Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Benchmarks - June, 2011

Campus Computing News

Hooper named Deputy CIO for the UNT System

By Claudia Lynch, *Benchmarks Online* Editor

John Hooper, who became Acting Vice President for Information Technology and Chief Information Officer of UNT last year, has assumed an additional role as Deputy CIO for the UNT System.

Read more BOOKMARK

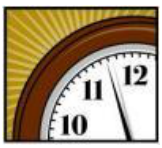
#Phishing

By **Allan Anderson**, Information Security Analyst

A __phish__ is a specific type of electronic communication. It is unsolicited, and usually contains a link or, more rarely, an attachment. The goal of these emails is to acquire a username and password, or other sensitive information from the person.

Read more BOOKMARK

Summer Hours



By Claudia Lynch, *Benchmarks Online* Editor

The 3W1 semester has ended and the SUM, 8W1, 5W1, and 10W classes are in full swing. Following are the hours for Computing and Information Technology Center-managed facilities during this time period and on through the summer.

Read more BOOKMARK

EIS Status Report

By Claudia Lynch, *Benchmarks Online* Editor

By the Numbers

Classroom support:

- 651,621 total classes/events scheduled
- Sept 2009-Aug 2010 -- 267 scheduled rooms with 208,343 classes-events scheduled
- Sept 2010-Aug 2011 – 267 schedule rooms with 195,059 classes-events scheduled

The latest issue of *theEnterprise*, Enterprise Information System Status Report, is now available.

Read more

EDUCAUSE Annual Conference: Still Time to Register!

By Claudia Lynch, *Benchmarks Online* Editor

There's still time to register for the EDUCAUSE 2011 annual conference. It is being held in Philadelphia, Pennsylvania this year (October 18–21; Online: October 19–21).

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home

Campus Computing News

Hooper named Deputy CIO for the UNT System

By Claudia Lynch, *Benchmarks Online* Editor

John Hooper, who became Acting Vice President for Information Technology and Chief Information Officer of UNT last year, has assumed an additional role as Deputy CIO for the UNT System. He will remain the CIO of UNT and continue to report to President Rawlins.

Mr. Hooper will become the day-to-day chief operating officer of the shared services organization in September, as we continue to move to a shared IT organizational structure. He will split the shared services duties with Michael Di Paolo, Associate Vice Chancellor and Chief Information Officer of the UNT System and will act as the System CIO in Di Paolo's abscence. See these articles in *inHouse* for further information about Hooper's new role and Di Paolo's recent appointment:

- DiPaolo to supervise information technology services
- Additional system technology duties assigned to Hooper

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » #Phishing

#Phishing

By Allan Anderson, Information Security Analyst

A __phish__ is a specific type of electronic communication. It is unsolicited, and usually contains a link or, more rarely, an attachment. The goal of these emails is to acquire a username and password, or other sensitive information from the person. Phishing is a big business. People often ask, 'Who actually clicks on these links?' Studies have proven that as long as at least a small percentage of links are clicked phishers can continue to operate. In 2007 it is estimated that 3.6 million adults lost $3.2 billion dollars due to phishing scams. Even here at the University we continue to see phishing emails hit our inbox from time to time. So lets go over what a phish is, how to recognize one, and what you can do about it.

Usually these types of emails are sent out to masses of people at once. The author will often pose as a bank, an IT Administrator, or sometimes as the University itself. To avoid filters, these messages are usually brief. The message usually states something like, 'There has been an error with your username and password, go here to change/update it.' A link will follow.

Now if this sounds like it might be something that you might actually receive from an IT Administrator, you may be wondering how you can go about distinguishing a phish email from legitimate emails. There are several giveaways to their techniques that you can look for. First look at the links provided. Now these can easily be made to appear legitimate, but some phishers are extremely lazy. For example look at this link:

https://www.unt.edu.lb.biz/password/accountchange.asp

It may appear that the first part of the link is good, after all it says unt.edu right there! However the __top level__ domain, the final part before the first single forward slash, is something suspicious like lb.biz. Certainly not a place that you want to go. Now as I said before, looking at the link in the email is not infallible, it is actually easy to make those links look like they are going to a trusted site, when in reality they are not even close. For example a link could be something like this in your email:

https://trusted.unt.edu/

Which looks fine! However the link might actually direct you to something like:

http://froYo.ca/pwnM3/n0Wz/

Not a place that you want to go! There are ways to check where a link actually goes without clicking on it. One trick is to hover your mouse cursor over it, and a small dialogue box will pop up with the actual link. This won't work with the new, popular link-shortening services, however. Another thing that you can check is the email address of the sender. If the sender is something like 'cutiegirlz3357@shadybiz.com' it is extremely unlikely that they are sending you important, legitimate emails. Of course there are other methods that these scammers can use. One giveaway is that the entire message is actually an image. To check this you can see that the text is not selectable by clicking and dragging across the text. The University spam filter actually does a good job of cleaning that type of stuff out. Obvious misspellings and vague signatures are also signs that the email may be a phish.

There is a small subset of phishes that are targeted specifically at a certain group of users. These happen when the scammer know (or think that they know) at least one thing about the user that they are targeting. For example, they may know that you work for the University, that you bank with a specific institution, or that you are a eBay user. They can then craft specific messages that target you. Sometimes they have even hacked a legitimate account at an organization, then the from field is even legitimate! Links in these emails often direct to sites that are made to look exactly like the websites of trusted organizations, all in an attempt to steal your credentials. Another popular trick is to send you to the actual website, but you get a fraudulent pop-up window that asks for your username and password, 'just to confirm', or 'for added security'. Of course, entering your username or password at any point results in them being stolen!

So you have looked over a specific email and decided that it is fraudulent. What should you do now? Deleting it is a

safe bet, but you can always forward it to spam@access.ironport.com to help us here at the University filter it for other users in the future. Those are your best bets if you know that an email is fake. Another measure that you can take is to _never_ click on links that are sent to you through email. It is always a good idea to type in links yourself into your browser. It may be a little more time consuming, but can save you some pain in the end. By working together and exercising a little caution, we can greatly reduce the amount of UNT credentials compromised in this way, as well as protect you from other scams looking for your bank accounts. Looking through your email carefully can pay great dividends in security. If you have any further questions, please email us at security@unt.edu .

About the title

If you go to Twitter and search on **#Phishing** you'll turn up lots of warnings about scams going around. Not a bad habit to get into, actually. - *Ed.*

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Summer Hours

By Claudia Lynch, *Benchmarks Online* Editor

The 3W1 semester has ended and the SUM, 8W1, 5W1, and 10W classes are in full swing. 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 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).

- 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:

July 4*: University Closed*
Semester Break*: August 12 @ 6 p.m. - 24, Re-open August 25*

- MUSIC:

  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:

  July 4*: University Closed*
  Semester Break*: August 12 - 24, Re-open August 25*

- CVAD (formerly SOVA):

  Closed:

  July 8 @ 5 p.m. - July 11
  Semester Break*: August 12 @ 5 P.M. - August 24, Re-open August 25*

  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:

  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 be closed:

July 4*: University Closed*
Semester Break*: August 12 - 24, Re-open August 25 @ 7 a.m.*

GAB 330

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

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

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

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**

Beginning June 6*: maintain normal hours, Monday through Saturday. Sunday:* ***Closed***

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

Summer Hours: June 6 - August 12:

Monday - Friday: 9 a.m. – 5 p.m.
Saturday: ***Closed***
Sunday: ***Closed***

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

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » EIS Status Report

EIS Status Report

By Claudia Lynch, *Benchmarks Online* Editor

The latest issue of *theEnterprise*, Enterprise Information System Status Report, is now available.Click on the link below to read about all the exciting things going on within EIS and their relationship to UNT and the UNT System. For example "Update on the Business Service Center and Shared Services," "Faculty Benefit from New Grade Roster Feature," "New Communication Feature greatly Improves International Recruitment Efforts," and many more informative articles..

http://eis.unt.edu/the-enterprise/2011/may2011.htm

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Network Connection

By Dr. Philip Baczewski, Director of Academic Computing and User Services

Something Old, Something New, Something Borrowed, Something Blue

June is bridal season, so I thought we'd follow that theme and explore some of the latest Internet news. Perhaps you will be reading this on the beach in between summer blockbusters. Or maybe you'll find respite from the early 100 degree summer heat in the cool nineties we're about to discuss. Read on.

Something Old

On June 1, news hit Slashdot that io.com also known as Illuminati Online would be shut down. Now aside from being one of the cooler site names seemingly ripped from the pages of a Neal Stephenson novel, Illuminati Online grew from a BBS system that that served as an online support community for a game called Illuminati produced by a company named Steve Jackson Games back in the 1990's. On March 1, 1990, the company's offices in Austin Texas were raided by the Secret Service in a theoretical hacker crackdown, and computers were confiscated including the one running the Illuminati BBS. But it turned out that the raid was totally unjustified according to a federal court and $50,000 in damages and $250,000 in attorney's fees later, the company got its equipment back and Illuminati Online was back online.

