Quantitative Methods Exam

2011 February, $4^{\underline{\text{th}}}$

1. An official bureau wants to estimate the average cash money per family in a given city. It is known that families can be included into two main groups: high level income group (4000 families) and low level income group (20000 families). It is assumed that usually the sample mean of cash money in high level groups is roughly 9 times more than the sample mean low level groups, and the standard sample deviations, S_h $(h = 1, 2)$, are proportional to the square root of the sample means of each group.

   Consider uniform and proportional allocations. Which of these allocations for a sample with $n = 1000$ is better in terms of variances?

   Assuming equal costs of sampling, what is the best allocation fractions for each group with minimum variance?

2. A researcher group form the *World Health Organization* (WHO) tries to estimate the proportion of infected people with *malaria* in a certain equatorial territory. As there is not an available list of people, the territory is divided into rectangular zones, except for two sylvatic areas and three deserted areas which have few people. The researchers decide that each of the rectangular zones can be considered a cluster, the two sylvatic areas can be considered a cluster and, finally, the deserted areas can be considered a cluster.

   The clusters are numbered from 1 to 60 and there is budget for sampling $n = 25$ clusters and to check every family within each cluster.

Cluster	**Number of people** (m_i)	**Number of infected** (a_i)	**Cluster**	**Number of people** (m_i)	**Number of infected** (a_i)
1	8	4	14	10	5
2	12	7	15	9	4
3	4	1	16	3	1
4	5	3	17	6	4
5	6	3	18	5	2
6	6	4	19	5	3
7	7	4	20	4	1
8	5	2	21	6	3
9	8	3	22	8	3
10	3	2	23	7	4
11	2	1	24	3	0
12	6	3	25	8	3
13	5	2			

Use the previous data to estimate the proportion of people who is infected by *malaria*. Write a bound of error on the error of estimation.

3. In a study to estimate the total radioactivity content of a shipment of fishes from the Bikini Atoll, it was taken a random sample of $n = 10$ animals which was analyzed and obtained:

Fish	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Rad. Content	0.0021	0.0030	0.0025	0.0022	0.0033	0.0027	0.0019	0.0021	0.0023	0.0025
Weight	4.0	4.8	4.3	4.2	5.0	4.6	3.9	4.1	4.2	4.4

The total weight of all fishes was 1800 Kg. Estimate, by a *ratio estimator*, the total radioactivity content of the fishes and a bound on the error of estimation (95%). Estimate, also by simple random sampling (SRS) the same total radioactivity content. Discuss which is the best procedure and why it is so.

4. Explain the basics and use of cross validation techniques in *Tree Models* methods. Indicate the similarities and differences between *k-means* and self organizing maps (*SOM*).