Lecture 7. Descriptive survey: cross-sectional and longitudinal data analysis

Longitudinal & cross-sectional data analysis is the point of interest for economists, social sciences specialists, and others. The interest results in attraction of statisticians, as the data are so „rich" that need „rich" families of statistical methods; besides this, new probabilistic models & methods are developed to melt the needs of practical specialists in this area.

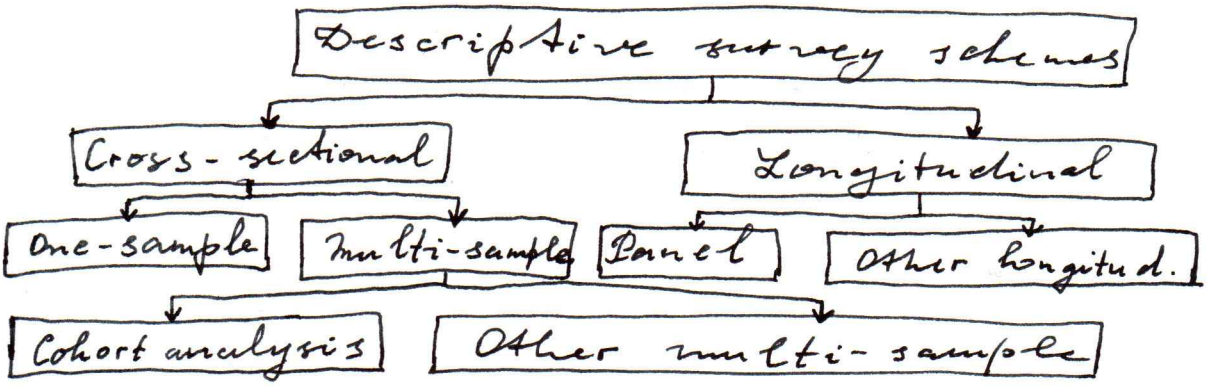

Fig. 1. Schemes for descr. surveys

With participants of so-called „panel" usually it is agreed in advance that some info from them will be available during some period.

Cross-sectional, acc. to Fig. 1, are splitted: 1-sample (single) & multi-sample (multiple). 1 sample or 2+ samples are taken out of the general data set correspondently. Diff. samples may be connected to different sets (diff. cities) or to diff. time moments. Although the same data set is analysed (big): country population, ...

Other example of multiple survey: polls at different districts: ∀ poll represents from n to m ~~subjects~~ districts; at ∀ district a sample of, say, 500 respondents is formed. A poll is performed at all districts simultaneously, with the same questions list, this makes possible to compare districts.

Sometimes, multi-sample survey data are studied with the so-called cohort analysis. A cohort is a group of people that dealed with some (certain) event during some fixed period. Classical example in US-textbooks: people participated in the Afganistan war. The other example of a cohort is a group of people born from 1971 – 1980. If there is same periodicity in the surveys, then (e.g. once in 5 years) there is a possibility for comparison: 1) Changes of the cohort characteristics during different time periods; 2) Characteristics of diff. cohorts.

In some situations, cohorts analysis prevents from errors in statistical inference; to illustrate this, let's look into the following example – average consumption of non-alcoholic drinks through different age groups; liters / year (table 1).

Respond. age	Year of taking the survey				Coh. not.
	1950	1960	1969	1979	
8-19	53	63	73	81	
20-29	45	61	76	76	C8
30-39	34	47	68	72	C7
40-49	23	41	59	68	C6
≥ 50	18	29	50	52	C5
Coh. not.		C1	C2	C3	C4

Table 1. An example of cohort analysis

Ad-hoc analysis: age increases $\Longrightarrow$ non-alc. drinks cons. decreases ?!

Cohort analysis: no one cohort decreases its level of consumption; but each next cohort consumes more, than the previous.

Consider now some models related to data of such types.

1. Basic model of non-homogeneity

$$y_{it} = \alpha_i + \beta' x_{it} + \varepsilon_{it}. \qquad (1)$$

In x_{it} there are k regressors; α_i – individual effect (invar), spec. for the i-th cross-sectional unit. If α_i is constant through all units, LSM $\longmapsto$ consistent & effective estimators for α & β.