1990 was when the Internet began gaining more traction among institutions of higher education. NSFNet had been established as an Internet backbone linking four major supercomputing centers in the U.S. and providing educational and research institutions (including UNT) with a path to communicate to a growing number of locations. Against this backdrop, the Steve Jackson Games incident served to focus attention on the growing importance of computing and network technology and helped spur growth of Internet civil liberties organizations like the Electronic Frontier Foundation. Even Neal Stephenson couldn't have made that up.

Something New

June 8 was World IPv6 Day. In case you forgot, IPv6 is a new way of generating numeric addresses for computers on the Internet, since we've already surpassed the billions of computers that IPv4 can support, and are already working on adding multiple trillions of devices. World IPv6 day was a chance for major Internet players like Facebook, Yahoo, and Cisco and government agencies and educational institutions to turn on their IPv6 routing and pretend like IPv4 never existed. But in spite of the exhaustion of IPv4 addresses, and the apparent benefits of IPv6, most sites are still not prepared for the migration to IPv6. So, what else is new?

Something Borrowed

In what seemed to be a play borrowed from Pennsylvania's playbook, North Carolina has passed a law that prevents municipalities from creating their own broadband networks for their citizens to access the Internet. It seems that to do so would be interfering with the "marketplace." Yet, you hardly have a marketplace of choices when trying to get high-speed Internet service for your home. Choices are usually limited to your cable TV provider or local telephone company and they both now seem more interested in selling you cable TV subscriptions than Internet access.

Perhaps the wireless marketplace will save the day, since all things wireless are rapidly increasing in popularity. But alas, word also comes that mobile carriers may stifle development in that arena with metered data plans and anti-competitive actions. We seem to be stuck on a giant Monopoly board, and instead of railroads, we have Verizon,

AT&T, Sprint, and T-Mobile on four sides. The Board Walk and Park Place are already off the market, so don't count on building your house there.

Something Blue

In case you missed it the .xxx domain was recently approved by the Internet Corporation for Assigned Names and Numbers (ICAAN) for use by adult and pornographic web sites. Some may remember that the previous U.S. administration blocked the establishment of the .xxx domain in 2007. Times change and you can now be sure as to where you can find porn on the Internet.

But wait, there's more. Just send a mere $185,000 to ICAAN and you can have your own personal domain suffix, provided you can justify a reason for having it and it doesn't infringe on anyone else's intellectual property. That's right, for only $185,000 you can pick what will appear at the end of your web site address. It really is going to be a long hot summer.

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » link-of-month

Link of the Month

UNT Staff Council Perks Program

One of the perks of working at UNT is, literally, perks! The Staff Council regularly negotiates with area businesses to bring discounts to UNT employees. Visit the Staff Council Perks page to find out what is currently available. Some recently added perks are (via *InHouse*):

- 20 percent discount at fromyouflowers/unt.com.
- HVAC discounted tune up in time for summer heat, offered by Air Solutions
- Car maintenance specials at Goodyear Tire and Auto Service Center, 521 University Drive.

You'll need to login with your EUID and password to see all the perks:

http://staffcouncil.unt.edu/content/perks

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » helpdesk-fyi

Helpdesk FYI

By Jonathan "Mac" Edwards, CITC Helpdesk Manager

EagleConnect Exchange ActiveSync Changes

The EagleConnect ActiveSync settings we have recommended, for connecting with various mail clients, have changed slightly. Previously we had recommended that the Server be set to **outlook.com**. This should now be set to **pod51000.outlook.com**.

iPhone Instructions:

You can easily configure your iPhone to check your EagleConnect Email using Microsoft Exchange Active Sync.

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: 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.**

Android Instructions:

Do to the wide variety of Android phones, mail applications, and ongoing issues* please use the general configuration settings below to configure your phone. If automatic configuration does not work please try manual settings.

Email: Your EagleConnect Address. Generally firstlast@my.unt.edu

Password: Your EUID Password.

Domain: Leave Blank.

Username: Your EagleConnect Address. Generally firstlast@my.unt.edu

Server: outlook.com

SSL Connection: Check if available.

Certificate: Accept if asked.

*At this time not all Android native mail clients are able to send EagleConnect email via Exchange ActiveSync. HTC phones appear to be the most likely to be affected. If you cannot send when using ActiveSync you may use a 3rd party app such as Touchdown, or use IMAP.

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » 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.**

July 20, 2010

Members present: Joe Adamo (CITC/AIS), Philip Baczewski (CITC & DCSMT), Michael Baggett (COVA), Jim Byford (CENG), Tim Christian (CAS), Matt Cooper (GSC), Yunfei Du (SLIS), Katy Gallahan (President's Office), Jane Himmel (LEPG), Elizabeth Hinkle-Turner (SCPG), John Hooper (CITC/AIS/EIS), Scott Jackson (Libraries), Kasey Close (GALMAC), William Moen (Academic Admn.), Jon Nelson (COM), Patty Palumbo (Admn. Affairs HRS), Patrick Pluscht (CLEAR), Ruthanne Thomas (RED), Will Senn (Decision Support), Kiseol Yang (SMHM) Members absent: Joe Adamo, Judith Adkison, Joel Arredondo, Tim Christian, Will Clark, Jim Curry, Jane Himmel, Abraham John, Alan Livingston, Jon Nelson, Patty Palumbo, Scott Warren, and Kiseol Yang. Joel Arredondo (SGA), Cengiz Capan (COB), Will Clark (UNT System Ctr.), Jim Curry (Classroom Support), Renee Drabier (IRT-HSC), Bruce Hunter (CAS/IAS), Abraham John (Student Development), Troy Johnson (Enrollment Management), Michael Monticino (CAS ITC Chair), Ramu Muthiah (CPACS), Charlotte Russell (CITC/AC), Joey Saxon (Finance Admn.), Scott Warren (Faculty Senate) Guest(s) Present: Paul Hons (for Judith Atkinson COE), Peter Sneehan (Decision Support), Renee Sims (for Joey Saxon(Finance Admn.)

Called to Order: 2:30 p.m. in GAB 309

The minutes of the previous UNT ITC meeting, September 21st, 2010 were approved with corrections.

Student Digital Communications Strategy Committee Report

Tim Christian opened the meeting for Dean Monticino and introduced Elizabeth Hinkle-Turner to discuss her "Student Digital Communications Strategy Committee Report". Elizabeth Hinkle-Turner requested answers for her business needs from IT, and submitted a chart which represents the general leaning of message types and discussed some items from the chart. Elizabeth Hinkle-Turner requested the members email their suggestions to achieve digital communications a success (eliminating email ambiguity, the enhancement of the existing communication resources for device technologies), and secondly making Eagle Connect more attractive (she noted Eagle Connect offers Office in the Clouds online, Skydrive offers 25 gigs of storage space; discussed the attributes of the Recommended Digital Communication Tools from the chart, some of which business units, and another group working on Advising Communications, Chaired by Mike McKay, are wanting to have greater proliferation of some of the EIS tools, one specifically noted is CRM that allows a 360 degree view of how the student is communicating and requires an action. CommGen a tool for postal mail and email communication which stores an archive of the letter that is sent, and also business units are wanting a phone call tool like Eagle Alert but she feels will dilute the effectiveness of Eagle Alert, also making links to Eaglenet more prominent. Elizabeth Hinkle-Turner has not received a response to an email she sent outlining these requests to IT people. Elizabeth Hinkle-Turner requested the members email if they find typos in her report. Joe Adamo suggested create a project list to CITC. Tim Christian suggested giving the newer chart to the Associate Deans and IT Shared Services; John Hooper agreed. Joe Adamo stated the units need to create a consensus of three prioritized solutions to use for aproject to present to CITC. Tim Christian suggested only accept cases that are likely to be successful. Paul Hons stated to streamline the current system based on the chart, prioritize the list, outcomes need to be assessed, and include plans for future communications (ex: iPad, text, etc…). Tim Christian summarized the discussion for everyone by stating the following:

- 1st objective – analysis, streamlining, communication plan, student outcome
- 2nd objective – cost efficiency
- 3rd objective – prioritize top three items from the list

- 4th objective – send the priorities to Christine to post online for members to review and for John to offer advice for the best path for sending to Shared Services for consideration

Lecture Capture System

Tim Christian introduced Jane Himmel to discuss voting results for the Lecture Capture System recommendation. Jane Himmel stated the results voted in favor of moving forward with a pilot on LCS and has begun working with Jim Curry to roll it out for Spring and to design a larger pilot for summer and fall; the second recommendation for ITC is to consider long-term funding for Lecture Capture on campus if the outcome of the pilot shows benefit to students and faculty, and CLEAR is willing to pay for hosting and support. Tim stated to move forward with the pilot project, requested an estimate of the cost of production, and asked for funding ideas suggesting Student Fees. Phillip Baczewski offered to learn who decides allocation of Technology Use Fees for Tim Christian.

Blackboard Learning Management System Upgrade

Jane Himmel announced that an upgrade to the Blackboard Learning Management System from Vista 8 to Learn 9.1 will take place over the next two years. Members of CLEAR and CITC's Distributed Learning Support (DLS) unit will meet with Blackboard representatives in early December, 2010 to discuss enhancements and to learn more about a product called Blackboard Mobile Learn.

