

Benchmarks

A green
light to
greatness.

UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [Issues](#)

Benchmarks - May, 2011

[Campus Computing News](#)



Our building is being renamed, but our services remain the same

By [Claudia Lynch](#), *Benchmarks* *Online* Editor

We here in Academic Computing and User Services have been informed that the Information Science Building will become **Sycamore Hall** effective June 1. The building designation ISB, therefore, becomes SYMR.

[Read more](#)    

[Summer Hours](#)



By [Claudia Lynch](#), *Benchmarks* *Online* Editor

The spring semester has ended and SUM, 3WK1, 8WK1 classes begin on Monday, May 16. Following are the hours for Computing and Information Technology Center-managed facilities during this time period and on through the summer.

[Read more](#)    

[Summer's Almost Here - Time to Hit the Road Again](#)



By [Dr. Elizabeth Hinkle-Turner](#), Assistant Director - Academic Computing and User Services*

The waning days of every spring semester finds us here in ACUS preparing to take our act on the road. By this, I mean that it is the season when the ACUS Student Computing, Helpdesk, and Documentation areas do our most event-based educational work.

[Read more](#)    

[IBM Missing Values 19 Now Available as an Add-On for PASW Statistics 19](#)

By the Numbers

Kinetic activity between May 1, 2010 and May 1, 2011:*

- Kinetic HelpCenter website visits: 82,819
- Remedy tickets submitted through the Kinetic system: 9,219 (29% of all Remedy tickets submitted during that time)
- Customers logged into Kinetic to check the status of their request 2,588 times.
- Customers logged into Kinetic and submitted an update to an existing request 326 times.

*Kinetic is a web-based interface to the Remedy system



By [Patrick McLoud](#), ACUS Host Systems Administrator

Academic Computing and User Services is pleased to announce the availability of the IBM Missing Values 19 Add-On for PASW Statistics 19. The Missing Values Add-On assists users of PASW Statistics 19 with analyzing and resolving missing values in their data.

[Read more](#)



Today's Cartoon

Click on the link above for an information age laugh.



Contact Us:

University Information Technology
1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.



UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#)

Campus Computing News



Our building is being renamed, but our services remain the same

By [Claudia Lynch](#), Benchmarks Online Editor

We here in Academic Computing and User Services have been informed that the Information Science Building will become **Sycamore Hall** effective June 1. The building designation ISB, therefore, becomes SYMR.

The following will NOT change:

- Building Number: #111
- The address: 307 S. Avenue B

As was noted in the [Link of the Month](#) in this issue, the name of the Science and Technology Library has already changed. It is now called the *The Eagle Commons Library*.

Facilities will be updating the signs and we will be updating websites and documentation when the time is appropriate. Look for further announcements on this topic in [InHouse](#) in the near future.



Contact Us:

University Information Technology

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.



UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [Summer Hours](#)

Summer Hours

By [Claudia Lynch](#), *Benchmarks* *Online* Editor

The spring semester has ended and SUM, 3WK1, 8WK1 classes begin on Monday, May 16. Following are the hours for Computing and Information Technology Center-managed facilities during this time period and on through the summer. The [Helpdesk](#) plans, at this point, to be open their normal hours, including on May 30 and July 4. They will be closed, however, to walk-ins on those days. The University is **officially closed** for [Memorial Day](#), May 30 and [Independence Day](#), July 4.

- [Data Management Services](#) will be closed when the University is closed and will otherwise maintain their normal hours over the summer.
- The **ACS General Access/Adaptive Lab** ([ISB 104](#)) will keep the following hours this summer:

Saturday: 10 a.m.-9 p.m.

Sunday: 1 p.m.-10 p.m.

Monday - Thursday: 9 a.m.-10 p.m.

Friday: 9 a.m.-9 p.m.

Hours for Other Campus Facilities

Check out the UNT Shuttle Summer Schedule here: http://www.unt.edu/transit/routes_sched.html

General Access Labs

- [WILLIS](#) (normal schedule is 24hr/7 days a week) **EXCEPT**:

UPDATED

May 16-May 27

Monday-Thursday: **Open** 7:30 a.m.- 8 p.m.

Friday: **Open** 7:30 a.m.-6 p.m.

Saturday: **Open** 9 a.m.-6 p.m.

Sunday: **Open** 1-9 p.m.

May 28-30 (University Closed May 30): **Closed**

May 31-June 2

Tuesday-Thursday: **Open** 7:30 a.m.-8 p.m.

June 3-5

Friday: **Open** 7:30 a.m.-1 p.m.

Saturday: **Open** 9 a.m.-6 p.m.

Sunday: **Open** 1 p.m.; resume 24hr schedule.

- [College of Information General Access Computer Lab \(CI-GACLab\)](#) (B205)::

Summer Hours May 16 - August 12:

Monday - Friday: 10 a.m. – 6 p.m.

Saturday: **Closed**

Sunday: **Closed**

Closed:

May 30 (Memorial Day): University Closed

July 4: University Closed

Semester Break: August 12 @ 6 p.m. - 24, Re-open August 25

- [MUSIC:](#)

3 Week 1 (3W1) May 16 - June 2:

Monday - Friday: 9 a.m. – 5 p.m.

Saturday: **Closed**

Sunday: **Closed**

Summer Hours: June 6 - August 12:

Monday - Thursday: 8 a.m. – 9 p.m.

Friday: 8 a.m. – 5 p.m.

Saturday: 10 a.m.-5 p.m.

Sunday: 1 p.m.-8 p.m.

- [PACS Computing Center](#) (College of Public Affairs and Community Service, Chilton Hall):

Summer Hours May 16 - August 12:

Monday - Thursday: 8 a.m. – 10 p.m.

Friday: 8 a.m. – 5 p.m.

Saturday: 8 a.m. - 8 p.m.

Sunday: Noon - 10 p.m.

Closed:

May 30 (Memorial Day): University Closed

July 4: University Closed

Semester Break: August 12 - 24, Re-open August 25

- [CVAD](#) (formerly SOVA):

Closed:

May 13 @ 5 p.m.-May 15, May 30, June 2 @ 5 p.m. - June 5, July 8 @ 5 p.m. - July 11

Semester Break: August 12 @ 5 P.M. - August 24, Re-open August 25

3 Week 1 (3W1) May 16 - June 2:

Monday - Thursday: 8 a.m. – 8 p.m.

Friday: 8 a.m. – 5 p.m.

Saturday: Noon - 5 p.m.

Sunday: Noon - 5 p.m.

10 Week 1 (10W1) - this includes 5 Week 1 (5W1) and 5 Week 2 (5W2) June 6 - August 12:

Monday - Thursday: 8 a.m. – 10 p.m.

Friday: 8 a.m. – 5 p.m.

Saturday: Noon - 5 p.m.

Sunday: Noon - 8 p.m.

- [COE:](#)

Maintain normal hours, Monday through Saturday. **Closed on Sundays.**

Closing: August 12; Reopen August 25 @ 7 a.m., resume normal hours.

- [COBA:](#)

All labs will be closed:

May 30 (Memorial Day): University Closed

July 4: University Closed

Business Lab (Downstairs – BA152)

Monday - Thursday: 8 a.m - 11:50 p.m.
Friday: 8 a.m. - 7:50 p.m.
Saturday: 8 a.m. - 7:50 p.m.
Sunday: Noon - 11:50 p.m.

General Access Lab (Upstairs – BA335)

Monday - Saturday: 8 a.m - 7:50 p.m.
Sunday: Noon - 7:50 p.m.

Curry Hall (Team Lab)

Monday - Thursday: 8 a.m - 11:30 p.m.
Friday & Saturday: 8 a.m. - 7:30 p.m.
Sunday: Noon - 11:30 p.m.

- [CAS:](#)

All labs will close: Friday, May 13 at 5 p.m.
Reopen: Monday, May 16; resume normal hours

All labs will be closed:

May 30 (Memorial Day): University Closed
June 3 @ 5 .m.-June 5 (Reopen June 6); **July 4:** University Closed
Semester Break: August 12 - 24, Re-open August 25 @ 7 a.m.

GAB 330

3 Week 1 (3W1) May 16 - June 2:

Monday - Thursday: 8 a.m. – 10 p.m.
Friday: 8 a.m. – 5 p.m.
Saturday: Noon - 8 p.m.
Sunday: Noon - 10 p.m.

