

GLOSSARY OF TECHNICAL TERMS

[Accompanying material for: Katrin Auspurg & Thomas Hinz (2015): Factorial Survey Experiments. Sage Series: Quantitative Applications in the Social Sciences No. 175. Thousand Oaks, CA: Sage]

Anchoring effects: Information provided in the question wordings, answer scales, or vignettes is used by respondents as a reference point for their judgments. For instance, with higher (lower) earnings presented in the vignettes, higher (lower) earnings are considered to be fair.

Anchoring vignettes: Standardized descriptions of hypothetical people or situations presented to all respondents. Using the judgments of anchoring vignettes as a reference point, a standardization of further judgments is achieved (e.g. an anchoring vignette presenting a third person's health description is followed by questions evaluating the respondent's own health status).

Block: → Deck.

Censored responses: When respondents have already chosen the end point on a limited response scale, they are no longer able to select an even more extreme answer, i.e. their answer is censored at the scale end point. This is often also called a ceiling effect and needs special data analyses for unbiased results like Tobit regressions.

Choice experiment (CE): Experimental survey method wherein respondents are asked to pick the most attractive alternative out of several alternatives presented in one joint *choice set*. The attributes of the alternatives are experimentally varied in their levels.

Conjoint analysis (CJA): Umbrella term for all decompositional survey methods or, in a more narrow sense, the experimental survey method used in consumer research where respondents rate or rank objects or services that are described with experimentally varied attribute levels.

Clustered robust standard errors: The estimation of standard errors for regression coefficients is adjusted to the data structure where the single observations are not independent of each other due to clustering (e.g. single respondents evaluate more than one vignette and therefore vignettes are clustered within respondents).

Computer-assisted personal / self-interview (CAPI / CASI): Survey based on a computer questionnaire, with interviewer assistance or in a self-administered mode (i.e. where respondents read and answer the questions themselves).

Construct validity: Extent to which operations and measurements reflect the theoretical concepts they are intended to measure.

Cross-elasticity: A trade-off between a change in a metric variable and a change in another (metric) variable, mostly expressed in percentage changes.

Cross-level interaction: Interaction term between variables of the lower and upper level, i.e. between L1 and L2 variables.

Deck or block or set: Selection of vignettes presented to single respondents.

Decompositional methods: Methods whereby respondents' overall evaluations of objects and situations are decomposed into the impact of the different dimensions underlying these objects or situations.

Design effect (deff): The design effect serves to evaluate the statistical efficiency of a sample. It is calculated as the ratio of the sampling variance to be expected under the given sampling technique in relation to the sampling variance of an analogous simple random sample (SRS).

Design efficiency: Measure of the goodness of an experimental design. Good designs provide maximum information (minimum variance and covariance) for intended parameter estimates. Common measures like *D*-efficiency are based on Fisher's information matrix (FIM).

Design resolution: The design resolution defines which effects, including possible interactions, are estimable. It is recommendable to use at least Resolution IV designs: With Resolution IV designs the researcher specifies some two-way interactions that are identifiable (while others might get confounded) and with Resolution V designs all two-way interactions are estimable independently of each other (and also of all main effects).

Dimensions or vignette variables: Characteristics of the vignettes that are experimentally varied in their levels. Dimensions should be meaningful for respondents' judgments and be varied according to a suitable experimental plan. FSs use several dimensions to realistically depict a situation.

Ecological validity: → Mundane realism.

Experimental design: Plan for running an experiment, e.g. experimental factors (dimensions) and levels to be used and their combinations. In a more specific sense, the experimental design denotes the sample of experiments or vignettes to be used in a study. In particular: ***mⁿ-design***: Vignette design with *n* dimensions showing *m* levels. For instance, a 2⁵-design means that there are five dimensions with two levels each. Similarly, additional dimensions or *experimental factors* can be added, e.g. a 2³3⁴-design means that there are three dimensions (factors) with two levels, and four dimensions (factors) with three levels.

