data set is partitioned, the output data set is sequential.)
- The SSELECT statement specifies that member A and K are to be included in the unload operation specified by the preceding COPY statement. Selected members are searched for in a Low-to-High (a - z) collating sequence, regardless of the order in which they are specified; however, they are copied in the same physical sequence in which they appear on the input partitioned data set. Once a member of a data set has been found, no further search is made for that member. When all members have been found, the copy or load step is terminated even if all input data sets have not been searched. For example: If A and B are specified, and A is found on the first of three input data sets, it is not searched for again. Then, if B is found on the second input data set, the operation is successfully terminated without

the third data set being searched, even if an A and B exist on the third. If, however, A is not found on the first input data set, then it (the first data set) will be searched for B before going to the second input data set to search for A.

- The third COPY statement indicates the start of the copy and load operations. The replace option is specified for the UPDATE data set; therefore, members in this data set replace identically named members on the output data set. The first INDD data set is an unloaded data set that is to be loaded. The second and third INDD data sets are partitioned data sets that are to be copied. (The input data sets are sequential and partitioned; the output data set is partitioned.)

For additional information on IBM utilities, check your *IBM OS/VS2 MVS Utilities* manual. The example came from *OS/VS2 MVS Utilities*, GC26-3902-1, p. 6-44.

Next issue: IEBGENER and IEBPTPCH. ■

## MICROCOMPUTERS



### Some PC Viruses Described

This is the final listing (the first appeared in the March 1991 issue of *Benchmarks*) of PC viruses described by Fridrik Skulason, author of the shareware antiviral program F-PROT (frisk@rhi.hi.is). These descriptions are as of January 1991 and were distributed with F-PROT 1.14. [There's still some good material here for social scientific research — ed.]

- **Old Yankee** — There is some confusion regarding the various "Yankee Doodle" viruses. They all originated in Bulgaria and play the tune "Yankee Doodle", but their structure is different. This is the first one. It only infects .EXE files, increasing their length by 1961 bytes. When an infected program is run, the virus will search for a non-infected file. When it has been infected, the virus plays the melody, before the original program is executed. At the very end of the virus, the word "motherf—er" appears.

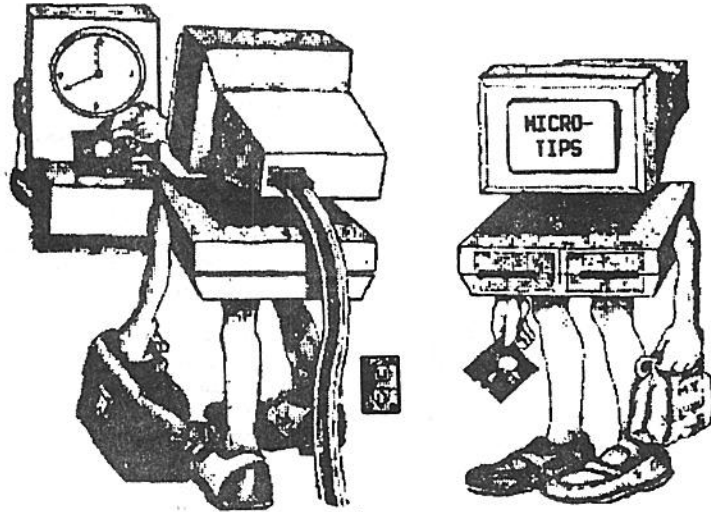
Another version of the same virus is also known. The main difference is the length - the variant is only 1624 bytes long.

- **Oropax** — This virus probably originated in W. Germany. It is not very harmful - when it activates it will just repeatedly play three melodies. The virus infects .COM files, other than COMMAND.COM. Infected files grow by 2756-2806 bytes, becoming a multiple of 51 bytes in length. This virus stays resident in memory, but it will not infect other programs when they are executed. Instead it will search for a file to infect when files are created or deleted, a sub-directory is created or the access mode of a file is changed. A few other functions may also trigger an infection.

The virus uses a random number generator to decide when to become

active, and if it does, it will start playing 5 minutes after an infected program is run.

- **Parity** — This is a simple, 441 byte virus, which will infect one .COM file when an infected program is run. The virus may emulate a parity error, displaying PARITY CHECK 2 on the screen and halting the computer.
- **Perfume** — A .COM infecting virus of German origin, that will sometimes ask the user a question and not run the infected file unless the answer is "4711", which is the name of a perfume. This virus will look for COMMAND.COM and infect it unless it is already infected. Infected files grow by 765 bytes. In the most common variant of the virus the questions have been overwritten with garbage.
- **Phoenix** — This is a family of viruses from Bulgaria, containing 1226, Proud (1302), Evil (1701) and Phoenix (1704). They all infect .COM files, and use a complicated encryption method, which complicates detection somewhat.
- **Piter** — This 529 byte Russian virus has not been analysed yet, but it infects only COM files, overwriting the beginning of the programs, and placing the original code at the end of the infected file.
- **Plastique (AntiCAD)** — This is a family of viruses from Taiwan. They are members of the Jerusalem family, perhaps based on the Fu Manchu variant, but are long - one 2900, one 3012 and three 4096 byte variants. One of the 4096 byte variants is also known as "Invader". The 4096-byte variants also contain code for infecting the boot sector. The reason the viruses are also known as AntiCAD is that they are targeted against the AutoCAD program. When a program named ACAD.EXE is run or sometimes when Ctrl-Alt-Del is pressed, the viruses will activate, overwriting data on floppy disks and hard disks,



*This column is intended to serve as a forum for sharing useful tips on making more productive use of microcomputers. If you have a tip that you feel may be of use to campus users, submit it to the Benchmarks editor for possible inclusion in a future issue.*

## Software License and Warranty Agreements

*This Micro-Tip was reprinted from an article by Debbi Buffamanti (MICDEB@UBVMS) that appeared in the State University of New York at Buffalo Newsletter, INTER-FACE (April/May 1990, page 19).*

**S**oftware, like books and musical recordings, is subject to certain copyright laws. In fact, the whole question of the ownership and distribution of intellectual property is an important one, and will probably become more crucial as times goes on. If developers can't be compensated for the research and development time they've invested in one particular software package, they will be a lot less likely to continue to develop new products tomorrow. In this article I'll be discussing the implications of copyright laws, and what you can expect if they are violated. First, let's see exactly what the new user may find inside their new software package.

