550.400: Mathematical Modeling and Consulting

Lecture Notes

Instructor:
Dr. N. H. Lee

JHU AMS 2012 FALL

Last Compiled on October 3, 2012

Outline

Adobe Connect

https://connect.johnshopkins.edu/meeting550400/

Announcement

- HW *SET* 1 due on Monday Oct 8
- Include your drawing in HW SET 1 as a figure
- Marked Work Statement is returned on Monday Oct 8
- Check the blackboard frequently for updates
- Ask non-sensitive questions on Blackboard Discussion Forum FAQ

Git FAQ

http://gitref.org/index.html

Vim FAQ

In Vim, how do you

- start, save, quit Vim?
- show numbers on the side?
- change the color theme?
- move around?
- spell check?
- find particular words?
- save typing while coding these LaTeX commands?
- install plugins?

Vim FAQ

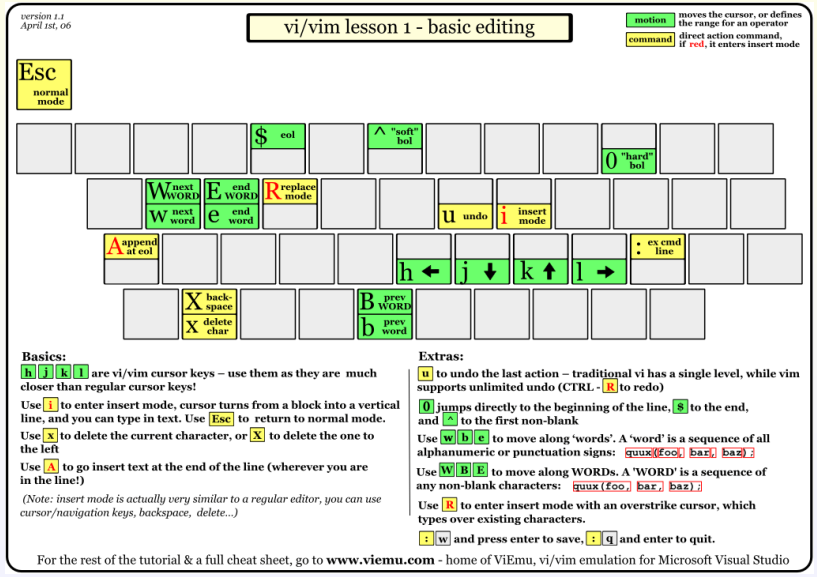
Vim is a *highly customizable* text editor

1. LaTeX, R, C/C++, Java, Python, Git and etc.
2. Regular expression, syntax coloring, auto-completion
3. <ESC>-mode
   - :-mode, aka., the last line mode
   - i-mode, aka., the insert mode

Vim FAQ

- Download & Install GVim or MacVim
- Download & Install tetris.vim
- Download & Install minibufexpl.vim
- Download & Install Gundo
- Download & Install Vim-LaTeX

Vim FAQ



LaTeX FAQ

How do you add a figure in LaTeX ?

```
\begin{figure}
    \caption{<+caption text+>}
    \begin{center}
        \includegraphics[width=<+number+>\textwidth]{<+filename+>}
    \end{center}
\end{figure}
```

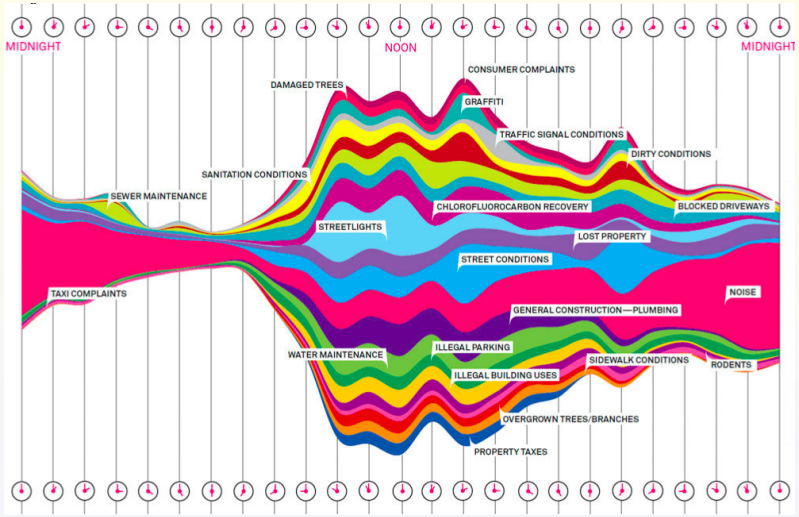
Here, <+...+> denotes a thing that you need to fill in.

