lecture 01

foundations

course overview // logarithms; derivatives; integrals; matrices; OLS

course overview

As an aspiring academic, you must establish a **reputation of excellence**.

What are our goals?

What are our goals?

Use, understand, and explain
the **tools** of modern statistical modeling.

~~paint mixing~~

engines	tools to fit models	ML, MM, Bayes, EM
distributions	tools to model the world	normal, bernoulli, Poisson
quantities of interest	tools to understand models	expected value, first difference
measures of fit	tools to evaluate fitted models	predictive distributions, LOOIC
tests and CIs	tools to evaluate hypotheses	Wald, LR, score
properties	tools to evaluate procedures	bias, consistency, coverage
computing	tools to compute estimates	glm(); {brms}

What do I ask of you?

What do I ask of you?

1. Read the notes and complete the exercises.
2. Write a good paper of “modest ambition” and “exceptional quality.”
3. Run a good workshop.
4. Talk to me about all this stuff.

do you understand a variety of tools?

midterm exam	20%	in-class, closed book exams based on the examples and exercises.
and		
final exam	25%	

---

can you explain these tools simply, clearly, and correctly?

workshop	15%	one-hour workshop on a particular statistical model

---

can you apply these tools well?

paper	30%	research project of modest ambition and exceptional quality

Exams

1. Based closely on exercises and examples.
   - Variants are fair game.
   - Make sure you can comfortably complete examples and exercises in a closed-book setting.
   - Make sure you are not relying on solutions or AI-help. Don't confuse "familiarity" with "understanding."
2. Exercises are "due" weekly, but I do not grade them.
   - I'll distribute detailed solutions.
   - Come talk to me often!
   - Collaborate as it helps you learn. Ideal model might be (1) make a best effort as an individual, (2) discuss difficult portions in groups of 2-3, (3) polish your work as an individual, and (4) review your work weekly.
   - Large group discussion?

Research Project

Your task this semester: **write a paper of modest ambition** and **exceptional quality** that makes a **descriptive argument** about politics.

“Descriptive Argument”

“A descriptive argument describes some aspect of the world. In doing so, it aims to answer ‘what’ questions (e.g., when, whom, out of what, in what manner) about a phenomenon or a set of phenomena. Descriptive arguments are about what is/was. For example: ‘Over the course of the past two centuries there have been three major waves of democratization.’” —Gerring (2012)


Mere Description

JOHN GERRING


A Case for Description

Carolyn E. Holmes
Meg K. Guliford
Mary Anne S. Mendoza-Davé
Michelle Jurkovich

ABSTRACT Descriptive research—work aimed at answering “who,” “what” “when,” “where,” and “how” questions—is vital at every stage of social scientific inquiry. The creative and analytic process of description—through concepts, measures, or cases, whether in numeric or narrative form—is crucial for conducting research aimed at understanding politics in action. Yet, our field tends to devalue such work as “merely descriptive” (Gerring 2012), subsidiary to or less valuable than hypothesis-drive causal inference. This article posits four key areas in which description contributes to political science: in conceptualization, in policy relevance, in the management and leveraging of data, and in challenging entrenched biases and diversifying our field.


Good Description

“Modest Ambition”

As a starting point, imagine 3,000 words that make a single, specific, narrow contribution.

Hint: imagine something that will take about 3 weeks to put together.

Carlisle’s AHH Scale

1. A body of work worthy of tenure at a major research university.
2. A book-length project.
3. An ambitious full-length article aimed at a top journal (e.g., AJPS or IO).
4. A full-length article aimed at a good subfield journal (e.g, POBE or JCR).
5. A research note aimed at a good journal (e.g., RAP).

“Exceptional Quality”

Hi Prof. Rainey,

I’ve attached a draft of my paper on [topic]. I’m thinking about submitting it to [journal] soon. Could you let me know if you have any suggested revisions prior to submission?

Grade	As a dissertation advisor, I’d say…
100%	“This seems ready to submit to a peer-reviewed journal. You should present it at a department workshop to help tie up any loose ends, but it seems ready to go.”
90%	“This seems *almost* ready to submit to a peer-reviewed journal. You should fix the minor issues listed below and then present it at a department workshop to help tie up any loose ends. But once you’ve fixed these minor issues, it seems ready to go.”
80%	“This is a good start. I see a couple of significant issues (and a few minor things) listed below that need work. Once you’ve addressed these issues, send me an updated version and I’ll give it another look.”
70%	“This paper isn’t adequately polished. Papers submitted to journals are significantly more polished than this—you should spend a lot of time revising and clarifying your paper before submitting. I’ve included some main areas to focus on below. In the future, it’s fine to discuss early work with me or ask me to review sketches of ideas, but it isn’t a good use of my time (or yours) to ask me for comments on a paper that you haven’t spend significant time making as good as you can. For drafts this rough, the arguments are tough to understand (and therefore critique and improve).

Papers that fail to demonstrate competence with the methods are graded less than 70%.

Workshop

Your task this semester: **design and run a 60-minute workshop on a particular statistical model**. (Ideally in teams of two.)