Other Course Hosting Possibilities

\Jane Himmel also shared that a CAS department has been considering hosting a few of its courses on Pearson Education LMS but stated that problems with accessibility compatibility with Windows 7 may be problematic for students. Jane Himmel will continue gathering information from the department to determine why the change is being considered. She also mentioned that another option is to use Pearson's Passport plug-in for Blackboard Learn. Bill Moen suggested potentially doing another learning management system evaluation project to identify why people are looking at different options before moving to Blackboard Learn 9. It was agreed that such a project would be too time-consuming to repeat frequently, but that gathering feedback from faculty and students on satisfaction with the learning management system would be advisable.

IT Shared Services

John Hooper updated members stating IT Shared Services has developed a Shared Service Catalog, there are 25 services provided by UNT, and are developing a cost for the services, and looking at service consolidation, what is the model for IT cost recovery and what is the formula for allocating cost with HSC and UNT; under current discussion is an IT governance model that covers all the shared services, and determine priority in resources allocations, changed the IT reporting structure, recruiting a CIO, and developing a business service center.

ITC on SharePoint

Tim Christian stated ITC is moving forward with migrating from the public site to the private SharePoint site, and it was agreed by the members for calendar appointment of the meetings that are schedule. Tim Christian stated it is time to verify membership of representatives and will be sending out memos to the Dean's and Chairs to select their members.The next UNT ITC meeting is scheduled for January 18th, 2011 in GAB 210M; will call for Agenda items two weeks before the meetings and request reports to be read before the meeting.Patrick Pluscht will send an invitation Technology Showcase on November 29th, 2010 at 2:00 p.m. with a cookout of free lunch.

Migration to Unified Messaging

Joe Adamo stated the migration to unified messaging is progressing well. He is meeting with HSC to set up test mailboxes, and his goal is to have system offices (900 on UNT campus) converted by the next quarter. Joe Adamo also stated a meeting with the vendor has been arranged to diagnose issues with wireless connectivity.

Meeting adjourned.

---

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

**DCSMT Minutes can be found here.

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

RSS Matters

Research and Statistical Support
University of North Texas

Examination of Cross Validation techniques and the biases they reduce.

Link to the last RSS article here: Cross Validation techniques in R: A brief overview of some methods, packages, and functions for assessing prediction models. -- Ed.

By Dr. Jon Starkweather, Research and Statistical Support Consultant

The current article continues from last month's brief examples of how to conduct some cross validation techniques in the R statistical programming environment. This month we focus more on what cross validation does and what problems it addresses. For a slight review of the basis of cross validation techniques, consider the following paragraphs.

When building a prediction model, be it a standard regression model or a more complex model, the model should not be initially fit to the same data on which prediction error estimates are also calculated. However, due to practical limitations, it is often not feasible to collect a new data set on which to evaluate the fit and / or predictive accuracy of a model. Using the same data to assess model fit / accuracy as was used to initially build the model carries with it the bias of over-fitting. Over-fitting essentially means the model requires more information than the data can provide. Over-fitting is one aspect of the larger issue of what statisticians refer to as shrinkage (Harrell, Lee, & Mark, 1996). When over-fitting occurs, the parameter estimates will be exaggerated and prediction error estimates will be downwardly biased; meaning, they will indicate less prediction error and better model fit than is really the case. This results from evaluation of predicted values against actual outcome values which were used to build the model (initial fit). One might refer to this as *double dipping.* For instance, in a standard regression analysis, the predicted values (y-hat or $\hat{y}$) are subtracted from the actual values of the outcome (y) in order to create and then evaluate the residuals. Bias is introduced because the residuals do not accurately reflect an estimate of prediction error with new data. Using the actual values of the outcome (y) for both model fit and in the evaluation of prediction error provides a bias view as to how the model will perform with new values of the predictor variables.

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 data used in this article was simulated and contains one continuous (interval/ratio) outcome variable (y) and seven other continuous (interval/ratio) variables (x1, x2, x3, x4, x5, x6, & x7). All 8 variables have an approximate mean of 10 and are (multivariate) normally distributed. Initially, a population was created (N = 1,000,000) and from it samples were randomly drawn by randomly selecting population identification numbers (p.id). The first sample (n = 100) was drawn and can be read in from the web (as will be shown below). The first sample contains an additional column for identifying each case of the sample (s.id).

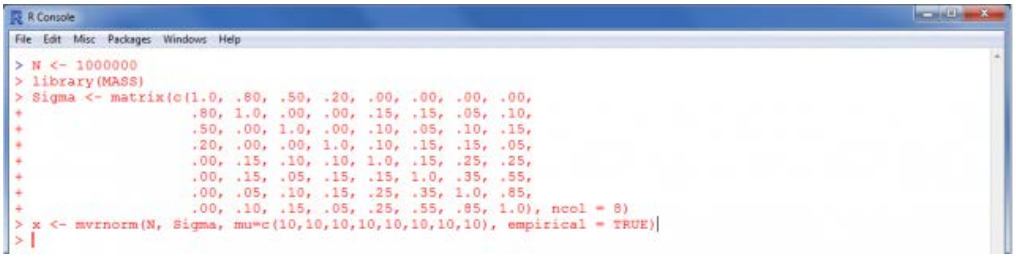
The general idea below is that we are working from a perspective of a researcher or research team with a theoretically

motivated study in which we *believe* seven measured variables (x1 - x7) predict the outcome (y). Our goal is to use cross validation to estimate the prediction error of our model and if warranted, identify a linear model (OLS regression) which offers the lowest prediction error estimate(s).

Of course, the *real* goal of this article is to show various approaches to cross validation (and *true* validation); as well as showing some of the things which can (and often do) compromise the validity of model fit estimates and prediction error estimates. Keep in mind throughout; we use a simple (OLS) linear regression as an example here, but the ideas conveyed here apply to other types of modeling (e.g. GLM, HLM, SEM, etc.).

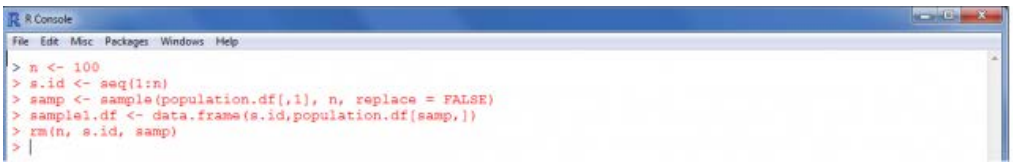
Creating the Data

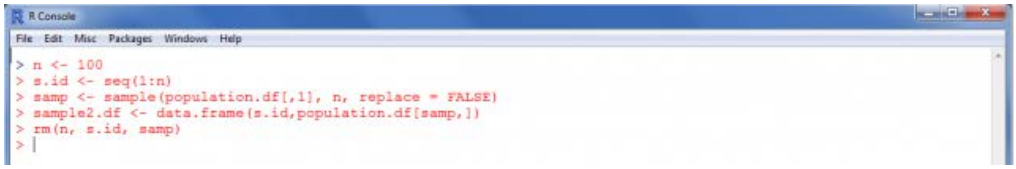
The population (*N* = 1,000,000) was created using the 'MASS' package (Venables & Ripley, 2002) and the 'mvrnorm' function to create a multivariate normal variance / covariance matrix with each of the eight variables centered on a mean of 10. Notice, the first variable was designated the outcome and the second, third, and fourth variables are the only *real* predictors – they have a relationship with the outcome while the other four variables do not. Also, there is no multicollinearity among the second, third, and fourth variables; whereas, there is multicollinearity among the other four variables.



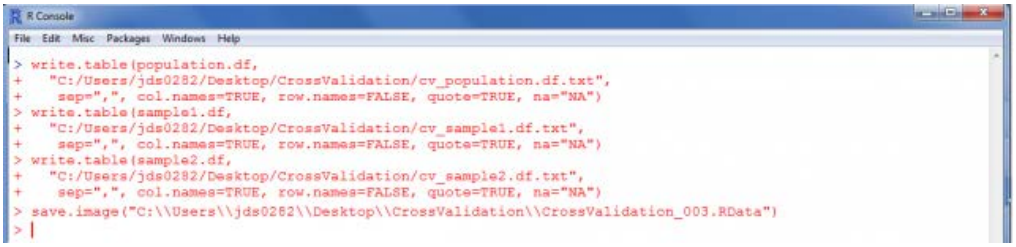
Then, a population identification variable (p.id) was created using the 'seq' function for sequentially labeling the cases from 1 to 1,000,000. Finally, the variables were put into a data frame and the variables were renamed.

Next, two samples were drawn by randomly picking population identification numbers.





Then, all three data frames and the workspace were saved out to the desktop and posted on the web so that they could be read in from the web and offer repeatable results.