10 Week 1 (10W1) - this includes 5 Week 1 (5W1) and 5 Week 2 (5W2) June 6 - August 12:

Monday - Thursday: 8 a.m. – Midnight
Friday: 8 a.m. – 5 p.m.
Saturday: Noon - 8 p.m.
Sunday: Noon - Midnight

GAB 550

3 Week 1 (3W1) May 16 - June 2:

Monday - Friday: 8 a.m. – 5 p.m.

10 Week 1 (10W1) - this includes 5 Week 1 (5W1) and 5 Week 2 (5W2) June 6 - August 12

Monday - Friday: 8 a.m. – 5 p.m.
Saturday: **Closed**
Sunday: **Closed**

Terrill 220

3 Week 1 (3W1) May 16 - June 2: Closed

10 Week 1 (10W1) - this includes 5 Week 1 (5W1) and 5 Week 2 (5W2) June 6 - August 12

Monday - Thursday: 8 a.m. – 8 p.m.
Friday: 8 a.m. – 5 p.m.
Saturday: **Closed**
Sunday: **Closed**

Wooten 120

3 Week 1 (3W1) May 16 - June 2:

Monday - Thursday: 8 a.m. – 8 p.m.
Friday: 8 a.m. – 5 p.m.
Saturday: **Closed**
Sunday: **Closed**

10 Week 1 (10W1) - this includes 5 Week 1 (5W1) and 5 Week 2 (5W2) June 6 - August 12

Monday - Thursday: 8 a.m. – 10 p.m.
Friday: 8 a.m. – 5 p.m.
Saturday: **Closed**
Sunday: **Closed**

• **UNT [Dallas Campus](#) - 155A**

3 Week 1 (3W1) May 16 - June 4:

Monday - Friday: 7 a.m. – 6 p.m.

Saturday: 8 a.m. - 4 p.m.

Sunday: **Closed**

Beginning June 6, maintain normal hours, Monday through Saturday.

• **[Engineering](#) General Access Lab (CENGAL, englab@unt.edu, Discovery Park, B129, 891-6733)**

3 Week 1 (3W1) May 16 - June 2:

Monday - Friday: 9 a.m. – 5 p.m.

Saturday: **Closed**

Sunday: **Closed**

Summer Hours: June 6 - August 12:

Monday - Friday: 9 a.m. – 5 p.m.

Saturday: **Closed**

Sunday: **Closed**

Closed:

May 30 (Memorial Day): University Closed

July 4: University Closed

Semester Break: August 12 @ 6 p.m. - 24, Re-open August 25

Remember:



[Get your alerts fast in case of inclement weather](#)

Visit the Emergency Management [website](#)

City of Denton Residents, [sign up](#) for the CodeRED Emergency Notification System



Contact Us:

University Information Technology

1155 Union Circle #310709

Denton, TX 76203 USA

Voice: 940-565-4068

Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338

<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.

unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.

UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [Summer's Almost Here - Time to Hit the Road Again](#)

Summer's Almost Here - Time to Hit the Road Again

By [Dr. Elizabeth Hinkle-Turner](#), Assistant Director - Academic Computing and User Services*

The waning days of every spring semester finds us here in ACUS preparing to take our act on the road. By this, I mean that it is the season when the ACUS Student Computing, Helpdesk, and Documentation areas do our most event-based educational work. Between Resident Assistant orientations, transfer student and freshman orientations and a variety of dorm-based events, university community members will seek ACUS personnel out and about and distributing and sharing all sorts of useful and vital student computing information. The summer fun usually culminates with our booth at the late August Mean Green Fling where we typically personally talk with over 500 individual students, faculty, and staff - answering any and all questions about computing on campus. Be on the lookout for us!



The Academic Computing and User Services (ACUS) Road Show at University Day 2011

Our educational presentations typically center around the information gathered at the following resources: [The Student Tour of Computing Resources](#), [The UNT Helpdesk Website](#), [The UNT General Access Computer Labs Website](#), and [The 'Is Your Computer Secure?' publication from the UNT CITC Security Team](#). Topics covered can include computer security, using online learning resources, getting software and hardware educational discounts, responsible use of social media and social networks, using EagleConnect - the student digital communication system, free and legal resources for entertainment file sharing, and using my.unt.edu. Presentations can be tailored towards students in the residence halls or surrounding apartment complexes, student orientations, staff orientations and any area where extensive knowledge of UNT student computing resources and opportunities is a must. ACUS personnel are ready and willing to hit the road, come to your area and do some educating!

Our schedule so far

In the meantime, look for us at every student orientation event occurring this summer. We typically have an information table at all of the 'resource fairs' associated with these events and at freshman orientations, half-hour-long information sessions are given. A complete schedule of orientation dates is given below:

- Transfer Student - June 3
- Transfer Student - June 10
- Freshman Orientation - June 15-17
- Freshman Orientation - June 19-21
- Freshman Orientation - June 26-28
- Transfer Student - June 24
- Freshman Orientation - June 29-July 1
- Freshman Orientation - July 11-13
- Transfer Student - July 8
- Transfer Student - July 22
- Freshman Orientation - July 17-19
- Transfer Student - July 29
- Transfer Student - August 6
- Freshman Orientation - August 17-18
- Transfer Student - August 19
- Transfer Student - August 24
and
- Other events as they are announced.

More information about these events can be found by going to the [Orientation and Transition Programs](#) website.

Would you like a presentation?

If your area would like a presentation about student computing services (or staff and faculty services), please do not hesitate to contact us here at CITC Academic and Computing Services. As you can see from this article, we are adept at taking our educational act on the road. Further information about scheduling such events can be gotten by contacting Claudia Lynch at lynch@unt.edu.

*This article is published yearly, with information updates where necessary.



Contact Us:

University Information Technology
1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green light to greatness.



Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [IBM Missing Values 19 Now Available as an Add-On for PASW Statistics 19](#)

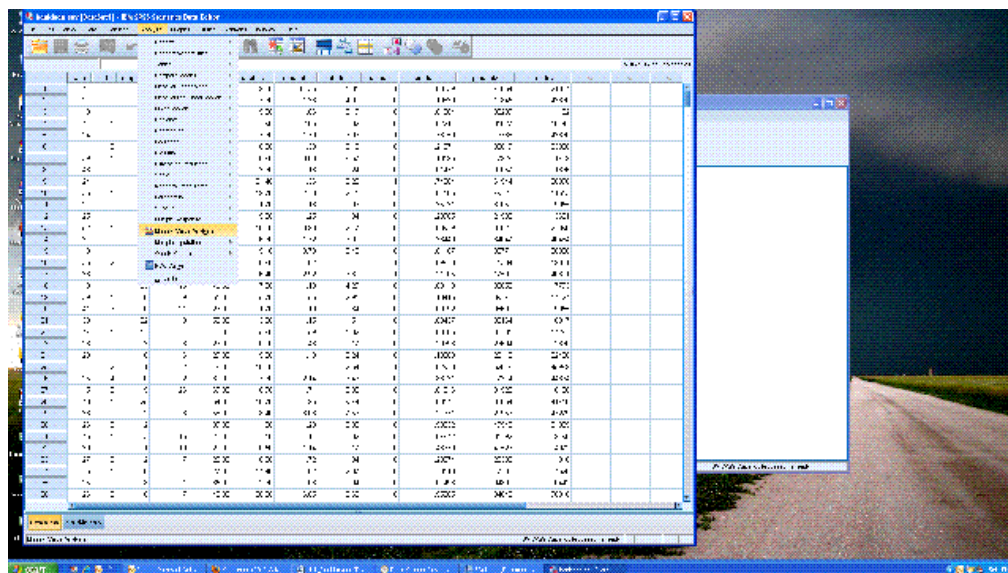
IBM Missing Values 19 Now Available as an Add-On for PASW Statistics 19