Imagine the joy of purchasing a new, very expensive and long-awaited piece of software; something you've always wanted, and finally mail-ordered. Then, upon opening the package, you're confronted with something resembling the following notice:

**IMPORTANT: READ CAREFULLY BEFORE OPENING** — By opening this sealed disk package, you indicate your acceptance of the following agreement:

The rest of this slightly foreboding document details the following sorts of things you may read on this enclosure; contractual things like . . .

- You have the right to use the software on a single terminal; or one copy on a licensed computer network.

Continued on page 27.

- The software is owned by the vendor of the product; and is protected by U.S. copyright laws, and international treaty provisions. **THEREFORE, YOU MUST TREAT THE SOFTWARE LIKE ANY OTHER COPYRIGHTED MATERIAL** (e.g. A BOOK OR MUSICAL RECORDING).
- You may make one copy of the product for backup or archival purposes.
- You may transfer the product to a single hard disk provided you keep the original for backup purposes.
- You cannot rent or lease the software.
- You may transfer the software, and the documentation, on a permanent basis (no rentals, loans, leases, etc.), provided you retain no copies and the recipient agrees to the terms of the license agreement.
- You may not reverse engineer, decompile, or disassemble the software.
- If the software package contains both 5.25" and 3.5" diskettes, you may only use the diskettes appropriate for the drives in your own computer.

Since this is a contractual agreement, the vendor also has certain obligations and responsibilities. These obligations include:

- That the software will perform "substantially" in accordance with the documentation for 90 days.
- Warranties, which are usually limited to 90 days from the date of purchase.
- Limited warranties, which may include the return of the purchase price or replacement of software at the vendor's option.
- The above is not valid outside the United States.
- Other warranties, which vary from state to state, may apply.
- In most cases, vendors will not be held liable for damages **WHATSOEVER** (i.e. loss of business data, profits or information arising out of the use/or inability to use the product).

**In short**, you are directly responsible for the actual use of the software, whether or not that software really does what the vendor's marketing department says it does, and that you are only allowed to use it in the fashion specified on your agreement.

**What happens if you make lots of unauthorized copies or decompile the software?** According to the Corporate Communications department of a major software house, the following conditions will apply:

- You should assume you are agreeing to a binding contract upon opening the software.
- The concept of "fair use" applies to software; this can include the effect upon the potential market for, or the value of, the copyrighted work.
- You are committing theft by illegal copying, and you may be subject to prosecution and fines which can (in some cases) exceed \$50,000.00 and up to 5 years in jail.
- U.S. copyright law is not clear whether the purpose of a back-up copy includes replacement in the event of theft; in some cases, back-ups can only be used if a replacement cannot be obtained at a fair price. Therefore, a replacement copy may need to be purchased.

**In short**, read those agreements! ■

## Viruses continued from page 26.

as well as garbling the contents of the CMOS.

- **Polimer** — The name of this virus is derived from a text message which infected programs may display when executed:

A le'jobb kazetta a POLIMER kazetta!  
Vegye ezt!

The virus is a simple, 512 byte .COM infector, which seems to be based partially on the Vienna virus.

- **Pretoria (June 16th)** — Pretoria is a direct-action, .COM infecting virus from South Africa. It overwrites the first 879 bytes of infected files with itself, and stores the original 879 bytes at the end of the file. This makes it impossible to fully restore programs shorter than 879 bytes, as their original length is not stored anywhere.

When an infected program is executed, the virus searches the entire current directory for .COM files to infect. As it uses a full-depth recursive directory search, this may take considerable time on an XT-class machine.

On June 16th the execution of an infected file will cause all entries in the root directory to be changed to 'ZAPPED'. June 16th is the day when the Soweto riots first broke out.

The virus uses a simple substitution encryption.

- **Prudents** — This is a Direct-Action Spanish virus, 1205 bytes long, which infects .EXE files. It is probably written by the same author as the PSQR variant of the Jerusalem virus, as it checks if that virus is present, when it is run. This virus will overwrite the last 32 bytes of any file it infects, possibly destroying the victim. The virus activates on May 1st - May 4th of any year, interfering with disk writes.
- **Saddam** — This virus is based on the Stupid virus, and is also from

Israel. It is 919 byte long. The most obvious effect of the virus is to display:

```
HEY SADDAM LEAVE QUEIT
BEFORE I COME
```

- **Shake** — Thake is a primitive 476 byte .COM infecting virus. It may infect the same program over and over, and infected programs may cause a reboot when executed. The name is derived from a string which is found inside the virus:  
Shake well before use!  
This message is occasionally displayed when an infected program is run, followed by termination of the program.
- **Slow** — This is a variant of the Jerusalem virus from Australia. Some reports indicate it may cause a slowdown of the system, but this has not been confirmed. The length is 1716 bytes, with extra 5 bytes added at the end of .COM files. The major difference between this virus and the standard Jerusalem virus is the use of a simple encryption in the Slow virus. A variant of Slow has been reported in California. This 2126 byte variant is known as "Scout's Valley".
- **South African "Friday 13."** — This is one of the oldest viruses around, but it is very rare. In fact, it may even be extinct, apart from copies in the hands of virus researchers. It is a "direct action" virus, that will seek out one or more programs to infect, every time an infected program is run. It will only infect .COM files, which grow by 415-544 bytes, depending on the variant in question. The original virus is 419 bytes long. Like the Lehigh virus it can be stopped simply by making .COM files read-only. Most other viruses are not bothered by this, however. A version of this virus, called "Virus-B", where the destructive part has been disabled has been distributed for demonstration purposes.