Experimental realism: Extent to which the participants take the experiments seriously or experience what happens in the experiment as real.

Experimenter demand effect: Trend whereby respondents react with a bias toward appropriate behavior that is “demanded” from them. In a more narrow sense, attempts on the part of respondents to fulfill researchers’ expectations (research hypotheses).

External validity: Extent to which results and causal mechanisms are generalizable to other settings, samples, or methods.

Feed-forward variables: → Preload variables.

Fixed effects (FE) regression model: Regression model mostly known from panel data analysis controlling for unobserved heterogeneity for invariant L2 characteristics (i.e. the respondent’s characteristics are constant for all evaluations one respondent makes). The model focuses on the variation of the dependent variable within subjects only.

Fractional factorial: Subset of the full factorial (or vignette universe).

FS or FSs: Abbreviation for factorial survey(s).

Full factorial: → Vignette universe.

Hierarchical data structure: Data structure wherein observations on a lower level (*level 1 or L1*) are nested within observations on an upper level (*level 2 or L2*). FSs where each respondent evaluated several vignettes show this data structure: Vignettes vary on L1, and are nested within respondents who vary on L2. (Sometimes there are even more levels, when respondents are nested in further units like households.)

Hierarchical linear model (HLM): The usual linear model is extended in response to a hierarchical data structure with different levels. L1 units are nested within L2 units. There is a structural equivalence to panel data analysis with points of measurement in time (L1) for different subjects (L2).

Internal validity: Extent to which the variation in the outcome variable is solely caused by variation in the independent variables or experimental treatments (i.e. vignette dimensions).

Levels: Values of the vignette variables (dimensions). Continuous variables (like earnings in Euros) are presented using a finite number of categories.

Level balance: All levels of single dimensions occur with same frequency across the design. In practice, this means that the dimensions show a maximum variance in their levels.

Likelihood ratio (LR) test: Serves to compare nested models (differing only with respect to additional parameters to be estimated). It is tested if the inclusion of the additional parameters contributes to a statistically significant increase in the log likelihood. The test informs researchers on the question of whether an extended model (with additional parameters) fits the data better than a parsimonious model.

Long data format: A data format wherein the single rows represent the different L1 observations (vignettes), and each L1 variable is stored in one column. (\leftrightarrow *wide data format*)

Magnitude scale: Answer scale wherein respondents are asked to match numbers (“number matching technique”) or mark the length of lines proportional to the perceived magnitude or intensity of stimuli. Magnitude scales allow respondents to express their feelings as freely as possible, and are assumed to reach interval scale level. In addition they are used to avoid the censoring of responses.

Mundane realism or ecological validity: Extent to which experimental situations and procedures are representative of situations and conditions in real life.

Number-of-levels effect: Dimensions with more levels than other ones achieve higher importance on evaluations only because of their higher variance in levels.

Order effects: Exist in relation to the order of vignettes, dimensions, or answer categories influencing the responses. Typical order effects are *carry-over effects*, whereby former vignettes, dimensions, or answer categories frame the evaluations of later ones, and *primacy (recency) effects*, whereby dimensions or response categories firstly (lastly) mentioned to respondents are better kept in mind and hence have more impact on evaluations than other ones. In addition, both learning and fatigue can cause order effects.

Orthogonal: A matrix is orthogonal if there are no correlations between the elements of each of the columns. In practice, this means that all effects can be estimated independently of each other.

Paper and pencil interview (PAPI): Surveys based on a paper questionnaire, with or without interviewer assistance.

Pilot study: \rightarrow Pretest.

Preload or feed-forward variables: Variables containing information that is specific to each participant and already known in advance. Those variables are frequently used in panel surveys to ask respondents if information from prior panel waves is still up to date. Similarly, those variables can be used to insert the vignette texts that are specific to single respondents.

