Benchmarks Online

Volume 5 - Number 7 * July 2002

Columns

Other Resources

Feature Articles

Campus Computing News

EagleMail Upgrades are the top of this month's "Campus Computing News."

New Software Available

The RSS office has recently acquired two new statistical packages : Stata and NUD-IST 6. Read all about it.

Important Summer Reading

Just in case you haven't been able to keep up with your Benchmarks Online reading lately, we've made a list of articles that you just shouldn't miss. Check it out.

Free Virus Protection for Home PCs

This is such important news, we're keeping it around for another month. Read all about it!

Today's Cartoon

Click on the title above for an information age laugh.

Don't forget to check out our monthly columns. This month's topics:

- **RSS Matters** -- "Using Robust Mean and Robust Variance Estimates to Calculate

About Benchmarks *Online*

Subscribe to Benchmarks *Online*

Robust Effect Size" The title says it all. :)

- **SAS Corner** -- "Bits and Bytes in Summer 2002" Will clue you in on the latest SAS news.
- **The Network Connection** -- "A New World of Spam" The topic everyone loves to hate!
- **Link of the Month** -- "Randy Cassingham's Spam Primer" - Speaking of Spam, more help is on the way.
- **WWW@UNT.EDU** -- "Getting Started with ColdFusion at UNT" - The place to start in your quest for ColdFusion information.
- **Short Courses** -- The Academic Computing Services (ACS) short courses are still going on this summer. Other training is also available. Check out this article for more information.
- **IRC News** -- Minutes of the Information Resources Council are printed here when they are available.
- **Staff Activities** -- New employees, people who are no longer employed at the Computing Center, awards and recognitions and other items of interest featured here.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Research and Statistical Support
University of North Texas

RSS Matters

The previous article in this series can be found in a previous issue of Benchmarks Online: Controlling the False Discovery Rate in Multiple Hypothesis Testing - Ed.

Using Robust Mean and Robust Variance Estimates to Calculate Robust Effect Size

By Dr. Rich Herrington, Research and Statistical Support Consultant

This month we demonstrate the calculation of robust effect sizes. The GNU S language, "R" is used to implement this procedure. R is a statistical programming environment that is a clone of the S and S-Plus language developed at Lucent Technologies. In the following document we illustrate the use of a GNU Web interface to the R engine on the "rss" server, http://rss.acs.unt.edu/cgi-bin/R/Rprog. This GNU Web interface is a derivative of the "Rcgi" Perl scripts available for download from the CRAN website, http://www.cran.r-project.org (the main "R" website). Scripts can be submitted interactively, edited, and be re-submitted with changed parameters by selecting the hypertext link buttons that appear below the figures. For example, clicking the "Run Program" button below creates a vector of 100 random normal deviates; displays the results; sorts and displays the results; then creates a histogram and a density plot of the random numbers. To view any text output, scroll to the bottom of the browser window. To view any graphical output, select the "Display Graphic" link. The script can be edited and resubmitted by changing the script in the form window and then selecting "Run the R Program". Selecting the browser "back page" button will return the reader to this document.

```
x<-rnorm(100)
y<-sort(x)
y
hist(x)
density(x)
```

Run Program

Introduction - Calculating Power and Effect Size

Power analysis involves the relationships between four variables involved in statistical inference: sample size (N), a significance criterion (α), the population effect size (d_{cohen}), and statistical power. For any statistical inference, these relationships are a function of the other three (Cohen, 1988). For research planning, it is most useful to determine the N necessary to have a specified power for a given α and d_{cohen} . The statistical power of a test is the long term probability of rejecting H_0 (null hypothesis) given a specified α criterion and sample size N. When the effect size is not equal to zero, H_0 , is false, so a failure to reject H_0 is a decision error on the part of the researcher. This is called a type II error (β) and is related mathematically to power. The probability of rejecting the null if it needs to be rejected (power) is one minus the type II error ($1-\beta$). Figure 1. below is a graphical representation of the relationship between the null distribution, the alternate distribution, and the critical scores under the null distribution. The area underneath the H_1 distribution (the alternate distribution), past the critical score of the left tail of H_0 , and past the critical score of the right tail of H_0 , represents the power of the statistical test being performed (the shaded area).

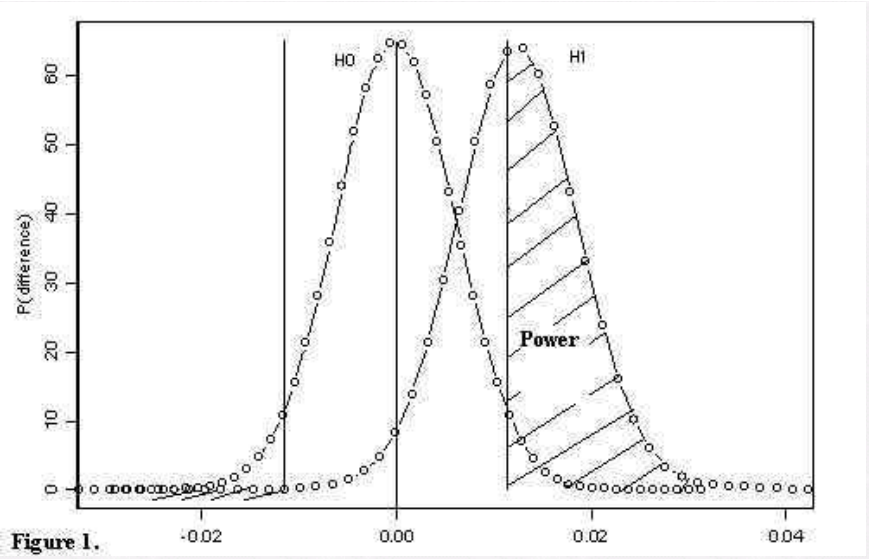


Figure 1.

Effect size is the degree to which H_0 (null hypothesis) is false and is indexed by the discrepancy between the null hypothesis and the alternate hypothesis. Power analysis specifies a non-centrality parameter to quantify this discrepancy. The noncentrality parameter for the difference between means is:

$$\delta = \left| \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_{diff}} \right|$$

where the difference between estimated population means is scaled in $\hat{\sigma}_{diff}$ units (known as the estimated standard error of the difference between means):

$$\hat{\sigma}_{diff} = \sqrt{\hat{\sigma}^2_{pooled} \cdot \left(\frac{n_1 \cdot n_2}{n_1 + n_2} \right)}$$

where,

$$\hat{\sigma}^2_{pooled} = \frac{(n_1 - 1) \cdot \hat{\sigma}^2_1 + (n_2 - 1) \cdot \hat{\sigma}^2_2}{n_1 + n_2 - 2}$$

and $\hat{\sigma}$ is the sample estimate of the population standard deviation. The denominator of the non-centrality parameter represents the estimated standard deviation of the sampling distribution for the null hypothesis for differences between means. Usually, $\hat{\sigma}_{diff}$ is calculated on the basis of a formula that assumes normality in the population since the standard deviation of the null sampling distribution cannot be calculated directly on the basis of the observed data without normality assumptions. For robust measures of location (i.e. M-estimate), the numerator would be the difference between two M-estimates, and the denominator would represent the standard deviation of the null hypothesis re-sampling distribution for the difference between M-estimates. For robust estimates (as well as the sample mean), the standard error can be estimated directly by calculating the standard deviation of the bootstrap estimates of the differences between the robust estimates of location (see September 2001 issue of Benchmarks). An alternative effect size for group differences has been advocated by Cohen (1988). Cohen s d_{cohen} measure is based on the pooled estimated population standard deviation:

$$d_{cohen} = \left| \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_{pooled}} \right|$$

Cohen provides guidelines for interpreting the practical importance of an effect size based on d_{cohen} when no prior research is available to anchor d_{cohen} meaningfully. Cohen s rule of thumb for a small, medium and large effect size are based on a wide examination of the typical difference found in psychological data. A small effect size for d_{cohen} is .20; a medium effect size for d_{cohen} is .50; and a large effect size for d_{cohen} is .80 (Cohen, 1992). Equating δ and d_{cohen} using algebra, the expression for δ is:

$$\delta = d_{cohen} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

It is noted that d_{cohen} is not a robust measure of effect size. The pooled sample standard deviation, which is used to estimate the population standard deviation ($\hat{\sigma}$) will be inflated in the presence of outliers thereby biasing the effect size measure. Furthermore, d_{cohen} assumes a normal distribution in the calculation of power estimates.