But you will need this in the preamble part of your LaTeX:

```
\usepackage{graphicx}
```

To save yourself from unnecessary glitches, insert `png`, `jpeg`, `pdf` files only.

Cross Sectional Study Vs. Longitudinal Study



There were 34,522 complaints called in to 311 between September 8 and September 15, 2010. Here are the most common, plotted by time of day.

Can Riots Be Predicted?

A Tunisian protester holds a baguette while taking to riot police in January 2011

LaTeX FAQ

How can I code a beamer?

```
\documentclass[hyperref={colorlinks=false},handout,10pt]{beamer}
\usetheme{Singapore}
\usecolortheme{lily}
\usefonttheme[onlymath]{serif} % What does this do?
```

OR

```
\documentclass[hyperref={colorlinks=false},handout,10pt]{beamer}
\usetheme{Berlin}
\usecolortheme{wolverine}
\usefonttheme[onlymath]{serif} % What does this do?
```

For a more complete array of themes, go to:

http://www.hartwork.org/beamer-theme-matrix/

LaTeX FAQ

How can I code a beamer?: a single side with no block

```
\begin{document}
        \begin{frame} # one frame per one slide
            \frametitle{hello world} # optional but you want one
                \begin{itemize}
                    \item apple
                    \item orange
                \end{itemize}
        \end{frame}
\end{document}
```

How can I code a beamer?: a single side with one block

```
\begin{document}
        \begin{frame} # one frame per one slide
            \frametitle{hello world} # optional but you want one
                \begin{block}{hey world}
                    Bob!
                \end{block}
        \end{frame}
\end{document}
```


LaTeX FAQ

How can I code a beamer?: two slides

```
\begin{document}
        \begin{frame} # one frame per one slide
            \frametitle{hello world} # optional but you want one
            \begin{itemize}
                \item apple
                \item orange
            \end{itemize}
        \end{frame}
        \begin{frame} # one frame per one slide
            \frametitle{hello world} # optional but you want one
                \begin{block}{hey world}
                    Bob!
                \end{block}
        \end{frame}
\end{document}
```


LaTeX FAQ

How can I code a beamer?: a single frame with two columns

```
\begin{document}
        \begin{frame} # one frame per one slide
            \frametitle{hi world} # optional but you want one
            \begin{columns}
                \begin{column}{0.5\textwidth}
                    \begin{itemize}
                        \item Alice!
                    \end{itemize}
                \end{column}
                \begin{column}{0.5\textwidth}
                    \begin{block}{hey world}
                        Bob!
                    \end{block}
                \end{column}
            \end{columns}
        \end{frame}
\end{document}
```

LaTeX FAQ

How can I code a beamer?: with a table of contents

```
\begin{document}

    \begin{frame}
      \frametitle{Outline}
      \tableofcontents
    \end{frame}

    \section{Hello World}  # optional
        \subsection{hello world} # optional
            \begin{frame} # one frame per one slide
                \frametitle{hi world} # optional but you want one
            \end{frame}

    \section{Hello New World}
            \begin{frame} # one frame per one slide
                \frametitle{hi new world} # optional but you want one
            \end{frame}
\end{document}
```


LaTeX FAQ

SO, how to put a code in the slide? and it looks like codes?

```
\begin{lstlisting}
require(tikzDevice)
x = rnorm(100)
plot.ts(x)
dev.off()
\end{lstlisting}
```

```
require(tikzDevice)
x = rnorm(100)
plot.ts(x)
dev.off()
```


