Econ 2250: Stats for Econ

Fall 2022

What we will learn

- Summary stats (counts, average, standard deviation, correlation)
- Brief intro into Python for data science
- Brief intro into SQL
- Some basic graphing
- Brief intro into linear models
- **But most importantly: How to tell truthful and convincing stories with data**

The most important about any analysis is your voice. If you come away from this class believing your own voice 10% more than you do today it will have been a great success.

Repo: https://github.com/jimcrozier/gt_econ2250

Please fill out form

(https://forms.gle/jJraadLaswWveQqG8)

About Me

trash jones
@jzux

this stupid carbon monoxide alarm is making me dizzy and giving me a headache, does anyone know how to turn it off

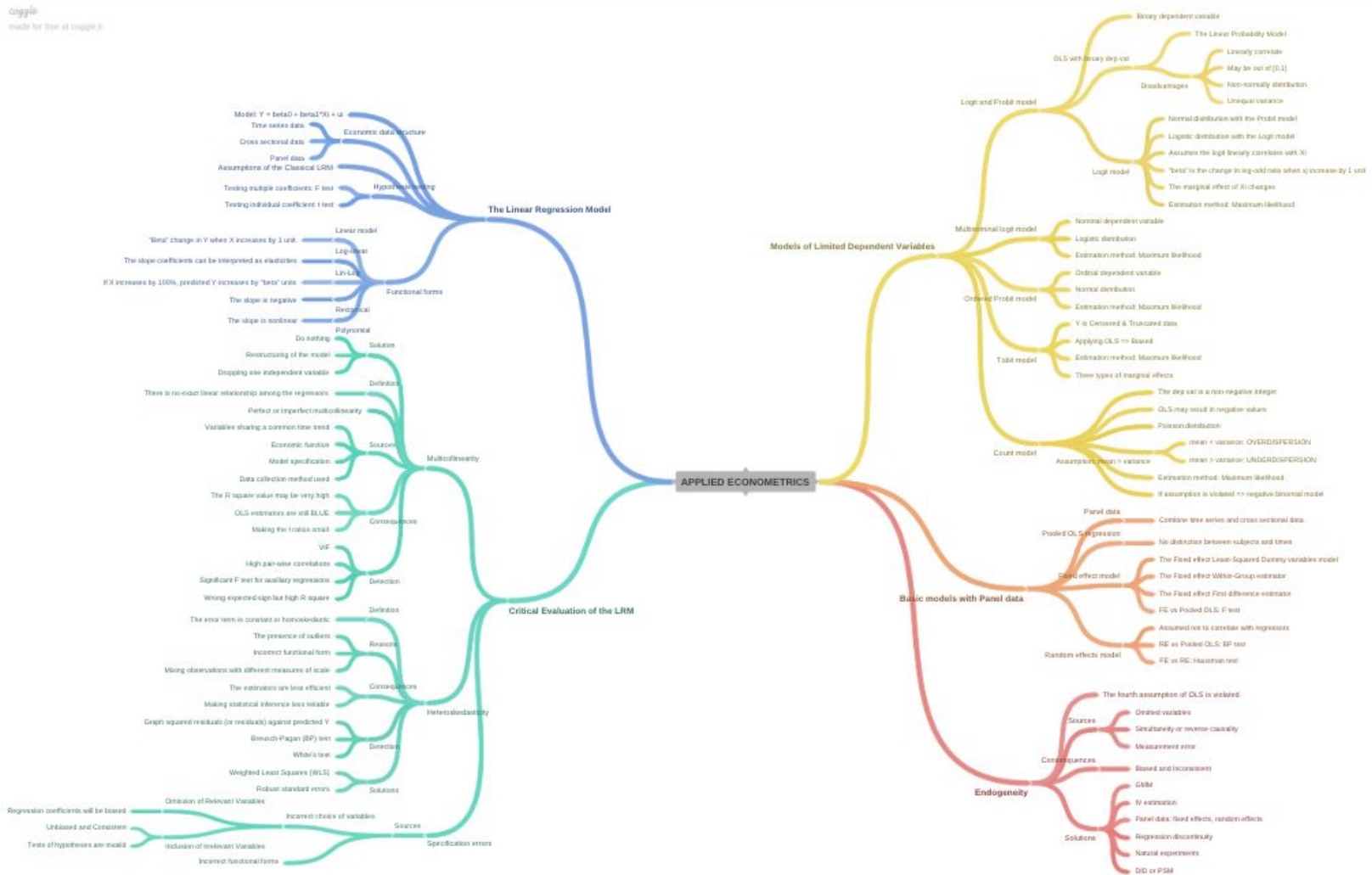9:35 AM · Aug 5, 2022 · Twitter for iPhone

596 Retweets   **20** Quote Tweets   **16K** Likes

No wrong answers here, this is for me to know what to focus on

(https://forms.gle/obwV7LyU8BTCNRFK9)

