Analysis of the 1933 German Election during the Weimar Republic

Prof. Ryan Weldzius

Due Date: 2025-04-16 by 11:59PM ET

Who voted for the Nazis? Researchers have attempted to answer this question by analyzing aggregate election data from the 1933 German election during the Weimar Republic. This exercise is based on the following article: Gary King, Ori Rosen, Martin Tanner, and Alexander F. Wagner (2008) "Ordinary Economic Voting Behavior in the Extraordinary Election of Adolf Hitler." *Journal of Economic History* vol. 68, no. 4, pp. 951-996.

We analyze a simplified version of the election outcome data, which records, for each precinct, the number of eligible voters as well as the number of votes for the Nazi party. In addition, the data set contains the aggregate occupation statistics for each precinct. The table below presents the variable names and descriptions of the data file `nazis.csv`. Each observation represents a German precinct.

Name	Description
`shareself`	Proportion of self-employed potential voters
`shareblue`	Proportion of blue-collar potential voters
`sharewhite`	Proportion of white-collar potential voters
`sharedomestic`	Proportion of domestically employed potential voters
`shareunemployed`	Proportion of unemployed potential voters
`nvoter`	Number of eligible voters
`nazivote`	Number of votes for Nazis

The goal of analysis is to investigate which types of voters (based on their occupation category) cast ballots for the Nazis. One hypothesis says that the Nazis received much support from blue-collar workers. Since the data do not directly tell us how many blue-collar workers voted for the Nazis, we must infer this information using a statistical analysis with certain assumptions. Such an analysis where researchers try to infer individual behaviors from aggregate data is called *ecological inference*.

To think about ecological inference more carefully in this context, consider the following simplified table for each precinct i:

Vote-Choice	Blue-Collar	Non-blue-collar	
Nazis	W_{i1}	W_{i2}	Y_i
Other parties or abstention	$1 - W_{i1}$	$1 - W_{i2}$	$1 - Y_i$
	X_i	$1 - X_i$	

The data at hand only tells us the proportion of blue-collar voters X_i and the vote share for the Nazis Y_i in each precinct, but we would like to know the Nazi vote share among the blue-collar voters W_{i1} and among the non-blue-collar voters W_{i2}. Then, there is a deterministic relationship between X, Y, and W_1, W_2. Indeed, for each precinct i, we can express the overall Nazis' vote share as the weighted average of the Nazis' vote share of each occupation:

$$Y_i = X_i W_{i1} + (1 - X_i) W_{i2}$$

Question 1 (5 points)

We exploit the linear relationship between the Nazis' vote share Y_i and the proportion of blue-collar voters X_i given in the equation above by regressing the former on the latter. That is, fit the following linear regression model:

$$E(Y_i|X_i) = \alpha + \beta X_i$$

Compute the estimated slope coefficient, its standard error, and the 95% confidence interval. Give a substantive interpretation of each quantity.

Answer 1

```
# Load packages you might need
library("tidyverse")
library("broom")
library("modelr")
```

Write answer here.

Question 2 (5 points)

Based on the fitted regression model from the previous question, predict the average Nazi vote share Y_i given various proportions of blue-collar voters X_i. Specifically, plot the predicted value of Y_i (the vertical axis) against various values of X_i within its observed range (the horizontal axis) as a solid line. Add 95% confidence intervals as dashed lines or shaded areas. Give a substantive interpretation of the plot.

Answer 2

Write answer here.

Question 3 (5 points)

Fit the following alternative linear regression model,

$$E(Y_i|X_i) = \alpha^* X_i + (1 - X_i)\beta^*$$

Note that this model does not have an intercept. How should one interpret α^* and β^*? How are these parameters related to the linear regression model given in Question 1?

Answer 3 (5 points)

Write answer here.

Question 4 (5 points)

Fit a linear regression model where the overall Nazi vote share is regressed on the proportion of each occupation. The model should contain no intercept and five predictors, each representing the proportion of a certain occupation type. Interpret the estimate of each coefficient and its 95% confidence interval. What assumption is necessary to permit your interpretation?

Answer 4

Write answer here.

Question 5 (Extra credit: 3 points)

Finally, we consider a model-free approach to ecological inference. That is, we ask how much we can learn from the data alone without making an additional modeling assumption. For each precinct, obtain the smallest value that is logically possible for W_{i1} by considering the scenario in which all non-blue-collar voters in precinct i vote for the Nazis. Express this value as a function of X_i and Y_i. Similarly, what is the largest possible value for W_{i1}? Calculate these bounds, keeping in mind that the value for W_{i1} cannot be negative or greater than 1. Finally, compute the bounds for the nation-wide proportion of blue-collar voters who voted for the Nazis (i.e., combining the blue-collar voters from all precincts by computing their weighted average based on the number of blue-collar voters). Give a brief substantive interpretation of the results.

Answer 5

Write answer here.