By [Patrick McLoud](#), ACUS Host Systems Administrator

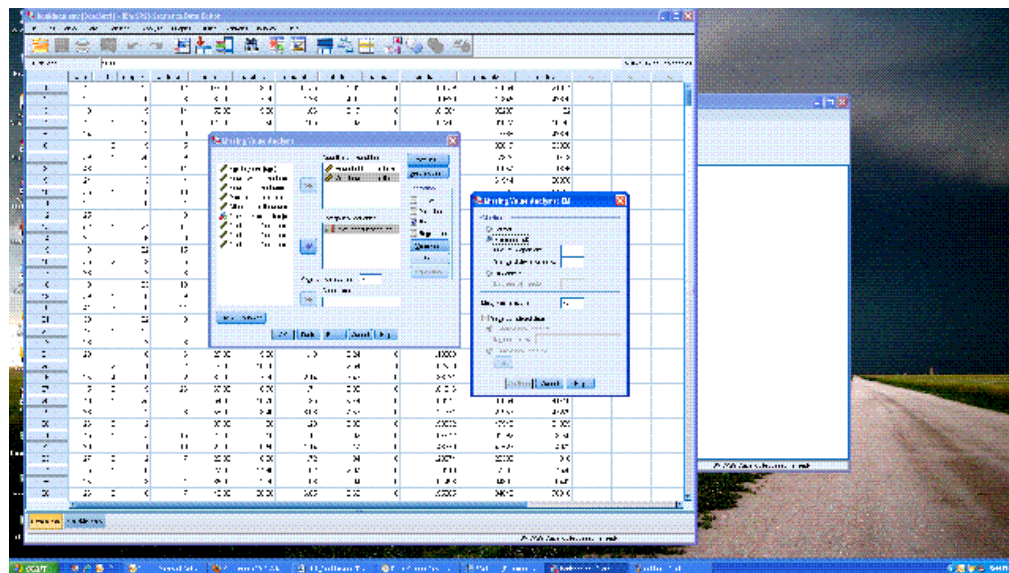
Academic Computing and User Services is pleased to announce the availability of the IBM Missing Values 19 Add-On for PASW Statistics 19. The Missing Values Add-On assists users of PASW Statistics 19 with analyzing and resolving missing values in their data. You can read more about the features available in the IBM SPSS Missing Values 19 Add-On here (<http://www.spss.com/software/statistics/missing-values/>).

At the current time, IBM SPSS Missing Values 19 Add-On is available with any network installation of PASW Statistics 19. It is specific only to version 19; it will not work with older network installations of SPSS 17 or PASW Statistics 18. It will also not work with single or site license installations of PASW Statistics 19. The Add-On is licensed through a license manager running on a university license server, so in order for a copy of PASW Statistics 19 to have the Missing Values Add-On enabled, it must be licensed from the same license server, as all network installations of PASW Statistics 19 are licensed.

The Add-On is simple to use and requires no additional license authentication. To use the Add-On as illustrated in the image below, click on the Analyze menu and then select the Missing Value Analysis option. Note the graphic to the left side of the Missing Value Analysis option; this lets you know that this particular option is licensed and available for use.



Once you've selected the Missing Value analysis option, subsequent windows will appear on your screen providing you with different options as shown in the image below.



We hope that this research tool will help you deal with missing values issues in your data. Happy computing!



Contact Us:

University Information Technology
1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.

UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [network-connection](#)

Network Connection

By [Dr. Philip Baczewski](#), Director of Academic Computing and User Services

Lordy, Lordy, FTP's Forty!

As if there isn't enough to make me feel old these days, I recently found out that [FTP](#), the Internet file transfer protocol, [turned 40 years old](#) last month. The [specification](#) for FTP was first published on April 16, 1971. FTP has long been one of the staple protocols in use on the Internet and was the primary method of exchanging files before the rise of the World Wide Web.

In the early days of the Internet, Mail, Telnet, and FTP were the primary communication tools. FTP was so important, I even wrote a book about it. Actually, it was a book [chapter](#) contributed to a volume with the silly title of "[Tricks of the Internet Gurus](#)" published in 1994. Having used and written about the Internet since 1987, I guess I was qualified at that time to have the title of Internet Guru and my wisdom included some of the more arcane features of using FTP on various multiuser operating systems.*



In the beginning

In 1994, most Internet communication was between large multiuser computers, with some gateways between the Internet and microcomputer networks, like [FidoNet](#). The World Wide Web was only 3 years old, and the wide scale commercialization of the Internet only started after 1995. FTP was a primary source of software and information before the advent of the Web. An archive of FTP sites, called [Archie](#) was the first large-scale Internet search engine, predating Google by almost a decade. (I also wrote a [book](#) chapter about Archie in long-ago times.)

The late 1980's through the 1990's saw an explosion in the development of computing technology which seems like a blur looking back on it today. At the same time, you had shareware, networking, Internet, cell phones, and personal computers all combining to quickly change the world in which we live. If you are looking for a current future with flying cars, you will be disappointed. It turns out that the most significant technology development was the [communicator](#). FTP played an important role in shaping the future that was to follow the early 1990's.

In most cases, when connecting from one host computer to another on the Internet, you would have to have an account on the remote computer in order to access and use it. One feature of an FTP server was to allow "anonymous" connections that used the word "anonymous" as the user name and accepted anything as a password, although most sites requested your e-mail address as the password and it was considered a courtesy to provide it. The anonymous feature of FTP allowed it to serve as the first large-scale one-to-many distribution system for online files.

One of the most important anonymous FTP sites on the early Internet was [WUArchive](#). A project of Washington University in St. Louis, WUArchive was a collection of software and files from many sources on the Internet and was accessed via FTP. It operated from 1988-2010. In it's heyday, it was one of the most popular sites on the Internet and the source for finding resources and software for the microcomputer systems that were available at that time (Apple, Microsoft, Amiga, Atari, etc.)

FTP today

FTP is still an important protocol underlying some kinds of Internet communication. But as a method for personal file transfer, it has fallen into disuse, partly because of concerns about security (the original FTP implementation sends all information in clear text) and partly because of its replacement by all that is Web. You might have to look closely to find an FTP client on your computer these days, although Linux and Mac OS systems have an ftp command as an

intrinsic part of the OS. Even in 1999 I [wrote](#), "Ftp and telnet remain unreplaced in the arena of Internet applications, however, they are not as commonly used now as they were originally. Our favorite World Wide Web browser has replaced our ftp client, and most people don't have the need to establish terminal sessions with remote computers anymore. Most of our current interaction with remote computers is done with our local personal computer acting as a specialized client, and most of our client activity is accomplished through our Web browser."

So, happy 40th to FTP. Maybe the current generation doesn't know you, but their world has been shaped by the large scale exchange of information that you first enabled. Maybe in another 40 years we'll have those flying cars.

*[Other authors](#) you may recognize include **Kevin Mullet** and **Billy Barron**.



Contact Us:

University Information Technology

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.



UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [link-of-month](#)

Link of the Month

Eagle Commons Library

The names, they are [a-changin'](#), and that goes for the name of the Science and Technology Library too. It is now called the *The Eagle Commons Library* and it's got some great new features. One new feature is the Collaboration and Learning Commons (CLC) which was designed specifically for group study. The CLC was featured in a *Bechmarks Online* [article](#) when it first opened in February of this year.

The CLC also has two instruction rooms that can seat up to 14 people each and teaching stations that give you access to Microsoft Office and NCast. Hardware is available to allow you to record your presentations. PC workstations are also available.

Visit the Eagle Commons Library website to find out all the wonderful new features of this re-designed campus resource:

<http://www.library.unt.edu/eagle-commons>



Contact Us:

University Information Technology

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.

UNT

[ABOUT BENCHMARK ONLINE](#) [SEARCH ARCHIVE](#) [SUBSCRIBE TO BENCHMARKS ONLINE](#)

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [helpdesk-fyi](#)

Helpdesk FYI

By [Jonathan "Mac" Edwards](#), CITC Helpdesk Manager

Configuring EagleConnect Email on your iPhone

You can easily configure your iPhone to check your EagleConnect Email using Microsoft Exchange Active Sync. The [Helpdesk](#) now recommends, when possible, using Exchange ActiveSync instead of IMAP.

1. On your iPhone select **Settings**, then choose **Mail, Contacts, Calendars**.
2. Choose **ADD Account**.
3. On the Add Account screen select **Microsoft Exchange**.
4. Enter your Exchange Account information
Email: Your EagleConnect Email address
Username: Your EagleConnect Email address
Password: Your EagleConnect Password
5. Exchange will attempt to detect, and automatically fill in, the rest of your settings. If this fails you will be asked to fill in additional information.
Email: Your EagleConnect Email address. Generally [firstlast@my.unt.edu](#).
Server: Either outlook.com or pod51000.outlook.com.
Username: Your EagleConnect Email address.
Password: Your EagleConnect Email password (should not require changes).
6. Select which parts of your account to Synchronize, then select **Sync**.