Measures of Robust Effect Size

Several problems exist with the d_{cohen} measure of effect size. The assumption of equal variances in the population is often dealt with by substituting a pooled variance estimate for σ . With data that appear to have unequal variances, questions arise about how to interpret H_0 . Another criticism of d_{cohen} is that both the location and scale (mean difference and sample standard deviation) of the sample are non-resistant measures. One strategy would be to replace the means and standard deviation with more resistant measures of location and scale. For example, one variation might be a difference of medians divided by MAD (median absolute deviation):

$$d_{median} = \frac{\mu_{M1} - \mu_{M2}}{MAD}$$

where $MAD = MED\left[\left|X_1 - M\right|, \ldots\ldots \left|X_n - M\right|\right]$ and M is the median of the scores in the control group. This effect size estimator does not seem like a good candidate since both the median and MAD are both known to be inefficient for Normal distributions compared to the mean and standard deviation.

Robust Effect Size based on M-estimators

Lax (1985) examined the performance of 17 different estimators of scale with heavy tailed distributions. Lax examined the performance of these scale estimators with the Normal distribution; a distribution with Cauchy tails (large kurtosis relative to the Normal The Slash dist.); and a mixture distribution of N(0,1) and N(0,100) for samples of size 20. The mixture distribution had 19 points sampled from N(0,1) and 1 point sampled from N(0,100) (One-Wild dist.). Lax combined the efficiencies (see July 2001 issue of Benchmarks) of the estimators for the three distributions into what was defined as triefficiency . The biweight midvariance (with c=9) estimator performed best, with favorable efficiencies across all scenarios: Normal (86.7%), One-Wild (85.8), and Slash (86.1). Following Wilcox (1997) the biweight midvariance can be calculated as follows. Setting (with c=9, M=sample median):

$$Y_i = \frac{X_i - M}{c \cdot MAD} \qquad \text{and} \qquad a_i = \begin{cases} 1, \text{ if } |Y_i| < 1 \\ 0, \text{ if } |Y_i| \geq 1 \end{cases}$$

the following is calculated:

$$\hat{\xi}_{bi} = \frac{\sqrt{n} \cdot \sqrt{\sum a_i (X_i - M)^2 (1 - Y_i^2)^4}}{\left|\sum a_i (1 - Y_i^2)(1 - 5Y_i^2)\right|}$$

The square of $\hat{\xi}_{bi}$ is called the biweight midvariance. It appears to have a breakdown point of approximately .5 (Hoaglin, Moesteller, & Tukey, 1983). Based on this robust variance, the following robust effect size can be calculated:

$$d_{robust} = \frac{\hat{\mu}_{Mest1} - \hat{\mu}_{Mest2}}{\hat{\xi}_{bi1}}$$

where, $\hat{\mu}_{Mest1}$ is the robust M-estimator for group 1 (using Huber objective function, with k=1.28 for both groups), $\hat{\mu}_{Mest2}$ is the robust M-estimator for group 2, and $\hat{\xi}_{bi1}$ is the square root of the biweight midvariance for group 1 (control group). The robust effects size d_{robust} does not assume equal variances among groups since only the robust variance for the control group is used (alternatively, a pooled estimated of both the control and experimental group biweight midvariances could be used, assuming equal variances).

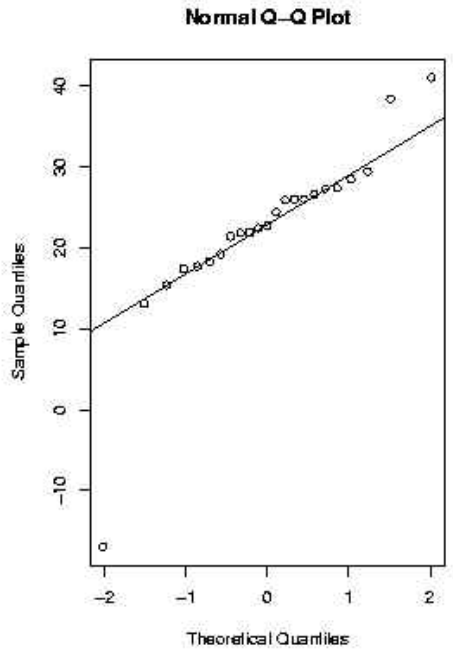
An Example Using GNU-S ("R")

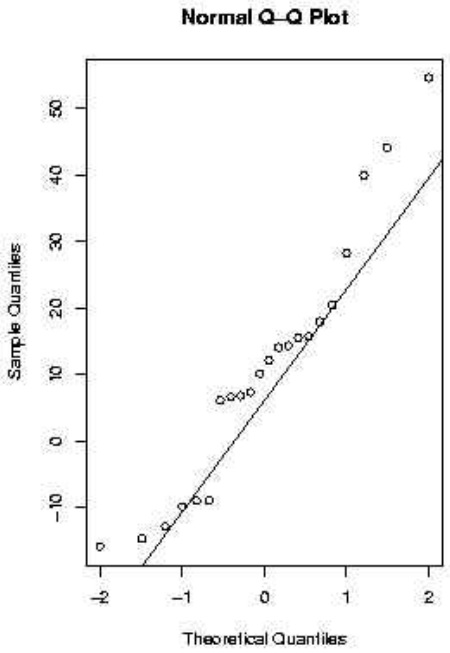
Doksum & Sievers (1976) report data on a study designed to assess the effects of ozone on weight gain in rats. The experimental group consisted of 22 seventy-day old rats kept in an ozone environment for 7 days (group y). The control group consisted of 23 rats of the same age (group x), and were kept in an ozone-free environment. Weight gain is measured in grams. The following R code produces quantile-quantile plots and non-parametric density plots of the two groups of data:

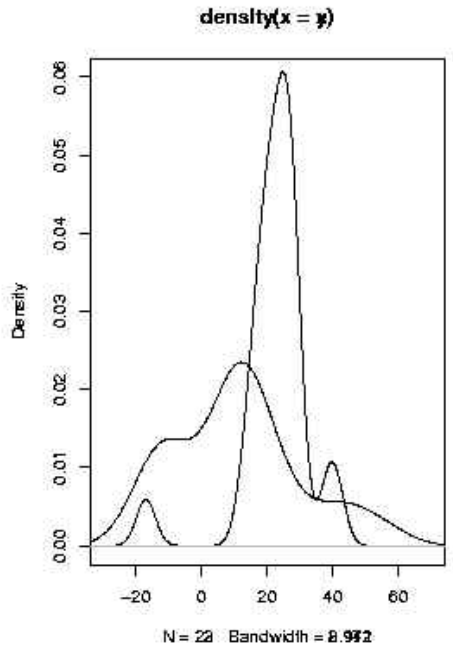
```
#### Assign data to vector x and y for the
#    two independent groups

library(eda)

#### Control Group (Ozone free environment)

x<-c(41.0,38.4,24.4,25.9,21.9,18.3,13.1,27.3,
    28.5,-16.9,26.0,17.4,21.8,15.4,27.4,19.2,
    22.4,17.7,26.0,29.4,21.4,26.6,22.7)

#### Experimental Group (Ozone environment)

y<-c(10.1,6.1,20.4,7.3,14.3,15.5,-9.9,6.8,
    28.2,17.9,-9.0,-12.9,14.0,6.6,12.1,15.7,
    39.9,-15.9,54.6,-14.7,44.1,-9.0)

summary(x)
summary(y)

par(mfrow=c(2,2)) #### Plot all graphs on same page in a
            #    2x2 grid

qqnorm(x)
qqline(x)

qqnorm(y)
qqline(y)

plot(density(x), ylim=c(0,.06), xlim=c(-30, 70))
par(new=T) #### Plot both graphs on same graph
plot(density(y), ylim=c(0,.06), xlim=c(-30, 70))
```