Examples

Read in the first sample data file (*n* = 100) from the web naming it "sample1.df" as below; and get the ubiquitous 'head' and 'summary' of the data to see what it looks like.

```
> 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
>
```

Fit the original sample 1 data to the linear model and calculate the Average Residual Squared Error for all cases in the sample (baseline [and biased] prediction error estimate: RSE.n).

```
> samp1.lm <- lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7, sample1.df)
> summary(samp1.lm)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7, data = sample1.df)

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

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.778977   0.169566 -16.389  < 2e-16 ***
x1           0.862742   0.008341 103.429  < 2e-16 ***
x2           0.547078   0.007593  72.050  < 2e-16 ***
x3           0.235641   0.008475  27.803  < 2e-16 ***
x4          -0.164290   0.009121 -18.013  < 2e-16 ***
x5          -0.140939   0.009475 -14.874  < 2e-16 ***
x6           0.018872   0.017262   1.093    0.277
x7          -0.081176   0.018421  -4.407 2.84e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07509 on 92 degrees of freedom
Multiple R-squared: 0.9942,     Adjusted R-squared: 0.9938
F-statistic:  2259 on 7 and 92 DF,  p-value: < 2.2e-16

> RSE.n <- (sum((sample1.df$y - samp1.lm$fitted.values)^2))/nrow(sample1.df)
> RSE.n
[1] 0.005187536
>
```

These initial estimates (generated above) indicate superb model fit (R-squared & Adj. R-squared) and an extremely small prediction error estimate (RSE.n); however, they are all compromised by over-fitting. 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.

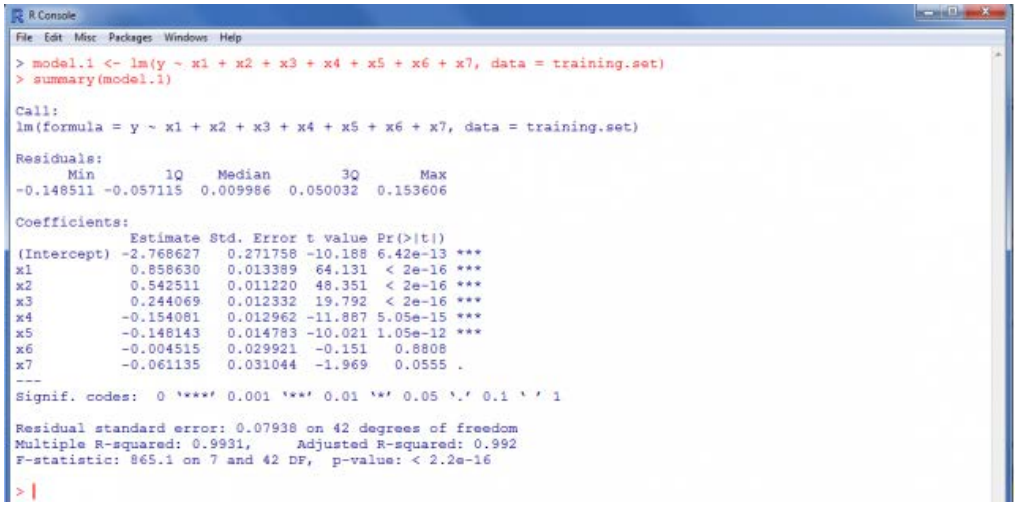
Split-half Cross Validation

Split-half cross validation (also called: split-sample or hold-out validation) involves simply dividing the data into two halves; one the training set, on which the model is fit, and one the testing set, on which the model is evaluated.

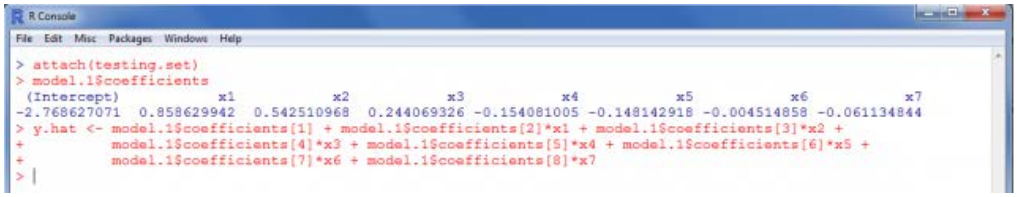
Divide the sample data into two halves, the training set and the testing set.

```
> nrow(sample1.df)
[1] 100
> training.set <- sample1.df[1:50,]
> head(training.set)
  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
> testing.set <- sample1.df[51:100,]
> head(testing.set)
   s.id   p.id        y        x1        x2        x3        x4        x5        x6        x7
51   51 183190  9.532576  9.803391  9.898863  9.591129  9.510771 11.179013 10.527221 11.149856
52   52 767464  9.171071  8.798694  9.751533 10.122447  9.713885  8.816620  9.676893  9.025118
53   53 144760 10.753408 10.584856 11.020712 11.188528 12.294643 11.255245 12.831842 11.743193
54   54 704173 11.021633 10.555549 11.191503 11.271303 10.580588 11.475620 10.903804 10.929445
55   55 679179 11.031378 10.494769 11.158354 10.454756 10.892952  9.890623  8.584673  9.004999
56   56 382746  9.847144 10.825381  8.316025 10.048572 10.086709 10.110882 11.503425 11.061784
>
```

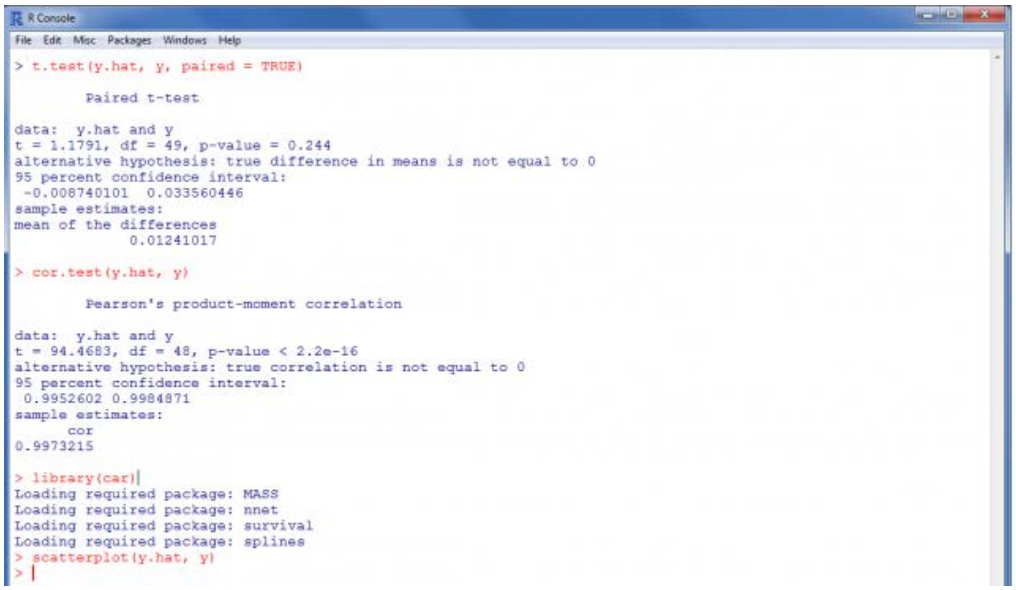
Specify/fit the model with the training set.

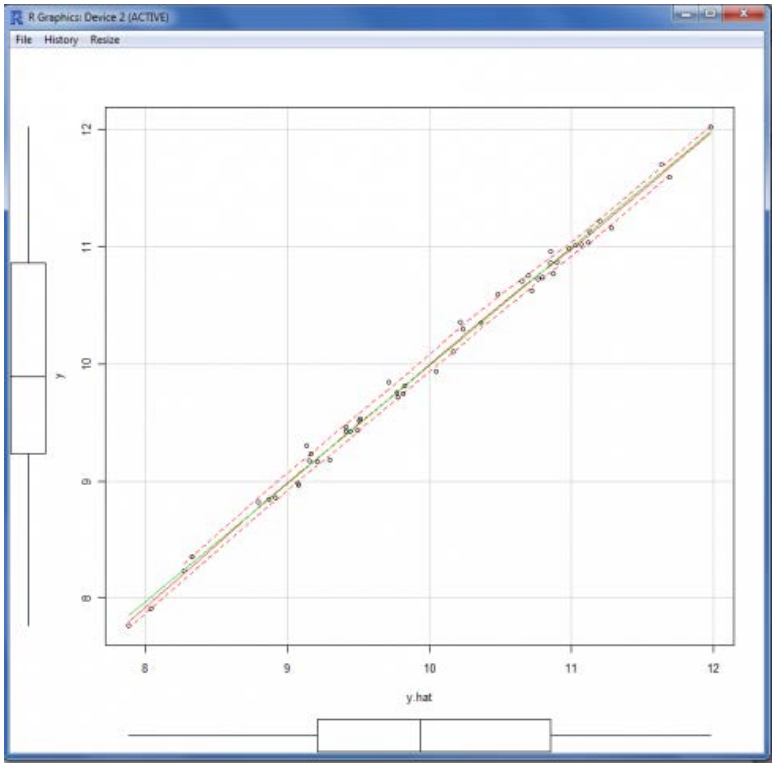


Apply the specified model (coefficients) to predictor values from the testing set to predict (model based) outcome values of the testing set.



Compare these predicted values to the actual values of the outcome in the testing set.





The results above look great; no difference between the two (paired) means, the correlation is virtually +1....perhaps a bit too good to be true?