Required materials

1. Take-home handout.
2. Slides.
3. A read-ahead document.

Workshop structure

1. 15-minute lecture.
2. 15-minute example.
3. 20-minute guided practice.

Aggressive Scaffolding

Date	Title	Exam	Research Project	Workshop Due
Aug 27	Review	–	–	–
Sep 03	ML	–	Optional	–
Sep 10	SE	–	Optional	–
Sep 17	Covariates	–	Optional	–
Sep 24	Bayes	–	Optional	–
Oct 01	MCMC	–	Prospectus	–
Oct 08	{marginaleffects}	–	–	–
Oct 15	–	Midterm	–	–
Oct 22	Hierarchical Models	–	First draft	Initial outline of handout
Oct 29	IRT	–	–	First drafts of all materials
Nov 05	MRP	–	–	Workshop rehearsal
Nov 12	Case Study: Fowler et al. (2023)	–	–	Hold workshop
Nov 19	Case Study: Fariss (2014)	–	–	
Nov 26	Thanksgiving Break	–	Second draft	
Dec 03	Wrap	–	Poster presentation	–
Dec 10	–	Final	Final draft	–

foundations

logarithms

Simplify $\log\left(\prod_{i=1}^{n} y_i^{a_i}\right)$.

derivatives

Let $\ell(\pi) = S\log(\pi)+(n-S)\log\big(1-\pi\big)$ for $0 < \pi < 1$, where S and n (with $0 \leq S \leq n$) are fixed numerical constants. Find $\dfrac{d\ell(\pi)}{d\pi}$.

matrices

Compute the product of the 2×2 matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and the 2×3 matrix $B = \begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix}$.

integrals, pmfs , and pdfs

Compute $\Pr(-1.96 \leq X \leq 1.96)$ for $X \sim \mathcal{N}(0, 1)$.

expected values and variances

Suppose the regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i.$$

1. Find the marginal effect of x.
2. Show that if the coefficient estimates are unbiased, then the plug-in marginal effect estimate is unbiased. (Would this generalize to, say $\frac{\beta_1}{\beta_3}$?)
3. Find the variance of the plug-in marginal effect estimate.

OLS and R

Starting with the code to the right, reproduce Clark and Golder's (2006) 1946-2000 Established Democracies model in Table 2 on p. 698. See also eq. 4 on p. 695 for their model specification. They use HC1 cluster-robust standard errors.

```
# run just once to get data package, then delete
devtools::install_github("carlislerainey/crdata")

# load packages
library(tidyverse)
library(sandwich)  # for robust SEs

# load Clark and Golder's data
cg <- crdata::cg2006  # from my data package

# quick look at variables and their names
glimpse(cg)
```


Table 2
The Strategic Modifying Effect of Electoral Laws

	Dependent Variable: Effective Number of Electoral Parties											
	Cross-Sectional Analysis								Pooled Analysis			
Regressor	1980s Amorim Neto & Cox Data[a]		1980s Amorim Neto & Cox Data[a]		1990s Whole Sample		1990s Established Democracies[b]		1946 to 2000 Whole Sample		1946 to 2000 Established Democracies[b]	
Ethnic			–0.05	(0.28)	0.06	(0.37)	–0.70	(0.68)	0.19	(0.13)	0.11	(0.14)
ln(Magnitude)			–0.08	(0.30)	0.51	(0.44)	–0.61	(0.59)	0.33*	(0.20)	0.08	(0.23)
UppertierSeats	0.04**	(0.01)	–0.07	(0.04)	0.01	(0.02)	–0.02	(0.06)	0.05***	(0.02)	–0.06*	(0.03)
PresidentCandidates			0.22	(0.27)	0.36	(0.26)	0.07	(0.22)	0.35**	(0.16)	0.26*	(0.15)
Proximity	–6.05***	(0.88)	–5.88***	(0.84)	–4.19***	(1.26)	–4.95***	(1.24)	–3.42***	(0.55)	–3.10***	(0.46)
Ethnic × ln(Magnitude)	0.39***	(0.07)	0.37*	(0.20)	–0.09	(0.17)	0.63*	(0.34)	0.08	(0.12)	0.26	(0.17)
Ethnic × UppertierSeats			0.07***	(0.02)	–0.005	(0.01)	0.01	(0.04)	–0.02**	(0.01)	0.06***	(0.02)
PresidentCandidates × Proximity	2.09***	(0.26)	1.84***	(0.43)	0.99**	(0.46)	1.42***	(0.44)	0.80***	(0.23)	0.68***	(0.23)
Constant	2.40***	(0.21)	2.60***	(0.51)	4.08***	(0.95)	5.15***	(1.32)	2.81***	(0.34)	2.92***	(0.35)
Observations	51		51		62		39		555		487	
R^2	.71		.77		.29		.48		.30		.40	

Note: Standard errors are given in parentheses for cross-sectional models; robust standard errors clustered by country are used for the pooled models.
a. See Amorim Neto and Cox (1997).
b. Established Democracies omits elections from countries that transitioned to democracy after 1989.
*$p < .10$. **$p < .05$. ***$p < .01$.