Run Program

Resulting qqnorm plots and density plots from R code above:







Both groups appear to have right and left tails which are "heavy". It appears as if the classical mean difference between the groups will be underestimated (smaller). The R code below estimates both the classical means, and robust means; classical estimated pooled standard deviation, and estimated pooled robust root biweight midvariance.

```
## Function to calculate pooled variance
std.dev.pooled<-function(std.dev1, std.dev2, n1, n2){
    pooled.std.dev<-(((n1-1)*std.dev1^2+(n2-1)*std.dev2^2)/
                    (n1+n2-2))^.5
}

## M-estimator function
mest<-function(x,bend=1.28){
#
#  Compute M-estimator of location using Huber's Psi.
#  The default bending constant is 1.28
#
if(mad(x)==0)stop("MAD=0. The M-estimator cannot be computed.")
y<-(x-median(x))/mad(x)  #mad in splus is madn in the book.
A<-sum(hpsi(y,bend))
B<-length(x[abs(y)<=bend])
mest<-median(x)+mad(x)*A/B
repeat{
y<-(x-mest)/mad(x)
A<-sum(hpsi(y,bend))
B<-length(x[abs(y)<=bend])
newmest<-mest+mad(x)*A/B
if(abs(newmest-mest) <.0001)break
mest<-newmest
```

```
## Function to calculate pooled variance
std.dev.pooled<-function(std.dev1, std.dev2, n1, n2){
   pooled.std.dev<-(((n1-1)*std.dev1^2+(n2-1)*std.dev2^2)/
                  (n1+n2-2))^.5
}

## M-estimator function
mest<-function(x,bend=1.28){
#
#  Compute M-estimator of location using Huber's Psi.
#  The default bending constant is 1.28
#
if(mad(x)==0)stop("MAD=0. The M-estimator cannot be computed.")
y<-(x-median(x))/mad(x)  #mad in splus is madn in the book.
A<-sum(hpsi(y,bend))
B<-length(x[abs(y)<=bend])
mest<-median(x)+mad(x)*A/B
repeat{
y<-(x-mest)/mad(x)
A<-sum(hpsi(y,bend))
B<-length(x[abs(y)<=bend])
newmest<-mest+mad(x)*A/B
if(abs(newmest-mest) <.0001)break
mest<-newmest
}
mest
}

hpsi<-function(x,bend=1.28){
#
#   Evaluate Huber`s Psi function for each value in the vector x
#   The bending constant defaults to 1.28.
#
hpsi<-ifelse(abs(x)<=bend,x,bend*sign(x))
hpsi
}


## Biweight midvariance function
bivar<-function(x){
# compute biweight midvariance of x
m<-median(x)
u<-abs((x-m)/(9*qnorm(.75)*mad(x)))
top<-length(x)*sum((x[u<=1]-m)^2*(1-u[u<=1]^2)^4)
bot<-sum((1-u[u<=1]^2)*(1-5*u[u<=1]^2))
bi<-top/bot^2
bi
}


#### Control Group (Ozone free environment)

x<-c(41.0,38.4,24.4,25.9,21.9,18.3,13.1,27.3,
    28.5,-16.9,26.0,17.4,21.8,15.4,27.4,19.2,
    22.4,17.7,26.0,29.4,21.4,26.6,22.7)

#### Experimental Group (Ozone environment)

y<-c(10.1,6.1,20.4,7.3,14.3,15.5,-9.9,6.8,
    28.2,17.9,-9.0,-12.9,14.0,6.6,12.1,15.7,
    39.9,-15.9,54.6,-14.7,44.1,-9.0)

## Classical Mean
mean(x)
mean(y)

## Robust Mean (M-estimator)
mest(x)
mest(y)

## Classical Pooled Variance
std.dev.pooled.x.y<-std.dev.pooled(var(x)^.5, var(y)^.5, length(x), length(y))
std.dev.pooled.x.y

## Pooled Robust variance
robust.std.dev.pooled.x.y<-std.dev.pooled(bivar(x)^.5, bivar(y)^.5, length(x), length(y))
robust.std.dev.pooled.x.y

## Cohens d (non-robust)
cohen.d<-(mean(x)-mean(y))/std.dev.pooled.x.y
cohen.d

## Robust Cohen's d
robust.d<-(mest(x)-mest(y))/robust.std.dev.pooled.x.y
robust.d
```

```
## Function to calculate pooled variance
std.dev.pooled<-function(std.dev1, std.dev2, n1, n2){
    pooled.std.dev<-(((n1-1)*std.dev1^2+(n2-1)*std.dev2^2)/
                (n1+n2-2))^.5
}

## M-estimator function
mest<-function(x,bend=1.28){
#
#  Compute M-estimator of location using Huber's Psi.
#  The default bending constant is 1.28
#
if(mad(x)==0)stop("MAD=0. The M-estimator cannot be computed.")
y<-(x-median(x))/mad(x)  #mad in splus is madn in the book.
A<-sum(hpsi(y,bend))
B<-length(x[abs(y)<=bend])
mest<-median(x)+mad(x)*A/B
repeat{
y<-(x-mest)/mad(x)
A<-sum(hpsi(y,bend))
B<-length(x[abs(y)<=bend])
newmest<-mest+mad(x)*A/B
if(abs(newmest-mest) <.0001)break
mest<-newmest
}
mest
}

hpsi<-function(x,bend=1.28){
#
#   Evaluate Huber`s Psi function for each value in the vector x
#   The bending constant defaults to 1.28.
#
hpsi<-ifelse(abs(x)<=bend,x,bend*sign(x))
hpsi
}


## Biweight midvariance function
bivar<-function(x){
# compute biweight midvariance of x
m<-median(x)
u<-abs((x-m)/(9*qnorm(.75)*mad(x)))
top<-length(x)*sum((x[u<=1]-m)^2*(1-u[u<=1]^2)^4)
bot<-sum((1-u[u<=1]^2)*(1-5*u[u<=1]^2))
```

Run Program

Results and Conclusion

```
> ## Classical Mean
> mean(x)
[1] 22.40435
> mean(y)
[1] 11.00909
>
> ## Robust Mean (M-estimator)
> mest(x)
[1] 23.14211
> mest(y)
[1] 9.687215
>
> ## Classical Pooled Variance
> std.dev.pooled.x.y<-std.dev.pooled(var(x)^.5, var(y)^.5, length(x), length(y))
> std.dev.pooled.x.y
[1] 15.36189
>
> ## Pooled Robust variance
> robust.std.dev.pooled.x.y<-std.dev.pooled(bivar(x)^.5, bivar(y)^.5, length(x), length(y))
> robust.std.dev.pooled.x.y
[1] 14.3583
>
```

The resulting M-estimators suggest that the population control group mean is downwardly biased (23.24 - robust; 22.40 - classical) and the experimental population group mean is biased upwardly (9.69 - robust; 11.01 - classical). Additionally, the robust pooled scale estimate is smaller than the classical pooled scale estimate (14.36 - robust; 15.36 - classical). Using these estimates to calculate Cohen's d measure indicates that the effect size is downwardly biased. Cohen's d based on classical estimators suggests a medium to large effect size (.74), whereas Cohen's d based on robust estimators suggests a very large effect size (.94). In terms of sample size planning for future experiments, the robust Cohen's d would suggest that a much smaller sample size would be needed to achieve the same power for a smaller effect size using non-robust estimators of location and scale - a considerable savings in terms of data that needs to be collected.

```
> ## Cohens d (non robust)
> cohen.d<-(mean(x)-mean(y))/std.dev.pooled.x.y
> cohen.d
[1] 0.7417875
>
> ## Robust Cohen's d
> robust.d<-(mest(x)-mest(y))/robust.std.dev.pooled.x.y
> robust.d
[1] 0.937081
>
```

References

Cohen J. (1992) A power primer. Psychological Bulletin, 112, 155-159.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd Edition. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Doksum, K.A. & Sievers, G.L. (1976). Plotting with confidence: graphical comparisons of two populations. Biometrika 63, 421-434.