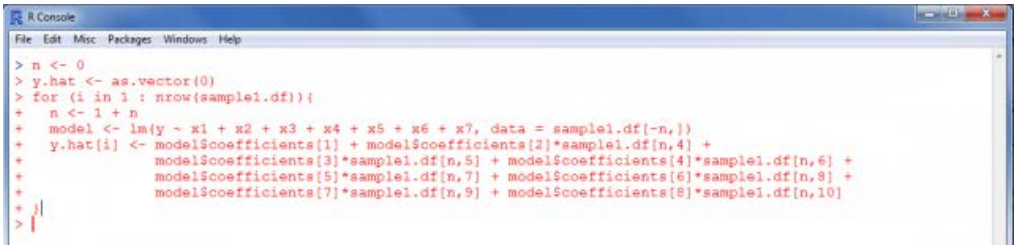
Calculating the average cross validation sums of squares.

Compare the 'cv.ss' to the Average Residual Squared Error (RSE.n). Notice, they are virtually the same, with RSE.n slightly biased (estimating less prediction error). The difference would be more pronounced with real data; here we have no measurement error and the effects sizes are massive.

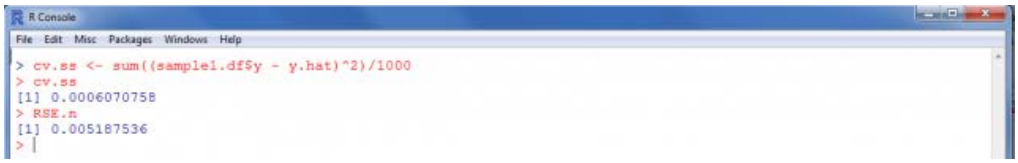
Leave One Out Cross Validation

The Leave One Out Cross Validation (LOOCV) strategy in its most basic form, simply takes one observation out of the data and sets it aside as the 'testing set' like what was done above. Then the model is applied to the training set of *n* - 1 cases (i.e. the data minus the single testing set case). The resulting coefficients are applied to the testing set case to produce a predicted value which in turn is then compared to the actual value (of y) of that single case. Below, we avoid the most basic form of LOOCV and instead iteratively conduct the procedure across all cases (i.e. each case is 'left out' once).

Setting up the initial conditions and creating an empty vector to store the values of y.hat for each iteration; then running the LOOCV iteratively with a 'for-loop'.



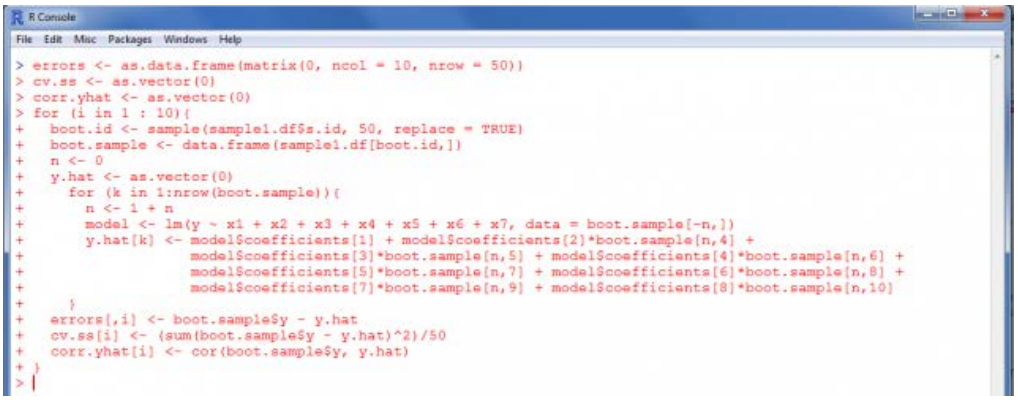
Calculating the average cross validation sums of squares and comparing it to our baseline RSE.n; again, both are very close to zero; with the cv.ss indicating an even smaller estimate of prediction error.



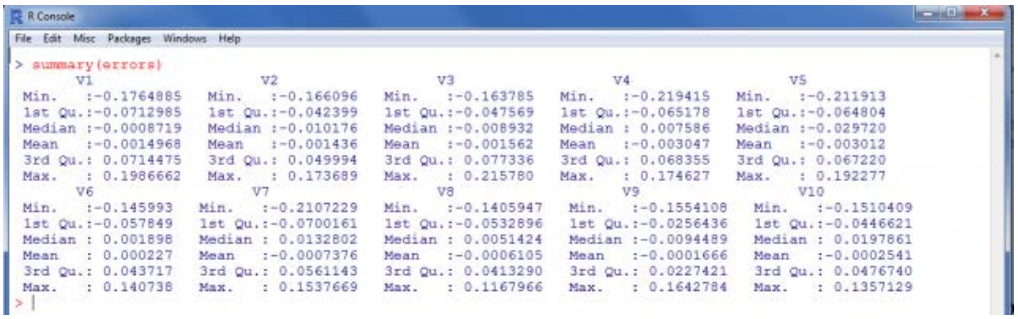
Bootstrapped LOOCV

Here, we create 10 bootstrapped samples (sample WITH replacement from the original sample) and apply the LOOCV from above to each of the 10 bootstrapped samples; each time saving the y.hat values and then calculating the errors (cv.ss) as well as the correlation between y.hat and the actual values of y in each bootstrapped sample. Typically, more than 10 bootstrapped samples would be required to reach a stable estimate; for example purposes, we use only 10 here.

Creating three empty objects for storing the output values; then run the for-loop.



Summary of the errors for each of the 10 iterations (loops); each vector corresponds to each loop/bootstrapped sample.



Comparison of the average cross validation sums of squares for each of the 10 loops (cv.ss) to the Average Residual Squared Error (RSE.n) from each iteration and the bootstrapped average of the cv.ss across iterations (an average of the averages).

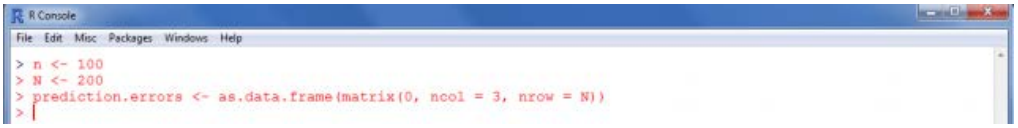
All 10 bootstrapped estimates (and their average) are lower than the RSE.n. The average of the averages across iterations provides the more robust estimate of prediction error.

Correlations between y and y.hat for each of the 10 iterations; all of which are all nearly 1.0.

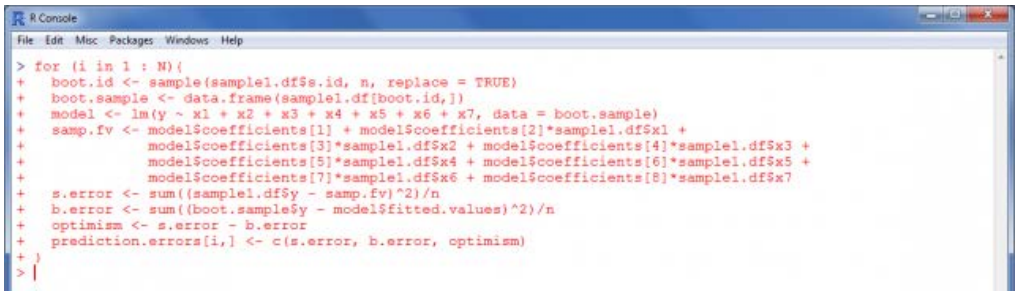
Bootstrapped (cross validation): Estimates of Prediction Error (Efron & Tibshirani, 1993).

The bootstrapped cross validation approach represents a robust method of estimating the prediction error for a model. Each bootstrapped sample (sampling WITH replacement from the original sample) has the model fit to it and the subsequent coefficients are then used to predict outcome scores of the original sample as well as the outcome scores of the bootstrapped sample. Naturally, the errors associated with the original sample estimates will be larger than the errors associated with the bootstrapped sample estimates; because, the bootstrapped sample was used to fit the model and generate the coefficients. Then, using the DIFFERENCE between the original sample errors and the bootstrapped sample errors provides us with an estimate of bias or "optimism" (Efron & Tibshirani, 1993, p. 248). Finally, the average optimism (average of each optimism from all the bootstrapping) can be added to the original sample RSE.n; which corrects it and provides a more accurate estimate of average prediction error.

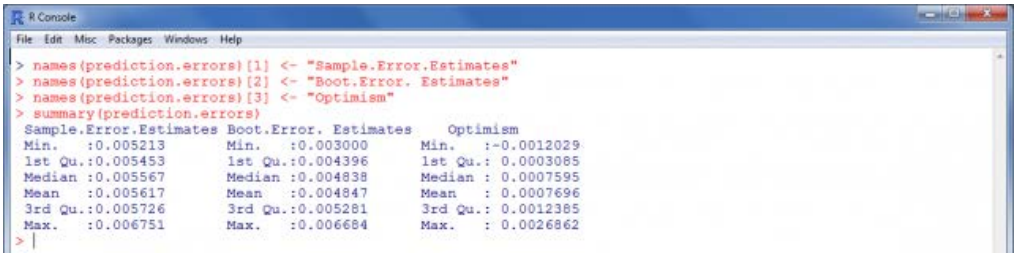
Setting initial conditions and creating an empty data frame to store results. The little 'n' sets the number of cases in each bootstrapped sample and the capital 'N' sets the number of bootstrapped samples to draw. The 'prediction.error' data frame stores the original sample errors, the bootstrapped sample errors, and the optimism value from each iteration or loop.



