

# Script: An Introduction to Factorial Survey Experiments (FSE), Part III

NCRM Online learning resource

Tamara Gutfleisch, Mannheim Centre for European Social Research (MZES), University of Mannheim

## Introduction

Welcome to Part III of this introductory course to factorial survey experiments. In Part I and II, we have learned about the principles underlying factorial surveys, and about the different steps in designing and conducting a factorial survey experiment. Now, I will show you some practical examples using Stata. Specifically, this part of the online course focuses on checking the data quality and how to analyse and interpret the data obtained from factorial surveys. It is not necessarily required that you have prior knowledge of Stata, but prior knowledge of any other statistical software (e.g., SPSS or R) would be desirable.

I will first give some background information about the data we are using, which is also listed here. This syntax file will be provided as supplementary material.

For the practical examples, we will use data from the EDYPOLU recruiter survey, which can be used for scientific purposes: <https://doi.org/10.17605/OSF.IO/ZTB6Y>. The EDYPOLU project aimed at examining the impact of certain applicant characteristics on recruiters' hiring intentions in Luxembourg. I jointly developed the experimental design together with my former PhD supervisor, Robin Samuel. Detailed information on the experimental design, the recruiter survey, and the codebook can be found in Gutfleisch and Samuel (2021). The vignette universe of our experimental design comprised 36 vignettes, which we all used. Thus, no vignette sample was drawn. However, we did use a D-efficient blocking technique to divide the 36 vignettes into 6 vignette sets, each including 6 vignettes. Three experimental variables were included: gender, unemployment, and nationality of applicants. The vignettes looked like hypothetical CV, like I showed you in Part I of this online course. Each respondent was randomly assigned to one vignette set or deck, and the order of vignettes across respondents was randomised. We used a "personalised" experimental set up, in which each respondent was randomly assigned to one unique vignette set before the data collection. However, as I already indicated, we will not focus on creating the vignette sample or merging the set up and respondent data, as this is out of scope of this short introductory course. You can find more practical examples that are not covered here in the supplementary material to the book by Auspurg and Hinz (2015) (→ [https://study.sagepub.com/auspurg\\_hinz/student-resources/ancillary-materials](https://study.sagepub.com/auspurg_hinz/student-resources/ancillary-materials)). In the authors' supplementary material, you will find example code of how to construct the vignette sample in SAS, examples of how to prepare the set-up data, and how to merge the set-up data and respondent data as well as more examples of how to analyse the data. This introductory course is meant to give you a first set of tools to help you understand what is important when analysing the data obtained from factorial survey experiments and how to interpret the data.

Relevant variables that we are going to use for these practical examples are:

- Respondent\_ID - Respondent ID
- sector - Occupational field
- vigid - Vignette ID (ranges from 1 to 36)
- vignet - Vignette set, individual vignette sets (i.e., with randomised order)
- deck - initial vignette sets (ranging from 1 to 6)

- vignr - Vignette position
- vigeval\_1 - Vignette evaluations
- vgndr - Vignette gender
- vuetim - Vignette unemployment
- vnat - Vignette nationality

## Getting started

Let's get started! After we have downloaded the data, we upload the scientific use file of the EDYPOLU recruiter survey into Stata.

```
* @0 Load the data for analysis
use "..\EDYPOLU_survey_SUF.dta", clear
```

The data already combines the experimental set up data and the respondent data. The EDYPOLU recruiter survey covered five occupational fields, which can be seen as separate experiments. For this course, I will focus on data from the catering field. However, the analyses could also be done in the full data set, but you would need to control for the occupational fields in the regression analysis.

```
* @1 Select sample
* Keep only one occupational field
keep if sector==4 // keep data from the catering sector
```

## Checking the characteristics of the vignette experiment

Once we have selected our sample, we can start checking the characteristics of the vignette experiment. This is an important step before starting any substantial analyses. For example, you want to check whether the randomisation has worked, meaning whether the correlations between vignettes dimensions are minimised and whether the vignette levels occur with more or less the equal frequency.

Let's have a look at the frequency of each deck first. In our case, there were six decks in total.

```
tab deck // frequency of initial vignette sets
```

As can be seen in the results (Table 1), the data set seems to be more or less balanced. There are relatively small differences in the frequency between each vignette deck, although deck 5 occurs less often than the other decks.