Hoaglin, D. C., Mosteller, F., & Tukey, J.W. (1983). Understanding robust and exploratory data analysis. New York: Wiley.

Lax, D.A. (1985). Robust estimators of scale: finite sample performance in long-tailed symmetric distributions. Journal of the American Statistical Association, 80(391), 736-741.

Wilcox, Rand R. (1997). Introduction to Robust Estimation and Hypothesis Testing. Academic Press, New York.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Research and Statistical Support
University of North Texas

SAS Corner

By Dr. Karl Ho, Research and Statistical Support Services Manager

Bits and Bytes in Summer 2002

- **New standalone SAS available for sale** - The SAS Learning Edition 1.0 is designed for beginners who want to learn SAS, either using the point-and-click interface or in the programming environment. The software is priced at $125 and comes with a book, *Getting Started with SAS Learning Edition*. However, it does not come without limitations. Since it is basically created for learning SAS, the software only reads 1,000 observations and it comes with no technical support (except an FAQ list). Plus, the license expires in December of 2006. The software is not available at the UNT Bookstore since a student version is for sale for UNT students. If you are interested in the SAS Learning Edition, check out: For more details, check http://www.sas.com/LE/.

- **Price cut: SAS student version at the UNT Bookstore** is available at $10. Remember the software will only be valid through February 1st, 2003.

- **Graphics to MS Office**: Ever having trouble in importing SAS graphics to a Word document? SAS has a very detailed article on how to export SAS graphics into different formats before inserting into MS Office documents. Check out: **An Introduction to Exporting SAS/Graph Output to Microsoft Office**

- **Importing delimited files**: we have discussed on working with delimited files in SAS in earlier issues. Here is an article from SAS with details and screenshots: **How to invoke the Import Wizard to read a CSV, TAB, or delimited file**

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Network Connection

By Dr. Philip Baczewski, Associate Director of Academic Computing

A New World of Spam*

Unsolicited E-mail seems to be more rampant than ever. I open my E-mailbox every day and find 5-10 unsolicited commercial E-mails (UCEs). The recurring themes urge me to reduce my personal debt, invest in stocks, use a miracle cure, or look at pictures of "beautiful" women. Occasionally, there is a more radical, outrageous, or downright offensive offer in the spam message. I have a very simple way to deal with spam. I delete it. I delete as soon as I recognize it as spam. Most of the time this means that I don't even read it, since the from address, to address, or subject gives it away.

In spite of blocking known spam sources and open relay hosts on our central mail relay host, the spam still comes. For one thing, there are more sources and more open relays. Eastern European networks have opened up (Russia and Romania, in particular) and Asian networks have grown (why anyone expects me to read messages in the international standard Korean font, I don't know -- those simply go into the trash). But the latest trick used by spam generation programs (you don't think anybody actually types those in, do you?) is to make up a fake address using the target address's mail exchange machine as the domain for the from address.

Why did I get this message?

I have received a number of inquiries along the lines of, "why is someone from UNT sending me such a message?" The only catch is that it never touched the UNT network until it was delivered. People either forget or don't know that the from address can easily be made up and does not even have to be a real address (if you ever have configured E-mail service in Netscape, remember that you are prompted to enter your E-mail address -- nothing stops you from making up a fake value). It appears that those who write spam generation programs think that a from address with the domain's mail exchanger as the domain is less likely to be rejected by a spam filter process.

Take the following example header from a spam I recently received:

> Received: from unknown (HELO symail.kustanai.co.kr) (182.155.161.240)
> by pet.vosni.net with smtp; 12 Jul 0102 12:09:56 -1000
> Received: from mta85.snfc21.pibi.net ([163.151.25.32])
> by rly-xr01.nihuyatut.net with smtp; 12 Jul 0102 02:06:32 -1100
> Received: from [134.91.242.118] by smtp4.cyberecschange.com with SMTP; 11 Jul 0102 15:03:08 +0900
> Received: from unknown (HELO rly-xw01.otpalo.com) (104.61.113.59)
> by mta21.bigpong.com with local; Thu, 11 Jul 0102 23:59:44 +0300
> Reply-To: <Lovely3616e58@hotbot.com>
> Message-ID: <033a47b55b6b$7423a0c2$5ca67dd8@lhutqn>
> From: <Lovely3616e58@hotbot.com>
> To: Lovely@mailhost.unt.edu
> Subject: The miracle drug is finally here!

This one is not from unt.edu, but the "to" address is to an ID on "mailhost.unt.edu". Mailhost.unt.edu is the primary mail host for unt.edu. Most mail bound to or from the UNT campus passes through that address. It's no secret. It's how the Internet works. That address is available from the Internet registry service so that any other site on the Internet will know where to direct mail bound for the unt.edu domain. This process is automated and happens unattended. That's what makes spam so easy. Nobody's watching because, given the volume of E-mail, nobody could.

The way you can tell the path of a message is to read the "Received" lines from the header. Most E-mail programs hide the header and just show you the basic from, to, and subject information, however, there's usually an option to view the whole header. If you look at the example above, you can see that it originated on otpalo.com and passed through Korea before it hit the UNT server.

I've seen similar examples with a from address which contains "unt.edu" but a quick glance at the header quickly shows that it is just a spammer trick.

Resigned to spam?

One of the reasons I am resigned to spam is that the alternative is to give up E-mail. I have been online for so many years that my E-mail address has appeared on many E-mail mailing lists, Web pages, and news groups. Any time you subscribe to a mailing list and especially, if you post your E-mail address or a mail-to on a Web page, your address is likely to end up on spam list. Even providing a legitimate commercial vendor with your address may lead to its inclusion on a spam list. Once the E-mail leaves your computer, your address is fair came for whoever finds it along the way.

So, before you jump to conclusions about who's responsible for spam, remember that you are. The more you disseminate your address through normal use of the Internet, the more likely it will end up on a spam list. The odd thing is that I can't imagine anyone spending money because they got a spam E-mail. Sometimes I think that if we ignore it, it just might go away. If we delete it without reading it, there's no reason for anybody to send it. You can always hope.

---

* Please see this month's Link of the Month for more ideas about dealing with Spam. We've written on the topic of Spam a lot through the years. "But is it Spam?" is another *Network Connection* article on this emotional topic.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Link of the Month

Each month we highlight an Internet, USENET Special Interest Group (SIG), or similar mailing list(s) or Website(s).

Randy Cassingham's Spam Primer

What's all this in my mailbox?!

This month's link is **The Spam Primer**. Just point your browser to http://www.SpamPrimer.com/ and you will find up-to-date information on fighting Spam. Randy Cassingham is, as stated on this Website, "the author and publisher of This is True, a weekly newspaper column reporting on bizarre-but-true news stories. To regularly receive True by e-mail for no charge, see Randy's web site. Also see True's sister publication, HeroicStories."

Randy says, "Please pass the URL for this site to others you think could benefit from the information here. The more people that truly understand spam, the harder it will make things for spammers. For a copy of this article that you can send to others, send a blank e-mail to a special e-mail autoresponder: Info@SpamPrimer.com. All other use requires specific permission; contact the author."

