

DATA PREPARATION AND DATA ANALYSES

In order to facilitate replication of the techniques explained in Chapter 5, the following material is provided (**§5_material.zip**):

- The folder **“Preparation”** contains the datasets **§5_population_sample.dta** and **§5_population_sample_setup.dta**, which can be used to practice the preparation steps in Section 5.1 (e.g. the merging of response data and setup data). The Stata do-file **§5_preparation.do** shows how to merge and prepare these data when working with Stata. The respondent data are still in the Stata wide format, while the setup data are in the long format. For data analyses, the data have to be merged, which can only be done when they are in the same format. In addition, data analyses require the long format.
- In addition, we provide the resulting data: **§5_data.dta**. These can be used to replicate the data analyses in Sections 5.2 and 5.3, as shown in the do-file **§5_analyses.do**. These data are already stored in the Stata long format (Stata version 11).

Before running the analyses files, we recommend readers first read through the following explanations.

5.1 General Information on the Data

As outlined in Section 5.2.3. in the textbook, the data used to illustrate the techniques of analyses were gathered in face-to-face interviews with a population sample in the city of Konstanz (southern Germany). We provide the original data of 250 respondents. Each participant evaluated 10 vignettes regarding the fairness of earnings for fictitious vignette characters on an 11-point rating scale ($-5 = \text{unfairly low}$ to $+5 = \text{unfairly high}$).

The vignette sample was generated as a *D*-efficient sample of 24 different decks of 10 vignettes each (*D*-efficiency of 90.4), excluding illogical cases (e.g. medical doctors without a university degree; see §1_earningsexample.pdf for more details). A Resolution IV design was used in which all of the main effects and some selected two-way interactions expected from social theories and prior studies (such as interactions between the gender and labor market characteristics of the vignette person) were orthogonalized. The decks were randomly allocated to the respondents. All of the data were collected using a paper-and-pencil interview (PAPI). After some initial questions, the FS module was employed as a self-administered module.

The fictitious employees were characterized by eight dimensions (including monthly gross earnings). The vignette variables were as follows: *vsex_f* (1 = *female*, 0 = *male*), *age* (in years, ranging from 25 to 60 years in 5-year intervals), *degree* (*vocational training* or *university degree*, with *no training* as the reference category), *children* (0, 1, 2, or 3), *job* (10 different occupations, from *unskilled worker* to *medical doctor*), *experience_d* (1 = *a lot of*, 0 = *only few*), and *tenure_d* (1 = *long*, 0 = *short*). Additionally, the prestige of the ten different vignette occupations is provided as a variable (magnitude prestige scores [MPSs]).

5.2 Codebook

The following list contains variable names and variable labels. VIGN refer to vignette characters, RESP to the respondents, and INT to the interaction terms of vignette characters and respondents.

Variable name	Variable label
<code>id_resp</code>	Respondent ID

vignr	Vignette Order
vig_eval	Vignette Evaluation (dependent variable for all regressions)
id_deck	Vignette Deck ID
v1	RESP: sex
v2	RESP: year of birth
v3	RESP: highest educational level
v4	RESP: difficulty to evaluate vignettes
v5	RESP: current employment status
v6	RESP: family status
v7	RESP: permanent partnership
v8	RESP: children (yes/no)
sex	VIGN: sex of vignette person
age	VIGN: age of vignette person (in years)
degree	VIGN: vocational/university degree of vignette person
prestige	VIGN: occupational prestige of vignette person's job
experience	VIGN: occupational experience of vignette person
job	VIGN: current job/occupation of vignette person
tenure	VIGN: tenure with current employer of vignette person
children	VIGN: number of children of vignette person
income	VIGN: monthly gross income of vignette person in EUR
educ_high	RESP: indicator of university degree (0/1)
log_income	VIGN: logged monthly gross income of vignette person in EUR
vsex_f	VIGN: indicator of female vignette person (0/1)
degree2	VIGN: indicator of vocational training degree of vignette person (0/1)
degree3	VIGN: indicator of university degree of vignette person (0/1)
experience_d	VIGN: indicator of a lot of experience of vignette person (0/1)

tenure_d	VIGN: indicator of long tenure of vignette person (0/1)
vsex_fXeduc_high	INT: RESP educ_high with VIGN female
degree2Xeduc_high	INT: RESP educ_high with VIGN vocational training degree
degree3Xeduc_high	INT: RESP educ_high with VIGN university degree
childrenXeduc_high	INT: RESP educ_high with VIGN number of children
prestigeXeduc_high	INT: RESP educ_high with VIGN occupational prestige
experienceXeduc_high	INT: RESP educ_high with VIGN experience
tenureXeduc_high	INT: RESP educ_high with VIGN tenure
log_incomeXeduc_high	INT: RESP educ_high with VIGN log_income
mean_income	VIG: mean income of vignette person

5.3. Data Preparation

Readers can easily merge the data from the population sample (the first 250 cases) **and the experimental setup** of the vignettes. The data files are as follows: §5_population_sample.dta and §5_population_sample_setup.dta (see folder “Preparation”).

The procedure is documented in the Stata do-file: §5_preparation.do.

This program contains some additional steps for recoding and defining useful variables (such as the prestige score and the logged income of vignette characters) that will be used in the analyses.

5.4 Replication of Analyses in the Textbook

To **replicate the data analyses in Sections 5.2 and 5.3** readers can use the Stata do-file §5_analyses.do. Please note that all of the variables in the regression models have already been created. After opening the data file §5_data.dta in Stata, all regression models should run.