time permitting…

Show that scalar notation

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \epsilon$$

is equivalent to the matrix notation

$$y = X\beta + \epsilon.$$

what we'll learn

extensions to linear regression model

several methods to fit the models

generic methods to compute quantities of interest

my view is modest

these tools are **not** magic

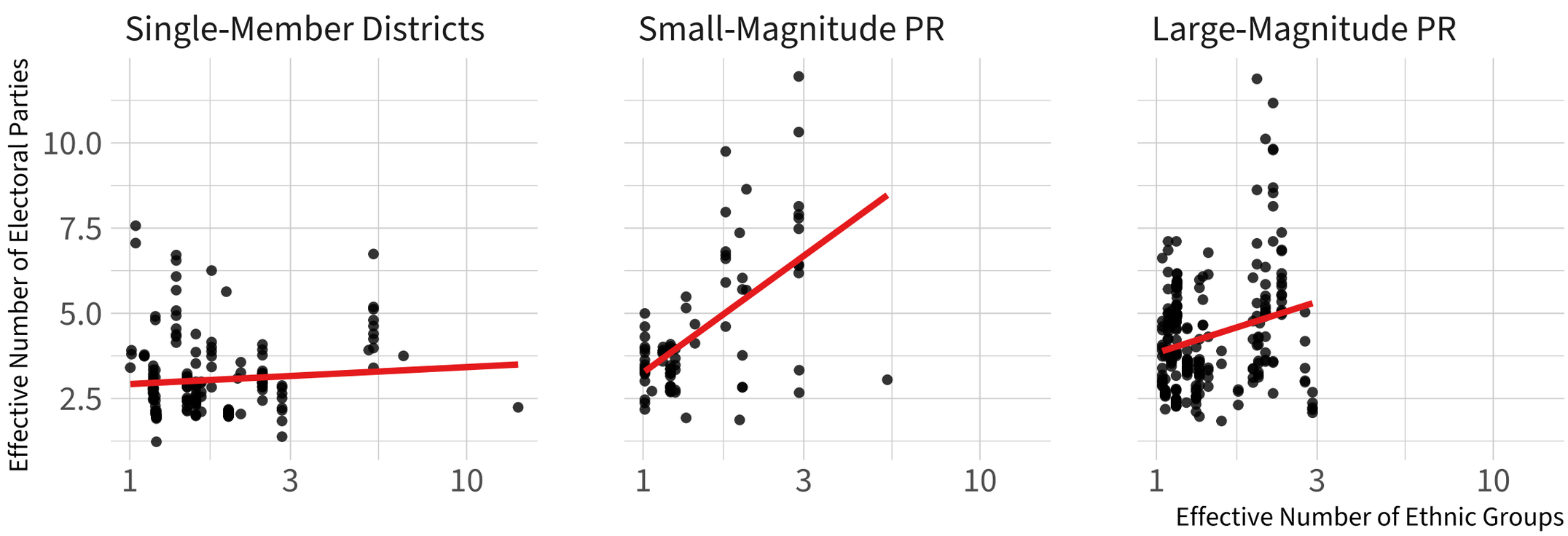
these tools are *modest* extensions
of scatterplots and histograms

We're almost always interested in $E(y \mid x)$
or $E(y \mid x = \text{hi}) - E(y \mid x = \text{lo})$.

Hypothesis 1: Social heterogeneity increases the number of parties, but only when electoral institutions are sufficiently permissive.[4]

A Simple Scatterplot

It's hard to do a lot better than this.



next time

$$y_i \sim f\left(y_i \mid \mu, \phi\right)$$

$$g(\mu) = \theta$$

Week 1 Exercises

Week 1 Exercises

I created some notes that review the concepts required for these exercises, but you should draw on familiar textbooks and notes where possible.

And this is an ambitious set of exercises. Make a serious effort to freshen your mathematical toolkit. Your effort will make for a smoother semester. If you find yourself struggling to finish all the problems, then focus on 4(f), 4(g), 5, 7, 8, 9, 10, 13(1), 14, 15, 20, 21, 22, 24(c), 24(g), 25(b), 25(e), 26(c), and 27. Use the other exercises as needed to brush up on weaknesses.

1 Fractions, Logarithms, and Exponents

Exercise 1 Some Practice with Fractions

Simplify each of the following:

a. $\frac{x}{y} + \frac{z}{y}$

b. $\frac{m}{n} - \frac{r}{n}$

c. $\frac{p+q}{r} + \frac{s}{r}$

Optional Early Assignments

- **Descriptive Research in Your Field** *(Optional; due by Sep. 3)* Reflect on what descriptive research looks like in your field and identify examples of high-quality description. Consider the following prompts as starting points:
    a. Engage with Gerring (2012), de Kadt and Grzymala-Busse (2025), and Holmes et al. (2024).[2] What parts of their arguments do you find most compelling?
    b. What is the value of description in your research area? Take one or two papers that you admire—the kinds of papers that you'd like to write—and ask yourself "what would this project have looked like if the authors had instead focused on carefully *describing* their critical concepts (rather than looking for their effects and/or causes)?"

[2] See also Part II of Gerring's *Social Science Methodology* for a supplement to Gerrying (2012).