[+ BOOKMARK](#) [📄](#) [📱](#) [📘](#) [...](#)



Contact Us:
University Information Technology



Email us:
Have questions on content or technical issues? Please contact us.



UNT System:
• [UNT Home](#)

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

unt.uit@unt.edu

- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.



UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [irc-news](#)

IRC News

Minutes provided by Christine Valenzuela Recording Secretary*

The IRC -- unofficially now known as the INFORMATION TECHNOLOGY COUNCIL (ITC) -- is currently undergoing a reorganization, see the May 20, 2008 [minutes](#) for more information.**>

No IRC/ITC minutes were available for publication this month.

*For a list of IRC Regular and Ex-officio Members click [here](#) (last updated 12/12/08). Warren Burggren is now the Chair.

**DCSMT Minutes can be found [here](#).



Contact Us:

University Information Technology

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [rss-matters](#)

RSS Matters

Research and Statistical Support

University of North Texas

Cross Validation techniques in **R**: A brief overview of some methods, packages, and functions for assessing prediction models.

Link to the last RSS article here: [Matching across two groups to isolate treatment effects.](#) -- Ed.

By [Dr. Jon Starkweather](#), Research and Statistical Support Consultant

This month's article focuses on an initial review of techniques for conducting cross validation in **R**. Next month, a more in-depth evaluation of cross validation techniques will follow. Cross validation is useful for overcoming the problem of over-fitting. Over-fitting is one aspect of the larger issue of what statisticians refer to as shrinkage (Harrell, Lee, & Mark, 1996). Over-fitting is a term which refers to when the model requires more information than the data can provide. For example, over-fitting can occur when a model which was initially fit with the same data as was used to assess fit. Much like exploratory and confirmatory analysis should not be done on the same sample of data, fitting a model and then assessing how well that model performs on the same data should be avoided. When we speak of assessing how well a model performs, we generally think of fit measures (e.g. R^2 , adj. R^2 , AIC, BIC, RMSEA, etc.); but, what we really would like to know is how well a particular model *predicts* based on new information. This really gets at the goals of science and how we go about them; observation yields description, experimentation yields explanation, and all of those utilize statistical models with the goal of explanation and/or prediction. When predictions are confirmed, evidence is born for supporting a theory. When predictions fail, evidence is born for rejecting a theory.

Fit measures, whether in the standard regression setting or in more complex settings, are biased by over-fitting – generally indicating better fit, or less prediction error than is really the case. Prediction error refers to the discrepancy or difference between a predicted value (based on a model) and the actual value. In the standard regression situation, prediction error refers to how well our regression equation predicts the outcome variable scores of new cases based on applying the model (coefficients) to the new cases' predictor variable scores. When dealing with a single sample, typically the residuals are a reflection of this prediction error; where the residuals are specifically how discrepant the predicted values (\hat{y} or \hat{Y}) are from the actual values of the outcome (y). However, because of over-fitting, these errors or residuals will be biased downward (less prediction error) due to the actual outcome variable values being used to create the regression equation (i.e. the prediction model). Cross validation techniques are one way to address this over-fitting bias.

Cross validation is a model evaluation method that is better than simply looking at the residuals. Residual evaluation does not indicate how well a model can make new predictions on cases it has not already seen. Cross validation techniques tend to focus on not using the entire data set when building a model. Some cases are removed before the data is modeled; these removed cases are often called the *testing set*. Once the model has been built using the cases left (often called the *training set*), the cases which were removed (testing set) can be used to *test* the performance of the model on the "unseen" data (i.e. the testing set).

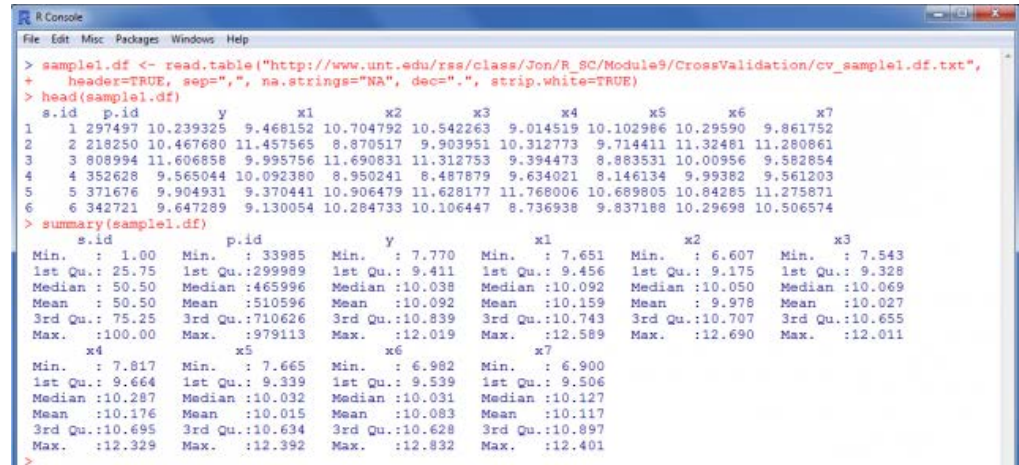
The examples below are meant to show how some common cross validation techniques can be implemented in the statistical programming language environment **R**. The examples below focus on standard multiple regression

situations using a sample drawn from a simulated population of *true scores*. Next month's article will show how the population was generated and how each sample was drawn, as well as a more in-depth exploration of how cross validation techniques address the over-fitting problem.

Example Data

The examples below were designed to be representative of a typical modeling strategy, where the researcher has theorized a model based on a literature review (and other sources of information) and has collected a sample of data. The setting for the examples below concerns a model with seven hypothesized predictors ($x_1, x_2, x_3, x_4, x_5, x_6,$ & x_7), each interval/ratio scaled, and one interval/ratio outcome variable (y). All variables have an approximate mean of 10. The sample contains two additional columns, one which identifies cases sequentially in the sample ($s.id$) and one which identifies cases sequentially in the population from which it was drawn ($p.id$). The sample contains 100 cases randomly sampled from a defined population of 1,000,000 individuals.