LaTeX FAQ

But, this requires the following in the preamble portion of your tex file:

```
\usepackage{listings}
\lstset{
basicstyle=\footnotesize\ttfamily,
numbers=left,
frame=bottomline,
framextopmargin=50pt,
}
```

You will also need `fragile` option for your frame:

```
\begin{frame}[fragile]
    \frametitle{hello world}
     \begin{lstlisting}
x = rnorm(100)
     \end{lstlisting}
\end{frame}
```


LaTeX FAQ

Where to get more help:

http://en.wikibooks.org/wiki/LaTeX/Presentations

R FAQ

```
for(itr in 1:8) {
    stanzaname = paste("stanza",itr,sep="")
    gitaddress = paste("git://github.com/nhlee/550400.",
                           stanzaname,".git",sep="")
    bashcommand = paste("git remote add ",
                           stanzaname," ",gitaddress,sep="")
    system(bashcommand)
}
```

- `1:8` creates a vector that ...
- `X = 1` assigns 1 to X
- `X <- 1` also assigns 1 to X
- lots of things are done through function
- `paste` and `system` are functions that ...
- functions has none or more arguments
- arguments are implicitly ordered but the order can be overridden

R FAQ

```
system(`ls -ld .*')
system(`cat .Rprofile')
system(`cat .bashrc')
system(`cat .gitignore')
system(`cat .vimrc')
```

- .xxx files are hidden
- ls -ld .* show the hidden files
- .Rprofile set up your R behavior
- .bashrc set up your bash behavior
- .gitignore set up your git behavior
- .vimrc set up you vim behavior
- these files are equivalent to Preference part of your GUI software

R FAQ

How to do software documentation (via R)

```
myfun <- function(x) {x^2}
package.skeleton(name='MYPAC',
                 list='myfun',
                 path='~/')
#Do the documentation
system('R CMD check ~/MYPAC')
system('R CMD build ~/MYPAC')
system('R CMD install MYPAC')
```

Assessing Causality (WMA, 527)

- Consistency of association:

  The association is observed in several different populations using different types of study design.

- Strength of association

  A bigger difference in outcomes between cases with and without the purported causal factor indicates a stronger association.

- Temporal relationship

  The cause preceded the effect. A correlation between two variables measured at the same time gives weaker evidence than one measuring the relationship between changes in the supposed cause and subsequent responses in the outcome.

- Mechanism

  There is a plausible means by which the alleged cause could affect the outcome.

Spurious Causality I

Is there a plausible means by which the alleged cause could affect the outcome?

Chocholate Consumption Vs. Electricity Production

```
cbe.loc<-'http://www.massey.ac.nz/~pscowper/ts/cbe.dat';
cbe <- read.table(cbe.loc,header=T);
plot(cbe[,1],cbe[,3]);
```

Euro & UK Pound Exchange Rate against US Dollar

```
xrate.loc <-'http://www.massey.ac.nz/~pscowper/ts/us_rates.dat';
xrates <- read.table(xrate.loc,header=T);
plot(xrates$UK,xrates$EU,pch=4);
```


Spurious Causality II

How about when there is no context goes with the variables? That is, you just have numbers.

Numerical simulation: presence of confounding variable

```
x <- y <- mu <- rep(0,1000);
for(i in 2:1000)
    mu[i] <- mu[i-1] + rnorm(1);
x <- mu + rnorm(1000);
y <- mu + rnorm(1000);
```

Numerical simulation: presence of "stochastic trend"

```
set.seed(10); x <- rnorm(100); y <- rnorm(100);
for(i in 2:100) {
    x[i] <- x[i-1] + rnorm(1);
    y[i] <- y[i-1] + rnorm(1);
}
```


Spurious Causality III

Working with a model under "stochastic trend" is a tricky business.

