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 imposrible by post, in the departments mailbox (outside the elevators) or handed 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 March 5

7. A person wishes to estimate the prices of cauliflower at the farmers' markets in a town. He has a list of 40 places where market trade is permitted. He selects five of these at random and visits them. At each place he first counts the number of stalls where cauliflower is sold. Three of these stalls are selected (or all if less than three). He buys one head at each site. When he gets home he weights the head and he notes the price per kilo.

Square (place)	Stalls with Cauliflower	Price 1	Price 2	Price 3	Mean price	Row variance estimate
7	4	7,89	8,12	7,95	7,99	0,014
14	1	11,30			11,30	-
21	3	9,20	9,75	10,02	9,66	0,175
23	2	9,40	9,55		9,48	0,011
31	7	8,85	9,30	9,56	9,24	0,129

a) Estimate the number of market stalls in the town that sells cauliflower. Calculate the standard error.
b) Estimate the average price at the market stalls in the town.
c) Give a formula for how the error of the average price can be calculated. Explain all notations carefully.

8. A question on the opinion on Health Insurance reimbursement is part of an ongoing study in which one quarter of the sample is replaced each quarter. The study is done with a simple random sample and the sample size is 1 000 individuals. The question is "Do you think that the current compensation is too high and that it should be reduced? Yes / No. ". (For simplicity, we assume that no non-response or "no opinion"-answers" occur. Furthermore, we assume that the population is large in relation to sample size)

a) The number of Yes-answers was in May a given year 230. Estimate the percentage of yes in the population and estimate the standard error, based solely on this survey.

b) In September, they were asked again. The responses on the two occasions were

	Yes in May	No in May	Only Sept	Total
Only May	60	190		250
Yes in Sept	155	35	50	240
No in Sept	15	545	200	760
Total	230	770	250	

Estimate the change between May and September by looking only at the changes in the common part and give the standard error of the estimate of change. (Hint: consider a variable which is -1 at the answers yes-no, and +1 for the answers' no-yes and 0 for yes-yes resp no-no).

c) What had the standard error been if the study was based on two independent samples, and only calculated the difference?

d) How can the estimate from b) be even more improved?

9. A common way to measure the proportion of "poor" in a country is the proportion of the population of households, whose yearly total income (corrected for household size) are below 60 % of the median income. A simple random sample of 1500 households have been asked about their incomes. The following table describe the 45 – 55 % percentiles in the sample.

%	Income	%	Income	%	Income
7.5	70 000	8.2	80 000	9.2	90 000
11.7	100 000	15.4	110 000	19.9	120 000
28.3	130 000	38.9	140 000	45	147 900
46	148 900	47	150 000	48	150 300
49	151 100	50	152 000	51	153 000
52	153 700	53	154 500	54	155 100
55	156 000				

a) Estimate the median income and state an uncertainty interval (You may have to interpolate in the table).

b) Suppose (in b) that it is known that the median income is exactly 150 000 SEK. Estimate the proportion of poor and give a standard error (You may have to interpolate in the table).

c) Estimate in the same way the proportion of poor by using the observed estimated median in a instead of 150 000 SEK. Do the same thing for the upper and lower bound from a.

d) If you used a 4s-interval in a) you may estimate the standard deviation as ¼ of the proportions corresponding to the upper and lower bound. Now combine the standard error from b) with this getting an approximate standard error for the estimated proportion of poor (assuming that the two error sources are uncorrelated). Give an uncertainty interval.