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

WWW@UNT.EDU

Getting Started with ColdFusion at UNT

By Claudia Lynch, *Benchmarks Online* Editor

There seems to be quite a bit of interest in ColdFusion these days so, since the Web folks are taking a break from writing this month, I thought I would further recap some resources available to you here at UNT to get you started using ColdFusion. Shannon Peevey, in UNT Central Web Support, started a series of articles called "Getting Started With ColdFusion at UNT" in September of 2001. You can read the series here:

1. Getting Started With ColdFusion at UNT
2. Your First ColdFusion Application
3. Using ColdFusion: Making a Connection to a Database
4. Dealing With Empty Variable Names In ColdFusion
5. The Quest for ColdFusion: Control Structures
6. The Quest for ColdFusion: Loop-Da-Loop

Shannon also teaches a Short Course called "Introduction to Macromedia ColdFusion." There are no classes left this semester, but it will be offered again in the fall. Please consult the ACS Short Course page for more information. The ColdFusion classes are listed at the bottom of the page. It is recommended that you have some knowledge of SQL before taking this class. Computer-based training on SQL is available via the UNT SmartForce Server.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Short Courses

By Claudia Lynch, *Benchmarks Online* Editor

The summer Short Courses are progressing. Please consult the Short Courses page to see the schedule. There are still classes available on S-Plus, Dreamweaver, and more. Please note also the other training opportunities listed below.

Customized Short Courses

Faculty members can request customized short courses from ACS, geared to their class needs. Other groups can request special courses also. Contact ACS for more information (ISB 119, 565-4068, lynch@unt.edu).

Especially for Faculty and Staff Members

In addition to the ACS Short Courses, which are available to students, faculty and staff, staff and faculty members can take courses offered through the Human Resources Department, the Center for Distributed Learning, and the UNT Libraries' Multimedia Development Lab. Additionally, the Center for Continuing Education and Conference Management has a new program for interdepartmental training in business computer literacy. These classes are offered for a fee but discounts are given to those associated with UNT, and Inter-departmental Orders are accepted.

GroupWise Training

If would like to have a Basic GroupWise seminar for your area, please contact Jason Gutierrez, Campus Wide Networks, jasong@unt.edu .

GroupWise 6 classes have already been scheduled for the fall semester. Here is the lineup:

Sept 17-19 - Introduction to GroupWise 6
Oct 22-24 - Basic GroupWise 6
Nov 19-21 - Intermediate GroupWise 6

All classes are from 10 am to 11:50 am in the Eagle student Services Center (ESSC), Room 152. For signup information, go to https://home.unt.edu/hr/training/treg.htm or E-mail Bhavna Vaswani at bvaswani@unt.edu

ProDirections Instructor-led Training

UNT has formed a partnership with ProDirections to offer instructor-led computer training on Microsoft Word, Excel, PowerPoint, and Access. Classes are $99+$42 for the book. Classes in a series (3 classes in the same series) are $99 for each class and the book is free. The Excel Series includes Basic Excel, Advanced Excel-part 1, and Advanced Excel-part 2. The Access Series includes Basic Access, Intermediate Access, and Advanced Access.

Upcoming workshops:

Basic Excel
July 23rd from 9 a.m.-2 p.m. (lunch provided)
August 27th from noon-5 p.m. (lunch provided)

Advanced Excel-part 1
July 24th from 9 a.m.-1 p.m. (lunch provided)
August 28th from 1-5 p.m.

Advanced Excel-part 2
July 25th from 9 a.m.-1 p.m. (lunch provided)
August 29th from 1-5 p.m.

Basic Access
August 13th from 9 a.m.-1 p.m. (lunch provided)

Intermediate Access
August 14th from 9 a.m.-1 p.m. (lunch provided)

Advanced Access
August 15th from 9 a.m.-1 p.m. (lunch provided)

To register, send E-mail to Bhavna Vaswani at bvaswani@unt.edu or call Human Resources at x4246. Payments can be made by either a check request or with a Purchasing Card and should go directly to ProDirections. Cancellations must be done 2 days prior to the workshop date to receive a refund.

For a description of each class please go to http://www.prodirections.com/ and click on "Corporate Workshops"

Center for Distributed Learning

The Center for Distributed Learning offers courses especially for Faculty Members. A list of topics and further information can be found at http://www.unt.edu/cdl/training_events/index.htm The center also offers a "Brown Bag" series which meets for lunch the first Thursday of each month at Noon in ISB 204. The purpose of this group is to bring faculty members together to share their experiences with distributed learning. One demonstration will be made at each meeting by a faculty member with experience in distributed learning. More information on these activities can be found at the Center for Distributed Learning Website.

UNT Libraries'

The UNT Libraries' Multimedia Development Lab has also offered free training to all University of North Texas faculty and staff in the basics of FrontPage and information architecture in the past. For more information see http://www.library.unt.edu/media/services.htm#Distributed.

Technical Training

Technical Training for campus network managers is available, from time to time,

through the Campus-Wide Networks division of the Computing Center. Check the CWN site to see if and when they are offering any training.

UNT Mini-Courses

These are a variety of courses offered, for a fee, to UNT faculty, staff and students as well as the general public. For additional information surf over to http://www.unt.edu/ccecm/cont_ed/Minicourse/Courses/UNT_Minicourse_Page.htm .

Alternate Forms of Training

Many of the General Access Labs around campus have tutorials installed on their computers. For example, the College of Education recently acquired some Macromedia Tutorials for Dreamweaver 4.0, Flash 5.0 and Fireworks 4.0.

The Training Web site has all sorts of information about alternate forms of training. Training tapes, Computer Based Training (CBT) and Web-based training are some of the alternatives offered. Of particular interest are courses available via SmartForce (formerly CBT Systems). See http://www.unt.edu/smartforce/ for more information.

There are also handouts for computer training on the following topics:

- GroupWise 5.2 Handout for Win95/NT
- FAQ for GroupWise 5.2
- Computers - Back to the Basics
- Introduction to Windows 95 /98/NT
- Introduction to Word 97
- Advanced Word 97 - MailMerge It Together
- Introduction to PowerPoint 97 (Creating a Slide Show)
- Introduction to Remedy (THE Call-Tracking Program)
- AND, the award winning Introduction to Excel 97

Adobe Acrobat Reader Format only for the following:

- Introduction to Microsoft Word 2000
- Introduction to Microsoft Excel 2000
- Creating a Slide Show with PowerPoint 2000
- Using Netscape Communicator & the UNT Home Page

Use the Internet to search for answers to Microsoft Office problems. See http://www.zdnet.com/zdhelp/filters/office/ December 1999's "List of the Month" offers links to free Microsoft Word and Excel information also.

Page One

Campus Computing

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

IRC News

Minutes provided by Sue Ellen Richey, Recording Secretary

IRC Regular and Ex-officio Voting Members: *Judith Adkison, College of Education; Ginny Anderson, Fiscal Affairs; Donna Asher, Administrative Affairs; Craig Berry, School of Visual Arts; Cengiz Capan, College of Business, GALC; Bobby Carter, UNT Health Science Center; Christy Crutsinger, Faculty Senate; Jim Curry, Academic Administration; VACANT, Student Association; Duncan Engler, University Planning Council; Don Grose, Libraries; Jenny Jopling, Instruction Program Group; Joneel Harris, EIS Project Group; Elizabeth Hinkle-Turner, Standards and Cooperation Program Group; Abraham John, Student Affairs; Christine Mitchamore, Graduate Student Council; Ramu Muthiah, School of Community Services; Jon Nelson, College of Music; Robert Nimocks, Director, Information Technology, UNTHSC; Patrick Pluscht, Distributed Learning Team; Kathleen Swigger, College of Arts and Sciences; Philip Turner, School of Library and Information Science and University Planning Council (Chair, IRC); Virginia Wheeless, Chancellor for Planning.* ***IRC Ex-officio Nonvoting Members:*** *VACANT, Telecommunications; Charles Andrews, GALMAC; Bill Buntain, Computing Center Networking; Jim Curry, Microcomputer Maintenance Shop; Richard Harris, Computing Center and University Planning Council; Coy Hoggard, Computing Center/Administrative; VACANT, UNT Health Science Center; Maurice Leatherbury, Computing Center/Academic; Sue Ellen Richey, Computing Center (Recording Secretary).*

There are no IRC minutes this month.