First, read in the sample data from the web, naming it 'sample1.df' ($df = \text{data.frame}$), and getting the ubiquitous 'head' and 'summary' to get an idea of what the data looks like.



```

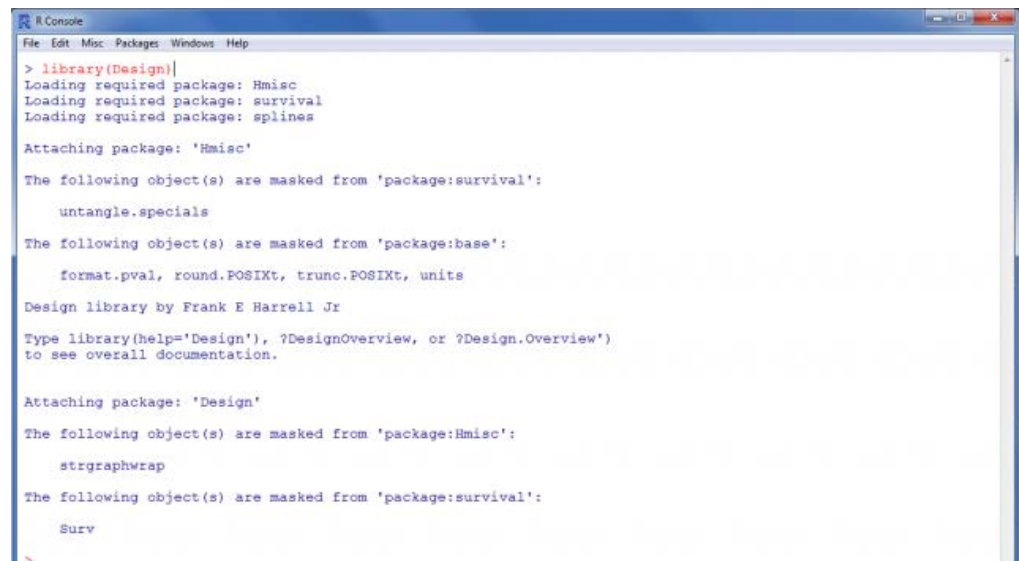
R Console
File Edit Misc Packages Windows Help
> sample1.df <- read.table("http://www.unt.edu/rss/class/Jon/R_SC/Module9/CrossValidation/cv_sample1.df.txt",
+ header=TRUE, sep=".", na.strings="NA", dec=".", strip.white=TRUE)
> head(sample1.df)
  s.id p.id      y      x1      x2      x3      x4      x5      x6      x7
1  1 297497 10.239325  9.468152 10.704792 10.542263  9.014519 10.102986 10.29590  9.861752
2  2 218250 10.467680 11.457565  8.870517  9.903951 10.312773  9.714411 11.32481 11.280861
3  3 808994 11.606858  9.995756 11.690831 11.312753  9.394473  8.883531 10.00956  9.582854
4  4 352628  9.565044 10.092380  8.950241  8.487879  9.634021  8.146134  9.99382  9.561203
5  5 371676  9.904931  9.370441 10.906479 11.628177 11.768006 10.689805 10.84285 11.275871
6  6 342721  9.647289  9.130054 10.284733 10.106447  8.736938  9.837188 10.29698 10.506574
> summary(sample1.df)
      s.id      p.id      y      x1      x2      x3
Min.   : 1.00   Min.   : 33985   Min.   : 7.770   Min.   : 7.651   Min.   : 6.607   Min.   : 7.543
1st Qu.: 25.75  1st Qu.:299989   1st Qu.: 9.411   1st Qu.: 9.456   1st Qu.: 9.175   1st Qu.: 9.328
Median : 50.50  Median :465996   Median :10.038  Median :10.092  Median :10.050  Median :10.069
Mean   : 50.50  Mean   :510596   Mean   :10.092  Mean   :10.159  Mean   : 9.978   Mean   :10.027
3rd Qu.: 75.25  3rd Qu.:710626   3rd Qu.:10.839  3rd Qu.:10.743  3rd Qu.:10.707  3rd Qu.:10.655
Max.   :100.00  Max.   :979113   Max.   :12.019  Max.   :12.589  Max.   :12.690  Max.   :12.011

      x4      x5      x6      x7
Min.   : 7.817   Min.   : 7.665   Min.   : 6.982   Min.   : 6.900
1st Qu.: 9.664   1st Qu.: 9.339   1st Qu.: 9.539   1st Qu.: 9.506
Median :10.287   Median :10.032   Median :10.031   Median :10.127
Mean   :10.176   Mean   :10.015   Mean   :10.083   Mean   :10.117
3rd Qu.:10.695   3rd Qu.:10.634   3rd Qu.:10.628   3rd Qu.:10.897
Max.   :12.329   Max.   :12.392   Max.   :12.832   Max.   :12.401

```

The 'Design' Package

Next, we specify the model. Typically, we would use the 'lm' function from the base 'stats' package to specify an Ordinary Least Squares (OLS) regression model. However, here we will use the 'ols' function in the 'Design' package (Harrell, 2009). So, first we must load the 'Design' package, which has several dependencies.



```

R Console
File Edit Misc Packages Windows Help
> library(Design)
Loading required package: Hmisc
Loading required package: survival
Loading required package: splines
Attaching package: 'Hmisc'
The following object(s) are masked from 'package:survival':
  untangle.specials
The following object(s) are masked from 'package:base':
  format.pval, round.POSIXt, trunc.POSIXt, units
Design library by Frank E Harrell Jr
Type library(help='Design'), ?DesignOverview, or ?Design.Overview)
to see overall documentation.
Attaching package: 'Design'
The following object(s) are masked from 'package:Hmisc':
  strgraphwrap
The following object(s) are masked from 'package:survival':
  Surv
>

```

Now, we can use the 'ols' function to specify the model and get a summary of it. Make sure to set the optional arguments 'x = TRUE' and 'y = TRUE' as these will save a design matrix of predictors and a vector of outcome values. These two objects will be used in the cross validation techniques below. If you are not familiar with the scientific notation of **R**, the 'e-00' refers to a negative exponent and the 'e+00' refers to a positive exponent. For example, $5.234e-03 = 0.005234$ and $5.234e+03 = 5234.00$.

```

R Console
File Edit Misc Packages Windows Help
> model.1 <- ols(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7, sample1.df, x = TRUE, y = TRUE)
> model.1

Linear Regression Model

ols(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7, data = sample1.df,
     x = TRUE, y = TRUE)

      n Model L.R.      d.f.      R2      Sigma
     100      515.3         7  0.9942  0.07509

Residuals:
      Min       1Q   Median       3Q      Max
-0.154608 -0.056157  0.004808  0.045365  0.186316

Coefficients:
              Value Std. Error    t Pr(>|t|)
Intercept -2.77898   0.169566 -16.389 0.000e+00
x1          0.86274   0.008341  103.429 0.000e+00
x2          0.54708   0.007593   72.050 0.000e+00
x3          0.23564   0.008475   27.803 0.000e+00
x4         -0.16429   0.009121  -18.013 0.000e+00
x5         -0.14094   0.009475  -14.874 0.000e+00
x6          0.01887   0.017262   1.093 2.771e-01
x7         -0.08118   0.018421  -4.407 2.836e-05

Residual standard error: 0.07509 on 92 degrees of freedom
Adjusted R-Squared: 0.9938
>

```

Next, we can begin exploring cross validation techniques. The 'validate' function in the 'Design' package "does resampling validation of a regression model, with or without backward step-down variable deletion" (Harrell, 2009, p. 187). Here, our examples focus on OLS regression, but the 'validate' function can hand a logistic model as well; as long as the model is fit with the 'lrm' function (Logistic Regression Model) in the 'Design' package. The key part of the output for this function is the 'index.corrected' measures of fit -- which corrects for over-fitting. We start with the default values/arguments for 'validate' which uses the 'boot' method (bootstrapped validation; Efron, 1983; Efron & Tibshirani, 1993). Bootstrapped validation takes B number of samples of the original data, with replacement, and fits the model to this training set. Then, the original data is used as the testing set for validation.

```

R Console
File Edit Misc Packages Windows Help
> val.boot <- validate(model.1, method = "boot", B = 40, bw = FALSE, rule = "aic",
+                    type = "residual", sls = 0.05, aic = 0, pr = FALSE)
> val.boot

      index.orig training      test      optimism index.corrected n
R-square 0.994216123 0.99423426 0.993705437 0.0005288253 0.993687298 40
MSE      0.005187536 0.00493656 0.005645568 -0.0007090077 0.005896543 40
Intercept 0.000000000 0.000000000 -0.009016284 0.0090162843 -0.009016284 40
Slope    1.000000000 1.000000000 1.000665178 -0.0006651782 1.000665178 40
>

```

Notice in the output above the index corrected estimates are all marginally worse in terms of fit and / or prediction error. In other words, the index corrected measures do not reflect the shrinkage caused by over-fitting. The "optimism" (Efron & Tibshirani, 1993, p. 248) is the difference between the training set estimates and the test set estimates and can be thought of as the amount of optimism of each initial estimate (e.g. how much the training estimates are biased).

Next, we can explore the 'crossvalidation' method, which uses B number of observations as the testing set (testing or validating the model) and the rest of the sample for the training set (building the model).

```

R Console
File Edit Misc Packages Windows Help
> val.cross <- validate(model.1, method = "crossvalidation", B = 40, bw = FALSE, rule = "aic",
+                    type = "residual", sls = 0.05, aic = 0, pr = FALSE)
> val.cross

      index.orig training      test      optimism index.corrected n
R-square 0.994216123 0.994228510 -0.06680937 1.061037880 -0.066821757 40
MSE      0.005187536 0.005175911 0.006538004 -0.001362129 0.006549665 40
Intercept 0.000000000 0.000000000 0.13240866 -0.132408663 0.132408663 40
Slope    1.000000000 1.000000000 0.98214115 0.017858848 0.982141152 40
>

