Midterm Exam

P.-L. Bourbonnais, N. Saunier and L. Sioui

October 25, 2012

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.

All documentation is permitted. Statistical tables are included at the end of the statement.

Exercise 1 (survey methods) 25 min (/ 4)

The Laurentides Intermunicipal Transport Council (CIT) is the transport company serving a suburban area north of Montreal. You wish to collect information concerning the profile of users of bus line 23, as well as their profile of use of this bus line. The route of the bus line connects the Sainte-Thérèse train station and the municipality of Sainte-Anne-des-Plaines, via the Collège Lionel-Groulx (CEGEP).

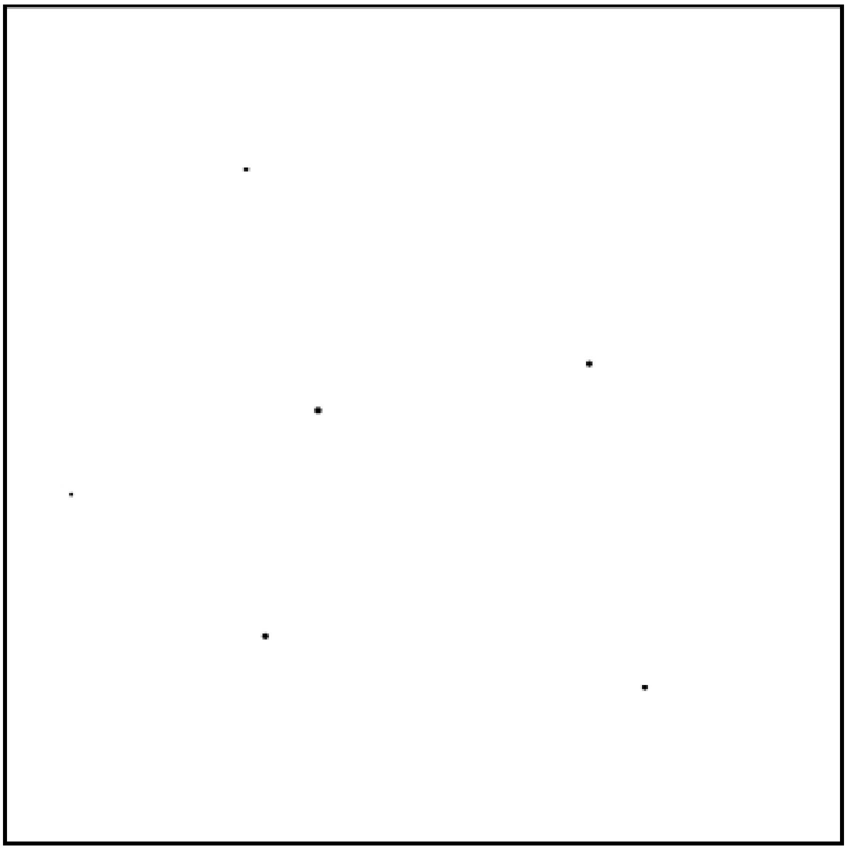
1. What is the reference population for this survey?
2. What data collection technique do you suggest? Explain how this collection works.
3. What will be the format of the questionnaire?
4. Identify an appropriate time frame for this survey as well as the time unit. To justify.
5. What is the minimum sample size you need to collect? You want a 95 % confidence level and you accept a 4 % margin of error. Based on smart card transaction data, you know that your benchmark population is 2,000 individuals, and you want to make sure you meet the 75 % proportion of students in the line customer base.

Solution

1. Users who use bus line 23
2. Interception, investigation aboard line 23 buses
3. Paper or iPad
4. A weekday in the fall (during the school and work calendar)
5. The minimum sample size is $n = (1.96^2 \times 0.75(1-0.75))/0.04^2 = 450$ (limited pop correction $n' = 450/(1 + 450/2000) = 367$ individuals)

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



Solution

Exercise 3 (data models) 45 min (/ 6)

We are interested in creating an information system for an airline. For this, it is necessary to model the different objects and concepts necessary for its operation. These entities are: airport, flight, plane, employee, type of employee, garage, passenger, ticket.

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

Solution

Exercise 4 (statistics) 45 min (/ 6)

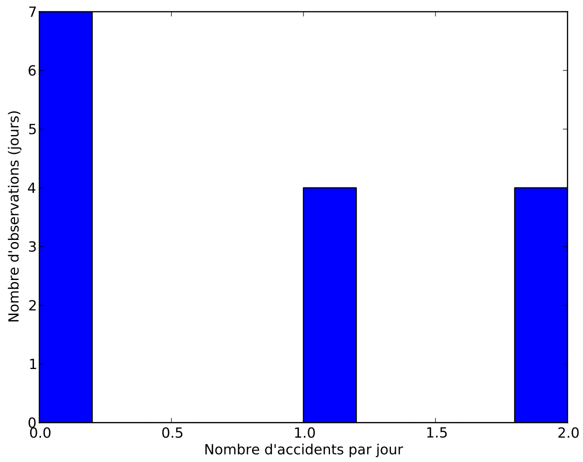
The number of accidents on a road is counted in the following table for 15 days (period 1):

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

1. Calculate the average number of accidents per day.
2. Calculate the 90 % and 95 % confidence intervals on the average number of accidents per day.
3. Calculate the number of days of observation necessary to obtain the average number of accidents per day with an accuracy of 0.2 for a confidence level of 90 and 95 %.
4. 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).
5. 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). Determine if the number of accidents has decreased with a risk of error of the first kind of 5 %.

