Final Exam

Due Thurs., December 18, 2025

Insert Name Here

Instructions

Answer the questions within this document. Keep the questions included below. Make sure to answer the questions in sufficient detail.

- Report all regression results in table form using `texreg`.
- Submit your rendered .pdf in Blackboard.

1. Unusual and Influential Data; Normality of Errors

For this problem, use the "influence" data.

A. Estimate the following linear regression model and report your results in a table using `texreg`:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

B. Use studentized residuals to assess which observations are outliers. Produce a graphical plot capable of identifying outliers, and identify outlying observations. Include your graph.

C. Use DFBETAS to assess which observations are highly influential. Produce graphical plots capable of identifying influential observations. Include your graphs. Report which observations are influential.

D. Though you have no theoretical background on this data, discuss one robustness check that might be appropriate for dealing with influential observations in this data.

E. Use the Jarque-Bera test to evaluate the normality of the residuals. What do the results suggest about additional data transformations or model specifications?

2. Binary Dependent Variables

Let's say you're interested in studying the individual-level predictors of *opposition to torture* for terrorist suspects. You have a random sample of American adults from the NES data. Open the "torture" data. Your dependent variable is:

- **tor2**: opposition to torture; 1=opposed to using torture, 0=not opposed (either in favor or moderately in favor)

You include the following independent variables:

- **pid**: Party identification; 7-point scale ranging from 0=Strong Republican to 6=Strong Democrat
- **Female**: 1=female, 0=male
- **educ4**: 1=HS or less, 2=some college or Assoc. degree, 3=B.A., 4=Grad or more

- **Race3**: 1=white, 2=black, 3=Hispanic; reminder: this is nominal variable, so you need to "dummy out" the categories and include two out of the three.
- **income5**: Income; 1=<=\$22.5k, 2=\$22.5-\$45k, 3=\$45-\$75k, 4=\$75-\$125k, 5=>\$125k
- **Age**: Age in years, ranging from 18-90

A. Estimate the model using both logit and the linear probability model (LPM, with robust standard errors). Report your results in tables using `texreg`. Report goodness-of-fit statistics for each model. Interpret the results in terms of sign and significance. Also interpret goodness-of-fit in each model.

B. For each model, show the effects of *party id*, *education*, and *race* in terms of predicted probabilities. For logit, use the "observed value" approach. Report your results in graphical form. Interpret those results.

C. In general, how does the LPM compare to the logit model?

D. A critic contends that education surely attenuates partisan divisions in opposition to torture. Estimate a logit model to test whether education attenuates the effect of party identification. Report your model results in a table using `texreg`, and also report a post-estimation analysis in graphical form. Use the observed value approach for predicted probabilities. Is the critic right or wrong?

3. Comp Style Question: Probability

Consider the contingency table (cross-tabulation) below using country-level data on the existence of judicial independence (coded as "yes" or "no") and the history of coups in a country since 1950 (coded as: 0 coups, 1 or 2 coups, or 3 or more).

Table 1: Coups and Judicial Independence

	Number of Coups			
Judicial Independence	0	1 or 2	3 or more	**Total**
No	38	26	32	96
Yes	52	7	6	65
Total	90	33	38	161

A. How many "joint probabilities" can one calculate from this table? How many "marginal probabilities?" How many "conditional probabilities?" For each answer, briefly explain how you got your answer.

B. Calculate P(JI=yes, no coups). Show your work.

C. Calculate all of the marginal probabilities. Use proper notation and show your work.

D. Calculate the following three probabilities: (1) P(JI=yes | coups=0), (2) P(JI=yes |

coups=1 or 2), and (3) P(JI=yes | coups=3 or more). First, what types of probabilities are these, and what is the general equation for calculating these probabilities? Use proper notation. Second, show your work for each calculation. Third, use those probabilities to evaluate the extent to which judicial independence and coups are related. Briefly discuss how this exercise (in part d) is connected to interpreting crosstabs more generally.

4. Comp Style Question: Probability and Distributions

Imagine we have measured support for free trade using a multi-item scale. The new scale ranges from 0 (low support) to 1 (high support). The variable follows a normal distribution (approximately) with mean=0.54, s.d.=0.26.

A. Approximately what percentage of respondents have extremely high support, i.e., values greater than 0.90? Show your work.

B. What's a sufficiently low score for free trade support such that just 10% of respondents are at or below that score? Show your work.

C. Under these conditions and assumptions, what's the median value of free trade support? Explain your answer.

5. Comp Style Question: Model Specification

Answer **one** of the following two questions.

A. Let's say you estimate a model with a nonlinear functional form for the effect of one of your variables (e.g., quadratic, diminishing marginal returns). A critic cries foul, claiming that "this model makes no sense at all since *linear* regression assumes *linearity*." How would you respond to the critic's claim?

B. A critic suggests that models with interaction terms ruin the elegance of the linear regression model. According to the critic, "The whole point of regression is *generalization*, which entails estimates of *average marginal effects*. What you get from interactions is *particularization*, which entails a bunch of conditional marginal effects that don't make any sense." Critically review this critic's arguments about interactions.