- **Stupid (Do-Nothing)** — The "Do-Nothing" or "Stupid" virus is not a well written one. In fact it is so badly written that in many cases it will simply cause the system to "hang", instead of properly infecting it. It seems that this virus was created by a lousy programmer, somewhere in Israel. It is not a serious threat. It will only work on machines with at least 640K of memory, because it always tries to hide itself at the same address, starting at 9000:0000. This virus infects .COM files, which grow by 583 bytes. The original infected program distributed by the author contained the string:

(c)Stupid 1989 Virushmock!

A variant of this virus, also from Israel, is somewhat longer - 919 bytes. It is known as the "Saddam" virus, because of the following text message:

```
HEY SADDAM
LEAVE QUEIT BEFORE I COME
```

- **Subliminal** — As this virus was found close to where the Dyslexia was discovered and as the viruses are clearly related, they are believed to have been written by the same person - "Subliminal" probably just being an earlier version of the same virus. The Subliminal virus derives its name from its activity - it will continuously display the text "LOVE, REMEMBER?" for a fraction of a second in a corner of the screen.
- **Suomi** — This is an encrypted 1008 byte virus from Finland. It infects COMMAND.COM as soon as an infected program is run. The virus has not been fully analyzed yet.
- **Superhack** — This is a 1077 byte virus, which infects .COM files in an ordinary way, but .EXE files are infected in a similar way, which prevents the virus from infecting .EXE files longer than 64K. The virus was written in Scotland and is known there as "Murphy".

- **SVC** — This is the first Russian "stealth" virus. It is 1689 bytes long and infects COM and EXE files. It has not been analysed yet, but it contains the text string:

(c) 1990 by SVC, Vers. 4.0

- **Sverdlov** — This is an encrypted, 1962 byte Russian virus, which was first discovered in Sverdlov. It is reported to display a moving banner across the screen, while playing the Russian national anthem. Afterwards drive C: is reported to be inaccessible and a low-level format is required to make it usable again.
- **Svir** — This is an unremarkable 512 byte direct-action .EXE file virus. Its effects (if any) are not yet known.
- **Sylvia** — This virus is a bit unusual, to say the least. It contains the following message:

```
This program          is infected
by a HARMLESS
Text-Virus V2.1
Send a FUNNY postcard to : Sylvia
Verkade,
Duinzoom 36b,
3235 CD Rockanje
The Netherlands.
You might get an ANTIVIRUS
program.....
```

It will display this message when an infected program is executed, but if the above text is tampered with, the following message, (which is stored in an encrypted form) will appear instead:

```
F--- YOU LAMER !!!!
system halted...$
```

Some people have a weird sense of humor...

As the text above indicates, the virus originated in the Netherlands. When an infected program is run, the virus will seek out up to 5 .COM files to infect. It will search drive C: and the current drive. The three system files, COMMAND.COM, IBMBIO.COM and IBMDOS.COM are not infected. The virus adds 1301 bytes to the beginning of

the files it infects (and also 31 bytes to the end), but does no other damage. The girl mentioned above exists, but she says that she has no idea who the author is. It is very likely that he knows her, though.

- **SysLock** — The SysLock virus infects .EXE and .COM files. It is a "Direct Action" virus that will search for files to infect when an infected program is executed. The virus will first modify the length of any program it infects, so it becomes a multiple of 16 bytes. Then the virus code, 3551 bytes is appended to the file.

One unusual "feature" of this virus is that it will search the disk for the

string "Microsoft" and change it into "MACROSOFT". The virus will not infect programs if the environment contains SYS-LOCK=@.

Four other variants of this virus are known. Two are called "Macho", since they both will change the string "Microsoft" into "MACHOSOFT". One is very close to the original virus, the other a bit different. The third variant, "Advent" will activate in December and then play "Oh, Tannenbaum". The last variant is 2232 bytes long and known as "Cookie", because it may display the message

I want a COOKIE!  
Taiwan

This virus seems to have appeared in Jan '90. It is a direct-action .COM infector, which activates on the 8th day of any month, overwriting the FAT and root directory of drives C: and D: Two variants are known, one is 708 bytes, but the other one is 743. Infected programs sometimes "hang", for some unknown reason.

- **Tenbyte** — This is a 1554 byte long .COM and .EXE infecting virus that was by accident posted to the V-ALERT electronic mailing list, which is intended for urgent mes-

sages regarding virus infections. Just like the "Stupid" virus, it will only work on machines with at least 640K memory. It activates on September 1st, and stays active until December 31st. While active, it will corrupt all disk writes, deleting the first ten characters of any data written to the disk and add ten "garbage" characters at the end. This will cause destruction of data files, as well as preventing the compilation of any programs.

- **Traceback** — The "Traceback" virus produces a screen display similar to that produced by the Cascade virus.

