MIDDLE EAST TECHNICAL UNIVERSITY
FACULTY OF ENGINEERING

CE 204 UNCERTAINTY and DATA ANALYSIS
Spring Semester 2012-2013
Homework 4- Date Due: May 27th, 2013 Monday till 17.00

IMPORTANT NOTICE:

- *You are allowed to collaborate with other students (or ask questions to your assistants/ instructors) on homework provided that you stay away from plagiarizing (according to dictionaries "to plagiarize" means to steal and pass off ideas and/or words/ solutions of another as one's own without citing the source). That is, collaboration is accepted if you write and give your own solutions. If you are caught on plagiarizing or cheating by handing in "too similar" homework, you will be graded by zero on this homework.*

1. Let X be a random variable with the following distribution
   $$f_X(x) = (\alpha + 1) x^{\alpha} ,   0 < x < 1$$
   $$f_X(x) = 0 ,   \text{otherwise.}$$

   Find the maximum likelihood function, L and the maximum likelihood estimator of the parameter α , based on a random sample of size n.

2. A survey in a representative area of a certain city showed that the standard deviation of the number of pipes with leaking water is 8 pipes. If we wish to determine the mean number of pipes with leaking to within ± 2 pipes with a 95% confidence what should be the sample size of our survey?

3. The mean and standard deviation of the compressive strengths of 24 concrete specimens are 50 and 5 N/mm^2 , respectively. Find 99% confidence limits for the true   mean and true standard deviation of the compressive strengths.

4. In a certain course two different sections 1 and 2 with class sizes $n_1 = 32$ and $n_2 = 40$ are found to have averages as $\bar{x}_1 = 70.2$ and $\bar{x}_2 = 76.5$ with assumed known standard deviations as $\sigma_1 = 7.4$ and $\sigma_2 = 14.6$, respectively. Find the 98% confidence interval on the difference in mean performance $\mu_1 - \mu_2$ of the two sections. Can we say that the average performance of the two sections are not differing significantly?

5. The ultimate strains (in 10^{-3}) of 10 steel reinforcing bars were measured , yielding the results shown below:

   19.4   16.0   16.6   17.3   18.4   17.9   17.8   18.8   19.8   19.9

   a) Draw a histogram and a cumulative distribution diagram for the strains of the above 10 bars using five class intervals from 16.0 x10^{-3} to 20 x10^{-3} .
   b) Perform chi –square (χ^2) test for the goodness of fit of uniform distribution in the interval from 16.0 x10^{-3} to 20 x10^{-3} for the ultimate strains of the bars   at a significance level of 1 % .

6. The final grades of the 140 students who took the CE 204 exam in the last Spring term is as follows:

Grade Intervals	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	3	2	7	4	17	27	33	28	14	5

    a) Draw the histogram and cumulative frequency diagrams for the grade distribution.
    b) Compute the mean, median, variance and standard deviation of the grades.
    c) Estimate the true mean and standard deviation of the grades (i.e. for all semesters) by using suitable point estimators.
    d) Compute 99 % confidence interval of the true mean of the grades.
    e) Compute 99% confidence interval of the population standard deviation of the grades.
    f) Find the probability that the sample mean of the grades for the last spring term will differ from its population mean (for all semesters) by less than 5.
    g) Set up 99 % upper confidence limit for the population mean of the grades.
    h) Set up 95 % lower confidence limit for the true standard deviation.

7. Consider the following data where x is the applied tensile force in kN and y is the corresponding elongation of aluminum specimens in mm.

X (kN)	5	10	15	20	25	30
Y (mm)	0.2	0.5	0.6	1.0	1.2	1.4

    a) Draw a scatter diagram for y and find the least squares regression line for estimating y in terms of x (Note : $E(Y/X) = y = a + b x$ (or $y = \alpha + \beta x$; and $\hat{y} = \hat{\alpha} + \hat{\beta}x$)
    b) Compute 95% confidence interval for b (or β, slope of the regression line).
    c) Compute 95% confidence interval for a (or α, y -intercept of the regression line).
    d) Compute the 95 % prediction interval for Y if X is observed to be 23 kN.
    e) Compute the coefficient of determination and correlation coefficient of X and Y,. State what they measure in regression analysis? Does a simple linear regression model seem appropriate for the given test results.
    f) Under zero load the elongation of the specimen must be zero. In this case what should the regression line (passing through the origin) and elastic modulus (slope of the regression line) of the specimen be?

8) The world populations (WP) and the average annual world population changes (WPC) between the years 2000 and 2049 are summarized below table as taken from the Drexel University Database.

Year	Population (Billions)	Avg Annual Pop Change (Millions)
2000	6.091	78.770
2001	6.170	78.373
2002	6.248	77.936
2003	6.326	77.489
2004	6.403	77.020
2005	6.481	76.720
2006	6.557	76.614
2007	6.634	76.492
2008	6.710	76.349
2009	6.787	76.121
2010	6.863	75.959
2011	6.939	75.824
2012	7.015	75.461
2013	7.090	74.928
2014	7.165	74.269
2015	7.239	73.674
2016	7.313	73.168
2017	7.386	72.579
2018	7.459	71.926
2019	7.531	71.200
2020	7.602	70.532
2021	7.672	69.921
2022	7.742	69.220
2023	7.811	68.468
2024	7.880	67.675
2025	7.948	66.988
2026	8.015	66.410
2027	8.081	65.773
2028	8.147	65.125
2029	8.212	64.477
2030	8.276	63.840
2031	8.340	63.178
2032	8.403	62.398
2033	8.466	61.573
2034	8.527	60.749
2035	8.588	59.898
2036	8.648	58.982
2037	8.707	57.921
2038	8.765	56.800
2039	8.822	55.715
2040	8.877	54.632

2041	8.932	53.515
2042	8.986	52.271
2043	9.038	50.979
2044	9.089	49.754
2045	9.139	48.552
2046	9.187	47.329
2047	9.234	45.982
2048	9.280	44.562
2049	9.325	43.218

a) Draw histograms and cumulative frequency diagrams for *the world populations (WP)* and *the average annual world population changes (WPC).*

b) Find the mean, median, mode, standard deviation and coefficient of variation for the variables WP and WPC.

c) Give your comments on the histograms and numerical descriptors that you have found in parts (a) and (b).

d) Are there any outliers in the data sets for WP and WPC? If there are, what would you suggest for a more robust statistical analysis?

e) Draw a scatter diagram for WP and WPC. (WP versus WPC) Are WP and WPC correlated? Find the correlation coefficient. Comment on your results.

Note : You may use any statistical software package for solving this problem. (e.g. Excel, Minitab, Matlab,…).