Periodic check

P.-L. Bourbonnais and N. Saunier

October 22, 2013

Note the scale (the total score is out of 20) and the indicative time to devote to each exercise. Pay special attention to the writing and definition of the notations you use.

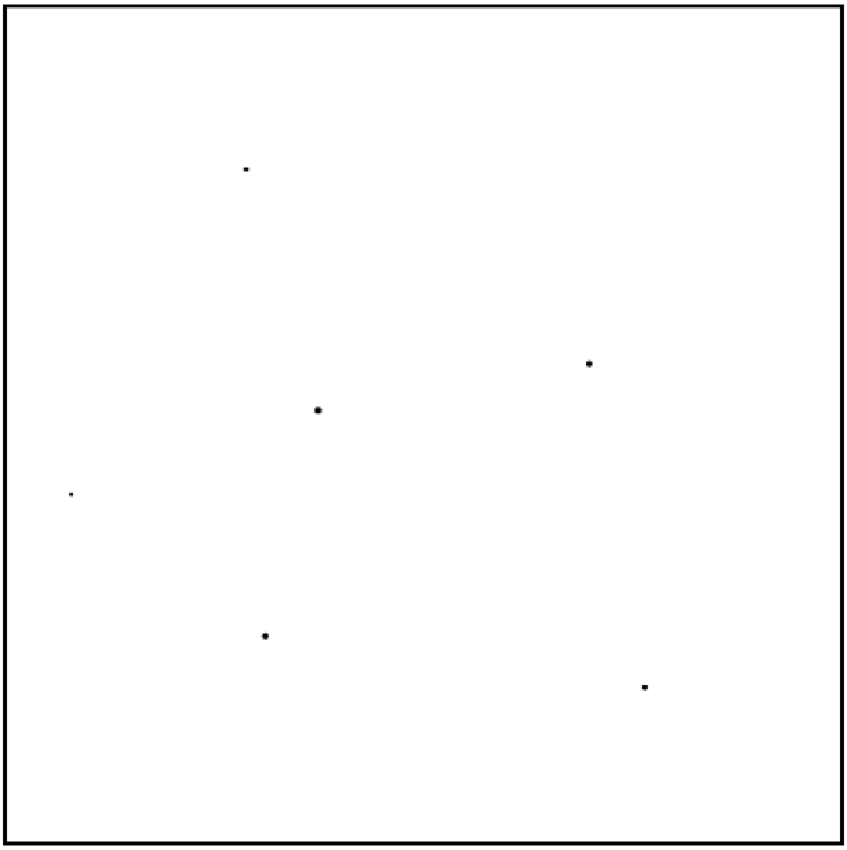
Statistical tables are included at the end of the statement.

Exercise 1 (survey methods) 20 min (/ 3)

Identify three different survey methods (survey techniques): for each, specify 2 advantages and 2 disadvantages.

Exercise 2 (GIS and spatial analysis) 35 min (/ 4)

1. Give an example of a centrality measure in spatial analysis and explain what this can be used for in a transport context. (1 point)
2. Give an example of a dispersion measurement in spatial analysis and explain what this can be used for in a transport context. (1 point)
3. What are the overall Moran I Index and the Global Geary C Index used for? In what context could they be used in transport? (1 point)
4. What is spatial interpolation and why do we use it? (1 point)
5. Explain in which case you would use MTM coordinates, UTM coordinates and WGS84 coordinates (latitude / longitude)? (1 point)
6. Is it possible to obtain a projection of the earth on a plane which preserves distances, shapes, angles and areas? Why? (1 point)
7. Draw what the Thiessen polygons would look like (as the crow flies) around the points below (don't measure exactly, just sketch). (2 points)



Exercise 3 (data models) 45 min (/ 6)
We are interested in creating an information system for a shipping company. For this, it is necessary to model the different objects and concepts necessary for its operation. These entities are as follows: port, route, ship, employee, type of employee, home port (single reference port for a ship), container (“ container ”).

1. Provide a data model in the form of an Entity / Association diagram involving all the entities listed above. Add attributes and associations between entities with their minimum and maximum cardinalities. (3 points)

2. Translate the Entity / Association schema into a relational schema. Clearly indicate primary and external keys, and suggest types for attributes. (3 points)

Solution
There are no n-m associations in this E / A diagram. A solution was presented during the course.

