SPEC REU R Resources: Visualizing & Analyzing Regressions Results – Groupwork

Gaea Morales and Claudia Salas Gimenez

March 2025

For this group work, we will continue working with the `IDC_training_2021.rds` dataset to explore the relationship between ethnic political power, political participation, and democratic outcomes.

The goal of this assignment is to practice constructing simple and multivariate regression models (both linear and binomial), generating preliminary visualizations to better understand the data, and formatting professional-quality tables. Your task will be to estimate four regression models—two simple regression models and two multivariate models—plot the results, and present them in a table using the `texreg` package.

Initial Setup

Before we begin, ensure that your working directory is set up correctly and that you have loaded the required libraries and the `IDC_training_2021.rds` dataset.

For reference, the IDC Powersharing Dataset was introduced in the research paper *"De Jure Powersharing 1975–2019: Updating the Inclusion, Dispersion, and Constraints Dataset"* by Ziff, Barnum, Abadeer, Chu, Jao, Zaragoza, and Graham (2024). You can find the research paper and the online appendix here.

```
# Set working directory
#setwd("YourFolderPath")

# Load required libraries
library(dplyr)
library(texreg)
library(ggplot2)
library(broom)
## We will use this package to tidy model results for plotting

# Load the dataset
IDC_training_2021 <- readRDS("IDC_training_2021.rds")
```

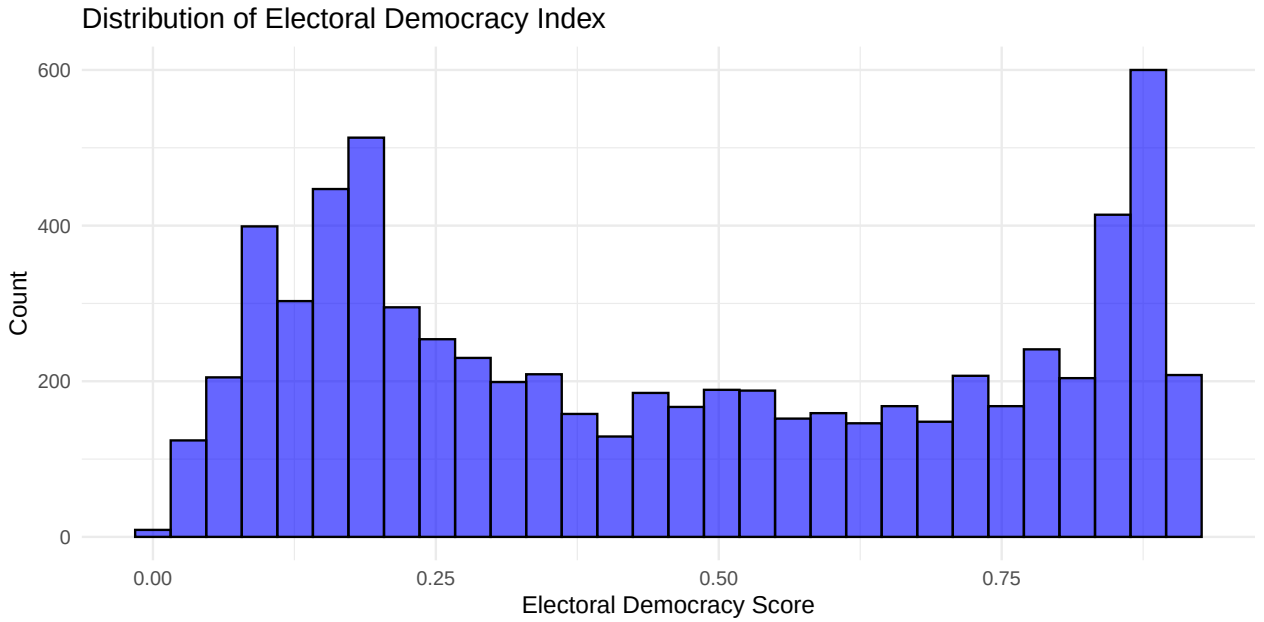
Exercise 1: Exploring the Relationship Between Ethnic Political Power and Democracy

How does ethnic minority political power influence democratic outcomes? In this group work, we will explore whether ethnic minority political power influences democratic outcomes, using two different measures for democracy as dependent variables: a continuous electoral democracy index and a binary classification of democratic status.