Table 1: Frequency of vignette sets

Initial deck ID	Frequency	Percent	Cumulative Proportion
1	78	17.57	17.57
2	78	17.57	35.14
3	72	16.22	51.35
4	72	16.22	67.57
5	66	14.86	82.43
6	78	17.57	100.00

Total	444	100.00
-------	-----	--------

Similarly, we can check the frequency of each vignette in our data. The frequencies are listed in Table 2 (only absolute frequencies). The results look good, each vignette occurs with more or less equal frequency.

```
tab vigid // frequency of vignettes
```

Table 2: Frequency of vignettes

Vignette ID	Frequency
1	13
2	13
3	13
4	13
5	13
6	13
7	13
8	13
9	13
10	13
11	13
12	13
13	12
14	12
15	12
16	12
17	12
18	12
19	12
20	12
21	12
22	12
23	12
24	12
25	11
26	11
27	11
28	11
29	11
30	11
31	13
32	13
33	13
34	13
35	13
36	13
Total	444

Next, we can check whether the randomisation of the order of vignettes across respondents has worked. One way of doing this, is to look at the bivariate correlations between the vignette position and the vignette ID and vignette set, respectively.

\*\* Check whether vignette position correlates with vignette ID or vignette sets

`pwcorr vignr vigid, star(0.05)`

`pwcorr deck vignr, star(0.05)`

The correlations are close to zero regarding the vignette IDs ( $r = 0.018$ ) and zero ( $r = 0.000$ ) regarding the vignette decks, both are not statistically significant on conventional levels. We can later control for the vignette position in our regression analyses.

Next, we can check for level balance. To do that, we look at the frequency of each level of the vignette dimensions.

\*\* Frequencies of levels of each dimension

`tab vgnr // levels of vignette gender`

`tab vuetim // levels of vignette unemployment`

`tab vnac // levels of vignette nationality`

Tables 3, 4, and 5 show the results. As we can see, each level occurs with equal frequency, indicating level balance in the data.

Table 3: Frequency of levels: Vignette gender

<b>Vignette: Gender</b>	<b>Frequency</b>
Male	222
Female	222
Total	444

Table 4: Frequency of levels: Vignette unemployment

<b>Vignette: Unemployment</b>	<b>Frequency</b>
No unemployment	148
One year unemployed after graduation	148
One year unemployed at time of applicat	148
Total	444

Table 5: Frequency of levels: Vignette nationality

<b>Vignette: Nationality</b>	<b>Frequency</b>
Luxembourgish native	74
Portuguese foreigner	74
Luxembourgish-Portuguese foreigner	74
French border worker	74
French foreigner	74
German border worker	74
Total	444

In the next step, we examine the bivariate correlations of vignette levels. First, we need to calculate dummy variables (with the values 1 and 0) for each level of our categorical vignette dimensions.

```
// Generate dummies of all levels of categorical vignette dimensions
tab vuetim, gen(vuetim_) // dummies for vignette unemployment
tab vnat, gen(vnat_) // dummies for vignette nationality
```

Next, we calculate the bivariate correlations (i.e., Pearson's correlation coefficient).

```
// Test bivariate correlations
pwcorr vgndr vuetim_1-vuetim_3 vnat_1-vnat_6, star(.05) // Pearson's correlation coefficient
```

Table 6 shows the results. Since we used the full factorial, all vignette dimensions and levels had a correlation of zero and perfect level balance in the experimental set up data. However, it is to be expected that we might observe deviations from perfect orthogonality and level balance in our case due to refusal by certain respondents to participate. This is because, as I have mentioned, the randomisation was done before data collection. Thus, in a small number of respondents, we might observe some correlations. Nevertheless, this does not necessarily result in low data quality. Similar to design where only a vignette sample is used, deviations from perfect orthogonality and level balance are expected. The deviations, however, should be relatively small in the ideal case.

As can be seen in the results, all bivariate correlations of the dimensions' levels are close to zero ( $r < 0.1$ ), and none of them are statistically significant. Of course, we see some correlations between the levels of one vignette dimension. All in all, our data therefore looks good, and we can conclude that the experiment has worked.

