CEG2002: Statistics for Civil Engineers
Exercises 1

Exercises1

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

1. The following data are the average monthly flows (in cubic metres per second) through the spillway of the Tombigbee Dam on the Alabama Gulf Coast over the last 18 months:

35	36	23	60	54	40	42	42	37
59	61	40	36	24	46	43	42	57

    a) Produce a stem-and-leaf plot for these data, and comment.
    b) Calculate the mean and standard deviation for these data.
    c) Find the modal flow through the Tombigbee Dam.

2. The Alabama Dam Authority wants to compare flows through the Tombigbee Dam with those through the Arklatex Dam. Flows through the spillway of the Arklatex for January – September 2006 are shown below:

39	60	64	48	26	81	65	58	42

    a) Calculate the mean and standard deviation for flows through the Arklatex Dam, and compare these with the numerical summaries for flows through the Tombigbee Dam found in question 1 part (a). Is it possible to tell which Dam has the highest average flow-rate?
    b) The average flow through the Arklatex for October 2006 was 175 cubic metres per second. If we include this latest observation, why should we be cautious about using the mean and standard deviation as measures of location and spread?
    c) Find the median and inter-quartile range for flows through the Arklatex, if you include the new observation.
    d) Construct a box-and-whisker plot for flows through the Arklatex Dam using the summaries obtained in part (c). Comment on your plot.

3. The following table shows the number of days on which various rainfall measurements were observed (in mm) at a location in the South West of England:

Rainfall (mm)	Frequency
$0 \leq x < 10$	4
$10 \leq x < 20$	15
$20 \leq x < 30$	22
$30 \leq x < 40$	25
$40 \leq x < 50$	20
$50 \leq x < 60$	16
$60 \leq x < 70$	8
$70 \leq x < 80$	0
$80 \leq x < 90$	0
$90 \leq x < 100$	1

    a) Estimate the mean and standard deviation for these data.
    b) Why are your answers in part (a) *estimates*?
    c) Find the modal class for these data.
    d) Construct a percentage relative frequency histogram for these data, and comment.

4. Suppose that 95% of bathing beaches pass E.U. hygiene regulations. In a random sample of 12 beaches, what is the probability that more than than 9 beaches pass the regulations?

5. A point D is chosen on the line AB, whose midpoint is C and whose length is b. If X, the distance from D to A, is a random variable having the uniform density $U(0, b)$, what is the probability that AD, BD and AC will form a triangle?

6. Suppose Z follows a standard Normal distribution, i.e. $Z \sim N(0, 1)$. Find:

   a) $P(Z < 1.75)$
   b) $P(Z > 2.3)$
   c) $P(-2 < Z < 2)$
   d) $P(-3 < Z < 3)$

7. Hourly maximum wind speeds observed at High Bradfield, West Yorkshire, are assumed to be Normally distributed with mean 40 mph and standard deviation 12 mph. Find the probability that, in any one hour period, the maximum observed wind speed at High Bradfield will:

   a) be less than 30 mph;
   b) exceed 55 mph, and
   c) lie between 15 mph and 45 mph.
   d) Find the wind speed which will be exceeded with probability 0.1.

8. Jars of river water collected along the Tyne may contain bacteria. Previous studies have shown that these bacteria occur at a rate of 5 per litre.

   a) Suggest an appropriate probability model for the number of bacteria per litre of river water.
   b) Using this model, find the probability that a randomly selected litre jar contains exactly 4 such bacteria.
   c) Using this model, find the probability that a randomly selected litre jar contains at least 1 bacteria.
   d) Find the probability that, in a five litre jar, exactly 27 bacteria are found.

Project A.1

You need to submit your solutions of the questions in this part.
All questions must be handwritten, and any graphs must be produced by hand. These questions are part of a statistical project. More questions and the deadline to submit the project will be given later.

1. The following data represent the running times (minutes) of films produced by a motion-picture company.

   $$81\ 165\ 97\ 92\ 87\ 114$$

   a) Produce a box-and-whisker plot for these data, and comment.

   b) Calculate the mean and standard deviation for the running times.

   c) Construct a 95% confidence interval for the mean running time.

2. Morag has purchased a pack of five fireworks. A manufacturer's warning on the box suggests that there is a 10% chance of a firework not lighting.

   a) What probability distribution can we use to model the number of fireworks that fail to light?

   b) What is the probability of all the fireworks lighting?

   c) Find the probability of exactly one firework failing to light.

3. A certain kind of sheet metal has, on the average, five defects per 10 square feet. If we assume a Poisson distribution, what is the probability that a 15-square-foot sheet of the metal will have at least six defects?

4. The actual diameter (in millimetres) of a rivet with nominal diameter 10 mm is a $N(10, 0.01)$ random variable. To be usable, a rivet must have a diameter in the range 9.8 to 10.2 mm. What proportion of rivets is usable?