IRC Meeting Schedule

The IRC generally meets on the third Tuesday of each month, from 2-4 p.m., in the Administration Building Board Room. From time to time there are planned exceptions to this schedule. This fiscal year, the December meeting was changed to December 11th, the April meeting was cancelled, and the May meeting changed to May 7th. All meetings of the IRC, its program groups, and other committees, are open to all faculty, staff, and students.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Staff Activities

Transitions

The following people no longer work in the Computing Center:

- **Motomi Hashiguchi**, ACS Lab monitor (part-time).

Awards, Recognition

- **Susan Bryant**, who worked in the Computing Center from 1986-2001 in Data Entry and Administrative Services was honored as a retiree recently. She is listed, along with the other retirees, in the June 21 *Inhouse*.

- Also honored in the June 21 *Inhouse* are **Luanne Linke**, Systems Programmer/Analyst in IBM Operating Systems Software Support, for 10 years of service to UNT and **Allen Bradley**, Campus Wide Networks Manager; **Duane Gustavus**, ACS UNIX Research Analyst; and **Virginia Richards**, Programmer Analyst on the UNT/HSC Human Resources Data Systems team were honored for 5 years of service.

The following people have been nominated as **Soaring Eagles** and will receive their award at the President's Staff Sack Lunch on October 15:

- **Charity Beck**, Computer Support Specialist, UNT Central Web Support, was recognized for providing needed FrontPage help.

- **Sue Ellen Richey**, Administrative Services, "took over duties for a fellow co-worker who was out on an extended maternity leave. What a team player!"

- **Bob Saringer**, CATV/Communications Technician, "consistently provides outstanding service by being positive and promoting goodwill with other departments!."

- **Ronnie Seay**, Production Control Specialist , was recognized for his work in modifying the automated job scheduling system to support WebBills.

- **Steve Voncelka**, Computer Operations Manager, was recognized in his efforts to "keep everyone updated on a serious air conditioning problem in the GAB. He even monitored this situation during Spring Break!"

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Campus Computing News

By Dr. Philip Baczewski, Associate Director of Academic Computing

EagleMail Upgrades

Some changes are being made to EagleMail which will hopefully not be visible, but may be sensed in better performance and increased reliability of the system. Previously, all mail coming into and going out through EagleMail was dependent on one computer. It has been a very reliable server, but reliance on that one system has made it very difficult to perform upgrades or make system changes which would cause EagleMail to become unavailable. At the heart of EagleMail is an IMAP mail system which not only services the EagleMail Web client, but is also available for access from clients like Microsoft Outlook, Netscape, and even Simeon. imap.unt.edu is also an outgoing mail server for those IMAP clients on the unt.edu network as well.

As of July 12, 2002, the EagleMail Web client, the IMAP server, and the outgoing mail server on imap.unt.edu are all running on sets of redundant servers which share the load of the network traffic and help to ensure that there is always a server available. This architecture also allows us to make upgrades by changing one computer out of a set at a time without having any down time for the E-mail service. We also will be moving E-mail mailboxes to a set of mailbox servers which will spread the load amongst four different systems and allow us to add servers as we need to to expand the number of users and maintain the best performance possible.

eaglemail.unt.edu

One side effect of these changes is that you might see the domain name "eaglemail.unt.edu" if you look closely in E-mail headers. When we started the student E-mail system all mail went to Jove. With various changes over the years, the centrality of the Jove system to the E-mail system has gradually lessened until Jove was finally removed from service in May. Most people use an address which ends in "@unt.edu". The unt.edu server is just a mail relay host which accepts mail from the "outside world" and routes it to the a mailbox system where your mail client can retrieve it. In the case of EagleMail, the destination is now eaglemail.unt.edu. However, we still recommend that you use "@unt.edu" in your EagleMail address and that is the address that we want to show up in all of your outgoing and inbound mail. Still, if you notice the eaglemail.unt.edu domain, remember that in any mail going out from UNT will have the address translated so that "@unt.edu" shows in the from address.

More upgrades on the horizon

We expect that with increases in enrollment, the need for more EagleMail accounts will also grow. Academic Computing Services is responding with a new scalable E-mail architecture which will allow us to grow the E-mail system to respond to the increasing demand. Any questions about EagleMail can be sent to Dr. Philip Baczewski, Associate Director of Academic Computing

(baczewski@unt.edu)

---

* For an example of looking closely at E-mail headers, see this month's Network Connection. - Ed.

