**Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**PLSC 309: Quantitative Political Analysis**

**Fall 2023**

**Final Exam (85 points)**

**Read the directions and question prompts CAREFULLY. This exam is open-book: you may use notes, code, google, lecture slides, or your textbook. You MAY NOT, however, communicate with anyone other than Professor Onder during the exam. Use R where calculations are necessary, and write your answers in the spaces below. Upload this file and your R codes to Canvas.**

**Question 1**: **Interpreting Multivariate Linear Regression (25 points)**

Table 1 below is taken from a study by Canelo, et al. about how people respond to controversial statements by celebrities. Sometimes, after saying something controversial, celebrities will defend their statements by saying their speech is protected by the First Amendment of the U.S. Constitution. This research examines whether making these First Amendment defenses of controversial statements make people more or less tolerant of criticism of the statement, economic consequences from the statement, and of the statement itself.

The authors conducted multivariate linear regressions. Each of the Tolerance variables is measured on a 1-5 scale, with 5 meaning very tolerant. Look at **Table 1, Column 3** (Tolerance of Initial Statement). Higher levels of the dependent variable in this model indicate a greater tolerance of a hypothetical controversial statement.

A screenshot of a cell phone

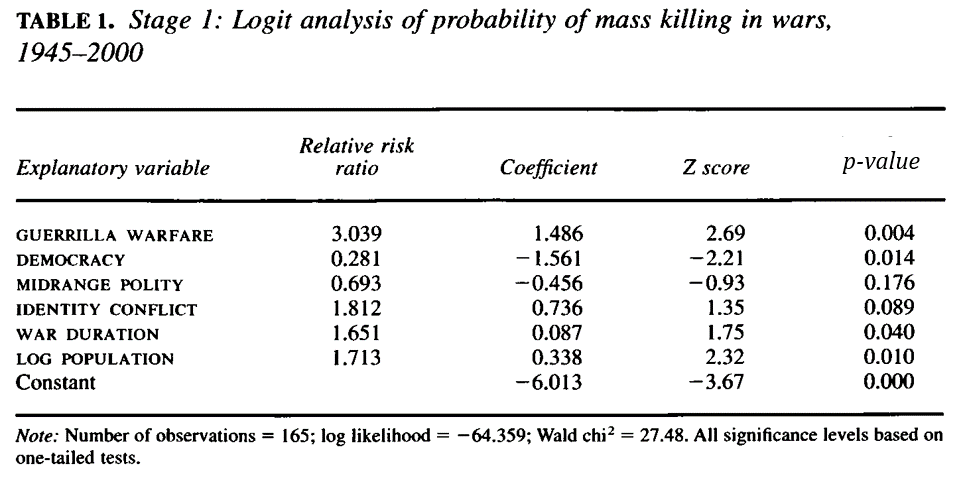
Description automatically generated

1. Which variables are statistically significant at the 95% confidence level (p<0.05)? **[5 points]**
2. First Amendment Knowledge is measured on a 0-4 scale (higher values = more knowledge). Interpret (in detailed, substantive terms) the effect of the First Amendment Knowledge variable. **[5 points]**
3. Interpret the R2 for this model. **[5 points]**
4. What is the effect of a three unit increase in Issue Agreement on the dependent variable? **[5 points]**
5. If you created a correlation matrix for the included independent variables and found that the correlation between First Amendment Knowledge and Education was 0.85. Are you concerned that multicollinearity might be affecting your results? Explain your answer? **[5 points]**

**Question 2: Interpreting Logistic Regression (17 points)**

The table below is taken from a study by Valentino, et al. about the incidence of mass killing by state forces during war. The dependent variable is coded with a 1 if the military of a government committed widespread mass killing of civilians and a 0 if not.

Here are descriptions of the other important variables:

* Guerrilla warfare: binary variable coded with a 1 if the enemy is using guerrilla warfare tactics and 0 otherwise
* Democracy: binary variable coded with a 1 if the country is democratic and 0 otherwise
* Midrange Polity: binary variable coded with a 1 if the country is an anocracy (somewhere between democracy and dictatorship)
* Identity conflict: binary variable coded with a 1 if the war is an identity (e.g., ethnic) conflict
* War duration: the number of years a war has gone on
* Log population: the population of the country (logged)

1. Why is logistic regression, rather than linear regression, the appropriate choice to model these data? What would be a downside or negative consequence of using a linear regression rather than a logistic regression in this case? **[7 points]**
2. Which variables are positively associated with mass killing **AND** statistically significant at the 95% confidence level? **[5 points]**
3. According to these results, is mass killing more likely to occur in democratic countries or dictatorships? How do you know? **[5 points]**