There are of course differences, since the viruses are totally unrelated. Every file infected with

"Traceback" contains the name of the file that infected it. This makes it possible to trace the path of the infection.

Another difference is that it is possible to make the characters "jump" back up, by pressing keys on the keyboard, after all the characters on the screen have fallen down.

There are two variants known of this virus, but the size is the only significant difference. The original virus is 3066 bytes long, but the variants are 2930 and 3031 bytes long.

- **TUQ** — This is a simple, non-remarkable, 453 byte long .COM infecting virus, probably written in Germany. It has no effects other than replication.
- **Turbo** — The name of the virus is derived from a text string contained inside it. The virus family contains two viruses, one 512 byte known as "Turbo Kukac", but the other as Turbo-448. Both viruses only infect COM files, and may cause program crashes. For example, COMMAND.COM will always crash if infected.
- **Turku** — This is one of two viruses from Finland, the other being the

"Suomi" virus. It will infect .COM and .EXE files. The length of the virus is either 1232 or 1472 bytes, depending on the file type. No interesting effects have been observed, other than occasional "repeats" of certain keys, when pressed.

- **Vacsina and Yankee Doodle** — A programmer in Bulgaria has written a number of viruses - 50 different variants or so. Two of the variants, number 5 and 39 "escaped" to the West in 1989. One of the features of virus in this family is that they contain a version number system, similar to that used in the "Den Zuk" virus. If a virus in the family finds a file infected with an older version of itself, it will remove the infection and re-infect with the new version.

A number of the variants play the tune "Yankee Doodle", but the viruses are not to be confused with the original "Yankee Doodle" virus, which is called "Old Yankee" by the F-FCHK program.

This family can be divided into two groups, one consisting of versions numbered below 38, but the other one versions 38 and upwards. The first group is identified as "Vacsina" variants by F-FCHK, and the second one as "Yankee Doodle" variants. However, this division is based on differences in the internal structure of the viruses - several of the "Vacsina" viruses also play "Yankee Doodle".

The "Vacsina" viruses seem to have been written originally to infect only .COM files. .EXE files are also infected, but that is done in two steps. First a short piece of code is added to the end of the file. Then a JMP command is added at the front of the file. This code seems to be based on the code used in FORMAT.COM and CHKDSK.COM in some versions of MS-DOS. When executed it will relocate the .EXE file. This makes the .EXE file structurally equivalent to a .COM file, so it can be infected as one.

The second group (versions 38 and upwards) infects .EXE files in a "ordinary" way.

Compared to most other viruses, these are fairly harmless. In the first versions a beep (BELL) is heard, every time a .COM-type file is successfully infected. As mentioned before, some of them play "Yankee Doodle", sometimes at 5 o'clock, but other variants play the tune when the computer is rebooted by pressing Ctrl-Alt-Del.

The latest versions of the viruses contain several advanced features - including self-correcting Hamming code, disabling of debugging tools, and the ability to search for and remove the Ping-Pong and Cascade viruses.

One related virus is also known, but it is shorter than any of the Yankee viruses, only 1049 bytes long. It seems to have been created by removing parts of the "Yankee" code.

- **Vcomm** — An .EXE infecting virus that came from Poland. It is not very well written, but easy to study because the commented source code was included in the sample that arrived from there. When an infected program is run, it will infect one .EXE file in the current directory. Infected programs are first padded so their length becomes a multiple of 512 bytes. Then the virus adds 637 bytes to the end of the file. It will also install a resident part that will intercept any disk write and change it into a disk read.
- **VFSI (Happy)** — This virus is believed to have been written by a student at the "Vish Finansovo-Stopanski Institut" in Bulgaria. It is a small, 437 byte, direct-action .COM-infecting virus. It can be stopped by making files read-only. When it activates, it may display the following messages.

HELLO!!! HAPPY DAY and  
SUCCESS from virus 1.1 VFSI-Svistov

- **Victor** — This is a 2442 byte .EXE and .COM virus from the USSR, at least according to the text found inside it:

Victor V1.0 The Incredible High Performance Virus  
Enhanced versions available soon. This program was imported from USSR.  
Thanks to Ivan

Little is yet known about its effects.

- **Vienna** — This virus, also called DOS-62, UNESCO and 648 will only infect .COM files. When an infected file is run, the virus will search for an uninfected file and infect it. One out of eight files infected is destroyed, by overwriting the first few bytes with instructions that will cause a restart when the program is run.

Infected files can be easily found because they contain an "impossible" value (62) in the "seconds" field of the time stamp.

Unfortunately the source code to this virus has been published in a book *Computer viruses: A High-Tech Disease*, so it will probably become very common in the future. This version was modified slightly, in order to make it a little less harmful - it would only infect files in the current directory.

The virus appends 648 bytes to the files it infects, but a 646 byte variant is also known.

One variant of this virus, "Lisbon", has been found in Portugal. It has clearly been modified and reassembled - probably in order to fool signature-type anti-virus programs. This virus overwrites the beginning of the programs it destroys with "@AIDS". The GhostBalls virus is also closely related. Several other variants have been reported in Bulgaria. Some of them are so different that they have received a new name, "New Vienna". The Bulgarian variants are similar to the original virus, but the changes include: Different length - 435,367,354 and 348

bytes. Different damage function - formatting of hard disk. Critical error handler added.

Other variants exist as well. One, named 'Violator' is 1055 bytes long, and it contains the following text strings:

TransMogrified (TM) 1990 by RABID N'tnl Development Corp.

Copyright (C) 1990 RABID !