APPLIED ECONOMETRICS

The Linear Regression Model
- Economic data structure
  - Model: Y = beta0 + beta1*X1 + u1
  - Time series data
  - Cross sectional data
  - Panel data
- Assumption of the Classical LRM
- Hypothesis testing
  - Testing multiple coefficients: F test
  - Testing individual coefficient: t test
- Functional forms
  - Linear model
    - "Beta" change in Y when X increases by 1 unit
  - Log-linear
    - The slope coefficients can be interpreted as elasticities
  - Lin-Log
    - If X increases by 100%, predicted Y increases by "beta" units
  - Reciprocal
    - The slope is negative
  - Polynomial
    - The slope is nonlinear

Models of Limited Dependent Variables
- Logit and Probit model
  - OLS with binary dep var
    - Binary dependent variable
    - The Linear Probability Model
    - Linearly correlate
    - Disadvantages
      - May be out of [0,1]
      - Non-normally distribution
      - Unequal variance
  - Normal distribution with the Probit model
  - Logit model
    - Logistic distribution with the Logit model
    - Assumes the logit linearly correlates with Xi
    - "beta" is the change in log-odd ratio when xj increase by 1 unit
    - The marginal effect of Xi changes
    - Estimation method: Maximum likelihood
- Multinominal logit model
  - Nominal dependent variable
  - Logistic distribution
  - Estimation method: Maximum likelihood
- Ordered Probit model
  - Ordinal dependent variable
  - Normal distribution
  - Estimation method: Maximum likelihood
- Tobit model
  - Y is Censored & Truncated data
  - Applying OLS => Biased
  - Estimation method: Maximum likelihood
  - Three types of marginal effects
- Count model
  - The dep var is a non-negative integer
  - OLS may result in negative values
  - Poisson distribution
  - Assumption: mean = variance
    - mean < variance: OVERDISPERSION
    - mean > variance: UNDERDISPERSION
  - Estimation method: Maximum likelihood
  - If assumption is violated => negative binomial model

Critical Evaluation of the LRM
- Multicollinearity
  - Solution
    - Do nothing
    - Restructuring of the model
    - Dropping one independent variable
  - Definition
    - There is no exact linear relationship among the regressors
    - Perfect or imperfect multicollinearity
  - Sources
    - Variables sharing a common time trend
    - Economic function
    - Model specification
    - Data collection method used
  - Consequences
    - The R square value may be very high
    - OLS estimators are still BLUE
    - Making the t-ratios small
  - Detection
    - VIF
    - High pair-wise correlations
    - Significant F test for auxiliary regressions
    - Wrong expected sign but high R square
- Heteroskedasticity
  - Definition
    - The error term is constant or homoskedastic
  - Reasons
    - The presence of outliers
    - Incorrect functional form
    - Mixing observations with different measures of scale
  - Consequences
    - The estimators are less efficient
    - Making statistical inference less reliable
  - Detection
    - Graph squared residuals (or residuals) against predicted Y
    - Breusch-Pagan (BP) test
    - White's test
  - Solutions
    - Weighted Least Squares (WLS)
    - Robust standard errors
- Specification errors
  - Sources
    - Omission of Relevant Variables
      - Incorrect choice of variables
        - Regression coefficients will be biased
        - Unbiased and Consistent
        - Tests of hypotheses are invalid
    - Inclusion of Irrelevant Variables
    - Incorrect functional forms

Basic models with Panel data
- Pooled OLS regression
  - Panel data
    - Combine time series and cross sectional data
  - No distinction between subjects and times
- Fixed effect model
  - The Fixed effect Least-Squared Dummy variables model
  - The Fixed effect Within-Group estimator
  - The Fixed effect First difference estimator
  - FE vs Pooled OLS: F test
- Random effects model
  - Assumed not to correlate with regressors
  - RE vs Pooled OLS: BP test
  - FE vs RE: Hausman test

Endogeneity
- The fourth assumption of OLS is violated
- Sources
  - Omitted variables
  - Simultaneity or reverse causality
  - Measurement error
- Consequences
  - Biased and Inconsistent
- Solutions
  - GMM
  - IV estimation
  - Panel data: fixed effects, random effects
  - Regression discontinuity
  - Natural experiments
  - DiD or PSM

Author: Do Huu Lua

Do Huu Luat

Matt Dancho (Business Science) @mdancho84 · 1h

One of my biggest mistakes when I began learning data science - I believed everything I heard.

Don't do this. 🧵

#rstats #datascience #excel

❤ 32

Dare Obasanjo
@Carnage4Life

There's a joke about engineers where a man in a hot air ballon asks a passerby "where am I?" as he is lost and the engineer's response is that he's in a hot air balloon. Engineers like to pride themselves on technical accuracy over usefulness. If they didn't, we wouldn't need PMs

10:12 AM · Aug 8, 2022 · Twitter for iPhone

Let's talk about that extra credit…

- This is about collaboration, and is in fact how teams build code together

- Demo a quick PR