## B. Exercises

You should be prepared to present the exercises and at least an attempt to solve them. A complete solution is not required but you should at least have read, understood and attempted to solve them before going to the lecture.

## For February 20

17. A company delivering letters has 5 000 mailman tours over the whole country. The number of letterboxes is known on every tour (x<sub>i</sub>; i=1, ..., 5000). In total, there are 5 000 000 letterboxes in the whole country ( $\Sigma_{i=1}^{5000}$  (x<sub>i</sub>) = 5 000 000). One hundred mailmen are selected with probabilities proportional the number of boxes on their tours. (Pareto  $\pi$ ps). Every mailman has to put a questionnaire in every tenth letterbox, where the person who usually empties the box is asked questions about the postal service and letter delivery. The questionnaire should than be posted to the mail company.

a) What is this sampling design called?

b) 10 000 questionnaires, with many questions, were handed out, We will only consider the question: "What is your opinion on the proposal to move all letter-boxes in apartment houses to the entrances? Good, Bad, No opinion. (Mark the answer that best corresponds to your opinion)". For every mailman tour the number of the three possible answers are counted (Data: ( $y_{iG}$ ,  $y_{iB}$ ,  $y_{iN}$ ); such that  $y_{iG}+y_{iB}+y_{iN} = z_i$  for i = 1, ..., 5000) (partial non response is counted as no opinion). (The number of questionnaires handed out is also known,  $z_i$  if there had been non response the  $z_i$  would have been equal to  $x_i/10$ ).

The company wants to estimate the proportion of households who consider this proposal as bad (out of all households in Sweden). Give and explain all formulas that you think they should use for estimation and their standard errors. State also all assumptions that you make)

c) Have you any comments on the sampling plan. Is it good in this case (for this question)? What do you think would happen if all answers were weighted equally?

18. A journalist wants to investigate the existence of "gender equality plans" in the companies owned by Swedish local authorities. (According to the Swedish law they need to have one). She selected 10 municipalities and two counties at random For the selected municipalities she found out the number of companies. She wrote to all those and asked for a copy of the plan. If she had not received a reply within a week, she called. When she published her article after a month, her data looked like this (they are not real)

	# companies	# received	# without plan	# no response	Sum no received plan
<u>Municipality</u>	•••••••••••••	p	P	100pono e	10001/00 Pian
А	12	7	4	1	5
В	7	2	0	5	5
С	4	4	0	0	0
D	17	2	9	6	15
Е	6	5	1	0	1
F	12	11	0	1	1
G	3	3	0	0	0
Ι	9	13	7	5	12
Н	25	3	3	3	6
J	12	4	2	6	8
Subtotal	107	54	26	27	53
Sum of squares	1537	422	160	133	521
County					
Ö	6	6	0	0	0
Ä	9	5	2	2	4
Subtotal	15	11	2	2	4
Sum of Squares	117	61	4	4	16

a) Assume that the sample of municipalities is an SRS from Sweden's 290! Estimate the percentage of municipal companies which have been able to present a plan after a month!

b) Specify the uncertainty! (You may use that the product sum between the first and last column is 818)

c) Same questions for the county municipal companies (there are 20 counties)!

d) Estimate the percentage of all municipal corporations and specify the uncertainty, (County and municipal together)!

19. Every Monday a company receives a shipment with about 20 boxes containing a special detail ("shranks") needed for the next week's production. Every box contains 560 details. Every week two boxes are selected randomly and from each box 10 details, also randomly. These are checked carefully for defective. This procedure is prescribed in detail in the contract.

1	2	3	4	5	6	7	8	9	10	11	12	13
14	20	21	24	22	20	22	19	14	16	20	24	22
0	0	0	2	0	0	1	4	0	0	0	1	0
3	0	1	0	1	0	0	3	1	0	0	0	0
	1 14 0 3	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$							

- a) Consider only the first week. Estimate the number and the proportion of defective in the first weeks delivery!
- b) Give the standard error!
- c) Now consider all 13 weeks. Estimate the proportion of defectives during the first quarter and state the standard error!

21. There is much debate in Sweden on the mathematical knowledge of nine grade students. Suppose that here is a internationally developed instrument (questionnaire) to measure the knowledge in mathematics on a scale between 0 and 100. The variance of the instrument (repeatability) is 25 and the variance between students over a population is said to be 400. The instrument is intended to be used in class of up to 50 persons in a class room surveilled by a specially trained person. The filling in the questionnaire takes about two hours (=three 40 minutes lectures). Someone has decided that this instrument should be used in Sweden and that the average knowledge should be estimated.

Your task is to construct a sampling plan. Your budget for data collecting is 2 MSEK and two months (counting costs but not time for training of personal). Suggest a plan. To do so you must determine some costs and variances which are not given here. Which costs and variances do you need other than those given? Suggest reasonable values for those and determine a reasonable sampling plan and give formulas for the estimation.