Activation Date: 08/15/90 - Violator Strain B (Field Demo Test Version)

\*NOT TO BE DISTRIBUTED\*

The text seems to indicate the existence of another version, which has not yet been reported anywhere.

Perhaps the most unusual variant is "Father Christmas" or "Choinka", which was discovered in Poland. It is 1881 bytes long, with most of the extra length devoted to a Christmas greeting. The Monxla or "Time" variant is 939 bytes long, and has different effects, depending on the exact time when it activates.

- **VirDEM** — This 1336-byte, direct action .COM-infecting virus was written in 1986, which makes it one of the oldest viruses in existence. It was written by R. Burger, the author of *Computer Viruses: A High-Tech Disease*. It will not spread unless modified, because the virus makes it quite clear that the program has been infected. VirDEM overwrites the first part of the program and appends the original code to the end of the file. At least two variants of this virus are known - one with all the text strings in German, but the other is an English language version. In addition, a modified 792 byte variant is also known, where all the text messages have been removed.
- **Virus-90** — The most interesting fact regarding this virus is that the author of it is known. He uploaded the virus to a number of BBS, saying that the source code was available for around \$20. The virus is a simple .COM infector that adds

857 bytes to any file it infects. It will only infect files on drive A: and B: but it would be easy to "fix" that. An infected program will display the message "Infected!" when it is executed, but otherwise the virus does nothing at all. The virus was uploaded for educational purposes, according to the author, but has now been removed.

A "new and improved" version, Virus-101 also exists. It infects .EXE files as well as .COM files, and is somewhat variable. Some attempts were made to make it difficult to disassemble or modify the virus, but this "protection" is easily defeated. The virus contains one unusual feature, it infects COM-MAND.COM by overwriting it, in the same manner as the Lehigh virus does, so no change in length is visible. Unlike Virus-90, this virus was not made publically available.

- **Voronezh** — This is a 1600 byte virus from Eastern Europe, which infects COM and EXE files. It overwrites the beginning of COM files, placing the original code in encrypted form at the end. EXE files are also infected in an unusual way - the original CS:PC is not changed, but the first 5 bytes of the program code are overwritten with a FAR CALL to the virus code. A related virus, 600 bytes long is also known, but it is only able to infect COM files.
- **VP** — First reported in April '90, this virus is of the direct-action .COM infecting kind. It contains one unusual feature - at the beginning of the virus a variable number of NOP instructions (0-15 in number) are added. This is probably done in order to confuse "on-the-fly" virus scanners. The virus then appends 909 bytes containing the virus code.
- **W13** — This is a rather primitive .COM infecting virus. Two variants are known, the first one is 534 bytes long, but the second, with some bugs

corrected, is only 507 bytes long. The variants are both of the "Direct Action" type and do nothing interesting. They are based on the Vienna virus, but mark infected files by setting the "month" field to 13, instead of setting the "seconds" field to 62. This virus originated in the Soviet Union.

- **Whale** — This is a recent, rather remarkable virus. It is long, 9216 bytes and able to infect COM and EXE files. The increase in file size is not visible though, while the virus is active in memory, as it uses several advanced "stealth" methods. Other effects of the virus are not known, but one infected program displayed the following message when run:

```
THE WHALE IN SEARCH OF THE 8
FISH I AM '~kzyvo)' IN
HAMBURG addr error D9EB,02
```

Most of the virus is devoted to encryption and code which moves blocks of virus code around. This overhead results in a considerable slowdown of infected systems.

- **Wisconsin** — Like the AntiPascal viruses, this one is targeted against Pascal programs. It even includes the string:

```
Death to Pascal — in encrypted form.
The virus is 825 byte long and only
infects .COM files.
```

- **XA1** — The XA1 virus overwrites the first 1539 bytes of infected .COM files with itself and stores the original code at the end of the file. On April 1st, a part of the virus will activate - overwriting the boot sector with code that will cause the computer to "hang" on next boot-up. The virus will also activate on December 21st and stay active until the end of the year. It will then display a Christmas tree, and the text:

```
Und er lebt doch noch: Der
Tannenbaum! Frohe Weihnachten
```

- **Zero Bug** — The "Zero Bug" will mark infected files in the same way as the Vienna virus, placing 62 in the

"seconds" field of the timestamp of the .COM files it infects. Apart from this, the viruses are very dissimilar.

This virus will search for COM-MAND.COM, using the value of the COMSPEC environment variable to locate the file. Then it will remain resident, hook INT 60 and infect every .COM file run. After some time has passed, a "smiley" (ASCII 1) will appear on the screen and "eat" any zero it finds on the screen.

The virus seems not too well written - containing some unreachable code, but it is unusual in some ways.

When it infects a file, it will add 1536 bytes in front of the original code, just like the Agiplan virus. It also contains one "feature" that will

probably be more used in the future - if the virus is active in memory and you look at a directory containing infected files, the virus will make the directory entries appear as they were before the infection. That is, you will not see any increase in file length. This method is also used by some of the latest viruses from Bulgaria.