Solution

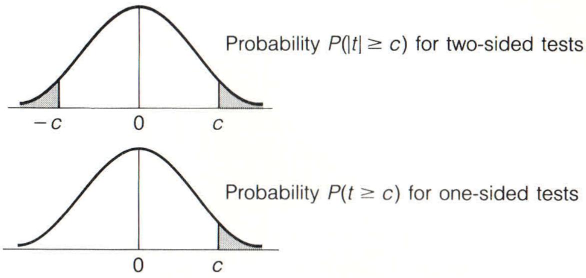
1. The average number of accidents per day is 0.80 accidents per day.
2. The expression $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ follows Student's law with 14 degrees of freedom, and the probability that such a variable is respectively in the interval $[-2.145, +2.145]$ and $[-1.761, +1.761]$ is 95 % and 90 %. The corrected standard deviation s is 0.86 and the confidence interval of the average number of accidents per day is therefore respectively $0.8 \pm 2.14\frac{0.86}{\sqrt{15}} = [0.32, 1.28]$ and $[0.41, 1.19]$ for confidence levels of 90 and 95 %.
3. We assume that the empirical standard deviation is close to the true standard deviation. The number of observations required is respectively $n = 1.64^2\frac{0.86^2}{0.2^2} = 49.7 \approx 50$ and $n = 1.96^2\frac{0.86^2}{0.2^2} = 71.03 \approx 72$ for confidence levels of 90 and 95 %.
4. Histogram of the distribution of the number of accidents per day:



5. 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 $\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = (0.8-0.45)/(0.86\sqrt{2/15}) = 1.11$ ($n_1 = n_2 = 15$, $s_1 = s_2 = 0.86$). 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 (ie 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 1.11 is between 0.25 and 0.1, which is far too high.

Table 3

Critical values for Student's *t* distribution

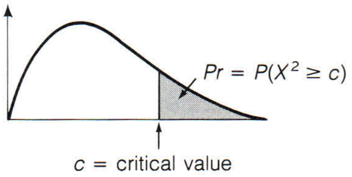


PROBABILITY

ν	.50	.20	.10	.05	.02	.01	.005	.002	.001	TWO-SIDED TESTS
	.25	.10	.05	.025	.01	.005	.0025	.001	.0005	ONE-SIDED TESTS
1	1.000	3.078	6.314	12.706	31.821	63.637	127.32	318.31	636.62	
2	.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598	
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924	
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	
5	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	
7	.711	1.415	1.895	2.365	2.998	3.499	4.020	4.785	5.408	
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	
10	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.537	
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	
13	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221	
14	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140	
15	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073	
16	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015	
17	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965	
18	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922	
19	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883	
20	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850	
21	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.257	3.189	
22	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792	
23	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767	
24	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745	
25	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725	
26	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707	
27	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690	
28	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674	
29	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659	
30	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646	
40	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551	
60	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460	

Table 2

Critical values for the chi-square distribution



	Pr							
ν	.500	.250	.100	.050	.025	.010	.005	.001
1	.455	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	1.386	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	2.366	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	3.357	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	4.351	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	5.348	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	6.346	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	7.344	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	8.343	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	9.342	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	27.79	29.82	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	20.34	24.93	29.62	33.67	35.48	38.93	41.40	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	26.34	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	69.33	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	79.33	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	89.33	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	99.33	109.1	118.5	124.3	129.6	135.8	140.2	149.4

Source: Abridged from Table 8 of *Biometrika Tables for Statisticians*, Vol. 1, edited by E. S. Pearson and H. O. Hartley (London: Cambridge University Press, 1962).

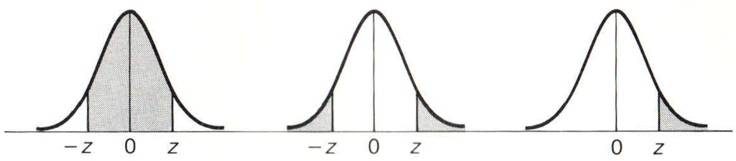


Table 1B

Critical values for the standard normal distribution

CONFIDENCE INTERVALS $P(\|Z\| \leq z)$	TWO-SIDED TESTS $P(\|Z\| \geq z)$	ONE-SIDED TESTS $P(Z \geq z)$	CRITICAL VALUE z
.10	.90	.45	.126
.20	.80	.40	.253
.30	.70	.35	.385
.40	.60	.30	.524
.50	.50	.25	.674
.60	.40	.20	.842
.70	.30	.15	1.036
.80	.20	.10	1.282
.90	.10	.05	1.645
.95	.05	.025	1.960
.98	.02	.01	2.326
.99	.01	.005	2.576
.995	.005	.0025	2.807
.999	.001	.0005	3.290
.9995	.0005	.00025	3.480
.9999	.0001	.00005	3.890
.99999	.00001	.000005	4.420
.999999	.0000001	.0000005	4.900

Source: D. B. Owen and D. T. Monk, *Tables of the Normal Probability Integral*, Sandia Corporation Technical Memo 64-57-51 (March 1957).