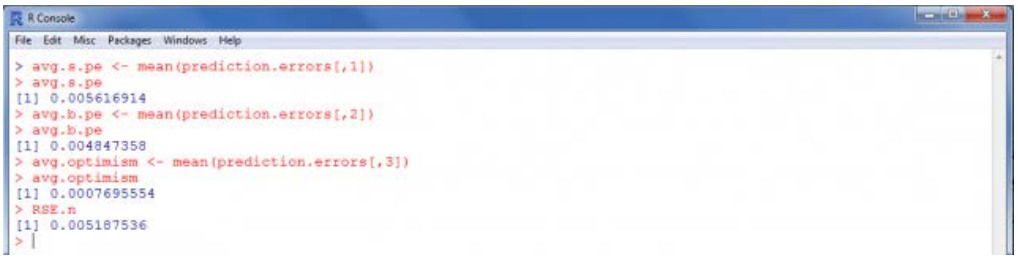
Run the bootstrapped cross validation loop.



The 'prediction.errors' data frame contains the following: the original sample-based prediction error estimate (average cross validation sums of squares), the bootstrapped sample-based prediction error estimate (average cross validation sums of squares), and the optimism estimate (difference between the previous two); 'prediction.errors' contains all 3 for each N = 200 iterations (bootstraps).



Averages of each and a comparison to the RSE.n.

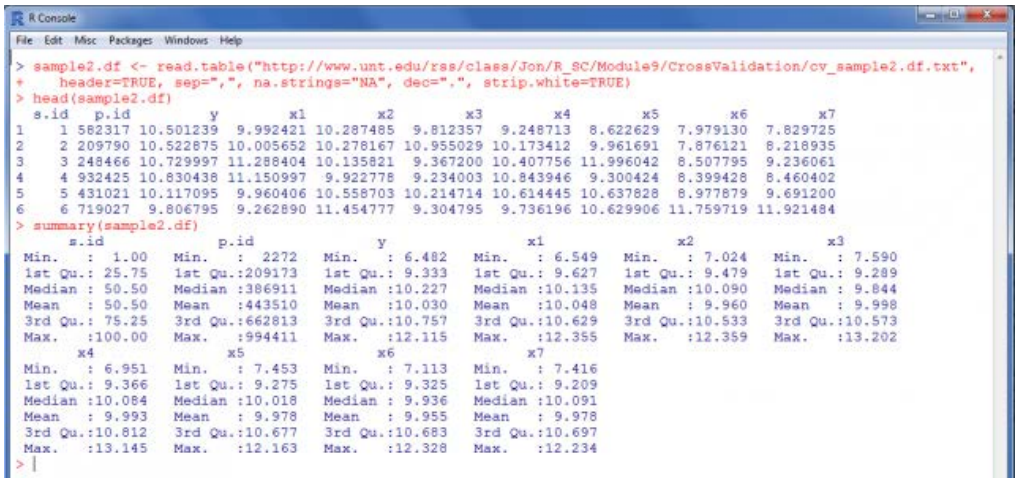


The improved bootstrapped prediction error estimate, which adds a bias correction to the original RSE.n; because, RSE.n is biased downwardly (i.e. predicted error estimate is smaller [less error] than it really should be due to overfitting). See Efron, 1993, p. 237 - 249.

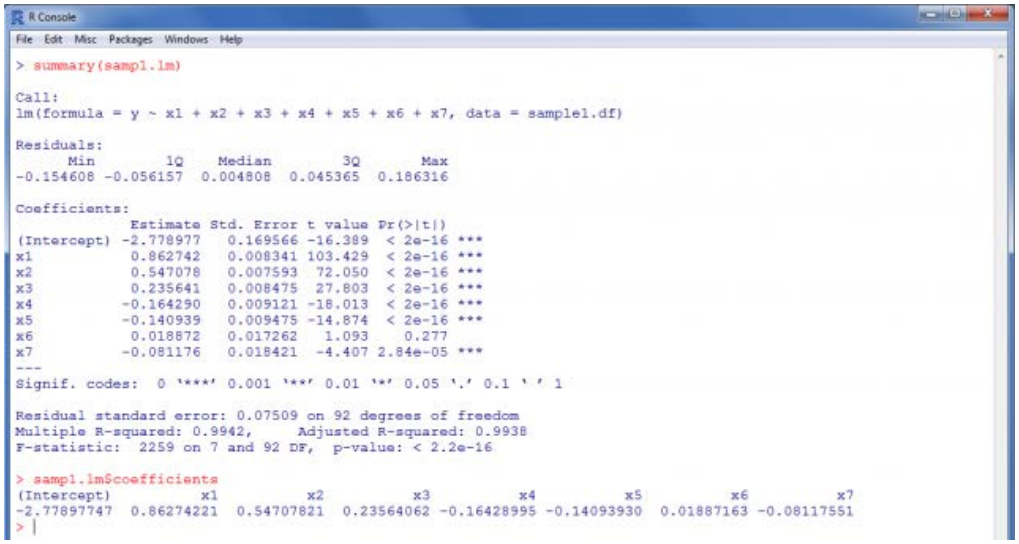
Real Cross Validation

'Real cross validation' = collecting another sample of data from the same population and using the new data (measured with the same instruments) to "validate" the model's accuracy. Of course, in actual research it is often impossible to do this, because of funding, time constraints, etc. Collecting new data: here, this is very easy because we have simulated the data by first generating a simulated population and then sampling from it to build the model. This is why our estimates are very 'unbiased' above; most of the prediction error estimates (regardless of cross validation technique) have produced very similar estimates (virtually the same near-zero estimates of prediction error).

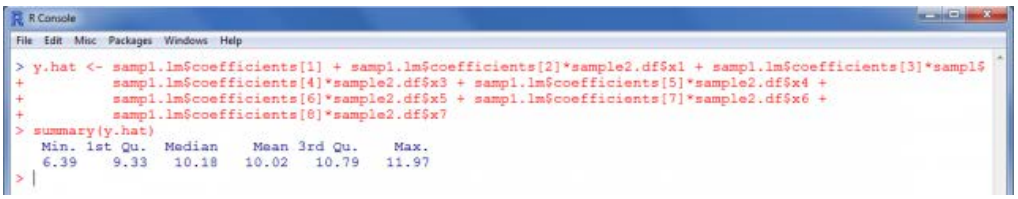
Collecting (drawing) a new sample from the population. Here; read in the second sample from the web, naming it "sample2.df".



Now we can use the original sample's linear model coefficients, applied to the new (2nd) sample's data (predictor variables), to 'predict' values on the outcome variable (y) of the new (2nd) sample data set. Then we can compare the 'predicted' values (y.hat) with the actual values (y) from the new data. Recall the first sample's linear model.



Applying the coefficients to the new predictor variable data to create the 'predicted' values of the outcome (y.hat).

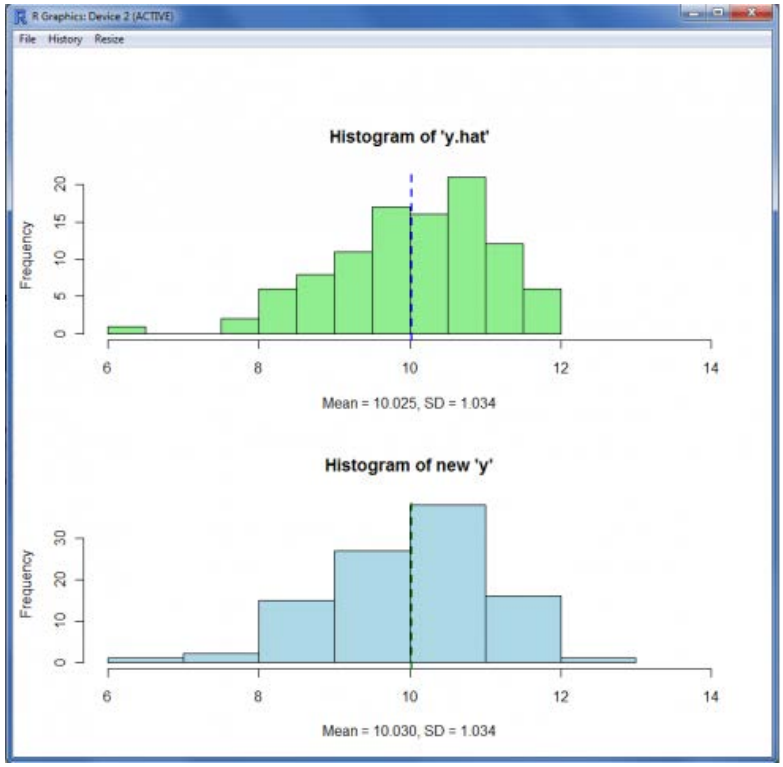


Comparison of the 'predicted' values (y.hat) to the actual (new) values of the outcome (sample2.df$y).

```
> summary(y.hat)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   6.39    9.33   10.18   10.02   10.79   11.97
> summary(sample2.df$y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  6.482   9.333  10.230  10.030  10.760  12.110
> mean(y.hat)
[1] 10.02455
> mean(sample2.df$y)
[1] 10.03038
> var(y.hat)
[1] 1.069014
> var(sample2.df$y)
[1] 1.069242
> sd(y.hat)
[1] 1.033932
> sd(sample2.df$y)
[1] 1.034042
> t.test(y.hat, sample2.df$y, paired = TRUE)

        Paired t-test

data:  y.hat and sample2.df$y
t = -0.8159, df = 99, p-value = 0.4165
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.020007830  0.008348458
sample estimates:
mean of the differences
          -0.005829686

> cor.test(y.hat, sample2.df$y)

        Pearson's product-moment correlation

data:  y.hat and sample2.df$y
t = 142.9945, df = 98, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9964470 0.9983956
sample estimates:
      cor
0.9976122

>
```