- **Zero Hunt (Minnow)** — This virus infects programs by overwriting them, but unlike most overwriting viruses it will generally not destroy the host program. This is because it will only overwrite unused areas within the program. The virus searches for a block of at least 416 zero bytes, and places itself there. As such programs are rare, this reduces the chances of the virus spreading. ■

## VAXCLUSTER USAGE STATISTICS

### February Top Ten Programs: CPU Time Used

Program	Description	CPU Time	Percent of Total
1. User programs	Compiled Programs	13 10:53:46.36	63.1
2. MAXCLAS	GIS Image Processing	1 07:52:26.28	6.2
3. NEWS	ANU News Utility	1 05:20:23.78	5.7
4. DISKEEPER	Disk Optimizer	0 21:25:59.05	4.2
5. XZZZY	Chat Utility	0 09:46:19.34	1.9
6. BACKUP	Disk Backups	0 08:55:37.11	1.7
7. NNTP_TCPWIN	News Transfer Utility	0 08:43:53.46	1.7
8. CLUSTR	GIS Image Processing	0 08:09:19.13	1.6
9. MAIL	VMS Mail	0 07:19:36.41	1.4
10. EDT	Editor	0 07:13:52.92	1.4
Total		21 07:22:45.53	

### February Top Ten Programs: Frequency of Runs

Program	Description	Number of Runs	Percent of Total
1. LOGINOUT	User login	88011	19.8
2. SET	VMS Utility	55817	12.6
3. DIRECTORY	VMS Utility	40042	9.0
4. User programs	Compiled Programs	34066	7.7
5. DELETE	VMS Utility	32855	7.4
6. SEND	BITNET message Utility	24553	5.5
7. SYSLOGIN	User Login	17259	3.9
8. MAIL SERVER	VMS Mail Server	16727	3.8
9. EDT	Editor	15296	3.4
10. MAIL	VMS Mail Utility	13091	2.9
Total		444252	

## Compressing Your VMS Mail File

Taken from an article by Tony McCracken that appeared in the January/February 1991 NewsBytes, the newsletter of Northern Arizona University Computer Services.

**F**requent users of the VAX MAIL utility can save a considerable amount of disk space by compressing their mail file. As old mail messages are deleted and new messages arrive, the file MAIL.MAI becomes fragmented. Compressing the mail file defragments the MAIL.MAI file and returns the reclaimed disk space to the user's account.

The people most frequently affected by mail file fragmentation are those who subscribe to listservers and/or receive and delete large quantities of mail. Users who

Continued on page 33

## The Best of the BBS



Edited by Ken Corey,  
VAX Programmer/  
Operator (BITNET:  
KWC@UNTVAX)

**W**elcome to the *Best of the BBS* column. This column highlights some of the more interesting and useful discussions on the UNT BBS. For those of you not familiar with the BBS, here is how to log into the UNT BBS.

- Sign-on by typing **CALL DEC** at the LAN prompt and then entering **BBS** as your Username at the VAX prompt.
- If you are already logged-on to the VAXcluster, type **BBS** at the \$ prompt.

The opinions expressed in this column do not necessarily reflect the views of Academic Computing Services or the Computing Center. Also, information in *Best of the BBS* has not been checked for accuracy.

### FILELIBRARY

#1565 5-FEB-1991 10:52:46.92

Subject : MS-Kermit 3.10

Hi,

I've uploaded the latest version (3.10) of MS-Kermit in the IBM.COMM area. The file name is KERM310.ARC. Included in this archive is a list of changes in the new version.

Billy

.....  
#1602 12-FEB-1991 14:15:28.07

Subject : New FPROT-114 Signature File

I've uploaded the latest FPROT-114 signature file into IBM.IMMUNE. It



is called F114SIG1.TXT. When you download it, rename it to SIGN.TXT (the old version is not needed any longer). The new file allows 18 more viruses to be detected.

Billy

**UTILITIES**

#1795 5-MAR-1991 09:41:19.25

Subject : 4DOS

I uploaded a COMMAND.COM replacement called 4DOS into the IBM.UTILITY area.

It replaces several utilities at once including CED and some of the directory utilities. It also fixes several irritating features of DOS like it supports wildcards on the TYPE command.

Billy

.....

#1800 Reply to #1795 6-MAR-1991 21:10:11.07

Subject : RE: 4DOS

It also allows larger programs to be run, by swapping out most of its code to disk or high memory when not being used.

It also allows wildcards such as \*K\*.DAT to work correctly.

It also has built in help for all MSDOS commands.

It also has a capability to GREATLY enhance batch files, both in speed and in color, with boxes, and all that.

—Another satisfied customer

**TRUE TYPE FONTS**

#1917 17-MAR-1991 21:19:22.15

Subject : TrueType Fonts availability.

'alo!

TrueType Fonts are now available from apple.com. Those that have FTP capable can anonymous login to apple.com and the files are in /pub/dts/sw.license directory.

truetype-fonts-1-0.hqx	665,868 bytes
mac-printing-tools.hqx	827,976 bytes
personal-lw-ls-install.hqx	468,007 bytes
stylewriter-install.hqx	405,537 bytes

You will need the first two files and one of the last two. It requires System 6.0.7 so if you don't have one, you can get them in the same directory (at least 4 of them: system, print, util 1, and util 2 and you also need diskcopy 4.1 —also in there).

VAXEN beware to have disk quota enough before 'get' them (hint: 'show quota' then multiply free spaces with 512). Sequent's users: well the /u disk was at 95% last Friday.....

On campus people can also get them from me. Best time for me is after work, i.e after 5 pm. Bring at least 3 disks for TrueType stuff and 4 more if you also want 6.0.7. Ask for 'mouse' at Matthews 322B.

Telephone people, sorry. I can't upload them to the BBS. It requires license, you know. When you copy System software from me, it assumes (by Apple) that you except the license but still does not give me right to upload..... Go bug ProSystem or apple dealers around here.

And yes, it does co-exist with ATM. There are only 4-5 (system) fonts in this release. As a freebie, it comes with Font/DA mover 4.1! (Whatever happens to 3.9 and 4.0?)

