CEG2002: Statistics for Civil Engineers
Exercises 2

(You don't need to submit your solutions. Some questions will be discussed in tutorial. The solutions of all questions will be published in the website later)

1. Jars of water are taken from the public supply in the Gothic Quarter of Barcelona and analysed for their lead content. A sample of 10 such jars gave the following results (in μg per litre):

   41 43 49 42 45 55 40 47 52 39

   Test the null hypothesis that the average lead content in this part of the city is 50 μg/litre, clearly stating any assumptions implicit in your test.

2. A similar survey was undertaken in the newer "Eixample" part of Barcelona; a sample of 15 jars of water gave the following results (in μg per litre):

   40 42 49 31 43 52 50 45 44 39 33 34 36 35 42

   Test the null hypothesis that there is no difference between the average lead content of water in these two areas of Barcelona, making sure you state any assumptions implicit in your tests.

3. Construct a 95% confidence interval for each of the samples in questions 1 and 2. How can these intervals be interpreted? How can they be used to support the outcome of the test you performed in question 2?

4. In a study of automobile collision insurance costs, a random sample of 80 body repair costs for a particular kind of damage gave

   $\bar{X}$ = £542 and S = £70.

   Construct a 90%, a 95% and a 99% confidence interval for the average cost of collisions of this kind. Obtain the corresponding intervals if the population standard deviation was known to be £70. How do your two sets of intervals compare. Why?

5. While performing a certain tasks under simulated weightlessness, the pulse rate of 12 American trainee astronauts increased on average by 26.4 beats per minute. The pulse rate of 15 such Russian astronauts increased on average by 23.5 beats per minute. From larger studies, the population standard deviation increase for American and Russian trainee astronauts has been shown to be 4.28 and 3.30 respectively. Test the null hypothesis that there's no difference in the average increased pulse rate between the two sets of astronauts.

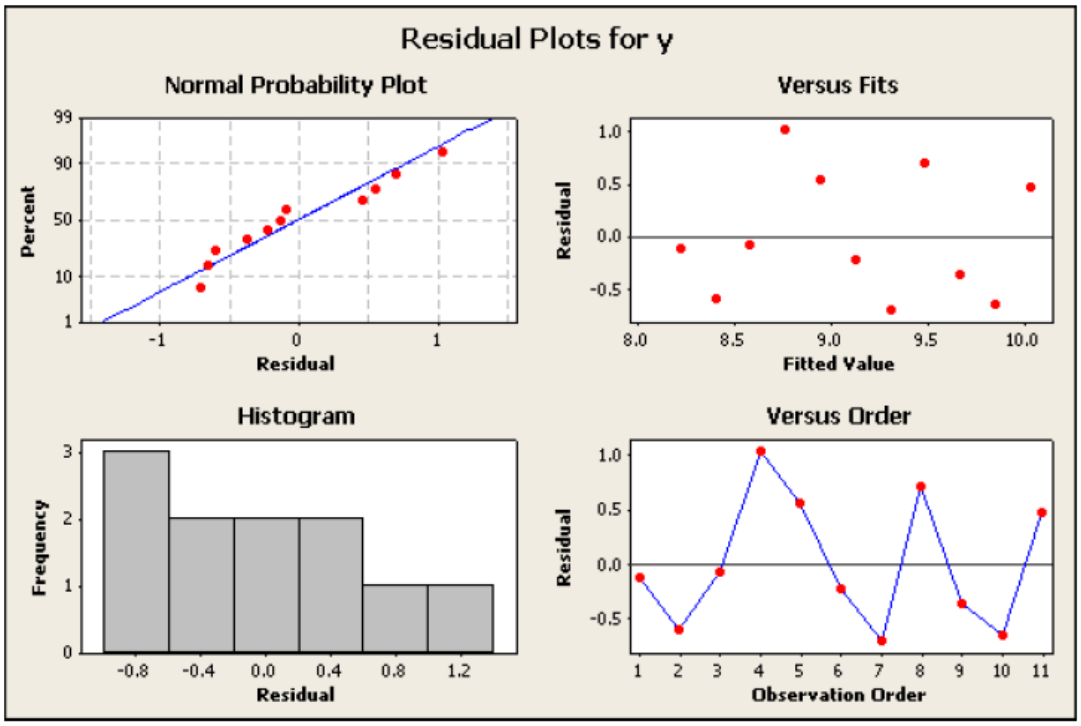
6. A study was made on the amount of converted sugar (y) in a certain process at various temperatures (x). The data were coded and recorded as follows.

   x: 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0
   y: 8.1 7.8 8.5 9.8 9.5 8.9 8.6 10.2 9.3 9.2 10.5

   a) Produce a scatter plot of y against x on graph paper, and comment.
   b) Perform a linear regression analysis by using a simple linear regression $y = \alpha + \beta x + \varepsilon$, and obtain the linear regression equation. Superimpose this line on your diagram in part (a). You may use the following summaries.

$$\sum x = 16.5, \sum y = 100.4,$$
$$\sum x^2 = 25.85, \sum y^2 = 923.58, \sum xy = 152.59$$

c) Using the output from **Minitab** below, would you say the assumptions underlying your regression analysis are valid? Explain and justify.
d) Calculate the predicted amount of converted sugar at a temperature of 1.15.



7. A construction engineer needs to select a steel pile length when the depth to rock is uncertain. The available actions are driving a 40- or 50-ft pile, and the possible states of nature are a 40- or 50-ft depth to bedrock. The consequences of any action-state pair can be given in a payoff table:

	a_1 (drive 40-ft pile)	a_2 (drive 50-ft pile)
θ_1, depth 40 ft	correct decision, no loss	cut off 10-ft piece of pile, 100-unit loss
θ_2, depth 50 ft	splice and weld 10-ft pile, 400-unit loss	correct decision, no loss

In this example, we can define either a loss function $l(a_i, \theta_j)$ or a utility function $u(a_i, \theta_j)$:

Loss function	Utility function
$l(a_1, \theta_1) = 0$	$u(a_1, \theta_1) = 0$
$l(a_1, \theta_2) = 400$	$u(a_1, \theta_2) = -400$
$l(a_2, \theta_1) = 100$	$u(a_2, \theta_1) = -100$
$l(a_2, \theta_2) = 0$	$u(a_2, \theta_2) = 0$

a) How to draw a decision tree?
b) Which decision you would make based on the minimax criterion?
c) If we know that the probability of a 40-ft depth to bedrock is 5/6, which decision you would make based on the expected values of loss function (or utility function)?