Student name:___________________________

CE 93 -- Engineering Data Analysis
First Midterm Examination
Thursday, March 3, 2005
9:10-10:00 AM, Sibley Auditorium

Work on all three problems. Write clearly and state any assumptions you make. The exam is closed books and notes. However, you can use one sheet of your own notes.

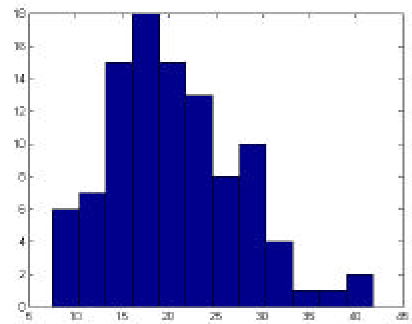
The problems have the following weights:

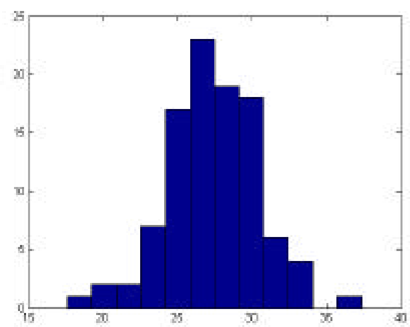
Problem 1 (35 points) _________________________
Problem 2 (35 points) _________________________
Problem 3 (30 points) _________________________

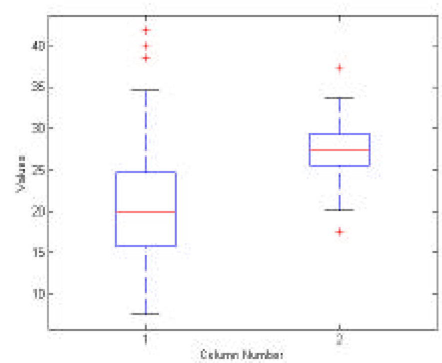
Exam grade (100 points) _______________________

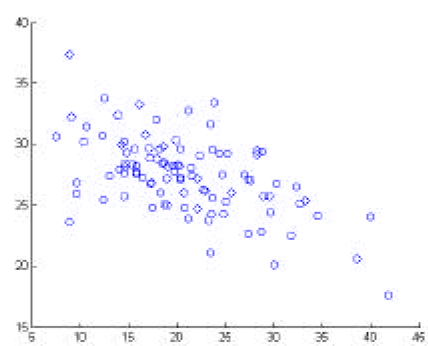
Problem 1. (7´5= 35 points)

Carefully examine the Matlab® session shown below. The resulting figures are shown in sequence on the right. Then respond to the itemized questions listed on the next page.









\>> whos

Name	Size	Bytes	Class
x	100x1	800	double array
y	100x1	800	double array

Grand total is 200 elements using 1600 bytes

\>> hist(x,12)

\>> hist(y,12)

\>> boxplot([x y])

\>> scatter(x,y)

\>> mean(x)

ans =

20.7627

\>> moment(x,2)

ans =

49.1303

\>> moment(x,3)

ans =

201.9081

\>> prctile(x,[10 50 90])

ans =

12.4305 19.9746 29.8058

Student name:__________________________________

Questions for Problem 1 (each question carries 5 points):

a) Approximately estimate the range and *IQR* of data **x**.

From histogram, range of data $\mathbf{x} \cong 42-8=34$

From box plot, *IQR* of data $\mathbf{x} \cong 25-16=9$

b) Describe the nature of skewness of the distributions of **x** and **y**.

Observing the histograms and the box plots, it appears that:
The distribution of **x** is moderately skewed to the right (positive skewness coefficient).
The distribution of **y** is slightly skewed to the left (negative skewness coefficient, but near zero).

c) How many outliers does each data set contain?

x data contains 3 outliers in the far right tail.

y data contains 2 outliers, one in each tail

d) What do you guess is the correlation coefficient between **x** and **y**?

Moderately negative, say around -0.50.

e) Determine an <u>unbiased</u> estimate of the standard deviation of data **x**.

Unbiased variance = $\frac{100}{99} \times 49.13 = 49.63$

Unbiased standard deviation = $\sqrt{49.63}$ = 7.046

f) Determine the median of data **x**.

$x_{0.50} = 19.975$

g) What fraction of data **x** lies in the interval 12.4 to 29.8?

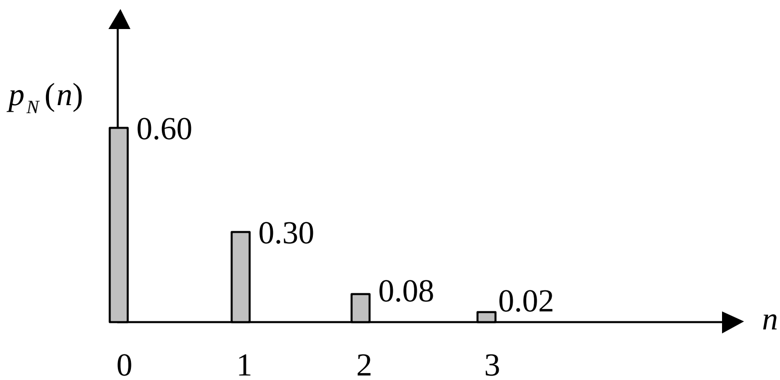
These two numbers are the 10% and 90% percentiles. Therefore, 80% of the data falls within this interval.

