ME104 Linear Regression Analysis: Problem Set 3
Inference, Intervals, and Hypothesis Testing

For this lab we will use the Comparative Study of Electoral Systems (CSES) Module 3: 2006-2011. CSES is a collaborative program of cross-national research among election studies conducted in over fifty states. Participating countries include a common module of survey questions in their post-election studies.

1. CSES asks the question "Where would you place yourself on this scale?" to tap the respondents position on a substantive left-right placement scale where 0 represents "Left" and 10 represents "Right". Since this is a 10 point scale, a mean above 5 shows a propensity towards the right of the ideological spectrum against a moderate value of 5. Let's suppose you want to know whether voters are increasingly collocated towards the "Right".

   (a) Define notation and state the null and alternative hypotheses for studying this.

   (b) For responses to the question "Where would you place yourself on this scale?", Stata shows the following result:

   ```
   One-sample t test

   Variable |     Obs        Mean     Std. Err.    Std. Dev.    [95% Conf. Interval]

   C3013NEW |   46813     5.477752    .0114345     2.474013     5.45534    5.500164

       mean = mean(C3013NEW)                                            t =  41.7815
   Ho: mean = 5                                          degrees of freedom =    46812

       Ha: mean < 5                   Ha: mean != 5                     Ha: mean > 5
    Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
   ```

   (c) Report the test statistic value, and show how it was obtained from other values reported in the table.

   (d) Explain what the P-value represents and interpret its value.

2. In response to the statement "On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the way democracy works in your country?", the response categories (Very satisfied, Fairly satisfied, Not very satisfied and Not at all satisfied) have scores 1, 2, 4, 5 and counts (5054, 28051, 15697 and 5618) [variable *C3019*]. Stata reports the following summary statistics:

Variable	Obs	Mean	Std. Dev.	Min	Max
C3019NEW	54420	2.793716	1.243297	1	5

   (a) Set up null and alternative hypotheses to test whether the population mean response differs from the neutral value, 3.

(b) Find the test statistic and P-value. Interpret and make a decision about the null hypothesis using $\alpha = 0.05$.

(c) Construct a 95% confidence interval for μ. Show the correspondence between the test and the confidence interval.

(d) Verify your result using Stata. Hints: before running the test you need to generate a new variable *C3019NEW* equal to *C3019* but without missing values.

3. In response to the statement "Would you say that any of the parties in your country represent your views reasonably well?", [variable $C3007_1$], 29722 people answered yes and 19869 answered no. Stata shows the following results to analyze whether a majority or minority of respondents would answer yes:

```
One-sample test of proportion          C3007_1n: Number of obs =    49591
```

Variable	Mean	Std. Err.	[95% Conf. Interval]	
C3007_1n	.4006574	.0022005	.3963445	.4049703

```
     p = proportion(C3007_1n)                               z = -44.2453
Ho: p = 0.5

    Ha: p < 0.5                Ha: p != 0.5                Ha: p > 0.5
 Pr(Z < z) = 0.0000      Pr(|Z| > |z|) = 0.0000       Pr(Z > z) = 1.0000
```

(a) Specify the hypotheses that are tested.

(b) Report and interpret the test statistic value.

(c) Report and interpret the P-value as a probability.

(d) Construct a 95% confidence interval for μ. Show the correspondence between the test and the confidence interval.

(e) In the words of this example, what would be a i) Type I error and ii) Type II error?

4. Additional work

(a) Using the Gallup et al. (1999) dataset `geodata.dta`, regress `gdp90` on `open6590` and `airdist`. Manually calculate the t-statistic of the `open6590` coefficient and check this test statistic against the critical value for t with $\alpha = 0.05$. Interpret the result.

(b) Manually calculate the F-statistic and R-squared values. Briefly discuss.