MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017/1.010 Computing and Data Analysis for Environmental Applications / Uncertainty in Engineering

---

Problem Set 7: Hypothesis Tests
Due: Thursday, November 13, 2003

Please turn in a hard copy of your MATLAB program as well as all printed outputs (tables, plots, etc.) required to solve the problem.

Problem 1: Large Sample Two-sided Hypothesis Tests for a Single Population

The NOAA National Climatic Data Center (NCDC) provides climatological records for many cities in the US. For example, you can find a record of annual precipitation totals at Boston at:

   Boston_precip.txt

For this exercise, you should use the NOAA data set to test the following hypothesis about the mean annual rainfall at Boston:

$$H0: a = E[annual\ rainfall] = 100\ cm$$

Derive a test statistic for three cases, using the first 10 years, first 40 years, and all 90 years of data, respectively. For each case, determine whether or not you can reject H0 at the 5% significance level (i.e. for α=0.5). Make a "large sample" assumption and use a two-sided test. Also report the p value for the test.

Problem 2: Large Sample Two-sided Hypothesis Tests for Two Populations

The Massachusetts Water Resources Authority (MWRA) conducts an ongoing program of monitoring bacterial levels in Boston Harbor and its tributaries. An example of some of the data collected is provided in the file MWRA1.txt . This file lists the coliform count C (reported in organisms per 100 ml) from the indicated stations and dates. The stations can be grouped as follows:

   Outer harbor (control): 104, 105, 118, 81, 82
   Inner harbor (affected): 19, 20

Check the null hypothesis that the means of the inner and outer harbor coliform levels are the same. Work with the transformed concentration $C_T = ln\ (C+1)$. Report the p level for a large sample two-sided test.

Some relevant MATLAB functions: normcdf,norminv,normplot, mean,var

Problem 3: Small Sample Two-sided Hypothesis Test for a Single Population, based on Stochastic Simulation

Suppose that you wish to test the hypothesis that the standard deviation of a lognormal population is equal to its mean (i.e. $H0$: $SD(x)$ - $E(x) = 0$).   You are given the following small sample:

$$[x_1, x_2, x_3, x_4, x_5] = [1.76 \quad 0.51 \quad 3.08 \quad 3.62 \quad 0.86]$$

Suggest an appropriate standardized small sample test statistic and use stochastic simulation to derive an approximate CDF for this statistic. Then determine the p value for a double-sided test of the hypothesis.

Problem 4: Small Sample Two-sided Hypothesis Test for a Single Population, based on the Chi-squared Statistic

Repeat the above problem, assuming that the population is normally distributed so that the Chi-squared statistic and CDF can be used. In this case, use the following small sample:

$$[x_1, x_2, x_3, x_4, x_5, x_6, x_7] = [0.81 \quad 1.72 \quad 0.41 \quad 3.18 \quad 0.86 \; 1.11 \quad 2.06]$$