A procedure of testing for confounding stochastic trend

```
require(tseries);
adf.test(x)$p.value; #this tests for stochastic trend in x
adf.test(y)$p.value; #so does this but in y
po.test(cbind(x,y)); #this tests for confounding factors in x and y
```

Are two exchange-rates confounded? "co-integrated"?

```
pp.test(xrates$UK)
pp.test(xrates$EU)
po.test(cbind(xrates$UK,xrates$EU))
ukeu.lm <- lm(xrates$UK ~ xrates$EU)
ukeu.res <- resid(ukeu.lm)
```


Apropos

Two non-stationary time series X_t and Y_t are *cointegrated* if some linear combination $aX_t + bY_t$, with a and b constant, is a stationary series.

- Have you heard of p-value?
- How about null and alternative hypotheses?
- Again, what do you mean by "stochastic trend"?
- What do you mean by "stationary processes"?

Hypothesis Test

`adf.test` & `pp.test`

- the null is that the time series has the stochastic trend
- the alt is that the time series is stationary

`po.test`

- the null is that two non-stationary series are not co-integrated
- the alt is that two non-stationary series are co-integrated

p-value

- a number between 0 and 1
- near zero means ...
- near one mean ...

Time Series Model

Is there a plausible means by which the alleged cause could affect the outcome?

A time series model is a *descriptive model*

- its primary goal is to describe quantitative relationship between variables,
- it need not provide the underlying mechanism/context,
- it need not be a generative model.

Time Series Model

A sequence $\{X_i : i = 0, \pm 1, \pm 2, \ldots\}$ of random variables taking values in $\mathbb{R}$ is said to be a "white noise" sequence if

- X_i and X_j are statistically independent,
- X_i and X_j are statistically identical,
- its mean is zero and its variance is posistive.

A white noise sequence is normal/Gaussian if its common likelihood function (aka. density) is normal/Gaussian, i.e.,

$$P(x \leq X_i \leq x + dx) \approx f(x)dx,$$

where

$$f(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Time Series Model

AR(p) model:

- AR stands for *autoregressive*
- in words, the current value is a function of past values plus some random noise
- for $p = 1, 2, \ldots$,

$$X(t) = \beta_0 + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + \varepsilon_t$$

- for example, X_t and Y_t defined below are AR(1) models,

$$X_t = X_{t-1} + \varepsilon_t,$$
$$Y_t = 0.5 Y_{t-1} + w_t$$

Mechanism of Causation?

> *The cause preceded the effect. A correlation between two variables measured at the same time gives weaker evidence than one measuring the relationship between changes in the supposed cause and subsequent responses in the outcome.*

Granger Causality: If $\{Y_t\}$ does not improve the forecasting performance of $\{Z_t\}$, then $\{Y_t\}$ does not Granger cause $\{Z_t\}$.

Exogeneity: $\{Z_t\}$ is exogenous (to $\{Y_t\}$ if it is not affected by the contemporaneous value of $\{Y_t\}$

Time Series Model

VAR(p) model:

- VAR(p) stands for *vector* autoregressive
- for example,

$$X_t = 0.5X_{t-1} + Y_{t-1} + \varepsilon_t,$$
$$Y_t = X_{t-1} + 0.5Y_{t-1} + w_t.$$

- more generally, for $p \times p$ matrix $\boldsymbol{A}$,

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{W}_t.$$

Time Series Model

Example on Page 222

```
require(tseries);

data(USeconomic);
myts = cbind(GNP,M1);
plot(myts);

fittedmodel = ar(myts, order.max=1, method ='ols', dmean = F, intercept = T);

print(fittedmodel);
```

- USeconomic contains a quarterly US economic series from 1954 till 1987
- GNP denotes the gross national product
- M1 denotes "real money", which means income adjusted by inflation

Time Series Model

Example on Page 224

```
require(vars);

fittedmodel <- VAR(myts,p=1,type='trend');

print(fittedmodel);
```

- Yet another way to fit a VAR(1) model in R
- VAR function from vars package is somewhat general than ar function in that you can have linear term, i.e., for `type='both'`, the RHS of the VAR(1) formula contains

$$\alpha_0 + \alpha_1 t$$