Research and Statistical Support
University of North Texas

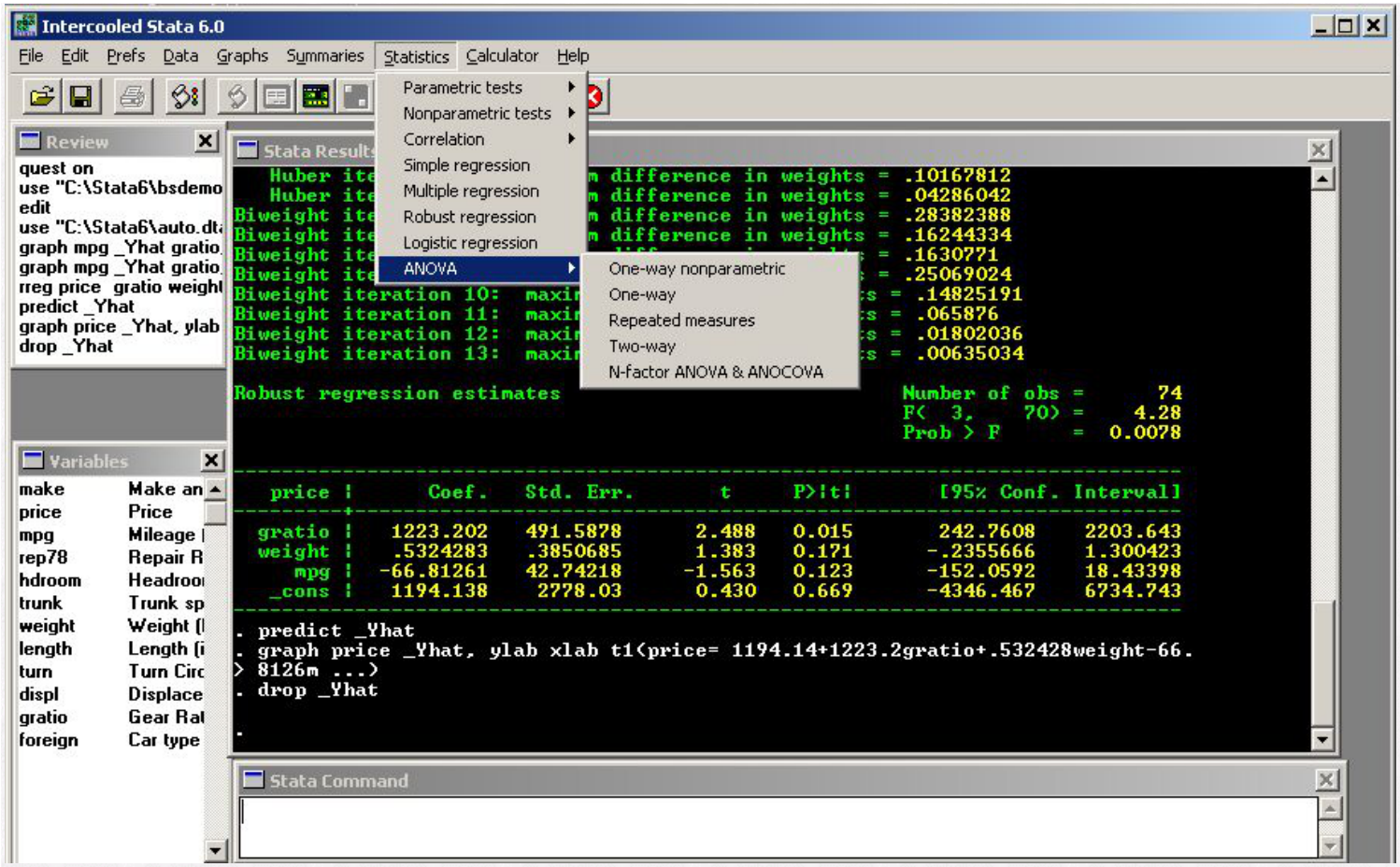
New Software Available

By Dr. Karl Ho, Research and Statistical Support Services Manager

The RSS office has recently acquired two new statistical packages: Stata and NUD-IST 6. Stata is a command-based, integrated Statistical package that covers a wide range of statistical procedures for researchers and educators.

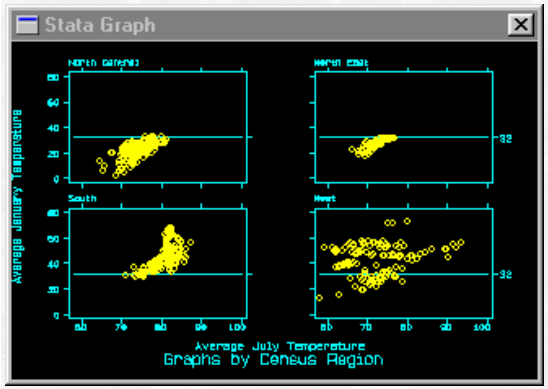
Like S-Plus and R, the beauty of Stata over other commercial proprietary packages is its constantly growing number of procedures provided by the vendor and Stata users. Stata is easily programmable and its modularity allows users to download "ado" files (ASCII files that contain Stata program) to perform statistical procedures programmed by other Stata users. You can easily search for and grab the Stata ado files from the internet for specialized procedures. Or the internet connectivity of the program allows you to update the software that carries the latest version and collection of the up-to-date ado files.

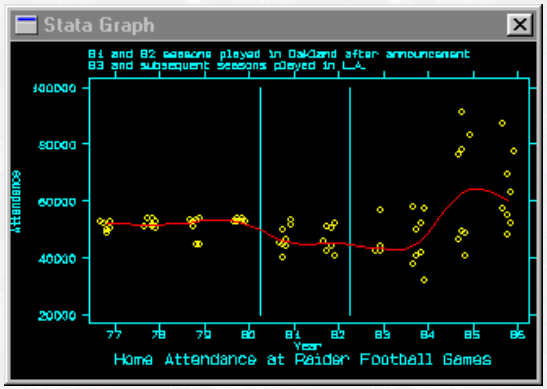
Stata is available for most popular operating systems including Windows, MacOS, UNIX and Linux. Despite that the software is command-based, i.e. users type in command and get output on the same screen, a free addition, StataQuest can be applied to add on a pull-down menu system allowing users to point and click:

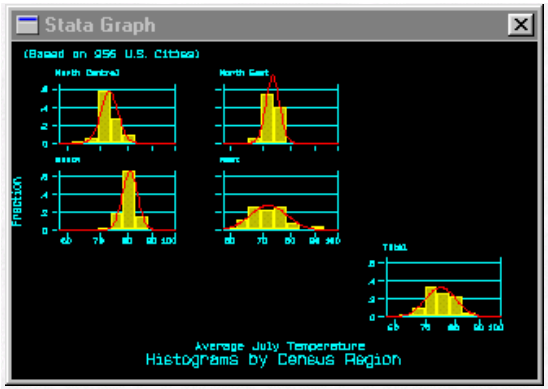


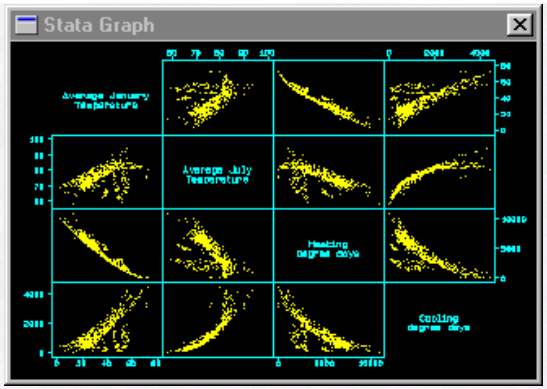
Currently, UNT has a special discount plan for faculty and student purchase. For details about the GradPlan pricing, please contact RSS office.

Below are some images produced by Stata.





