

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 (Solutions provided at end of each problem)
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:

$$H_0: a = E[\text{annual rainfall}] = 100 \text{ cm}$$

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 H_0 at the 5% significance level (i.e. for $\alpha=0.5$). Make a “large sample” assumption and use a two-sided test. Also report the p value for the test.

Problem 1 Solution

```
% Problem Set 7 -- Question 1
clear all
close all
load precip.txt
alpha=.05
rain=precip(:,2);
N=[10 40 90];
z1=norminv(alpha/2,0,1);
zu=norminv(1-alpha/2,0,1);
for i=1:3
    mu(i)=mean(rain(1:N(i)));
    sigma(i)=std(rain(1:N(i)));
    z(i)=(mu(i)-100)/(sigma(i)/sqrt(N(i)));
```

```

    pass(i)=(z(i)>=zl&z(i)<=zu);
    if z(i)<=0
        p(i)=2*normcdf(z(i),0,1);
    else
        p(i)=2*(1-normcdf(z(i),0,1));
    end
end
pass
p

```

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 2 Solution

```

% Problem Set 7 -- Question 2
clear all
close all
load mwra.txt
% edited the data file so that all outer values were first,
then inner
% values
outer=log(mwra(1:195,2)+1);
inner=log(mwra(196:373)+1);
mo=mean(outer);
mi=mean(inner);
vo=var(outer);
vi=var(inner);
No=length(outer);
Ni=length(inner);
z=(mo-mi)/sqrt(vo/length(outer)+vi/length(inner));
if z<=0
    p=2*normcdf(z,0,1)
else
    p=2*(1-normcdf(z,0,1))

```

end

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. $H_0: SD(x) - E(x) = 0$). 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 3 Solution

```
% Problem Set 7 -- Question 3
clear all
close all
nrep=10000;
data=[1.76 0.51 3.08 3.62 0.86];
% get parameter values for lognormal distribution
mu=mean(log(data));
sigma=std(log(data));
% generate random samples based on the data:
sim=lognrnd(mu,sigma,nrep,5);
% ahat = SD(x)-E(x)
ahat=std(sim,0,2)-mean(sim,2);
stdahat=std(ahat);
zhats=(ahat-0)/stdahat;
figure
cdfplot(zhats)
xlabel('z')
ylabel('CDF of z')
zhatdat=(std(data)-mean(data)-0)/stdahat
```

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:

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

Problem 4 Solution:

```
% Problem Set 7 -- Problem 4
data=[0.81 1.72 .41 3.18 .86 1.11 2.06];
% If we take:  $sx^2/mux^2 * E[x]^2/sigma^2$ , we can say that
%  $mx$  approximates  $E[x]$ 
%  $H_0$  says that  $E[x]^2/sigma^2=1$ 
% If  $H_0$  is true  $sx^2/mux^2$  must be Chi-Squared distributed:
% Calculate  $sx^2/mux^2$ 
varx = var(data);
mux = mean(data);
z=varx/mux^2*(length(data)-1)
% from the chart in the text, find P corresponding to  $z=2.5649$ 
% this is much lower than 10.64, which corresponds to  $p/2=0.1$ 
% so P is greater than 0.2.
% This means that we cannot reject the hypothesis that
%  $SD(x)-E(x) = 0$  or  $SD(x)/E(x) = 1$ 
```