Pretest or pilot study: Small-scale preliminary study to test the feasibility, comprehensibility, time, and possible methodological problems of a study or questionnaire design.

Psychological realism: Extent to which those psychological processes occurring in experiments are the same as those happening in real life.

Random effects (RE) regression model: Regression model mostly known from panel data analysis that accounts for variance in components of the error term. A RE model accounts for the fact that the units of analysis cannot be considered as independently drawn from a random sample. An (often questionable) assumption is that the independent variables are uncorrelated with the error term. RE

models estimate within- and between-subject variations. Invariant characteristics of L2 units (like respondent characteristics) can be included.

Random intercept (RI) regression model: A simple random effects (RE) model. Compared to OLS, only one additional parameter for the variance of the intercept is added. This accounts for the variation of the error term across units.

Random slope (RS) regression model: Like RI, RS models are RE models as well. Here, however, not only the intercept but also slope parameters are assumed to randomly vary across units.

Random utility theory (RUT): Theory on preferences based on general economic theories on utility maximization, which provides a theoretical motivation for the choice task and data analysis techniques used in CEs and might also be used to motivate FSs on behavioral intentions.

Resolution: → Design resolution.

Run: Each run in an experiment indicates a distinct experimental stimulus. In FS studies, the number of runs represents the number of vignettes in the sample.

Semi-partial R^2 -values: These values measure the isolated relevance of independent (vignette) variables in explaining respondents' evaluations. They express the proportion of explained variance R^2 that goes back to the single variables.

Set: → Deck.

Setup data: Data that include the numerical codes of the single dimensions and the complete vignette texts to be presented to respondents. In addition there are unique IDs that define the vignettes that respondents get presented with and their order.

Simple random sample (SRS): Sample wherein all units are selected at random and each unit from the population has exactly the same probability of being sampled.

Social desirability bias (SDB): Bias caused by respondents' tendency to over-report socially desirable attributes or actions and under-report socially undesirable ones.

Statistical power: Probability that a statistical test will reject the null hypothesis (H_0) when actually the research hypothesis is true. The statistical power is the probability of not committing a *type II* error (or false negative rate) in statistical testing. With this error being denoted as β , the power equals $1-\beta$. The higher the power, the more efficient a design is in a statistical sense.

Status-quo option: Option in CEs stating that there would be no changes to the current state (sometimes also called the "no-choice option").

Tobit regression model: A regression model that takes (left and/or right) censoring of the dependent variable into account (i.e. the dependent variable cannot be observed below and/or above certain

thresholds). The model relies on strong assumptions on the normality and homoscedasticity of error terms to provide adequate estimations.

Trade-off: The multidimensional design forces the respondents to consider the dimensions simultaneously, which allows the calculation of how willing respondents are to forgo part of one dimension in order to get more of another dimension. In technical terms, a trade-off represents the cross-elasticity between the respective dimensions.

Validity: Most important criteria for the quality of measurements in the social sciences. One distinguishes *internal validity* (the confidence with which the variation of the dependent variable is solely caused by the variation of the independent variables or experimental stimuli), *construct validity* (the extent to which the instrument measures what it purports to measure), and *external validity* (the extent to which results from the measurement can be extended to other kinds of measurements, settings, or groups of participants). (→ see also mundane or experimental realism).

Vignettes: (Mostly verbal) descriptions of hypothetical situations or objects that represent the experimental stimuli or treatments.

Vignette universe or **full factorial:** All possible combinations of levels and dimensions.

Vignette variables: → Dimensions.

Wide data format: A data format wherein the single rows (observations) represent the different L2 observations (respondents) and the n_d L1 variables (outcomes, vignette IDs, etc.) are stored in separate columns (↔ *long data format*).

Willingness to pay (WTP), willingness to avoid (WTA): Willingness to pay (avoid) attractive (unattractive) options or attributes, estimated in monetary terms or other metrics like waiting time.