The following creates two histograms (in one graphics window); showing y.hat and y.

```
> oldpar <- par(oma=c(0,0,3,0), mfrow=c(2,1))
> hist(y.hat, col = "lightgreen", xlim = c(6,14), main = "Histogram of 'y.hat'", xlab = "Mean = 10.025, SD = $
> abline(v = mean(y.hat), col = "blue", lwd = 2, lty = "dashed")
> hist(sample2.df$y, col = "lightblue", xlim = c(6,14), main = "Histogram of new 'y'", xlab = "Mean = 10.030,$
> abline(v = mean(sample2.df$y), col = "darkgreen", lwd = 2, lty = "dashed")
> par(oldpar)
>
```



So, clearly our model (based on the original sample) is doing an excellent job of predicting new values of 'y' based on new values of all the predictors. This is due to a few advantages of simulating or creating the data in the manner we did: the variables have zero measurement error and the cases were randomly sampled from a defined (and constant) population (i.e. zero sampling error/bias); although, because this data is simulated, there is a slight chance the same

case(es) may appear in multiple samples. However, that was not the case here; meaning, each case is unique to sample 1 or sample 2. This is largely due to the overwhelmingly large population size (*N* = 1000000) in relation to sample sizes (*n* = 100).

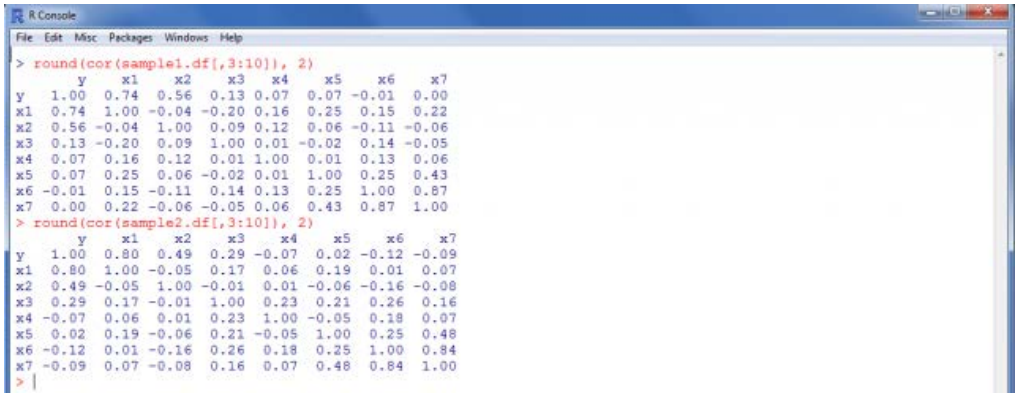
Furthermore, our samples have greater than a 10 to 1 ratio of cases to variables. Simply put, our sample sizes (*n* = 100 each), though not large, are adequate because; for each of our 8 variables we have at least 10 cases -- it would obviously be better to have a 20 to 1 (or even higher) ratio; but, 10 to 1 is about the lowest acceptable ratio. However, one should also remember that the sample size in relation to the population size is important. The smaller the sample (in comparison to the population), the more likely the sample, even a random sample, will contain bias.

In a 'real-world' setting, it is not uncommon to have violations of one, or more of the above conditions -- any one of which would bias the estimate of RSE.n; meaning the model would have more prediction error than the RSE.n would indicate.

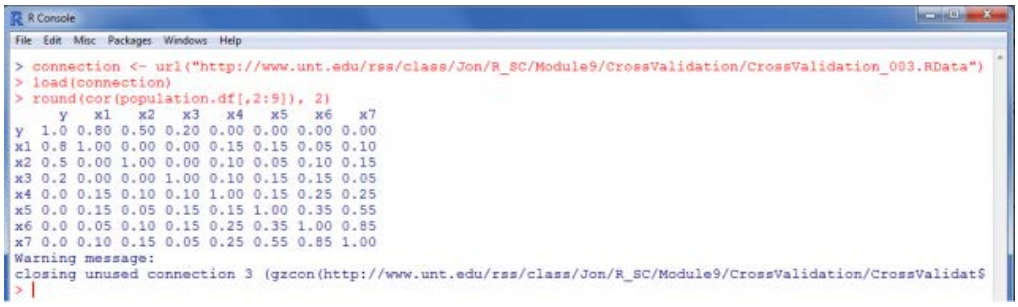
Two other threats to the accuracy of estimating prediction errors for a model are model specification error and model search error. Model specification error can happen in three ways; (1) using the wrong model 'form' (e.g. using a linear model rather than a quadratic model when the quadratic more accurately represents the data), (2) errors of omission (leaving crucial variables out of the model), and (3) errors of inclusion (including non-essential variables in the model). Model search error is related to the first type of model specification error mentioned above. Model search error refers to the bias associated with the method of selecting a model form. In other words, model search error refers to the uncertainty of picking the model form you did; which can contribute to the prediction error. The term model search error is used as a reflection of the *search* for the best model or best set of models (e.g. Bayesian Model Averaging). It is important to note that cross validation techniques do not address the problem of model search nor do they address model specification error of the first type (i.e. using the wrong model form). Errors of inclusion can increase multicollinearity and cause bias in model fit estimates, prediction error estimates, and individual parameter estimates (i.e. coefficients). To clarify this last point, notice that throughout the above results, our model has performed almost perfectly with respect to model fit (R-squared & Adj. R-squared), small residuals (errors), and very small predicted error estimates. However, closer inspection of the sample(s) data (and the population data) will reveal two related faults with our model.

First, we have model specification errors of inclusion; four of our predictors are not related to the outcome (AT ALL!), which means they have no business being included in the model. The only thing they do is artificially increase fit measures (R-squared & Adj. R-squared). Second, these four variables are even more disruptive because they bring with them multicollinearity (intercorrelations among themselves and to a lesser extent with the actual 3 predictors); a condition which decreases the validity of the coefficients in terms of interpreting variable importance.

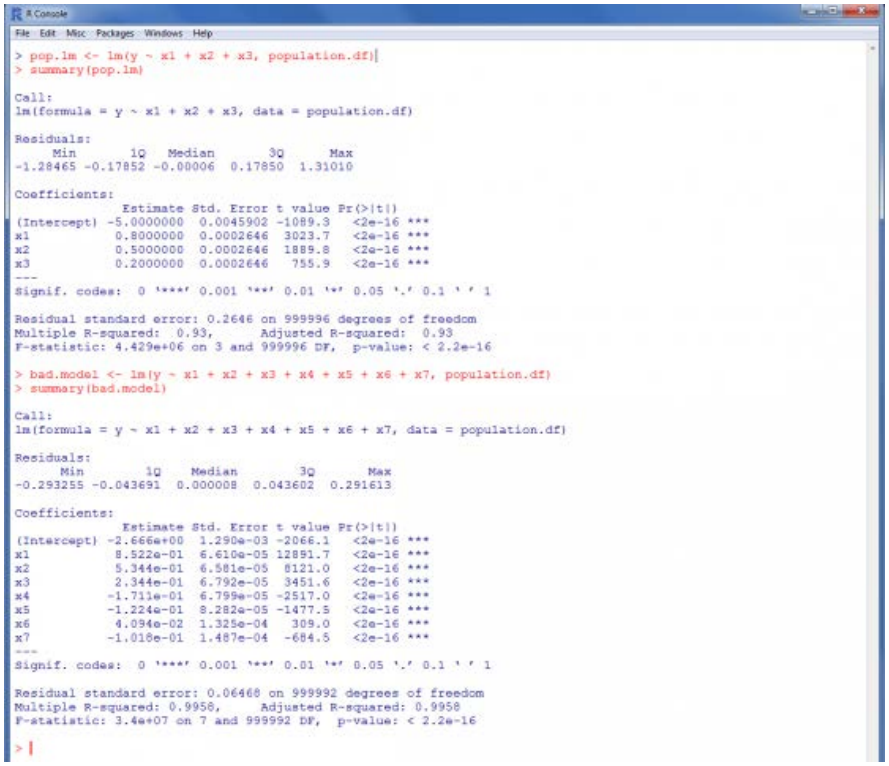
To 'see' the relationships referred to above, simply take a look at the correlation matrix of 'y' and all the 'x' variables. Below, both sample correlation matrices are displayed with rounding two places after the decimal (if interested in seeing the matrices without rounding use this: cor(sample1.df[,3:10]) and this: cor(sample2.df[,3:10])



If we examine the true population values we see that indeed, only x1, x2, and x3 are related to y; the other four variables (x4, x5, x6, & x7) do not contribute to y and they do add multicollinearity to our model (further confusing the results). However, in order to review the population, we must read in the original work space (because the internet connection will time out if attempting to read in the data directly due to the large number of cases; *N* = 1,000,000).