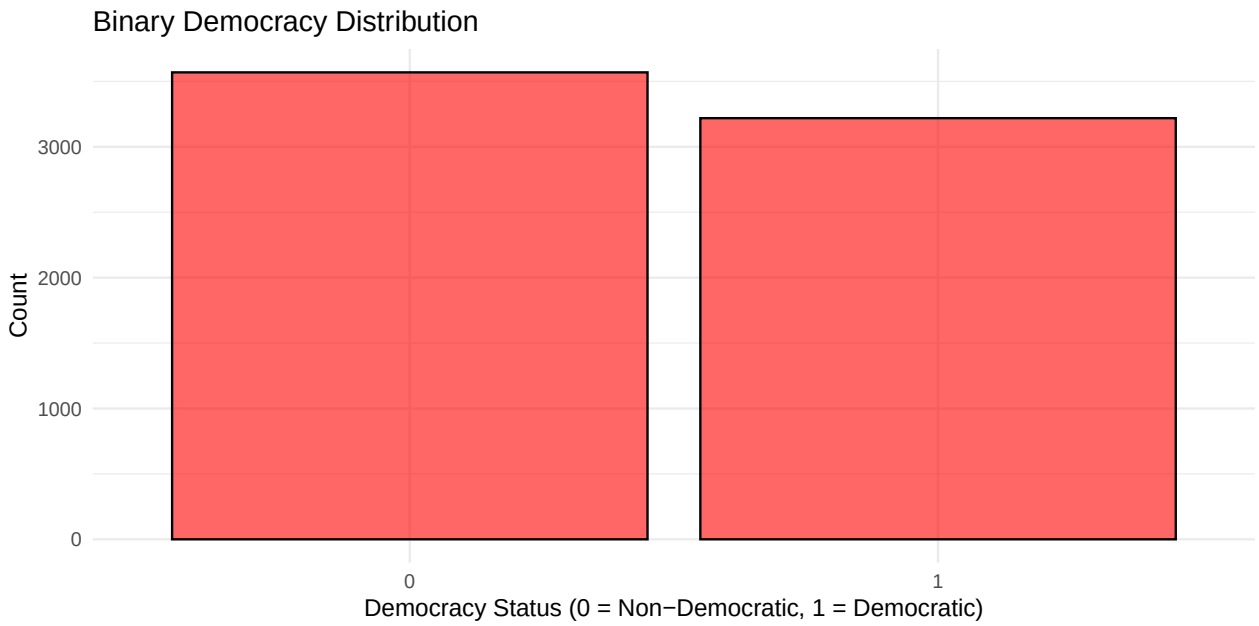
Exercise 1.1: Preliminary Plots

Before running any regressions, plot the distribution of electoral democracy scores and the frequency of democratic (1) and non-democratic (0) observations. These preliminary visualizations will help identify skewness, outliers, and patterns that could impact the regression results.

```
ggplot(IDC_training_2021, aes(x = v2x_polyarchy_VDEM)) +
  geom_histogram(bins = 30, fill = "blue", alpha = 0.6, color = "black") +
  theme_minimal() +
  labs(title = "Distribution of Electoral Democracy Index",
       x = "Electoral Democracy Score",
       y = "Count")
```



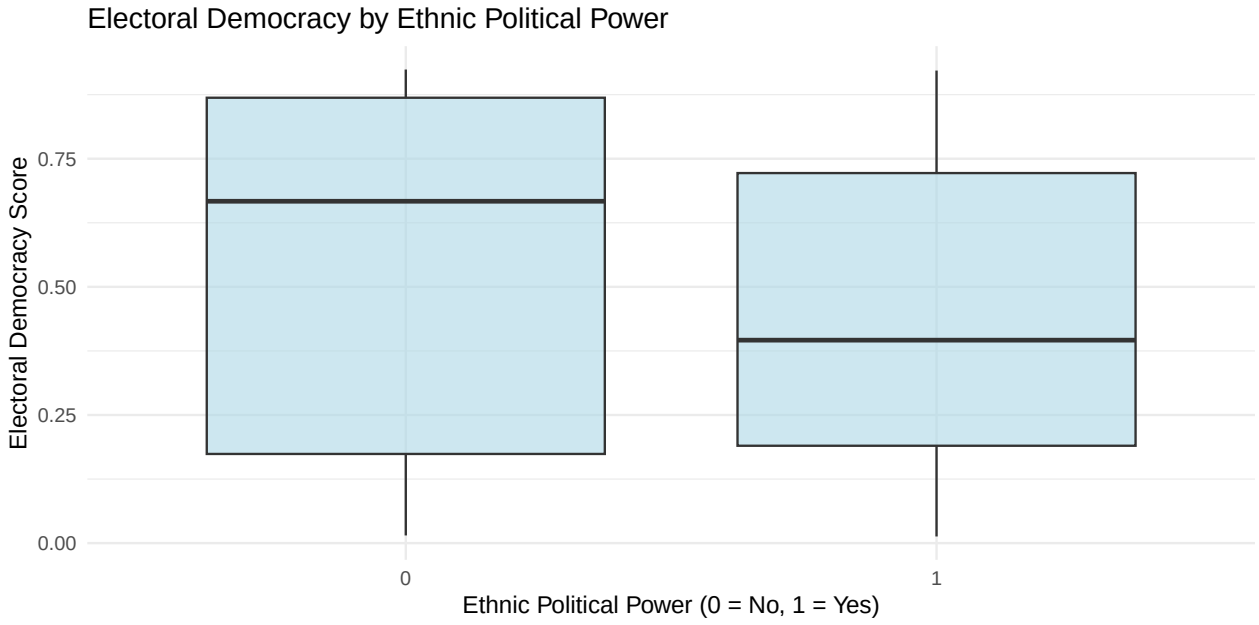

```
ggplot(IDC_training_2021 %>% filter(!is.na(democracy_BX)),
         ## Remove NA values from democracy_BX before plotting
       aes(x = as.factor(democracy_BX))) +
  geom_bar(fill = "red", alpha = 0.6, color = "black") +
  theme_minimal() +
  labs(title = "Binary Democracy Distribution",
       x = "Democracy Status (0 = Non-Democratic, 1 = Democratic)",
       y = "Count")
```



