

## Problem Set 5: Interpreting Bivariate Statistics and Plotting Political Variables

For this assignment, I'm asking you to perform some basic analysis of some data on vote shares for California House incumbents from the 2004 election. The data set for this assignment can be found on the website (right below the assignment listing). There are two versions of the data: Stata and CSV. You will need to read the data into R in a fashion identical to what you did for problem set 2 and/or 4. Once you attach the data in R, please answer the following questions/do the following tasks (this assignment is due the day of the final exam):

The main variables you will work with are: **demvote** which is the vote share for Democratic incumbents; **repvote** which is the vote share for Republican incumbents; **balavs** which is the vote share for incumbents from the Bay Area/Los Angeles area; and **nonbalavs** which is the vote share for incumbents not from the Bay Area/LA area.

Please answer the following questions:

1. Provide a boxplot of the Democratic vote share and Republican vote share. Interpret the plot. (10 points)
2. A researcher hypothesizes that because Californians tend to, on average, be more liberal than individuals from other states, Democratic incumbents will, on average, tend to garner a larger share of the two-party vote than when compared to Republicans.
  - a. State the appropriate null and alternative hypotheses. (6 points)
  - b. Provide a  $t$ -test of this hypothesis (show your R code). What is the estimated  $t$ ? (5 points)
  - c. Given this  $t$ -statistic what is the critical  $t$ -value? Assume your alpha level is .05. Consult the table below to determine the correct critical  $t$ . (2 points)
  - d. Given the critical  $t$  and your  $t$ -statistic, can you reject the null hypothesis? Why or why not? (6 points)
  - f. What is the  $p$ -value for this test (as given in your R output? What is the interpretation of this number (4 points)
  - e. What is your overall interpretation of the relationship between vote shares and party? (7 points)
3. Provide a boxplot of the BALA vote share and non-BALA vote share. Interpret the plot. (10 points)
4. A researcher hypothesizes that because voters in districts in the two major metropolitan regions (the Bay Area and Los Angeles) tend to be more homogeneous than compared to non-

BALA districts, vote shares for incumbents from BALA districts will on average be higher than incumbents from non-BALA districts.

- a. State the appropriate null and alternative hypotheses. (6 points)
- b. Provide a  $t$ -test of this hypothesis (show your R code). What is the estimated  $t$ ? (5 points)
- c. Given this  $t$ -statistic what is the critical  $t$ -value? Assume your alpha level is .05. Consult the table below to determine the correct critical  $t$ . (2 points)
- d. Given the critical  $t$  and your  $t$ -statistic, can you reject the null hypothesis? Why or why not? (6 points)
- f. What is the  $p$ -value for this test (as given in your R output? What is the interpretation of this number (4 points)
- e. What is your overall interpretation of the relationship between vote shares and region? (7 points)

**Table 1. Critical t-values for 52 degrees-of-freedom (two-tailed):**

df	.10	.05
52	1.6747	2.0066