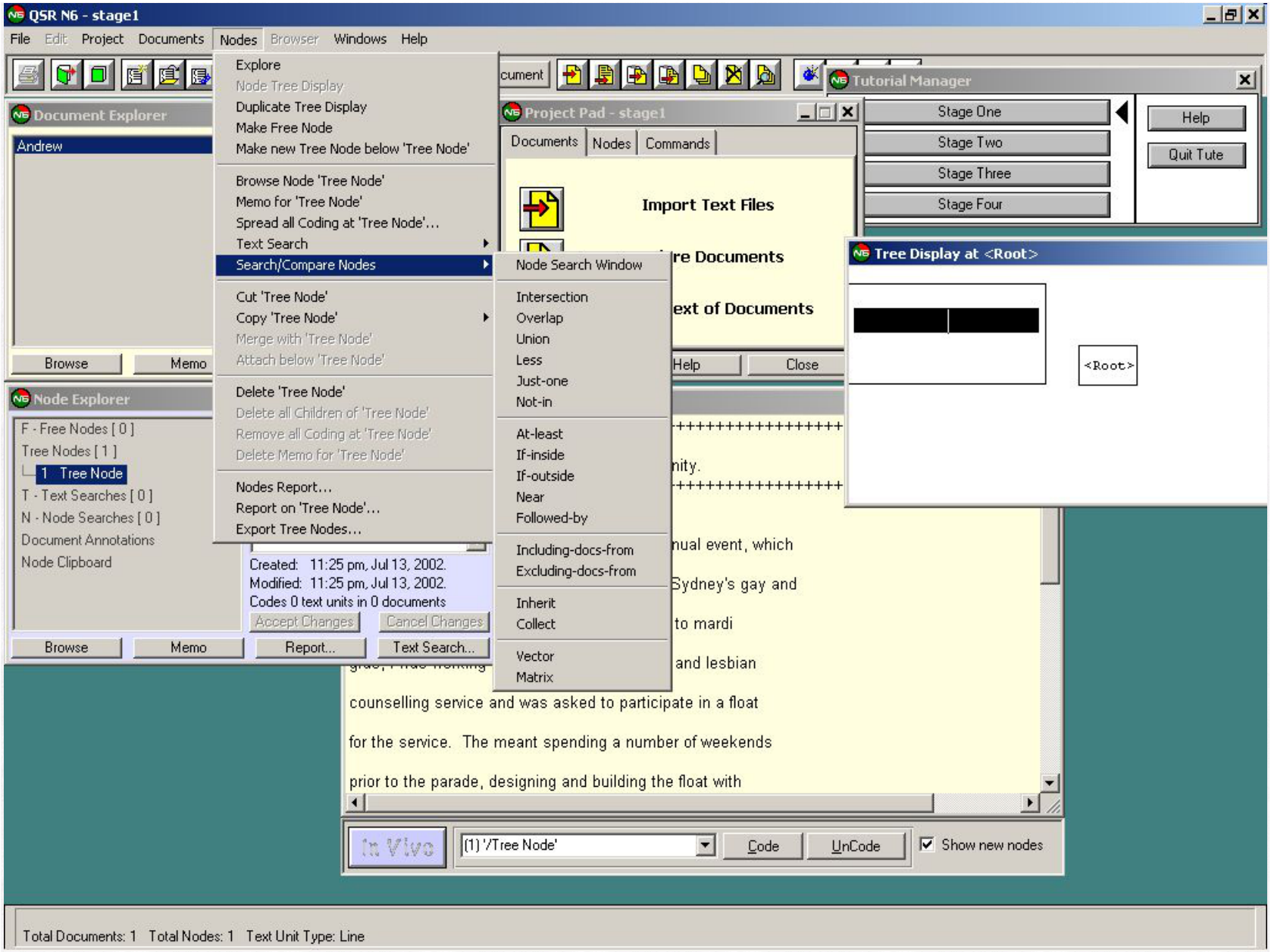




NUD-IST

N6 is the latest version of NUD-IST, a package for code-based qualitative analysis. It has an efficient management of "Non-numerical Unstructured Data" such as notes, paragraphs from newspaper, on which the program performs content analysis.

¡@



¡@

N6 is for Windows only, but an older version (N4) is available for MacOS PowerPC.

In the upcoming *Benchmarks* issues, we will give more details on these two packages. Stay tuned.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Important Summer Reading

By Claudia Lynch, *Benchmarks Online* Editor

The summer is a time when people are on the go. They take off for vacation, they leave school for the summer etc. Sometimes it is even possible that they don't read every issue of *Benchmarks Online*! :) For that reason, we have provided you a list of articles that you just shouldn't miss:

General Information

- Summer Hours
- Academic Mainframe Services to be Terminated in 2003
- Appropriate Use of Personal Web Page Publishing on People.unt.edu
- New Telecommunications Infrastructure Fund Grant Announced
- The Force is better and stronger than ever!
- Remedy Enhanced: Introducing Rem-Mail

GroupWise

- Automatically Archiving Your GroupWise E-mail
- And, speaking of GroupWise, don't forget the rules about sending *Everyone* mail!

Research and Statistical Support

- Create SAS Maps on the Web

Network Connection

- Adventures in Wireless Networking
- Napster is Dead

Link of the Month

- Ask a Librarian: Reference Assistance at the UNT Libraries
- Center for Distributed Learning

WWW@UNT.EDU

- All these articles, and more, are re-caped in this month's column.

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Free Virus Protection for Home PCs

By Travis Brown, Campus Wide Networks Computing Team

UNT is now offering free antivirus software for all its faculty, staff, and students to use on their home computers! Good deal? No it's Great! By installing McAfee's VirusScan Home Use Option v.6.0 on your home PC (other versions for MAC and Linux), you can have the security of knowing that you have taken the most important step to protect yourself from malicious viral attacks that plague the internet.

Features

McAfee antivirus software includes:

- **System Scan** (VShield) - scans memory to stop viruses before they can be run
- **Download and E-mail Scan** - scans files as they are downloading (Includes Microsoft Outlook Express, Eudora, AOL, Netscape and others)
- **Internet Filter** - blocks malicious Internet sites
- **Automatic Updates** - updates virus definitions automatically to help catch the latest viruses

Availability

You can download the software for free at http://cwn.unt.edu/virus/dist.html. Remember to have you EUID and password handy. The file is 23 MB in size, so it'll take a few hours to download over a modem, but it's well worth it. We may have a CD available shortly for a nominal distribution fee; keep watching UNT's virus information page (http://cwn.unt.edu/virus/) for details.

Virus Activity

Computer viruses are spreading faster than ever. "Iatro" (Greek for doctor) is UNT's GroupWise E-mail scanner* and stops about 3,000 infected E-mail messages each month (see http://cwn.unt.edu/virus/vstatlg.html for mind boggling details). It stopped a record number of 14,900 infected E-mail messages in May. Since new viruses appear every weak, it is ***CRITICALLY IMPORTANT*** that every computer is continually updated with the latest definition files. All the information you'll need and simple configuration instructions can be found at http://cwn.unt.edu/virus/.

Happy and safe computing!

---

* EagleMail virus scanning is now available also. According to this article in the January 2002 issue of *Benchmarks Online*, "An attachment virus scanning

option has been added for messages that are received through the EagleMail interface. For message composition, attachments are automatically tested for viruses. If the attachment tests positive, the request to attach is refused. For messages received, there is now a link name 'Virus Scan' located next to each attachment. Click this link to determine whether the E-mail sent to you is infected. Please be aware that virus scanning is not foolproof and new undetectable viruses are created everyday, so please be cautious with all attachments."

Page One

Campus Computing News

New Software Available

Important Summer Reading

Free Virus Protection for Home PCs

Today's Cartoon

RSS Matters

SAS Corner

The Network Connection

Link of the Month

WWW@UNT.EDU

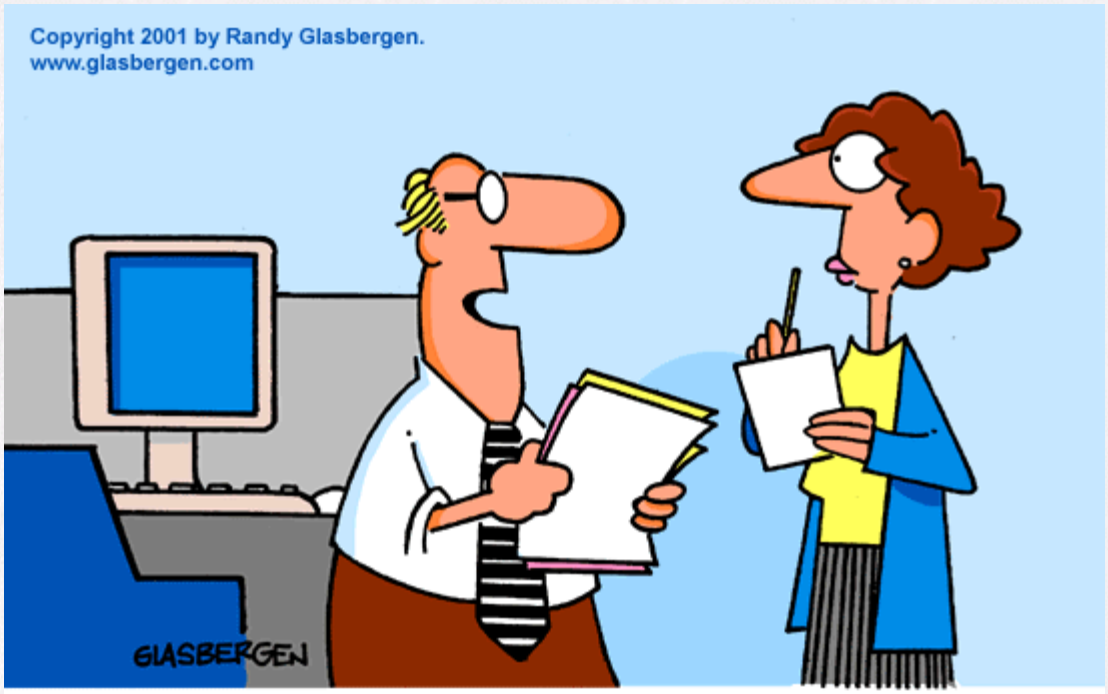
Short Courses

IRC News

Staff Activities

Subscribe to Benchmarks *Online*

Today's Cartoon



"We've got 57 team managers, 36 project coordinators, and 63 concept implementors—not bad for a company with only 18 employees!"

From "Today's Cartoon by Randy Glasbergen", posted with special permission. For many more cartoons, please visit www.glasbergen.com.