Bonus Exercise

As a bonus question, visualize the distribution of electoral democracy scores across countries based on whether ethnic minority groups have political power (`cntrl_relevance_binary_EP`). The goal of this plot is to reveal potential differences in democracy levels based on ethnic political power and provide an initial understanding of the relationship before running regressions.

```
ggplot(IDC_training_2021 %>% filter(!is.na(cntrl_relevance_binary_EP)),  # Remove NAs
       aes(x = as.factor(cntrl_relevance_binary_EP), y = v2x_polyarchy_VDEM)) +
  geom_boxplot(fill = "lightblue", alpha = 0.6) +  # Boxplot with transparency
  theme_minimal() +
  labs(title = "Electoral Democracy by Ethnic Political Power",
       x = "Ethnic Political Power (0 = No, 1 = Yes)",
       y = "Electoral Democracy Score")
```



```
## The median electoral democracy score is higher for countries where ethnic minority
## groups do not have political power (0).
## Countries where ethnic minorities do have political power (1) tend to have lower
## democracy scores on average and a slightly more compressed distribution.
## The interquartile range (IQR) is larger for cases where minorities lack political
## power, suggesting greater variability in democracy scores within this group.
```

Exercise 1.2: Running Simple Regressions

Now, to begin the regression analysis, estimate two simple regression models, one for each measure of democratic outcome, using the binary measure for whether ethnic minority groups have political power as the independent variable. Save the output of these models as objects with clear and appropriate names.

Given the different types of dependent variables (continuous and binary), what statistical models are appropriate?

```
# Estimate models
## DV: electoral democracy
mod1 <- lm(v2x_polyarchy_VDEM ~ cntrl_relevance_binary_EP,
           data = IDC_training_2021)

summary(mod1)
```

```
##
## Call:
## lm(formula = v2x_polyarchy_VDEM ~ cntrl_relevance_binary_EP,
##     data = IDC_training_2021)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52875 -0.26679 -0.02879  0.29221  0.47321
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.543746   0.008609   63.16   <2e-16 ***
## cntrl_relevance_binary_EP -0.094953   0.009428  -10.07   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2886 on 6763 degrees of freedom
##   (751 observations deleted due to missingness)
## Multiple R-squared:  0.01478,    Adjusted R-squared:  0.01463
## F-statistic: 101.4 on 1 and 6763 DF,  p-value: < 2.2e-16
```

```
## DV: binary measure of democracy
### Remember to use a logit model using glm() with link = "logit" function when the
### dependent variable is binary
mod2 <- glm(democracy_BX ~ cntrl_relevance_binary_EP,
            family = binomial(link = "logit"),
            data = IDC_training_2021)

summary(mod2)
```

```
##
## Call:
## glm(formula = democracy_BX ~ cntrl_relevance_binary_EP, family = binomial(link = "logit"),
```

```
##     data = IDC_training_2021)
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 0.26423    0.06247   4.230 2.34e-05 ***
## cntrl_relevance_binary_EP -0.42353    0.06816  -6.213 5.19e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8932.4  on 6452  degrees of freedom
## Residual deviance: 8893.5  on 6451  degrees of freedom
##   (1063 observations deleted due to missingness)
## AIC: 8897.5
##
## Number of Fisher Scoring iterations: 3
```

Exercise 1.3: Estimating Multivariate Models

Next, since we are interested in the effects of ethnic politics on democracy, let's deepen the analysis by introducing control variables that might shape the relationship between ethnic politics and democratic outcomes. Two potentially relevant factors are institutional constraints on ethnic political participation (whether rules exist banning the formation of ethnically-based political parties) and the economic development level of a country (measured by GDP per capita).