```

Next, we can take a look at the ".632" bootstrapped method which corrects for the bias in prediction error estimates "based on the fact that bootstrap samples are supported on approximately .632n of the original data points" (Efron, 1983; Efron & Tibshirani, 1997, p. 552).

```

R Console
File Edit Misc Packages Windows Help
> val.632 <- validate(model.1, method = ".632", B = 40, bw = FALSE, rule = "aic",
+                    type = "residual", sls = 0.05, aic = 0, pr = FALSE)
Weights for .632 method (ordinary bootstrap weights 0.025)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0188 0.0229 0.0255 0.0250 0.0269 0.0315
> val.632

      index.orig training      test      optimism index.corrected n
R-square 0.994216123 0.994488283 0.992860602 0.0008566896 0.993359434 40
MSE      0.005187536 0.004761727 0.006551268 -0.0008618789 0.006049415 40
Intercept 0.000000000 0.000000000 0.014447216 -0.0091306407 0.009130641 40
Slope    1.000000000 1.000000000 0.998938150 0.0006710891 0.999328911 40
>

```

The 'DAAG' package

Another package which is capable of performing cross validation is the Data Analysis And Graphing ('DAAG') package

(Maironald & Braun, 2011) which also has several dependent packages.

```
R Console
File Edit Misc Packages Windows Help
> library(DAAG)
Loading required package: MASS
Loading required package: rpart
Loading required package: randomForest
randomForest 4.6-2
Type rfNews() to see new features/changes/bug fixes.

Attaching package: 'randomForest'

The following object(s) are masked from 'package:Hmisc':

  combine

Attaching package: 'DAAG'

The following object(s) are masked from 'package:MASS':

  hills

The following object(s) are masked from 'package:Design':

  vif

The following object(s) are masked from 'package:survival':

  lung

> |
```

The 'DAAG' package contains three functions for k – fold cross validation; the 'cv.lm' function is used for simple linear regression models, the 'CVlm' function is used for multiple linear regression models, and the 'CVbinary' function is used for logistic regression models. The k – fold method randomly removes k – folds for the testing set and models the remaining (training set) data. Here we use the commonly accepted (Harrell, 1998) 10 – fold application.

```
R Console
File Edit Misc Packages Windows Help
> val.daag <- CVlm(df = sample1.df, m = 10, form.lm = formula(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7))
Analysis of Variance Table

Response: y
  Df Sum Sq Mean Sq F value    Pr(>F)
x1   1  48.7    48.7  8631.8 < 2e-16 ***
x2   1  30.8    30.8  5471.0 < 2e-16 ***
x3   1   5.1     5.1   900.1 < 2e-16 ***
x4   1   1.8     1.8   312.4 < 2e-16 ***
x5   1   2.5     2.5   438.3 < 2e-16 ***
x6   1   0.2     0.2   41.2 6.0e-09 ***
x7   1   0.1     0.1   19.4 2.8e-05 ***
Residuals 92    0.5     0.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

fold 1
Observations in test set: 3 8 16 18 39 42 45 62 63 100
Predicted 11.5141 9.9671 9.3781 8.8386 9.4042 9.7155 10.0094 9.277 9.079 9.4779
y          11.6069 9.9490 9.3040 8.8825 9.3647 9.6782 10.0593 9.178 8.966 9.4293
Residual  0.0928 -0.0182 -0.0741 0.0439 -0.0395 -0.0373 0.0499 -0.099 -0.113 -0.0487

Sum of squares = 0.047   Mean square = 0.0047   n = 10

fold 2
Observations in test set: 4 9 10 19 30 47 55 78 82 93
Predicted 9.5047 11.785 10.0616 10.3095 10.7081 10.776 11.0915 8.2624 8.021 8.7906
y          9.5650 11.829 10.0821 10.3614 10.6913 10.612 11.0314 8.2295 7.909 8.8210
Residual  0.0604 0.044 0.0206 0.0519 -0.0168 -0.163 -0.0601 -0.0329 -0.112 0.0304

Sum of squares = 0.054   Mean square = 0.0054   n = 10

fold 3
Observations in test set: 35 54 56 73 74 77 87 88 89 99
Predicted 11.287 11.0791 9.713 11.2339 9.395 9.108 10.8626 9.7615 11.293 9.7556
y          11.1669 11.0216 9.847 11.2197 9.421 9.301 10.7678 9.7507 11.155 9.7173
Residual -0.118 -0.0575 0.134 -0.0142 0.026 0.193 -0.0948 -0.0108 -0.139 -0.0383

Sum of squares = 0.1    Mean square = 0.01    n = 10

fold 4
Observations in test set: 2 21 24 41 50 64 71 75 80 95
Predicted 10.5363 10.00998 10.274 8.8239 10.88 9.4221 9.8130 10.1583 10.3260 11.0276
y          10.4677 10.01757 10.417 8.7842 10.84 9.4606 9.7438 10.1024 10.3483 11.0132
Residual -0.0686 0.00768 0.142 -0.0267 -0.064 0.0284 -0.0649 -0.0646 0.0222 -0.0144
```

Some output (folds) has been omitted.

```
fold 9
Observations in test set: 6 11 22 34 37 44 51 53 65 79
Predicted 9.6250 11.57 9.265 9.8877 10.9623 8.97 9.5057 10.6938 11.6039 9.8000
y          9.6473 11.70 9.154 9.9035 11.0038 9.14 9.5326 10.7534 11.6992 9.8147
Residual  0.0223 0.13 -0.111 0.0158 0.0415 0.17 0.0269 0.0596 0.0953 0.0147

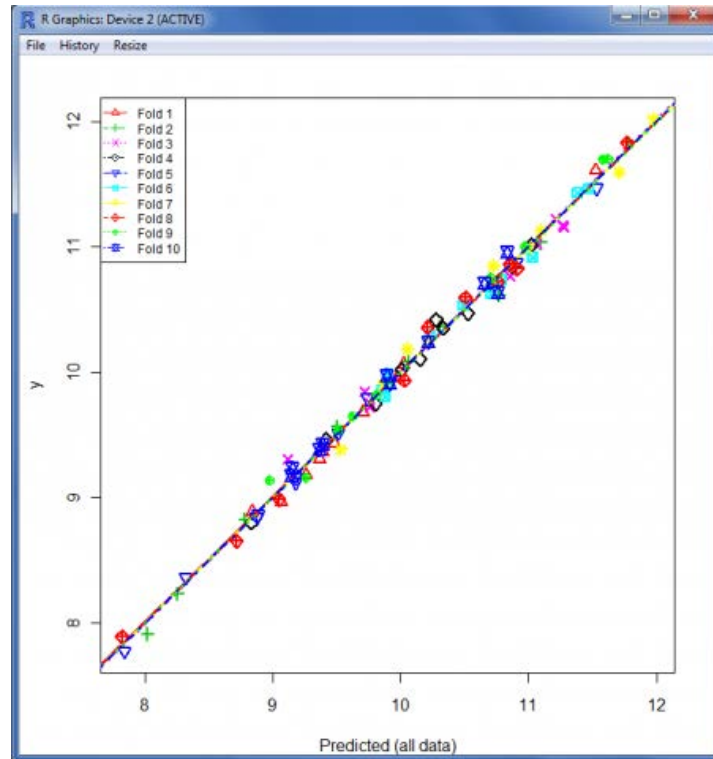
Sum of squares = 0.074   Mean square = 0.0074   n = 10

fold 10
Observations in test set: 1 20 29 32 38 52 69 72 81 91
Predicted 10.2170 10.765 9.353 9.8876 9.90237 9.1390 9.1459 9.3839 10.6522 10.824
y          10.2393 10.632 9.379 9.9687 9.90626 9.1711 9.2300 9.4207 10.7054 10.955
Residual  0.0223 -0.133 0.026 0.0811 0.00389 0.0321 0.0841 0.0369 0.0532 0.131

Sum of squares = 0.055   Mean square = 0.0055   n = 10
Overall ms
0.00624
>
```

Here, at the bottom of the output we get the cross validation residual sums of squares (Overall MS); which is a corrected measure of prediction error averaged across all folds. The function also produces a plot (below) of each

fold's predicted values against the actual outcome variable (y); with each fold a different color.



