HSE and University of London Double Degree Programme in Data Science and Business Analytics

Statistical Methods for Market Researh, 2023-2024

Classteacher: Ksenia Kasianova.

Class 6: Variance Reduction

Task 1

Consider an estimator of $E(Y)$:

$$\widehat{Y}_{cv} = \bar{Y} - \theta\bar{X} + \theta E(X),$$

(a) Find the variance of this estimator

(b) Minimise the variance by θ

(c) Derive the variance of $\widehat{Y}_{cv}$ under optimal θ

Task 2

Let X is binary with values 1 and 0. Let $p = E(X)$. Show that the following estimates are identical:

$$\widehat{Y}_{strat} = p\bar{Y}_1 + (1-p)\bar{Y}_0,$$
$$\widehat{Y}_{cv} = \bar{Y} - \hat{\theta}\bar{X} + \hat{\theta}E(X)$$

Task 3

How to choose the sample sizes $n_1, n_2, ..., n_k$ so that the available resources are used in an effective way?

There are two aspects of choosing the sample sizes:

– Minimize the cost of survey for a specified precision.

– Maximize the precision for a given cost.

(a) Can both of these goals be achieved?

(b) Consider the stratified sample variance estimator fith finite population correction:

$$\text{Var}\,(\bar{y}_{st}) = \sum_{i=1}^{k} w_i^2 \,\text{Var}\,(\bar{y}_i)$$

where

$$\text{Var}\,(\bar{y}_j) = \frac{N_i - n_i}{N_i n_j} S_j^2.$$

And let the cost function be

$$C = C_0 + \sum_{i=1}^{k} C_i n_i$$

where

C : total cost

C_0 : overhead cost, e.g., setting up the office, training people etc

C_i : cost per unit in the i^{th} stratum

$\sum_{i=1}^{k} C_i n_i$: total cost within the sample.

Find minimal variance by n_i under this cost function. (Hint: consider the Lagrangian function).

(c) Minimize variability for fixed cost $C = C_0^*$.

(d) Minimize cost for a given variability $V = V_0$.