
# Statistically optimal sample size allocation in food composition tables

National Food Research Institute,
Yoshiki Tsukakoshi, Akemi Yasui

## Codex issues on sampling

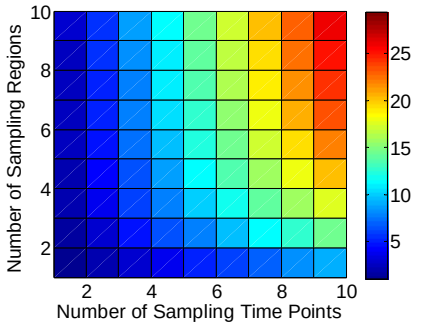
- **Codex Committee on Method of Analysis and Sampling (30th session @Balatonalmádi, Hungary)**
  - For some specific foods and chemicals, sampling methods has been adopted.
  - ISO 17025 (adopted by reference) requires estimating uncertainty
- **General Guidelines on Sampling (CAC/GL 50-2004)**
  - Mainly for inspection (accepting/rejecting a lot)
- **Eurachem Guide(also submitted to CCMAS by UK etc.)**
  - **"Measurement uncertainty arising from sampling"**
  - Detailed document on sampling uncertainty
  - No remark on sample size allocation

8th International Food Data Conference
October 1-3, 2009
Bangkok, Thailand

# Guideline on measurement uncertainity arising from sampling


102 pages

- Concept of error
- Estimation of error

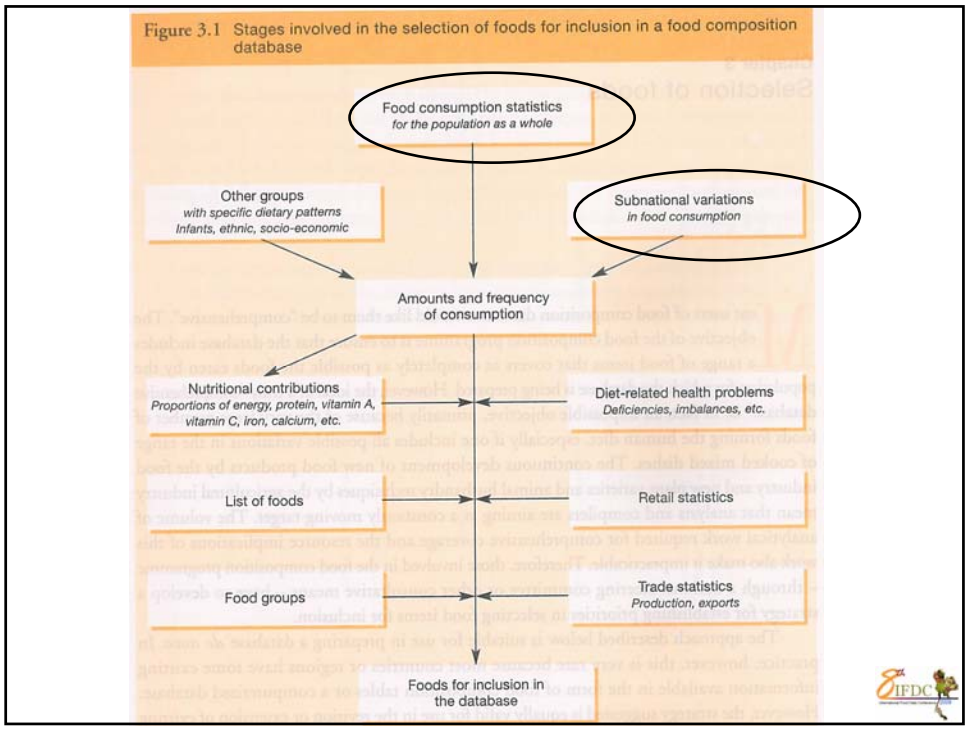



# Sample size allocation

Three sweet potatoes and One Irish Potatoes?

- Important part in designing sampling plan for food composition database
- Usually, more than one sample can be collected.
- The total number of samples is limited due to resources.

One sweet potatoes and Three Irish Potatoes?




## What is sample allocation problem

- $Y = \sum n_i c_i$
  - $N_i$:Consumption of food item i.
  - $C_i$:Composition/Concentration in food item i.
- $c_i$ : has sampling error $e_i$
- How to minimize error of Y?

## Proportional Allocation

- Sample size N proportional to "quantity"
- Sample size N is as proportional to $c_i n_i$ (concentration times production)
- If you intake great quantity of nutrients from the food, sample size is big
- If you consume many, the concentration is small, then sample size is small

## Neyman Allocation

- Sample size proportional to sample quantity times sample variance
  - Variable foods are allocated with more sample
- $N \propto cNi\,\sigma i$
- Optimal to reduce total error
- $\sigma i$ is not always known
  - Could be found from previous study or food composition database of other country.

# Statistics for quantity and their availability in Japan

- *Food consumption statistics – weight record*
  - ○Best statistics for intake estimate
  - ×Not published for each item
  - ×Seasonality not considered.
- *Food production statistics*
  - – Indirect but approximates intake
  - ○Rich information as is needed to develop economic policies
  - ×Not for foods(for other industrial)
- *Household survey*
  - – Indirect but approximates intake
  - ×Many food items are shown as “other” category.
- *Market statistics*

# Example: Potato group

| | Contribution in carbohydrate intake (n) | Coefficiet of variation (v) | Proportional Allocation (n×c) | Neyman Allocation (n×v×c) |
|---|---|---|---|---|
| Sweet potato (carbhydr:31%) | 2% | 8% | 0.07 | 0.16 |
| Irish potato (c:18%) | 3% | 18% | 0.09 | 0.6 |
| Taro potato (c:13%) | 0.2% | 14% | 0.01 | 0.03 |
| Yam (c:27%) | 0.4% | 5.5% | 0.01 | 0.03 |
| Corn starch (c:81%) | 15% | 0.8% | 0.43 | 0.13 |
| Potato starch (c:86%) | 1.3% | 0.04% | 0.02 | 0.00 |

# Example: Potato group: Other composition(minerals etc.)

| | Neyman Allocation for carbohydrate | Neyman Allocation for Protein |
|---|---|---|
| Sweet potato | 57 | 11 |
| Irish potato | 17 | 30 |
| Taro potato | 3 | 3 |
| Yam | 2 | 2 |
| Corn starch | 12 | 3 |
| Casava starch | 1 | 0.2 |

- Take maximum sample size

or

- Use different sample size for different substance

# Conclusion

- To consider sample allocation, Production statistics will be suitable for Japan
- Neyman allocation seems reasonable than proportional allocation.
  - For Neyman allocation, variability data is required
  - Some can be obtained if you can access the raw analytical data.
  - Some can be obtained from US database
- Good starting point will be major food
  - No statistics available for minor food and usually they are too small

Problems

- Not enough data for cooked items (baked, boiled,…)
- Statistics for food mostly used for industrial purpose (Starch)
- Optimization by allowing different sample size for different substance will be very complicating