Table 6: Bivariate correlations of vignette levels

	vgndr	vuetim_1	vuetim_2	vuetim_3	vnat_1	vnat_2	vnat_3	vnat_4	vnat_5	vnat_6
vgndr	1.00									
vuetim_1	0.02	1.00								
vuetim_2	0.00	-0.50*	1.00							
vuetim_3	-0.02	-0.50*	-0.50*	1.00						
vnat_1	0.00	0.02	0.00	-0.02	1.00					
vnat_2	0.02	-0.01	0.00	0.00	-0.20*	1.00				
vnat_3	-0.01	-0.01	-0.01	0.02	-0.20*	-0.20*	1.00			
vnat_4	-0.01	0.00	0.00	-0.01	-0.20*	-0.20*	-0.20*	1.00		
vnat_5	0.00	0.02	-0.02	0.00	-0.20*	-0.20*	-0.20*	-0.20*	1.00	
vnat_6	0.00	-0.02	0.02	0.00	-0.20*	-0.20*	-0.20*	-0.20*	-0.20*	1.00

\*  $p < 0.05$

Lastly, we can inspect the distribution of vignette evaluations. In this case, the vignette evaluations are the recruiters' hiring intentions, measured on an 11-point scale ranging from 0 (practically zero) to 10 (excellent). So far, we did not consider missing values in the data because we wanted to check whether the experiment has worked. Respondents who did not participate in the present survey were still assigned a vignette set, and it is therefore useful to include these respondents when assessing whether the experiment has worked. However, the vignette evaluations have some missing observations. From the codebook, we know that value -9 indicates non-response. We first code these observations as missing values.

```
replace vigeval_1=. if vigeval_1=-9 // define missing values
```

We can show a frequency table (Table 7) or, respectively, draw a histogram to inspect the distribution of vignette ratings in our sample.

```
tab vigeval_1 // show frequencies
hist vigeval_1, d // draw histogram of vignette evaluations
```

First, we look at the frequency table. We can see that we have 300 valid vignette evaluations (Table 7), and that respondents used the whole answer scale, which is what we want. The histogram shows that the evaluations tend towards higher values on the rating scale and thus is slightly left skewed.

Table 7: Frequency table of vignette ratings

<b>Vignette ratings: Hiring intentions</b>	<b>Frequency</b>
0. Practically zero	4
1	4
2	8
3	12
4	12
5	33
6	57
7	64
8	60
9	29
10. Excellent	17
<b>Total</b>	<b>300</b>

Figure 1: Histogram of vignette ratings (recruiters' hiring intentions)



### Data analysis and interpretation of results

After checking whether the experiment has worked, we can start with the analysis of our data. As we have learned in Part II of this online course, with mixed designs, the data is hierarchically structured. The vignette evaluations of applicant characteristics (i.e. the vignette dimensions) comprise our observations on the level 1 (the “individual level”). These vignette evaluations are nested within respondents (level 2). Thus, the vignette evaluates within one respondent are not independent from

each other, which is, however, the assumption underlying ordinary least squared (OLS) regression. The standard errors of standard OLS regression would therefore be biased and we need to account for this clustering in our data. This can be done by estimated an OLS regression using robust clustered standard errors or by estimating multilevel regression analysis (which are essentially random effects models). Multilevel modelling separates the error term of the regression analysis in two parts (one for level 1 and one for level 2) to account for the clustering (Hox, 2010). Due to the experimental design, the assumption of this type of models that the variables in the model need to be uncorrelated with unobserved characteristics is true for the level 1 variables by design. Therefore, you can either choose OLS regression with robust clustered standard errors or multilevel modelling to analyse your data. We will do both in this course.

Let's start with the OLS regression. We include all our experimental variables in the model and also control for the vignette position.

\*\* Ordinary least squared regression with robust clustered standard errors

```
reg vigeval_1 i.vuetim i.vnat vgnr i.vignr, vce(cluster Respondent_ID)
```

Table 8 shows the results of the OLS regression. We see for each vignette level their effect on the vignette ratings. For example, we see that one year of unemployment at the time of the application significantly reduces recruiters' hiring intentions by 1.8 scale points. We also see that female applicants have on average a slightly higher likelihood of being considered for a job in the catering field, since being a female applicant increases recruiters' hiring intentions (i.e., the vignette ratings) by 0.5 scale points. We also see that the vignette ratings are on average lower for French and German applicants compared to Luxembourgish applicants. Luckily, the vignette order does not seem to have a statistically significant effect on the vignette ratings in most cases.

