*Ecn 102 - Analysis of Economic Data*
University of California - Davis
John Parman

January 13, 2011

# Problem Set 2

This problem set will be not be graded and does not need to be turned in. However, it will provide a good review for material that will appear on the first midterm and good practice with Excel skills that you will continue to use later in the course. You are strongly encouraged to work through the entire problem set.

---

1. **Formulating a Hypothesis** For each scenario below, write down the most appropriate null and alternative hypothesis you would use for statistical inference on the population mean. Also write down the critical value you would use for your hypothesis testing if you want to use a five percent significance level.

   (a) It has been determined that mercury levels higher than two parts per billion in drinking water are unsafe. You are responsible for determining whether Davis drinking water is safe based on 100 water samples taking from different parts of Davis.

   (b) A cookie company has determined that it is most profitable to have 15 chocolate chips in each cookie. If there are more than 15, the cookies become too expensive to produce. If there are less than 15, the company loses customers. They want to test whether they are operating at the efficient level by looking at a sample of 500 cookies.

   (c) A pollster wants to determine whether a majority of voters would vote for a constitutional convention for California. The pollster has the opinions of 1000 randomly sampled voters.

2. **Hypothesis Testing with Continuous Data** For this question, use the Yolo county census data available on Smartsite (cens00-yolo.csv). These data are a 5% sample of the Yolo county population aged 16 to 65 from the year 2000. HOURS gives the individual's average number of hours worked per week in the previous year. INCTOT gives the individual's total annual income for the previous year.

   (a) Drop any observations for which hours worked last week are zero or total income is zero.

   (b) Using the remaining observations, test the following hypothesis about the mean annual income $\mu$ using a significance level of 10%:

   $$H_o : \mu \leq \$35000$$

   $$H_a : \mu > \$35000$$

(c) If you want to test whether the mean Yolo county annual income is equal to \$34,000, what would your p-value be?

(d) Based on your answer to part (c), what is the lowest significance level at which you would reject the null hypothesis that the mean annual income is equal to \$34,000?

(e) Calculate a 90% confidence interval for the population mean of annual income.

(f) How would you expect your answers to the previous questions to change if you had a 10% sample of Yolo county residents rather than a 5% sample?

3. **Hypothesis Testing with Proportions Data** For this question, use the Gallup poll data on whether or not people favored statehood for Alaska and Hawaii (statehood-poll.csv). This was a poll taken in the 1950's and was meant to be a random sample of the adult population of the United States.

(a) Suppose that the government decided to grant statehood only if 70% of the population favored statehood. Based on the poll data, use an upper one-tailed test to determine whether Alaska should receive statehood. At what significance levels would Alaska be granted statehood?

(b) Do the same for Hawaii. Are there any significance levels at which Hawaii would be granted statehood?

(c) Calculate a 95% confidence interval for the percentage of people who favor statehood for both Hawaii and Alaska. (Hint: You may need to construct a new variable to do this.)

4. **Type I and Type II Errors**

(a) Describe a situation in which a researcher would be very concerned about Type I errors. Would the researcher choose a large or small value for $\alpha$ in this situation?

(b) Describe a situation in which a researcher would be very concerned about Type II errors. Would the researcher choose a large or small value for $\alpha$ in this situation?

(c) Draw a graph showing the probability of a Type II error when $\mu_0$ is 100 but the true population mean is 110.

5. **Univariate Data Transformation**

(a) Using the GDP data from Problem Set 1, construct a graph that shows the growth rate of real GDP over time and a three-year moving average of the growth rate of GDP over time. Does the three-year moving average effectively smooth the graph? What happens if you switch to a seven-year moving average?

(b) Suppose that we have data on daily temperatures for a period of six months. The temperatures are given in degrees celcius. The data have a mean of 15 degrees celcius, a standard deviation of 4 degrees celcius and a range of 18 degrees celcius.

Suppose that we convert the data into degrees fahrenheit. What will the new mean, standard deviation and range be? Are there any of the measures of central tendency and dispersion that we've discussed that would not change?