The 'boot' package

Lastly, we can use the 'boot' package (Ripley, 2010) for cross validation of generalized linear models (e.g. binomial, Gaussian, poisson, gamma, etc.). Bootstrapping can be used to correct for some of the bias associated with the other cross validation techniques.

```
R Console
File Edit Misc Packages Windows Help
> library(boot)
Attaching package: 'boot'
The following object(s) are masked from 'package:survival':
  aml
> |
```

First, we must fit the model. Our example below is *really* an OLS regression model, but if we specify 'family = gaussian' then it is the same as using 'lm'. If we had a logistic model, then we would specify 'family = binomial(link = logit)' to fit the logistic model.

```
R Console
File Edit Misc Packages Windows Help
> model.2 <- glm(y ~ x1 + x2 + x4 + x5 + x6 + x7, sample1.df, family = gaussian)
> summary(model.2)

Call:
glm(formula = y ~ x1 + x2 + x4 + x5 + x6 + x7, family = gaussian,
    data = sample1.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5173 -0.1635  0.0071  0.1343  0.4584

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.6058      0.4589   -1.32  0.19006
x1           0.8186      0.0250   32.78 < 2e-16 ***
x2           0.5777      0.0229   25.21 < 2e-16 ***
x4          -0.1748      0.0278   -6.29 1.0e-08 ***
x5          -0.1014      0.0286   -3.55 0.00061 ***
x6           0.2171      0.0479    4.53 1.8e-05 ***
x7          -0.2745      0.0520   -5.28 8.6e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.0524)

Null deviance: 89.6896 on 99 degrees of freedom
Residual deviance: 4.8774 on 93 degrees of freedom
AIC: -2.269

Number of Fisher Scoring iterations: 2
> |
```

The 'cv.glm' function "estimates the k -fold cross validation prediction error for generalized linear models" (Ripley, 2010). If k -fold is set to the number of cases (rows), then a complete Leave One Out Cross Validation (LOOCV) is

done. The LOOCV method is intuitively named; essentially, one case is *left out* as the testing set and the rest of the data is used as the training set. If this process is repeated so that each case is given a chance as the testing case, then we have the complete LOOCV method. The 'cv.glm' function returns a 'delta' which shows (first) the raw cross-validation estimate of prediction error and (second) the adjusted cross-validation estimate. The adjustment is designed to compensate for the bias introduced by not using leave-one-out cross-validation. The default for 'cv.glm' is complete LOOCV.

First, we run the common 10 – fold cross validation. Below, the majority of seed information is cut off the end of the figure.

```
R Console
File Edit Misc Packages Windows Help

> val.10.fold <- cv.glm(data = sample1.df, glmfit = model.2, K = 10)
> val.10.fold
$call
cv.glm(data = sample1.df, glmfit = model.2, K = 10)

$K
[1] 10

$delta
      1      1
0.0572 0.0567

$seed
 [1] 403 100 1118745441 1531885299 2763106 2140198868 1789174487 -845059891
 [9] -1092788124 -695153078 -320332875 1768072975 1625781270 388718272 642795651 1378608289
[17] -456324464 -949634706 376543001 762448811 -707143270 1098218844 1309676639 -302363595
[25] -1069787188 -1394539742 237698429 1290899063 -814237794 -266332584 2139062203 -2033086119
[33] 25004776 -705146090 -1756965743 2102593859 -1873443758 -1694296348 -1047834329 -1442872963
[41] -707344268 2118199898 966616741 -1075465025 229050182 559293904 -1608157197 -1169305839
[49] -16478400 -1680050402 1531439081 -2105537253 -1848449238 -1822459828 699682831 -1586685115
```

Next, we run the complete LOOCV method, specifying k as the number of rows in the sample data (nrow). Again, below the majority of the seed numbers have been left off the figure.

```
R Console
File Edit Misc Packages Windows Help

> val.loocv <- cv.glm(data = sample1.df, glmfit = model.2, K = nrow(sample1.df))
> val.loocv
$call
cv.glm(data = sample1.df, glmfit = model.2, K = nrow(sample1.df))

$K
[1] 100

$delta
      1      1
0.0567 0.0567

$seed
 [1] 403 200 1118745441 1531885299 2763106 2140198868 1789174487 -845059891
 [9] -1092788124 -695153078 -320332875 1768072975 1625781270 388718272 642795651 1378608289
[17] -456324464 -949634706 376543001 762448811 -707143270 1098218844 1309676639 -302363595
[25] -1069787188 -1394539742 237698429 1290899063 -814237794 -266332584 2139062203 -2033086119
[33] 25004776 -705146090 -1756965743 2102593859 -1873443758 -1694296348 -1047834329 -1442872963
[41] -707344268 2118199898 966616741 -1075465025 229050182 559293904 -1608157197 -1169305839
[49] -16478400 -1680050402 1531439081 -2105537253 -1848449238 -1822459828 699682831 -1586685115
[57] -1104835332 1457453906 -1691345715 2114509959 -1991874706 -1192748728 -549462709 1936425129
[65] -1067540876 -370121370 1601017505 367705155 180404 1750183540 -266762717 -667706435
```

Obviously the delta numbers match because we used the LOOCV method. Recall, the first delta value is the raw cross validation estimate of prediction error and the second is the adjusted cross validation estimate; which is supposed to adjust for the bias of not using the LOOCV method.

Conclusions

Three packages were employed to demonstrate some relatively simple examples of conducting cross validation in the **R** programming language environment. Cross validation refers to a group of methods for addressing the some over-fitting problems. Over-fitting refers to a situation when the model requires more information than the data can provide. One way to induce over-fitting is by specifying the model with the same data on which one assesses fit or prediction error. The examples here were conducted using simulated data. Rather strikingly, you may have noticed, the estimates of prediction error were not terribly different from the full sample (over-fitted) estimates, even though this sample was considerable small ($n = 100$) in comparison to its parent population ($N = 1,000,000$). These results might lead one to think cross validation and over-fitting are not things one needs to be concerned with. However, there are a few reasons our estimates here were not more starkly different than the full sample estimates and you might be surprised to find that some of our predictor variables are not at all related to the outcome variable. Next month's article will reveal the *secrets* behind those statements. However, cross validation and over-fitting are serious concerns when dealing with real data and should be considered in each study involving modeling.

References & Resources

Chernick, M. R. (2008). *Bootstrap methods: A guide for practitioners and Researchers* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Efron, B. (1983). Estimating the error rate of a prediction rule: Some improvements on cross-validation. *Journal of the American Statistical Association*, 78, 316 - 331.

Efron, B, & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The .632+ bootstrap method. *Journal of the American Statistical Association*, 92(438), 548 - 560.

Harrell, F., Lee, K., & Mark, D. (1996). Tutorial in Biostatistics: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15, 361 – 387. Available at: <http://www.yaroslavvb.com/papers/steyerberg-application.pdf>

Harrell, F. (1998). Comparisons of strategies for validating binary logistic regression models. Available at: <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/RmS/logistic.val.pdf>

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag, Inc.

Harrell, F. E. (2009). Package 'Design'. Available at CRAN: <http://cran.r-project.org/web/packages/Design/index.html>

Maindonald, J., & Braun, W. J. (2011). Package 'DAAG'. Available at CRAN: <http://cran.r-project.org/web/packages/DAAG/index.html>

Moore, A. (2008). Cross-Validation: tutorial slides. Available at: <http://www.autonlab.org/tutorials/overfit.html>

Ripley, B. (2010). Package 'boot'. Available at CRAN: <http://cran.r-project.org/web/packages/boot/index.html>

Schneider, J. (1997). Cross validation. Available at: <http://www.cs.cmu.edu/~schneide/tut5/node42.html>

Wikipedia. (2011). Cross-validation (statistics). Available at: http://en.wikipedia.org/wiki/Cross-validation_%28statistics%29

Tune in next time, *Same bat channel, same bat time...*



Contact Us:

University Information Technology

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.



UNT

ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [short-courses](#)

Short Courses

Instructor-led courses are still on hold. Please contact an [RSS member](#) or [Claudia Lynch](#) if you are interested in taking such a class or wish to have someone offer a class for your students. **SPSS and SAS courses; they are now offered online only.** RSS staff will be still be available for consultation on those topics, however. Another class available online is [Introduction to R](#).

Surf over to the [Short Courses](#) page to see instructions for accessing the SPSS and SAS online learning and other training that is available to you. You can also see the sorts of instructor led courses that have been offered in the past.

Special classes can always be arranged with the RSS staff. See "Customized Short Courses" below for further information. Also, you can **always** contact the RSS staff for one-on-one [consultation](#). **Please read the [FAQ](#) before requesting an appointment though.**

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](#) (they have a new comprehensive training curriculum), and the [Center for Learning Enhancement, Assessment, and Redesign](#). Additionally, the [Center for Achievement and Lifelong Learning](#) offers a variety of courses, usually for a small fee.

EIS training is [available](#). Questions or comments relating to EIS training should be sent to EISTCA@unt.edu.