Problem 2. (20+15= 35 points)

The number of earthquakes to occur during the life of a building, denoted N, is expected to be 0, 1, 2 or 3 with the probabilities indicated in the PMF shown below. At each event, there is a probability 0.05 that the building will be damaged. Assume the building is immediately repaired after each damaging event, and that the states of the building after successive earthquake events are statistically independent. Determine:

a) The probability that the building will be damaged due to earthquakes during its lifetime.

b) If the building is known to have sustained earthquake damage during its lifetime, what is the probability that more than one earthquake occurred during this period?



a) Define D = damage due to earthquake. By total probability theorem,

$$P(\overline{D}) = \sum_{n=0}^{3} P(\overline{D} \mid N = n)P(N = n)$$

$$P(\overline{D} \mid N = 0) = 1, \quad P(\overline{D} \mid N = 1) = 1 - 0.05 = 0.95$$

$$P(\overline{D} \mid N = 2) = (1 - 0.05)^2 = 0.9025, \quad P(\overline{D} \mid N = 3) = (1 - 0.05)^3 = 0.8574$$

$$P(\overline{D}) = 1 \times 0.60 + 0.95 \times 0.30 + 0.9025 \times 0.08 + 0.8574 \times 0.02 = 0.974$$

$$P(D) = 1 - P(\overline{D}) = 1 - 0.974 = 0.0257$$

Alternative solution:

$$P(D) = \sum_{n=0}^{3} P(D \mid N = n)P(N = n)$$

$$P(D \mid N = 0) = 0, \quad P(D \mid N = 1) = 0.05$$

$$P(D \mid N = 2) = 1 - P(\overline{D} \mid N = 2) = 0.0975, \quad P(D \mid N = 3) = 1 - P(\overline{D} \mid N = 3) = 0.1426$$

$$P(D) = 0 \times 0.60 + 0.05 \times 0.30 + 0.0975 \times 0.08 + 0.1426 \times 0.02 = 0.0257$$

b) If damage occurred, then there must have been 1, 2 or 3 earthquakes.

$$P(N > 1 \mid D) = 1 - P(N = 1 \mid D) = 1 - \frac{P(D \mid N = 1)}{P(D)} P(N = 1) = 1 - \frac{0.05}{0.0257} 0.30 = 0.416$$

Student name:___________________________________

<u>Problem 3. (7+10+8+5= 30 points)</u>

The concentration of a pollutant in the effluent of a chemical plant is a random variable X, bounded between 0 and 2 units, and has the probability density function

$$f_X(x) = ax^2(2-x) \qquad 0 < x < 2$$

a) Determine the value of parameter a.

b) Determine the mean and coefficient of variation of X.

c) Determine the CDF of X.

d) What is the probability that the concentration is greater than 1.5?

a) The area underneath the PDF must equal 1. Therefore,

$$\int_0^2 ax^2(2-x)dx = a\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 = a\left[\frac{16}{3} - \frac{16}{4}\right] = \frac{4a}{3} = 1$$

$$a = \frac{3}{4}$$

b) $$\mathrm{E}[X] = \int_0^2 \frac{3}{4}x^3(2-x)dx = \frac{3}{4}\left[\frac{2x^4}{4} - \frac{x^5}{5}\right]_0^2 = \frac{3}{4}\left[\frac{32}{4} - \frac{32}{5}\right] = \frac{6}{5} \quad \underline{\text{mean}}$$

$$\mathrm{E}[X^2] = \int_0^2 \frac{3}{4}x^4(2-x)dx = \frac{3}{4}\left[\frac{2x^5}{5} - \frac{x^6}{6}\right]_0^2 = \frac{3}{4}\left[\frac{64}{5} - \frac{64}{6}\right] = \frac{8}{5}$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - E^2[X] = \frac{8}{5} - \frac{36}{25} = \frac{4}{25}$$

$$\sigma = \sqrt{4/25} = 2/5 = 0.4$$

$$\text{c.o.v.} = \frac{2/5}{6/5} = \frac{1}{3} \quad \underline{\text{coefficient of variation}}$$

c)

$$F_X(x) = \int_0^x \frac{3}{4}x^2(2-x)dx = \frac{3}{4}\left(\frac{2x^3}{3} - \frac{x^4}{4}\right) \qquad 0 < x < 2$$

$$= 1 \qquad 2 \le x$$

d) $$P(X > 1.5) = 1 - F_X(1.5) = 1 - \frac{3}{4}\left(\frac{2 \times 1.5^3}{3} - \frac{1.5^4}{4}\right) = 0.262$$