A. Take home assignments, Sampling and estimation, winter semester 2012

Can be sent preferably by e-mail Daniel.Thorburn@stat.su.se. If this is impossible by post, in the departments mailbox (outside the elevators) or handed in personally to him or Nicklas Pettersson. The assignments must have reached DT or NP before 17.00.

The solutions must be your own. You are allowed to discuss the problems with each other on a superficial level but two solutions that resemble each other too much will get less credit (e.g. the same calculation errors or paragraphs with the same wordings). Note also that these assignments will influence the marks on the course. Some of them are difficult in order to be able to differentiate between you. Do not expect to be able to solve everything perfectly.

For Monday February 13

5. (from Lohr) In academic life it is important to publish research papers in refereed scientific journals. In an American college a simple random sample of 50 faculty members were asked about their publications. For the respondents the faculty was noted.

Stratum, h	Number of members, N _h	Numbers in sample n _h
Biology	102	7
Physics	310	19
Social science	es 217	13
Humanities	178	11
Total	807	50

The frequency table of the number of publications is as follows

Number of publ	Biology	Physics	Social s	Humanities
0	1	10	9	8
1	2	2		2
2			1	
3	1	1		1
4		2	2	
5	2	1		
6		1	1	
7	1			
8		2		

- a) Estimate the total number of refereed publications by faculty members in the college and give the standard error. Treat as an SRS without stratification
- b) Estimate the total number of refereed publications by faculty members in the college and give the standard error. Use poststratification.
- c) Did poststratification improve precision? Explain why or why not?

5. In a study of investment plans (Y) one has drawn a simple random sample of 100 establishments among 1570. In the frame one has access to auxiliary information on the number

of employees (X) from the previous year. One got $\Sigma_{S}(Y_{i})/100 = 35,7$ Mkr, $\Sigma_{S}(X_{i})/100 = 219, \Sigma_{U}(X_{i})/1570 = 187$ and $s_{Y} = 57$ Mkr, $s_{X} = 205$ and $r_{X,Y} = 0,24$.

- a) Estimate the total amount of planned investments with a ratio estimator and derive the standard error.
- b) About 60 % of the firms lack investment plans. If the amount is plotted in a scatter plot the figure will look something like the following (those on the x-axis lack plans. Comment on the choice of a ratio estimator. Are there other possible choices? (You are not restricted to the given quantities, but you are free to use everything known from the sample and the frame)

Investment plans





- c) Suggest another sampling plan (other than SRS) which might be good).
- d) In the frame there are other data from previous years e.g. total balance, turnover, solidity, investments previous years. Mention som other estimation methods where also this auxiliary information might be used (for sampling and/or estimation).

6. A company has 2400 employed. It has four working places with 800, 700, 500 and 400 employees. Now the company wants to make a stratified sampling study of the employees' satisfaction.

- a) They intend to estimate the proportion of satisfied (in percent) at the four working places separately and they want all standard errors to be below 5 percent units (regardless of the proportions). How large must the sample size (at least) be at each of the four working places?
- b) How large will the standard error of the proportion in the whole company be (at most) if the proportions of satisfied are estimated from a stratified sample of this size?
- c) The employees may also be classified into three groups (blue collar workers, white collar and higher officials according to the following table.

Working place	Blue collar	White collar	Officials	Total
Kiruna	500	270	30	800
Säffle	500	180	20	700
Tyresö	250	200	50	500
Malmö	-	200	200	400
Total	1250	850	300	2400

The company is also interested in the proportion of satisfied for these three groups (but only as an average over the whole company). How large standard errors can be maximally expected in the three groups? You may assume that the allocation to groups within places is made proportionally.