Microsoft E-Learning

Microsoft E-Learning courses are now available for **faculty and staff** via our UNT-Microsoft Campus Agreement. Please contact Claudia Lynch at lynch@unt.edu for instructions on accessing this training. If you haven't accessed the training since last year you will need to get a new access code.

Microsoft Outlook Tutorials and much more

The Enterprise Messaging and Directory Services Group has all sorts of useful information on their [website](#), including tutorials and FAQs. The home page displays a list of their newest tutorials with tutorial topic pages displaying the most accessed pages. You can search the site for whatever you're interested via a Search Box on the left-hand side of the page.

Central Web Support

Consult Central Web Support for assistance in acquiring "Internet services and support." As described on their [website](#):

CWS provides Internet services and support to UNT faculty, staff and students. Services include allocating and assisting departments, campus organizations and faculty with web space and associated applications. Additionally, CWS assists web developers with databases and associated web applications, troubleshooting problems, support and service.

CLEAR (was Center for Distributed Learning)

[CLEAR](#) offers courses especially for Faculty Members. A list of topics and further information can be found [here](#).

The center also offers a "Brown Bag" series which meets for lunch one **Wednesday** a month (recently changed from the first Thursday of each month) at Noon in Chilton 245. 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 [CLEAR Website](#).

Ed2go

Ed2go are courses that are offered, for a fee, to UNT faculty, staff and students as well as the general public. According to the CALL [website](#):

CALL has partnered up to provide online learning on a variety of topics. From standardized test preparation to database programming to training for libraries and their staff, there's a variety of areas from which to choose in online learning.

The online minicourses, provided in conjunction with Ed2go, are standardized 12-lesson modules released over a six week period. (Courses are active for eight weeks to provide some flexibility). Each module features a quiz. Lessons are instructor-led and course participants and instructor communicate through a course discussion board. Lessons can be downloaded and saved. At the end of the course there is a final quiz. A passing grade opens a window that allows students to print out a course completion certificate.

All courses are \$89, and UNT faculty, staff and students may receive a \$10 discount.

For additional information surf over to <http://www.ed2go.com/unt/>

Information Security Awareness

The UNT Information Security team offers Information Security Awareness [courses](#) to all UNT faculty and staff. Topics to be covered will include workstation security, sensitive data handling, copyright infringement issues, identity theft, email security, and more.

It is a policy requirement that ALL staff take an information security course at least once a year.

Please contact [Allan Anderson](#) in CITC Information Security if you have any questions, or would like more information about the online training. **Either attending a live class or going through the online training will count towards your training requirement.** You can also request a customized course to be taught for your department.

Alternate Forms of Training

Many of the General Access Labs around campus have tutorials installed on their computers.

See <http://www.gal.unt.edu/> for a list of labs and their locations. The Willis Library, for example, has a [list of Tutorials and Software Support](#). The Library Instructional Unit also offers workshops and training, including "tech skills" training. Visit their websites for more information: <http://www.library.unt.edu/library-instruction>

The [Training Website](#) has all sorts of information about alternate forms of training. Computer Based Training (CBT) and Web-based training are some of the alternatives offered, although due to the rising costs of training, shrinking budgets and changing technology, computer-based training at UNT is in a state of transition. For up-to-date information on CBT at UNT, see the CBT [website](#).

Gartner Research Services

Way back in 2006 we announced [Gartner Core Research Services Now Available to the UNT Community](#). Our subscription for Gartner services has always included **all** UNT faculty, students, and staff. All you need to do to access the subscription is to **log into the UNT Gartner portal page** at <https://gartner.unt.edu/>. Once you have logged in, you can view upcoming webinars: <http://www.gartner.com/webinars/> and listen to Gartner podcasts here: http://www.gartner.com/it/products/podcasting/asset_137461_2616.jsp.

State of Texas Department of Information Resources

Another possible source of training for staff and, perhaps, faculty members is the Texas Department of Information Resources. A look at their Education and Training [website](#) reveals some interesting possibilities.

New Horizons Computer Learning Centers

New Horizons is a DIR vendor, which means that state agencies, like UNT, get special pricing for their services negotiated at the State level (click [here](#) for more information about DIR vendors). [New Horizons](#) offers courses at their own facilities in Dallas and Fort Worth, but will arrange for onsite training as well.



Contact Us:

University Information Technology

1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green
light to
greatness.



UNT

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#) » [issues](#) » [2011-05](#) » [staff-activities](#)

Staff Activities

Transitions

New Employees:

- **Brenton Micheal Copeland**, Business Services Support/Student Development (AIS).
- **Kathy Cooley**, IT Specialist, UNT Administration Support, Departmental IT Support.

No longer working in the Computing and Information Technology Center:

- **Susan Pierce**, Manager, Developer, UNTranet-SharePoint. Retired!
- **Georgia Baldwin**, Administrative Specialist, Telecom Services.

Changes, Awards, Recognition, Publications, etc.

Service to UNT

Congratulations to **Ginger Boone**, IT Manager, Business Services Support/Student Development, Departmental IT Support (AIS) on her **20 years** of service to UNT. Congratulations also to these folks who are celebrating:

15 years of Service

- **Saeid Parivash**, IT Programmer Analyst, EIS Tools Support, AIS Tools & User Services.
- **Monty Slayton**, IT Manager, Storage & Backups, Enterprise Systems Technical Services (AIS)

10 years of service

- **Thomas (T.J.) Adamowicz**, IT Specialist, Enterprise Messaging and Directory Services Group, Communications & Collaboration Services.
- **Danja Franklin**, IT Programmer Analyst, EIS Application Infrastructure Management, AIS Tools & User Services.
- **David George**, IT Technician, EIS Application Infrastructure Management, AIS Tools & User Services.
- **Silvester Montalvo**, Communications Specialist, Infrastructure Engineering & Installation, Communications & Collaboration Services.

They were all [recently recognized](#) in *InHouse*.

Congratulations to the Graduates

We have a couple of distinguished Helpdesk employees who have just graduated:

- **Andrew Neiford:** Graduated with a BA in Communication Studies with a Marketing minor. Was recently accepted into Texas Wesleyan School of Law for Fall 2011.
- **Trayton Oakes:** Graduated with a BA in Political Science and Economics (double major), Magna Cum Laude, Honors College Recognition Award.

We're proud of you too!

These folks deserve a round of applause also:

- **David Heflin**, Helpdesk Consultant -- David has recently received honors for some of his Advertising and Design work: Fort Worth ADDYs: 1 Silver, 2 Golds (advanced to District 10 ADDYs), and a Best of Show Award. District 10 ADDYs: 2 Golds, both advanced to National ADDYs. National ADDYs: 1 award to be announced June 4th. DSVC national Student Show: "Best Pro-bono Ad Campaign." *Print Magazine* 2011 Hand Drawn Competition: Finalist, Some work may appear in the June issue.
- **Monty Slayton**, IT Manager, Storage & Backups, Enterprise Systems Technical Services (AIS) - was re-elected Mayor of The City of Lake Bridgeport for another term. He has also been selected for the Board of Wise County Committee on Aging/Meals on Wheels. Slayton has also, in the past year, been selected to participate in the Compellent Customer Council (C3). This is a group of select Compellent accounts that meet to give the user's viewpoint on current and future company directions.
- **Allan Anderson**, Information Security Analyst, was one of the keynote speakers at a recent seminar, "*Security in Emerging Technologies*," offered by the Center for Decision and Information Technologies here at UNT. It was a part of their 2010-2011 Seminar Series on Information Security and Privacy.



Contact Us:

University Information Technology
1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

Site last updated on April 22, 2016

[Disclaimer](#) | [AA/EOE/ADA](#) | [Privacy Statement](#) | [Web Accessibility Policy](#) | [State of Texas Online](#) | [Emergency Preparedness](#)

Benchmarks

A green light to greatness.



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

[Home](#)

Today's Cartoon



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



Contact Us:

University Information Technology
1155 Union Circle #310709
Denton, TX 76203 USA
Voice: 940-565-4068
Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338
<http://it.unt.edu/benchmarks/>



Email us:

Have questions on content or technical issues? Please contact us.
unt.uit@unt.edu



UNT System:

- [UNT Home](#)
- [UNT System](#)
- [UNT Dallas](#)
- [UNT Health Science Center](#)

