Dis 1: General Info; Get Started with Stata

1 General Info

1. **Contact Me**

   You can reach me by **sending me an email** or **attending my (virtual) office hours**.

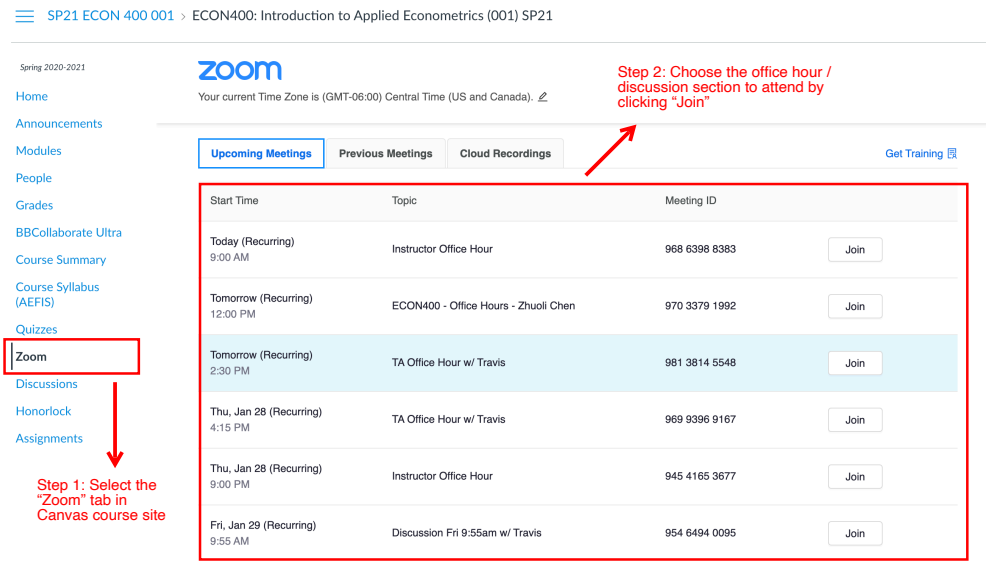
   - Email me at **zchen596@wisc.edu** (please start the subject line with "Econ 400").
   - Virtual office hours take place on Zoom at the following times:
     - Mondays, noon - 1:00pm
     - Wednesdays, noon - 1:00pm
     - Or by appointment

2. **Discussion Sections**

   - Attendance
     - Live attendance is not required, but strongly encouraged.
     - Recordings will be posted (I tape both of my sections, but only the best one will be posted; both sections cover the same material)
     - Sections will take place on Zoom at the following times:
       * Fridays, 1:20 - 2:10pm
       * Fridays, 2:25 - 3:15pm

       Feel free to attend the one that works the best with your schedule.
   - Handouts
     - Handouts (and dataset used) will be available by Thursday midnight on Travis's (the other Econ 400 TA) website.
     - Solutions and section recordings will be posted by Friday night. Same place.

3. **Where to Find These Zoom Sessions?**



2 Get Started with Stata

2.1 What is Stata?

Stata is a statistical software that we use in this course to analyze data and perform econometric analysis.

Side note: Stata is a proprietary software, and it has limitations. For the purpose of this class, Stata is more than sufficient. But for more challenging projects that you may encounter in the future, you might want to look into some open-source programming languages that are friendly for data analysis (such as Julia, R, Python; UW stats department has a course on R, and the econ department offers course on machine learning using Python).

2.2 How to Access Stata?

You can access Stata using either one of the two following methods:

- **Installing it onto your personal laptop**:

  Visit Campus Software Library for installation guide and license & activation key. The Stata version licensed by the university is Stata/SE.

- **Logging into Winstat (i.e. a remote server)**:

  To use Winstat, you must install a software package called **Citrix Receiver** on your computer. The following link has more information on Citrix Receiver and Winstat: `https://kb.wisc.edu/sscc/using-winstat`

2.3 Use Help

- **Google** is always a good starting point
- Alternatively, try using the `help` function in Stata. For example, if you want to know how to use the `describe` function, try

```
help describe
```

2.4 Tips to Help You Master Stata

- Start off by working in the command window and typing individual commands to test out whether the command is doing the right thing for you.
- When you settled on the right series of commands, use a do-file to record them, so you can easily re-run this series of commands (a do-file is the same as a script you've encountered in other programming languages / software).
- Make sure to log or to export your result, so that you have something to refer to at a later date (& to turn in for your problem sets).
- Google and the `help` function are always your friend.

3 Problems

1. For our first exercise, let's try loading different types of data into Stata:
   (a) Download this discussion's dataset, and unzip them to some place on your computer.
   (b) Before we can load any data file, we have to first tell Stata where to look for. To do this, we need to change our working directory. Let's go through two different ways on how to do that.

   (c) Import `gss2014.xls` into Stata (hint: try `import`)

   (d) Import the first row of data from `gss2014.csv` into Stata

   (e) Save the imported data as `gss2014.dta`

   (f) Load the saved `gss2014.dta` (hint: try `use`)

   > For any file of Stata's data format (`.dta`), `use` (or its variant) is used to load the data.
   > For other data formats (such as `.csv`, `.xls`), we first need to `import` it into Stata.

   (g) Stata also has built-in dataset that you can use. Try

   ```
   sysuse "census.dta", clear
   ```

   See https://www.stata-press.com/data/r16/g.html for the list of dataset included in Stata.
   (h) Stata can also directly access `.dta` files via a url. Try

   ```
   use "http://www.stata-press.com/data/r13/sp500.dta", clear
   ```

2. Load `panel2007.dta` into Stata. Let's now do some basic data analysis.

   (a) Browse the dataset in Stata using `browse`. Is there anything jumping out to you in this dataset?

   (b) Obtain a review of the variables that are included in your dataset using the `describe` command.

   (c) Based on your observation in (a) & (b), convert the `id` column data into the correct data format.

   (d) Replace the original `panel2007.dta` file with the current one correcting the `id` variable.

   (e) Obtain some summary statistics for our data (such as mean, minimum, maximum of each variable) by using the `summarize` command.

3. Continue from question 2. Now let's try to clean and merge `panel2007.dta` with `panel2008.dta`

   (a) Since the currently loaded data, `panel2007.dta`, are from 2007, let's make this clear by creating a variable called `year`.

   (b) Again, since Stata can only have one set of data present in its memory, we need to first save the 2007 panel data before merging it with the 2008's.

   (c) Now let's append the 2007 data to the 2008's. As in (a), create a `year` variable for the 2008 data prior to appending. Save the merged data as `appended_data.dta`

   (d) Another way to combine these two datasets is to merge them together (i.e. horizontally instead of vertically appending). To do so, take our 2007 and 2008 datasets and rename the score variables to score2007 and score2008. Then, use the `merge` command to horizontally merge the two datasets by their `id`. Save the merged data as `appended_data.dta`

   (e) Last step: Put everything we did in question 2 and 3 into a do-file, and wrap around your commands in a log.

   (f) * For people interested in making their Stata log file fancy, or if they want to export regression table results into LaTeXor some other format, feel free to check out the following resources:

      i. Stata Markup and Control Language (SMCL)
      ii. estout: Making regression tables in Stata