**Question 3: Estimating a Logistic Regression (25 points)**

Using a modified version of Fearon and Laitin’s (2003) dataset on civil wars, you will estimate a logistic regression that models the probability that a civil war starts in a country in a given year. You can read in the dataset using the following R code (which is also supplied in the provided RScript: FinalExam.R):

logitdataset<-read.csv("https://raw.githubusercontent.com/ilaydaonder/PLSC309FinalExam/main/FL2003.csv?token=GHSAT0AAAAAACEJT6GU2PJVF2AASLBRRPCEZEXBLUQ")

Here are the important variables:

* onset- A binary variable coded with a 1 if a civil war started in that country-year
* priorwar- A binary variable coded with a 1 if a country had a recent civil war
* gdpen- GDP/capita (logged)
* pop- Population of the country (logged)
* mountains- The percentage of a country covered by mountains (logged)
* oil- A binary variable coded with a 1 if the country is a major oil exporter
* newstate- A binary variable coded with a 1 if the country is a new country
* dem- A binary variable coded with a 1 if the country is a democracy
* ef- Ethnolinguistic Fractionalization (a continuous measure of ethnic diversity)
* relfrac- Religious Fractionalization (a continuous measure of religious diversity)

1. Using R, estimate a logistic regression with the *onset* variable as your dependent variable. Include all other variables listed above as the independent variables. Create a stargazer table and paste it below (You DO NOT need to edit the table to make it “pretty,” just the results are fine). **[8 points]**
2. Based on your results, which variables reduce the risk of civil war onset and are statistically significant at the 95% confidence level? **[5 points]**
3. Imagine a country that has 1) recently experienced a prior war, 2) is not a major oil producing state, 3) is a new state, 4) is a democracy, and 5) has the mean level of population, GDP/capita, mountains, ethnolinguistic fractionalization, and religious fractionalization. What is the predicted probability of a civil war in that country? (HINT: when calculating the means, don’t forget to include na.rm=TRUE) **[5 points]**
4. The Iranian Revolution began in 1978 and is classified as a civil war due to the level of violence. In 1978, Iran:
   1. did not have a recently experienced prior war,
   2. was a major oil producing state,
   3. was not a new state,
   4. was not a democracy,
   5. had a logged population of 10.5,
   6. had a logged GDP/capita of 5.2,
   7. had a logged mountainous terrain value of 3.7,
   8. had an ethnolinguistic fractionalization score of 0.67, and
   9. had a religious fractionalization score of 0.2.

Based on this model, what was the predicted probability of a civil war in that case? In your view, did the model do a good job of predicting that civil war? **[7 points]**

**Question 4. Estimating Multiple Linear Regression (18 points)**

We political scientists get in the news every two years because the media wants us to predict election outcomes. In reality, political scientists often do so with simple regression analyses. These models typically predict the two-party popular vote based on the state of the economy, the incumbent’s popularity, and whether one of the candidates is an incumbent president.

You can read in a dataset with this information since 1948 dataset with this R code (which is also supplied in the provided RScript):

electionsdataset<-read.csv("https://raw.githubusercontent.com/ilaydaonder/PLSC309FinalExam/main/elections.csv?token=GHSAT0AAAAAACEJT6GVNLSAIYQVYTA4KPEQZEXBQBQ")

You’re going to predict the result of the 2020 election, even though it has already happened! Here are the variables you need:

* incumbent\_vote: the percentage of the two-party vote received by the current president’s political party (in 2020, the incumbent President was Donald Trump, a republican)
* ch\_gdp: Percent change in real disposable income per capita
* approval: Incumbent president’s approval rating in the final Gallup Poll in June
* unemployment: Seasonally adjusted employment rate in June before the election
* incumbent: Presence (1) or absence (0) of a first-term incumbent in the race

1. Estimate three linear regressions, all with incumbent\_vote as the dependent variable.
   1. For the first model, include only approval as independent variable.
   2. For the second model, include approval, ch\_gdp, incumbent, and unemployment as independent variables.
   3. For the third model, include (a) ch\_gdp, (b) incumbent, (c) unemployment, and (d) the interaction between incumbent and unemployment as independent variables.

Paste the stargazer table presenting the results below (again, no need to make it pretty). **[8 points]**

1. Which of the three models fits better? How do you know? **[5 points]**
2. Predict the 2020 election! As of early December, President Trump’s approval rating is 43% (put in 43, not 0.43), and the percent change in real disposable income per capita is 0.84. President Trump is an incumbent. The unemployment rate in June was 11.1% (put in 11.1, not 0.111). Using the **second** model, what is Trump’s expected vote share if the election were held today? **[5 points]**
3. Look at Model 3. What is the effect of being an incumbent on their vote share in a hypothetical situation where unemployment rate was 0, all else equal? **[2 points Extra Credit]**