Table 8: Results of OLS regression predicting vignette ratings (recruiters' hiring intentions)

	Coefficient	Robust clustered Standard errors
<i>Vignette: Unemployment (ref.: No unemployment)</i>		
One year unemployed after graduation	-0.209	(0.203)
One year unemployed at time of application	-1.797***	(0.357)
<i>Vignette: Nationality (ref.: Luxembourgish)</i>		
Portuguese foreigner	-0.486	(0.329)
Luxembourgish-Portuguese foreigner	-0.549	(0.341)
French border worker	-0.876**	(0.293)
French foreigner	-0.694*	(0.279)
German border worker	-1.135**	(0.416)
<i>Vignette: Female (ref.: Male)</i>	0.472*	(0.184)
<i>Vignette position (ref.: Position 1)</i>		
Position 2	0.378	(0.310)
Position 3	0.527	(0.317)
Position 4	0.763*	(0.339)
Position 5	0.148	(0.451)
Position 6	0.424	(0.311)
Constant	7.249***	(0.336)

Robust clustered standard errors in parentheses. N = 300 observations from 50 respondents. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Finally, we will have a look at the results from a linear multilevel regression analysis. In Stata, this can be done using the 'mixed' command.

**\*\* Multilevel analysis**

```
mixed vigeval_1 i.vuetim i.vnat vgnr i.vigr || Respondent_ID: , var // indicating that the data is clustered at the respondent level
```

Table 9 shows the results. As in Table 8, we see for each vignette level their effect on the vignette ratings. As we can see, the results hardly differ between the linear multilevel regression model and the OLS regression. A key difference is that in the output of the multilevel regression, we now see two variances, one for the vignette level and one for the respondent level. As I have explained, this is because multilevel models separate the error term in two parts to account for the clustering in the data and to compute correct standard errors. However, this is mostly of interest if you want to analyse the variance at each level. I would assume that in most cases, you are interested in interpreting the effects of vignette levels on the vignette ratings.

Table 9: Results of linear multilevel regression predicting vignette ratings (recruiters' hiring intentions)

	Coefficient	Standard errors
<i>Vignette: Unemployment (ref.: No unemployment)</i>		
One year unemployed after graduation	-0.209	(0.215)
One year unemployed at time of application	-1.797***	(0.218)
<i>Vignette: Nationality (ref.: Luxembourgish)</i>		
Portuguese foreigner	-0.486	(0.308)
Luxembourgish-Portuguese foreigner	-0.549	(0.309)
French border worker	-0.876**	(0.305)
French foreigner	-0.694*	(0.310)
German border worker	-1.135***	(0.305)
<i>Vignette: Female (ref.: Male)</i>	0.472**	(0.176)
<i>Vignette position (ref.: Position 1)</i>		
Position 2	0.378	(0.305)
Position 3	0.527	(0.321)
Position 4	0.763*	(0.305)
Position 5	0.148	(0.308)
Position 6	0.424	(0.311)
Constant	7.249***	(0.351)
Variance: respondent level	1.133	(0.152)
Variance: vignette level	2.267***	(0.101)

Standard errors in parentheses. N = 300 observations from 50 respondents.

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

This was the last part of this introductory course on factorial survey experiments. Of course, some aspects of the research process could not be discussed in detail within the scope of this course, but I hope that I was able to provide with some good starting points to create your own research projects



around factorial survey experiments. As I said, you can find additional references on some of the issues we have discussed in this course and beyond in the recommended reading list in the supporting material of this course.

Thank you for your attention and have a good day.

## **References**

Auspurg, K. and Hinz, T. (2015) *Factorial Survey Experiments*. Los Angeles: Sage.

Gutfleisch, T. and Samuel, R. (2021) *EDYPOLU – Recruiter Survey. Scientific Use File. Data Documentation*. doi: <https://doi.org/10.17605/OSF.IO/7BP5E>.

Hox, J. J. (2010) *Multilevel Analysis: Techniques and Applications*. 2nd edn. New York: Routledge.