\_\_Mouse.o()~ x-4435 COE:mouse  
mouse@vaxb.acs.unt.edu

x-4437 COE2:mouse  
mouse@mouse.coe.unt.edu ■

---

Compress continued from page 32

have had their accounts for a long time but have never compressed their mail will also receive a substantial benefit.

Compressing your mail not only saves disk space, it also improves the overall performance of the mail utility by speeding up access to the MAIL.MAI file. An unfragmented mail file is processed faster by the MAIL utility.

Compressing your mail takes five steps (these examples assume you don't have a mail subdirectory):

- ① Enter the VAX MAIL Utility.

\$ mail

- ② Enter the compress command at the MAIL> prompt.

MAIL> compress

- ③ Exit the MAIL utility.

MAIL> exit

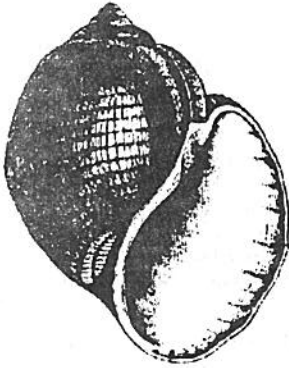
- ④ Reset the file protection on the MAIL.OLD file to allow deletion.

\$ set protection=o:rwed mail.old

- ⑤ Delete the file called MAIL.OLD. The MAIL.MAI file will be much smaller and easier to process.

\$ delete mail.old;

If you have any questions regarding compressing the MAIL.MAI file, contact the VAX operators (Voice: 565-4161, VAX MAIL: operator). ■



## The UNIX Shell

By Marc St.-Gil, UNIX Systems Programmer (mstgil@sol.acs.unt.edu)

**H**ello! And welcome back to The UNIX Shell. This month's column is a short note about using folders in the 'elm' electronic mail package. 'Elm' is available on sol and also on ponder. 'Elm' can be installed on almost any UNIX system. If you want to use 'elm' on your UNIX system, have your system administrator contact me for more information.

Now then, the question has arisen, how do I access 'folders' in 'elm'? First of all you need a directory to keep your folders in. Your account's home directory should already have a subdirectory named 'mail.' This is where your folders should reside. 'Elm's concept of a folder is simply a file containing one or more mail messages. 'Elm' doesn't care where you keep a folder, but it does have the idea of a default location. The default location for folders in 'elm' is the 'mail' directory in your home directory. 'Elm' offers you a shortcut to specifying this location by beginning the folder name with an '=' character. For example, the folder '=sent' refers to a folder file named '\$HOME/mail/sent'. 'Elm' also understands the idea of a current working directory. This means that the folder named 'junk' will be in the directory you were currently working in when you invoked the 'elm' program.

When you are in 'elm', at the index or any of the read mail prompts, you may 's'ave the current message to a folder,

'd'elece the current message, or 'C'opy the current message to a folder. The 's'ave command is invoked with the character 's.' This command will save your current mail message to the folder you specify and mark that message to be discarded from your current folder when you exit 'elm' or change to a new folder. The 'd'elece command is invoked with the character 'd.' This command marks a mail message to be discarded from your current folder when you exit 'elm' or change to a new folder. The 'C'opy command is invoked with the character 'C' (note that this is a capital C). This command is the similar to 's'ave except it does not mark the current message to be discarded. At the index menu, you may also issue the 'c'hange folder command with the character 'c.' This command will prompt you for a new folder to work in. By default, 'elm' always starts up in your incoming mail folder. This folder resides in a special directory with everyone else's incoming mail folders. Your incoming mail folder will be called '/var/spool/mail/userid,' where userid is your user ID (i.e. ac01).

**NOTE:** Any mail you leave in your incoming mail folder will show up in the unread message count displayed by the finger command. This is just one more good reason not to leave mail there for any length of time.

You may want to review some messages that you have saved with 'elm' in other folders. This is where the 'c'hange folder command becomes useful. You may manipulate any folder with all the same commands that you use in your incoming mail folder.

For more information on using the 'elm' electronic mail package, use the 'man elm' command. The Usenet news group comp.mail.elm also has interesting information. A handout on using 'elm' will be available soon in the Computing Center (ISB 119). Be sure to check for it sometime in the next month. ■

## UNIX News

### Documentation

- The *GNU Emacs Manual* has been placed, for reference purposes, in the following locations: ISB 110, 5th floor GAB Lab, College of Business Computer Center.
- Vi Reference material is now available in ISB 119 in the form of a handout. "An Introduction to Editing With Vi," is free for the asking.

### Software

The following software has been installed on the Solbourne and is available for use:

- C
- FORTRAN 77
- Perl
- TCP/IP
- E-mail
- NFS
- Kermit
- Emacs
- csound

Software planned to be installed in the near future:

- SAS
- Gaussian 90
- USENET Newsreader
- cmix
- C++
- TPU

## UNIX Users Group

**T**he Metroplex UNIX Users Group meets the first THURSDAY of every month at the Hewlett-Packard facility in Las Colinas. Contact Marc St.-Gil (565-2324, (mstgil@sol.acs.unt.edu) for more information).

# COMPUTING TECHNICAL SERVICES

## Mainframe Performance Statistics

### Operating Systems Performance Statistics for February

CPU	SYSTEM	Planned Production Hours	Production Hours Achieved	System Uptime
ACAD	VM/XA2	672.00	668.74	99.5%
ACAD	MUSIC/SP	649.43	645.30	99.4%
ACAD	MVS/JES2	672.00	665.96	99.1%
ACAD	COMPLETA	661.81	655.43	99.0%
ADMN	MVS/JES2	672.00	669.25	99.6%
ADMN	COMPLETA	261.00	259.72	99.5%
ADMN	ADABASA	644.04	639.16	99.2%

- The ACAD CPU achieved 100% uptime in February.
- The HDS/7360 DASD achieved 100% uptime in February.
- The HDS/7380 DASD achieved 100% uptime in February.
- The ADMN CPU achieved 100% uptime in February.
- The HDS/7360 DASD achieved 100% uptime in February.
- The HDS/7380 DASD achieved 100% uptime in February.
- The EMC Solid State Disk achieved 92.4% uptime in February.

