

1. The following data is a random sample of brands of vanilla yogurt and their calorie content. A diet guide claims that you will get 120 calories from a serving of vanilla yogurt. Is there sufficient evidence that the mean number of calories from vanilla yogurt is different from the diet guide's stated claim? List hypotheses and use PROC TTEST in SAS and `t.test()` in R to carry out the analysis.

160	120	200	220	230	120	180
140	130	170	190	80	100	170

2. The dataset called `economy.csv` contains data from the 2008 *Economic Report of the President*. Six variables are included in the dataset:

Year: 1960-2012; the data starts in 1960, as a sort of general beginning of the "modern" government/budget/economic era.

Unemployment: Civilian unemployment rate, in percent.

GDP Growth: percent change, over the previous year, in the real gross domestic product.

Inflation: percent change in special consumer price index (CPI), December to December.

Surplus: surplus (positive) or deficit (negative) in the federal budget, as a percent of GDP.

Party: Political party of the president responsible for that budget year. R = Republican and D = Democratic.

One of the most important things that a president does to influence the economy is prepare/propose/haggle and (hopefully) eventually sign a federal budget. The federal budget in the first year of a president's term is set by the president in the previous year, so this was used as a time lag. For instance, 1960 was a republican year; Kennedy took office in January of 1961, but he had effectively no budgetary influence until the new federal fiscal year began in the following October. Thus, 1961 is assigned as a Republican year.

The idea for this dataset came from Michael Kinsley, writing for the *Los Angeles Times*, in a paper published in August 2004 (he has also reprised his analysis in Slate.com). He took some of these variables and calculated and compared means only; no consideration for any sort of variation, let alone at any attempt at valid statistical inferences. [This problem and its description was borrowed from Dr. Brian Dennis ☺] Use both R and SAS to do the analyses.

- a. Obtain summary statistics for each of the 4 economic variables (don't do Year and Party will be your grouping variable) BY PARTY. Include sample size, mean, sd, variance, and 5 number summary.
- b. Create boxplots (in pairs by Party) of each of the variables.
- c. Create histograms of all variables by party.
- d. Perform a 2-sample t-test on each of the 4 economic variables, using Party as the categorical variable. Perform the tests with 2-sided alternative hypotheses, using  $\alpha = 0.10$ . For each test, list all hypotheses and obtain the output table with the numerical analyses (including the F test for equality of variances to determine which 2-sample t-test is the appropriate test).