The sample plan for the Norwegian index of food and non-alcoholic beverages

Pia.Ronnevik@ssb.no

Statistics Norway

Background

- Statistics Norway started in 2005 with scanner data in "full scale" for the index of food and non-alcoholic beverages.

- Before 2005, the sample of stores was selected together with the other stores in HICP/CPI.

  - selected by region and industry

- Since 2005, the selection of stores in the index of food and non-alcoholic beverages is following a separate sample plan.

- Due to big structural changes in the market the last 5 years, a new sample plan was established in 2010.

  - combined with a routine for annual updating

The sample plan in the period 2005 - 2009

- The sample consisted of 150 stores and kiosks.

- The draft population was stratified by:
  - the 4 big chains in Norway
  - the chain concepts; kiosk, low price, nearby market and supermarket.

- Neyman allocation was used to find out the number of stores that should be selected in each stratum.

- The sample in each stratum was randomly selected without replacement.

The new sample plan in 2010

- How should we best stratify our draft population?

- How big should the size of the sample be?

- Which allocation method should we use to find out how many stores should be selected in each stratum?

- How should we select the stores in each stratum?

- How do we make an annual routine that update the sample each year?
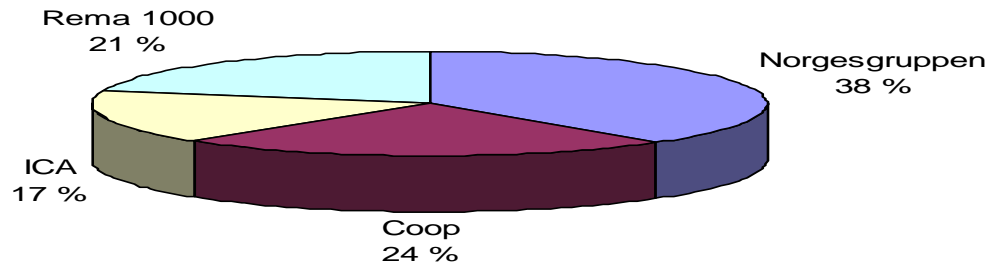
The draft population as of 2010

- Selected from the CRE (Central Register of Establishments and Enterprises), defined as the industries in 47.11 and 47.12 (SIC 2007).

- Consist of 4171 stores that are able to provide scanner data and have figures of turnover in the CRE.

- Excludes small local meat, fish and vegetable stores.

- Consists of 94% of the total turnover in the industries.

# Comparisons of old/new draft population

2005:
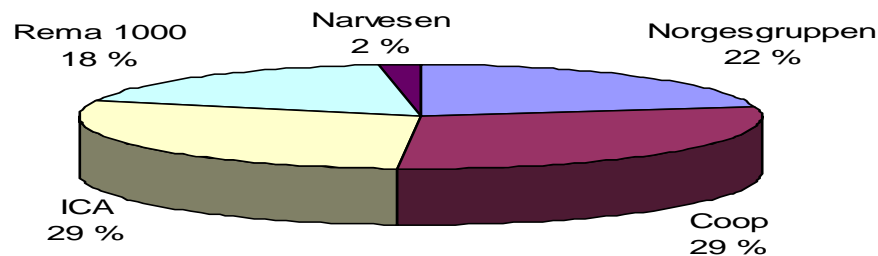
2010:

Stratification

- Dividing the draft population into non-overlapping subpopulations called strata.

- Observations within each stratum should be homogeneous in terms of what is measured.

- Often done to obtain a more representative sample, and reduce the uncertainty in the estimates.