### Key Causes Of Lost Productivity In February: ACAD CPU

#### Miscellaneous

1. VM/XA systems software development . 4.85 HOURS
  2. MVSSP systems software development 0.85
  3. Undetermined causes for systems restarts. 0.73
- TOTAL 6.43 HOURS

### Key Causes Of Lost Productivity In February: ADMN CPU

#### Miscellaneous

1. Systems software development. 3.24 HOURS
  2. ADABASA system failure. 1.64
  3. COMPLETA system failure. 0.23
- TOTAL 5.11 HOURS

## ACADemic (HDS) Program Hit Parade

### February Top Ten Programs : Frequency Of Runs

Program	Description	#of Runs	% of Total
1. IEWL	Linkage Editor	13422	14.4
2. IEBGENER	IBM Utility	13189	14.2
3. PGM=* .DD	Compiled Program	13023	14.0
4. ADARUN	ADABAS Utility Module	12158	13.0
5. SPCHLCOB	COBOL2 Report Writer	6953	7.5
6. IGYCRCTL	VS COBOL2 Compiler	6559	7.0
7. ADASQL	ADABAS SQL precompiler	6318	6.8
8. SASLPA	SAS	4211	4.5
9. IDCAMS	VSAM Utility	2805	3.0
10. IEBPTPCH	IBM List Utility	2259	2.4

### February Top Ten Programs: CPU Seconds Used

Program	Description	CPU Seconds	% of Total
1. SAS370	SAS Version 6.06	130887	40.3
2. SASLPA	SAS Version 5.18	45155	13.9
3. SPCHLCOB	COBOL2 Report Writer	41629	12.8
4. PGM=* .DD	Compiled Program	32496	10.0
5. ADARUN	ADABAS Utility Module	11678	3.6
6. SPSS	SPSS Version 4.0	11210	3.5
7. COMPLET4	Academic COM-LETE	10930	3.4
8. IEWL	Linkage Editor	8233	2.5
9. IGYCRCTL	VS COBOL2 Compiler	8029	2.5
10. SSS4001	Operations Automation	3221	2.5

## Disk Backup Schedules

SYSTEM	BACKUP	DESCRIPTION
Administrative MVS/SP	Daily	Monday - Friday around 7 p.m. (after COM-LETE is shut down) & on Saturday & Sunday if COM-LETE has been up that day.
	Weekly	Full pack dumps taken each Sunday morning.
	Monthly	Full pack dumps taken on the first day of each month.
Academic MVS/SP	Daily	Monday - Sunday during the early hours of the morning.
	Weekly	Full pack dumps taken each Sunday.
	Monthly	Full volume dumps taken on the first day of each month.
MUSIC/SP	Daily	Wednesday - Monday starting at 4 a.m. and lasting about 30 minutes.
	Weekly	Tuesday mornings at 3 a.m., these last about 2 hours.
	Semester	Once a semester, a permanent backup is taken.
VM/XA	VM Weekly	Early every Wednesday morning.
	CMS mini-disks	Daily backup performed early every morning. Weekly backup every Tuesday starting after Midnight.
	Semester	Once a semester, a permanent backup is taken.
VAXcluster	Daily	Incremental backups are performed Monday -Thursday at 6 p.m. Saturday & Sunday at 5 p.m.
	Weekly	Full backups are performed every Friday beginning at 8 a.m. generally last all day.
	Monthly	A "stand alone" backup is performed monthly. Dates and times are given in the system log-on message.
	Semester	Once a semester, a permanent backup is taken.

A full description of the system backup procedures can be found by typing **HELP BACKUP** on MUSIC or the VAXcluster.



**University  
of North  
Texas  
Computing  
Center**

Richard A. Harris, Associate Vice President for Computing  
 Steve Minnis, Director of Computing Technical Services  
 Dave Molta, Director of Academic Computing  
 Coy Hoggard, Director of Administrative Computing  
 Claudia Lynch, *Benchmarks* Editor  
 Dr. Philip Baczewski, *Benchmarks* Associate Editor

## Get a Subscription to *Benchmarks*

*Benchmarks* is a vital link between the UNT Computing Center and the users of our facilities. It is important for all users of the computing facilities to maintain a file of these newsletters because they contain materials which will periodically update existing documents as well as information and suggestions on uses of OS/MVS, MUSIC/SP, the VAXcluster, Microcomputers, and other resources available to UNT students and faculty. To facilitate the dispersal of *Benchmarks*, \*\*\* FREE \*\*\* subscriptions are available. To receive yours, send the following information to us either by snail mail (the post office or campus mail), FAX (817) 565-4060, or through electronic mail, to the UserID AS04 on MUSIC, VMS, or CMS.

Name: \_\_\_\_\_

Mailing Address: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**PLEASE GIVE A CAMPUS ADDRESS (NOT BOX) IF POSSIBLE! - It's Cheaper !!**

**Renewal**     **Change of Address**     **Cancel Subscription**



**Academic Computing Services**  
**The Computing Center**  
**NT Box 13495**  
**University of North Texas**  
**Denton, TX 76203**  
**FAX 817-565-4060**

---