To compare the 'true' population model with the mis-specified model, run both on the population. Be advised, this takes a minute or so due to the 1,000,000 cases.



```
> pop.lm <- lm(y ~ x1 + x2 + x3, population.df)
> summary(pop.lm)

Call:
lm(formula = y ~ x1 + x2 + x3, data = population.df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.28465 -0.17852 -0.00006  0.17850  1.31010

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.0000000  0.0045902 -1089.3   <2e-16 ***
x1           0.8000000  0.0002646  3023.7   <2e-16 ***
x2           0.5000000  0.0002646  1889.8   <2e-16 ***
x3           0.2000000  0.0002646   755.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2646 on 999996 degrees of freedom
Multiple R-squared:  0.93,      Adjusted R-squared:  0.93
F-statistic: 4.429e+06 on 3 and 999996 DF,  p-value: < 2.2e-16

> bad.model <- lm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7, population.df)
> summary(bad.model)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7, data = population.df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.293255 -0.043691  0.000008  0.043602  0.291613

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.666e+00  1.290e-03 -2066.1   <2e-16 ***
x1           8.522e-01  6.610e-05 12891.7   <2e-16 ***
x2           5.344e-01  6.581e-05  8121.0   <2e-16 ***
x3           2.344e-01  6.792e-05  3451.6   <2e-16 ***
x4          -1.711e-01  6.799e-05 -2517.0   <2e-16 ***
x5          -1.224e-01  8.282e-05 -1477.5   <2e-16 ***
x6           4.094e-02  1.325e-04   309.0   <2e-16 ***
x7          -1.018e-01  1.487e-04  -684.5   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06468 on 999992 degrees of freedom
Multiple R-squared: 0.9958,     Adjusted R-squared: 0.9958
F-statistic: 3.4e+07 on 7 and 999992 DF,  p-value: < 2.2e-16

>
```

Based on the results of the 'bad.model' it appears superior; smaller residuals, larger R-squared and Adj. R-squared, and smaller residual standard errors. But, we 'know' those are biased because the correlation matrix(-ces) shows us what is *truly* related to the outcome (y) and what is not. Of course, the p-values cannot be relied upon due to the enormous number of cases.

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/twiki/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. Availabe at: http://www.cs.cmu.edu/~schneide/tut5/node42.html

Venables, W. N., & Ripley, B. D. (2002). Package 'MASS'. Available at CRAN: http://cran.r-project.org/web/packages/MASS/index.html

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

validation_%28statistics%29

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Short Courses

Instructor-led courses are currently not being scheduled. 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 Horizins 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.

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » staff-activities

Staff Activities

Transitions

New Employees:

- **Nicole Dao**, IT Programmer Analyst, UNT Student Records Systems Development Team (AIS).
- **Randall Crane**, DCS Support student assistant (part-time).
- **Mark Hurtado**, UNT Administration Support student assistant (part-time).
- **McCall Clark**, CITC Helpdesk Consultant (part-time).
- **Farcia Mascarenhas**, CITC Helpdesk Consultant (part-time).
- **Laura Hernandez**, CITC Helpdesk Consultant (part-time).
- **Timothy Cannon**, Operations Student Technician (part-time).
- **Laurie Holtzclaw**, Operations Student Technician (part-time).
- **Noah Kindervag**, CSS Tech, Classroom Support Services (part-time).
- **Athanasios (Nassos) Galiopoulos**, IT Manager, UNTranet-SharePoint, Enterprise Collaboration Services (AIS).
- **James Collins**, CSS Tech, Classroom Support Services (part-time).
- **James A Gould**, CSS Tech, Classroom Support Services (part-time).
- **Sita Periathiruvadi**, RSS Intern, Academic Computing and User Services (part-time).
- **Yung-Chia Hao**, RSS Intern, Academic Computing and User Services (part-time).
- **Austin Roberts**, *Telecom* Student Employee (*part-time*).

No longer working in the Computing and Information Technology Center:

- **James Journey**, DCS Support student assistant (part-time).
- **Trayton Oakes**, CITC Helpdesk Consultant (part-time).
- **Aaron Glover**, Operations Student Technician (part-time).
- **Arthur Pichon**, Operations Student Technician (part-time).
- **Mary Yingst**, CSS Tech, Classroom Support Services (part-time).
- **Evan Pritchard**, CSS Tech, Classroom Support Services (part-time).
- **Anantha Gorthy**, IT Programmer Analyst, Payroll/Human Resources Services (AIS)

- **Aaron "Hunter" Childress**, CSS Tech, Classroom Support Services (part-time).
- **Caitlin Clark**, CSS Tech, Classroom Support Services (part-time).
- **Carl Etheridge**, CSS Tech, Classroom Support Services (part-time).
- **Venkata "Teja" Yarlagadda**, AIS Training- Web Developer (part-time).
- **Kevin Treadwell**, Communications Specialist, Infrastructure Engineering & Installation, Communications Services.
- **Brandon Potter**, IT Technician, Computer Operations.
- **Masha Aziz,** Programmer Analyst, UNT Student Records Systems Development Team (AIS).

Changes, Awards, Recognition, Publications, etc.

Changes/Reorganizations

CCAIM

Robert Jones, AIS Tools and User Services Director, informs us:

I have asked **Luanne Linke** to take on primary responsibility for project management for the variety of projects the EISTUS area works on. Over the coming months a large portion of her time will involve project management for two major efforts (The EIS hardware refresh and Blackboard Analytics implementation). This change will involve reassigning some of her day-to-day duties to **Billy Huber**. Billy is already heavily involved in those efforts and is a capable and respected leader within AIM. This reassignment will allow Luanne to focus more completely on project management which is one of Luanne's many strengths. Billy and **Philip Brooks** will continue to report to Luanne. The rest of the AIM team will work more closely with Billy on day-to-day assignments and service delivery. This means that you may be working more directly with Billy when AIM leadership is required.

Here's a brief summary of the change:

Luanne remains as official Team Manager:

- Providing direct supervision of Billy Huber and Phillip Brooks
- Acting on formal personnel responsibilities
- Developing project plans for most EISTUS managed projects. These tend to be technical project plans in support of larger efforts like upgrades or software implementations.
- Managing projects and conducting project meetings

Billy will act as an Assistant or Alternate Team Leader:

- Provide day-to-day assignments and guidance for AIM service delivery and support
- Lead weekly team meetings
- Attend other meetings (Chg. Mgmt., etc) on AIM's behalf
- Monitor and manage problem tracking as well as system availability and performance for AIM supported products both PeopleSoft and non-PeopleSoft

Phillip will continue as Production Control Lead

Provide day to day assignments and guidance to the Production Control team members

- Provide AppWorx support and scheduling to programming staff
- Produce weekly report of outstanding problem tickets
- Provide Change management migrations Tues and Thurs eve
- Provide expedited migrations as approved
- Produce agenda for the AIS Change management meeting

I appreciate both Luanne and Billy's openness to this arrangement and believe we'll all benefit in the end. I appreciate too your working with them as this change begins to take place. Please let me know if you have any questions.

Enterprise Collaboration Services

As reported at DCSMT recently, **Jason Myre** announced a reorganization in his area, Enterprise Collaboration Services, with **Yancey Yeargan** managing the Directory Services group, **Jason Gutierrez** managing the Messaging group, **Adam Jensen** managing Central Web Services, and **Nassos Galiopoulos** managing the Untranet/Sharepoint group.

Training Cerfitication

Congratulations to **David Wright**, IT Specialist, Enterprise Services, CITC Infrastructure & Technical Services. He was recently recognized in the *HR Connections* newsletter.

New Babies!

Congratulations, again, to **David Wright**. He and his wife Lilliana welcomed a daughter, **Vivienne Adele Wright**, on May 13, 2011 (6 lbs 10oz, 20.25" long). Congratulations are also in order for **Brandi Burns**, IT Specialist, and her husband Adam, on the birth of their son **Ash Stevenson Burns**, on May 30, 2011 (6 lbs 4 oz, 20.5" long).

Retirees Honored

Recent retirees were honored at the annual Retirement Luncheon on May 4. The luncheon is held annually to honor retiring faculty and staff for their dedicated service. Those from CITC who were among the honored retired or were on modified service from 2010-11 are:

- **Charles Fairall**
- **Nancy Fisher**
- **Jennifer Lafleur**
- **Maurice Leatherbury**
- **J. Silvestre Montalvo**

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2011-06 » EDUCAUSE Annual Conference: Still Time to Register!

EDUCAUSE Annual Conference: Still Time to Register!

By Claudia Lynch, *Benchmarks Online* Editor

There's still time to register for the EDUCAUSE 2011 annual conference. It is being held in Philadelphia, Pennsylvania this year (October 18–21; Online: October 19–21).

Click on the banner below to find out more information and/or to register. Early-bird rates are available through **September 20**.

Unable to attend EDUCAUSE 2011 in Philadelphia this year?

Take advantage of the online event instead: http://www.educause.edu/E2011/Program/Online

BOOKMARK

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

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home

Today's Cartoon

"I wanted more definition in my glutes, so I got the dictionary tattooed on my butt."

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

BOOKMARK

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