Hierarchical classification

Chain_x — Chain

Kiosks — Low price — Nearby — Supermarket — Concept

Chain profile
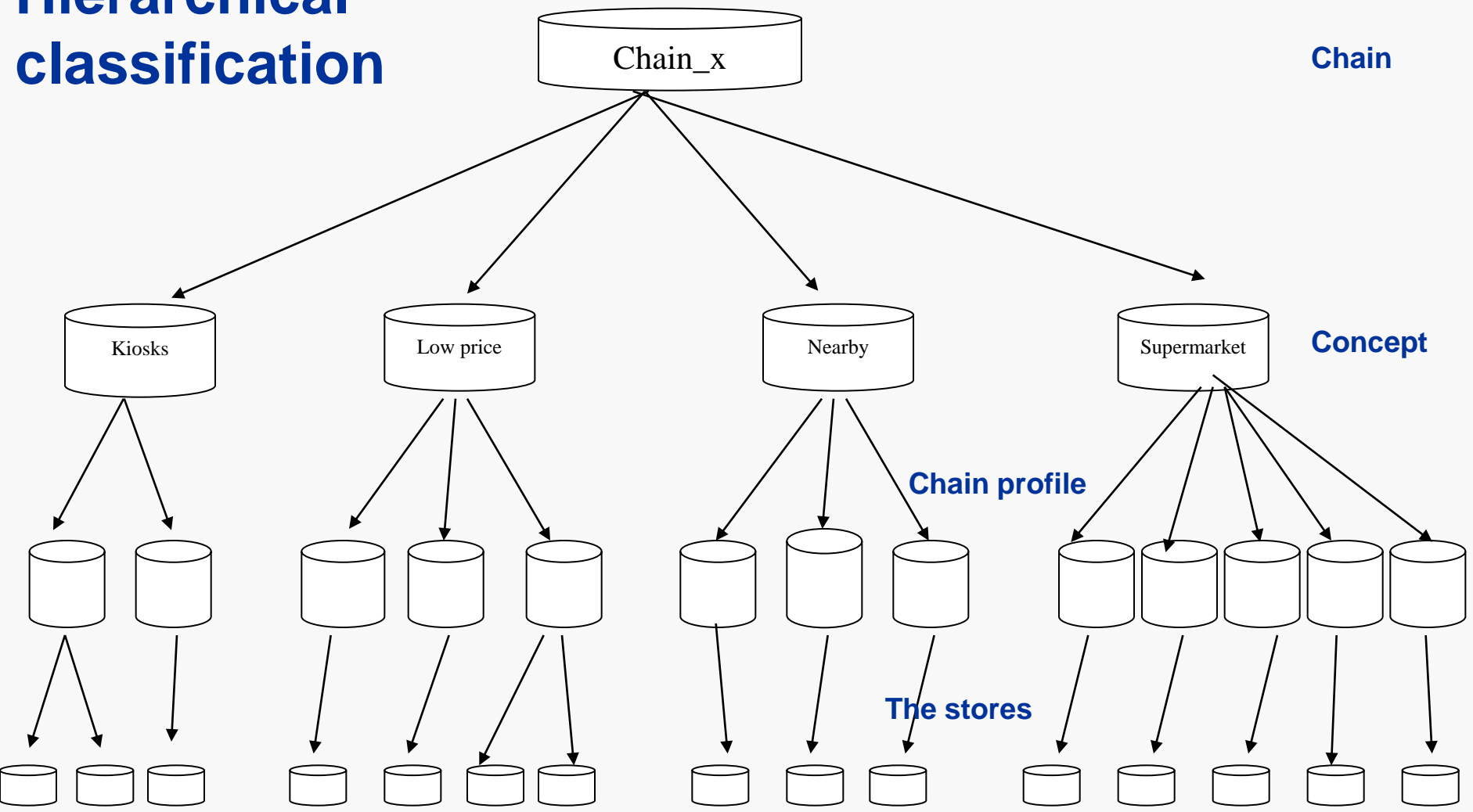
The stores

The sample size

- A result of trade-offs between theoretical considerations and some practical limitations.

- 180 stores, 30 more than the previous sample.

- Larger draft population, due to several new concept and chain profile.

- More strata require a bigger sample, due to the fact that no strata can be empty.

- The sample can't be too large due to resource constraints.

How many stores should be selected in each stratum?

Neyman allocation:

- $n_l = n * (W_l S_l)/(\sum W_k S_K)$

   $= 180 * (W_l S_l)/(\sum W_k S_K)$

- Takes into account strata's size and their internal variation as for instance in turnover.

- Minimizes the uncertainty range of estimates.

Proportional allocation:

- $n_l = n (N_l/N)$

   $= n * W_l$

- The method divides the strata by the same proportion as the draft population is divided into strata.

- Proportional allocation gives the units in all strata same probability of being selected.

Arguments for choosing proportional allocation

- In 2010 the strata are more homogeneous with regard to the turnover, so we don't have to emphasise the strata that have a larger spread in turnover.

- The stores turnover indicates the size of the product range. Therefore strata with high turnover should be represented with more stores in the sample.

- Neyman allocation implies a larger sample size to ensure enough stores in each stratum. This is not desired due to resource constraints.

The selection procedure

The stores in each stratum are selected randomly without replacement.

- All the stores in each stratum have an equal chance of being selected, but the inclusion probability is different between the various strata as a result of the allocation.

- It is not necessary to use systematically selection for example after turnover, since the spread within the different strata in terms of turnover are not of significance.

Comparison of old/new selection

2005

- 150 shops

- Stratified by chain and concept

- Neyman allocation

- Randomly selection without replacement

2010

- 180 shops

- Stratified by chain, concept and profile

- Proportional allocation

- Randomly selection without replacement

The annual routine

- **First we establish the draft population from the CRE-registry.** We remove the stores that do not have chain association or turnover in the CRE-registry.

- **Stratification.** We stratify after chain, concept and chain profile.

- **Allocation.** We use proportional allocation.

- **The roll.** We find out the differences between the new and old sample. We then add or remove stores from the old sample so we have corrected any imbalances.

- **The selection procedure.** The stores in each stratum are selected randomly without replacement.

- **Comparison of the new sample with the old.** This should be done to see how much the market of grocery and convenient stores has changed since the last sample was drawn for the index of food and non-alcoholic beverages.

The sample plan and calculation method

- Changed methodology as of January 2013.

  - From a superlative index to an unweighted geometric index at elementary level combined with a exclusion of the unimportant items.

- Due to equal weighting the sample plan even more important.