Adjust your models in Exercise 1.2 to include these controls and assess whether they alter the initial findings.

Helpful Hint Instead of using GDP per capita directly, transform it using the natural logarithm to reduce skewness, improve linearity in regression models, and better capture diminishing returns in economic effects on democracy.

```r
# Estimate multivariate models
## DV: electoral democracy
mod3 <- lm(v2x_polyarchy_VDEM ~ cntrl_relevance_binary_EP + partynoethnic_IDC +
             log(gdppc_WDI_PW), data = IDC_training_2021)

summary(mod3)
```

```
##
## Call:
## lm(formula = v2x_polyarchy_VDEM ~ cntrl_relevance_binary_EP +
##     partynoethnic_IDC + log(gdppc_WDI_PW), data = IDC_training_2021)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.83094 -0.18356  0.07579  0.18056  0.47999
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.320750   0.022115 -14.504  < 2e-16 ***
## cntrl_relevance_binary_EP -0.012027   0.008985  -1.339  0.18076
## partynoethnic_IDC         -0.025809   0.009361  -2.757  0.00585 **
## log(gdppc_WDI_PW)          0.098169   0.002191  44.806  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.2435 on 6244 degrees of freedom
##   (1268 observations deleted due to missingness)
## Multiple R-squared:  0.286,  Adjusted R-squared:  0.2856
## F-statistic: 833.6 on 3 and 6244 DF,  p-value: < 2.2e-16
```

```
## DV: binary measure of democracy
mod4 <- glm(democracy_BX ~ cntrl_relevance_binary_EP + partynoethnic_IDC +
              log(gdppc_WDI_PW), family = binomial(link = "logit"),
            data = IDC_training_2021)

summary(mod4)
```

```
##
## Call:
## glm(formula = democracy_BX ~ cntrl_relevance_binary_EP + partynoethnic_IDC +
##     log(gdppc_WDI_PW), family = binomial(link = "logit"), data = IDC_training_2021)
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -4.662332   0.211195 -22.076  < 2e-16 ***
## cntrl_relevance_binary_EP   0.003141   0.085414   0.037    0.971
## partynoethnic_IDC          -0.533545   0.088613  -6.021 1.73e-09 ***
## log(gdppc_WDI_PW)           0.571876   0.021436  26.678  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8318.1  on 6000  degrees of freedom
## Residual deviance: 7214.7  on 5997  degrees of freedom
##   (1515 observations deleted due to missingness)
## AIC: 7222.7
##
## Number of Fisher Scoring iterations: 4
```

Exercise 2: Creating a Publication-Grade Table

Now that we have estimated our four models, present all the results in a well-structured table in HTML format.

Note: Make sure you save your output file in the appropriate folder on your device with a clear filename.

```
# Create a publication-grade table
htmlreg(list(mod1, mod3, mod2, mod4),
        file = "democracyregs_gw_22.html",
        caption = "Ethnic Politics and Democracy",
        caption.above = TRUE,
        custom.header = list("Electoral Democracy" = c(1, 2),
                             "Binary Democracy" = c(3, 4)),
        custom.coef.map = list("cntrl_relevance_binary_EP" = "Ethnic Political Power",
                               "partynoethnic_IDC" = "Ethnic Party Ban",
                               "gdppc_WDI_PW" = "GDP per capita",
                               "(Intercept)" = "Intercept"),
        stars = c(0.001, 0.01, 0.05),
```

```
        custom.note = "%stars. Polyarchy index is from vDem; binary democracy is
        from Boix, Miller, and Rosato.")
```

Exercise 3: Interpreting Results

Lastly,, in a few sentences, interpret the results displayed in your table. What do the coefficients suggest about the relationship between ethnic political power and democracy? How do the results change when adding control variables? Do these results align with our initial predictions?

```
## In Model 1 (electoral democracy) and Model 2 (electoral democracy with controls),
## the coefficient for thnic political power is negative but loses significance when
## controls are added.
##
## In Model 3 (binary democracy) and Model 4 (binary democracy with controls), the
## negative effect of ethnic political power is strong and significant in Model 3
## (-0.42***) but disappears in Model 4 after adding controls.
##
## The boxplot initially suggested that countries where ethnic minorities have political
## power tend to have lower democracy scores, which aligns with the significant negative
## coefficients in Models 1 and 2.
```

Conclusion

In this group assignment, we explored the relationship between ethnic political power and democracy using both simple and multivariate regression models. The preliminary visualizations helped us identify patterns in the data, while the regression analysis allowed us to evaluate the role of ethnic political power in shaping democratic outcomes.

In the next homework assignment, we will continue practicing these concepts, refining our regression models and professional-quality tables, and further improving our interpretation of statistical results.