Exercise 4 (statistics) 50 min (/ 7)
The number of accidents on a road is counted in the following table for 15 days (period 1):

Day	Number of accidents
1	1
2	1
3	1
4	0
5	2
6	0
7	0
8	0
9	2
10	0
11	1
12	1
13	1
14	1
15	1

1. Write the algorithm for calculating the median of a set of n real numbers x_i. (1 point)
2. Calculate the average and the median of the number of accidents per day. (1 point)
3. Calculate a 95 % confidence interval for the average number of accidents per day. (1 point)
4. Calculate the number of days of observation necessary to obtain the average number of accidents per day with an accuracy of 0.15 for a confidence level of 90 and 95 %. (1 point)
5. Draw the histogram of the distribution of the number of accidents per day (and not the time series of the number of accidents as a function of the day). (1 point)
6. To improve road safety, a policeman is placed visibly on the side of the road for 15 days. During this period 2, the average number of accidents per day is 0.45 and the empirical standard deviation has not changed (it is assumed that the variances are the same for periods 1 and 2 and that the number of accidents per day follows a normal law). Determine if the number of accidents has decreased with a risk of error of the first kind of 5 %. (2 points)

Solution

1. Here is an algorithm (assuming an existing $sort$ sorting function on real numbers, current view):

   input: n real numbers x_i
   output: the median of the n real numbers x_i

   start
   $sorted\_list = sort(x_i)$
   if n even

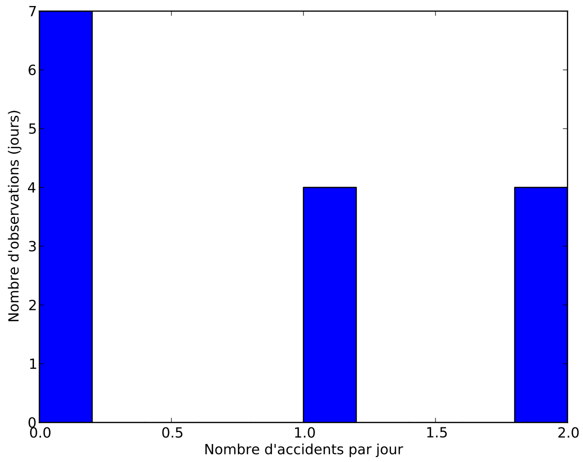
return the element in position $n/2$ of $sorted\_list$
otherwise
return the element in position $(n-1)/2$ of $sorted\_list$
end

2. The mean and median of the number of accidents per day are 0.80 and 1 accidents per day respectively.

3. The expression $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ follows Student's law with 14 degrees of freedom, and the probability that such a variable is respectively in the interval $[-2.145, +2.145]$ and $[-1.761, +1.761]$ is 95 % and 90 %. The corrected standard deviation s is 0.68 and the confidence interval of the average number of accidents per day is therefore respectively $0.8 \pm 2.14\frac{0.68}{\sqrt{15}} = [0.42, 1.18]$ and $[0.49, 1.11]$ for confidence levels of 95 and 90 %.

4. The number of observations required is respectively $n = 1.64^2\frac{0.68^2}{0.15^2} = 55$ and $n = 1.96^2\frac{0.68^2}{0.15^2} = 79$ for confidence levels of 90 and 95 %.

5. Histogram of the distribution of the number of accidents per day:



6. We test the hypothesis H_0: the average number of accidents has not changed against H_1: the number of accidents has decreased. The test statistic is $\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = (0.8 - 0.45)/(0.68\sqrt{2/15}) = 1.41$ ($n_1 = n_2 = 15$, $s_1 = s_2 = 0.68$). The statistic follows Student's law at $n = 15 + 15 - 2$ degrees of freedom. The cut-off value of the distribution for a risk of the first kind of 0.05 is 1.701 (i.e. the probability that a variable according to Student's law with 28 degrees of freedom is greater than or equal to 1.701 is 0.05). We cannot therefore reject H_0, the number of accidents does not seem to have been affected. We can find in the table that the value p (or risk of the first kind) for 1.41 is between 0.10 and 0.05, which could be accepted with a confidence level of 90 %.