

Not the First Digit!

Using Benford's Law to Detect Fraudulent Scientific Data

Other Research Data

Author(s):

Diekmann, Andreas

Publication date:

2007-05-16

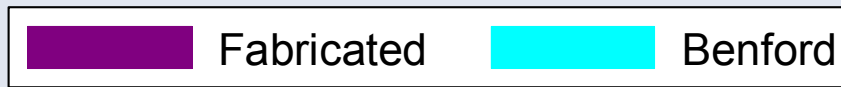
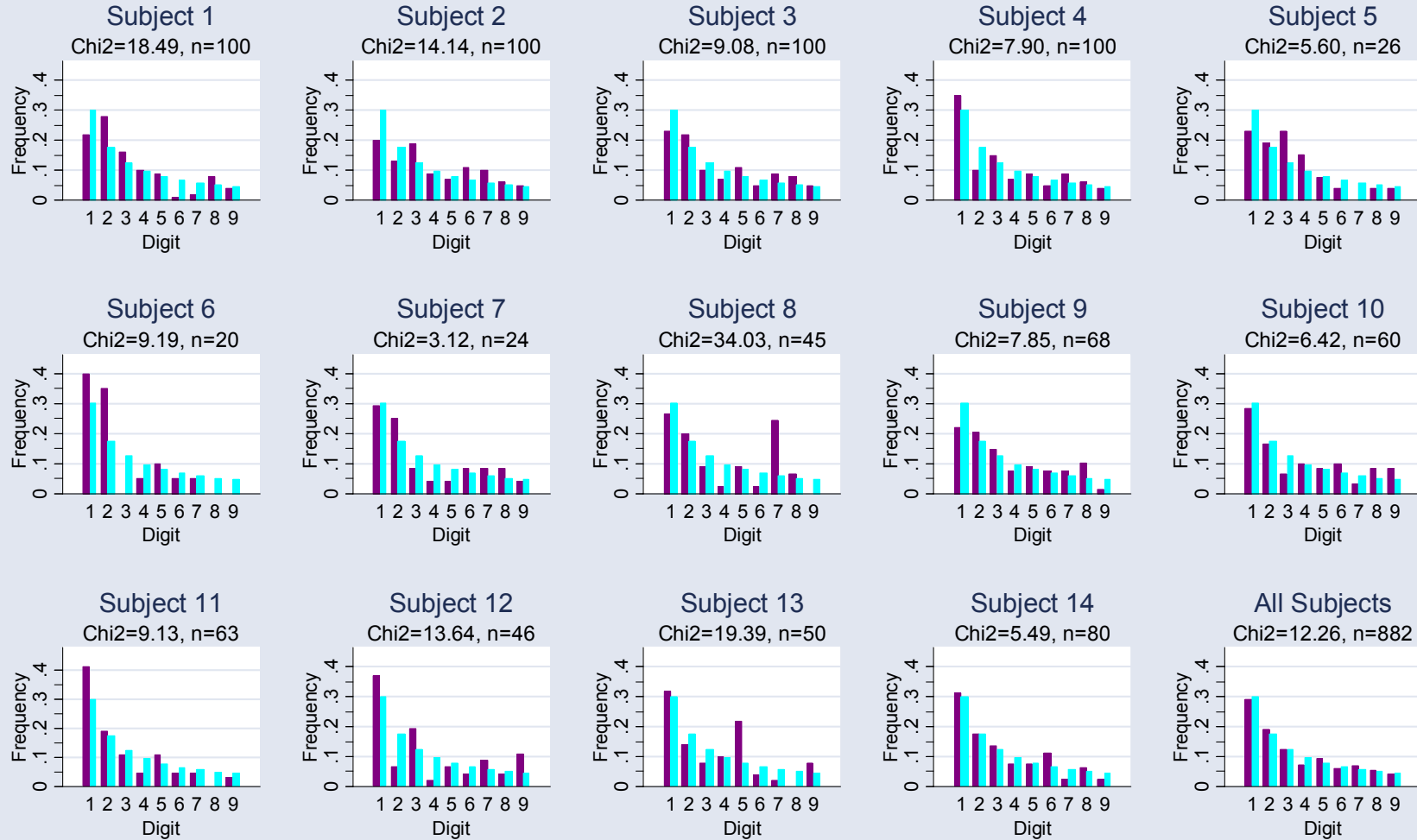
Permanent link:

<https://doi.org/10.3929/ethz-b-000310246>

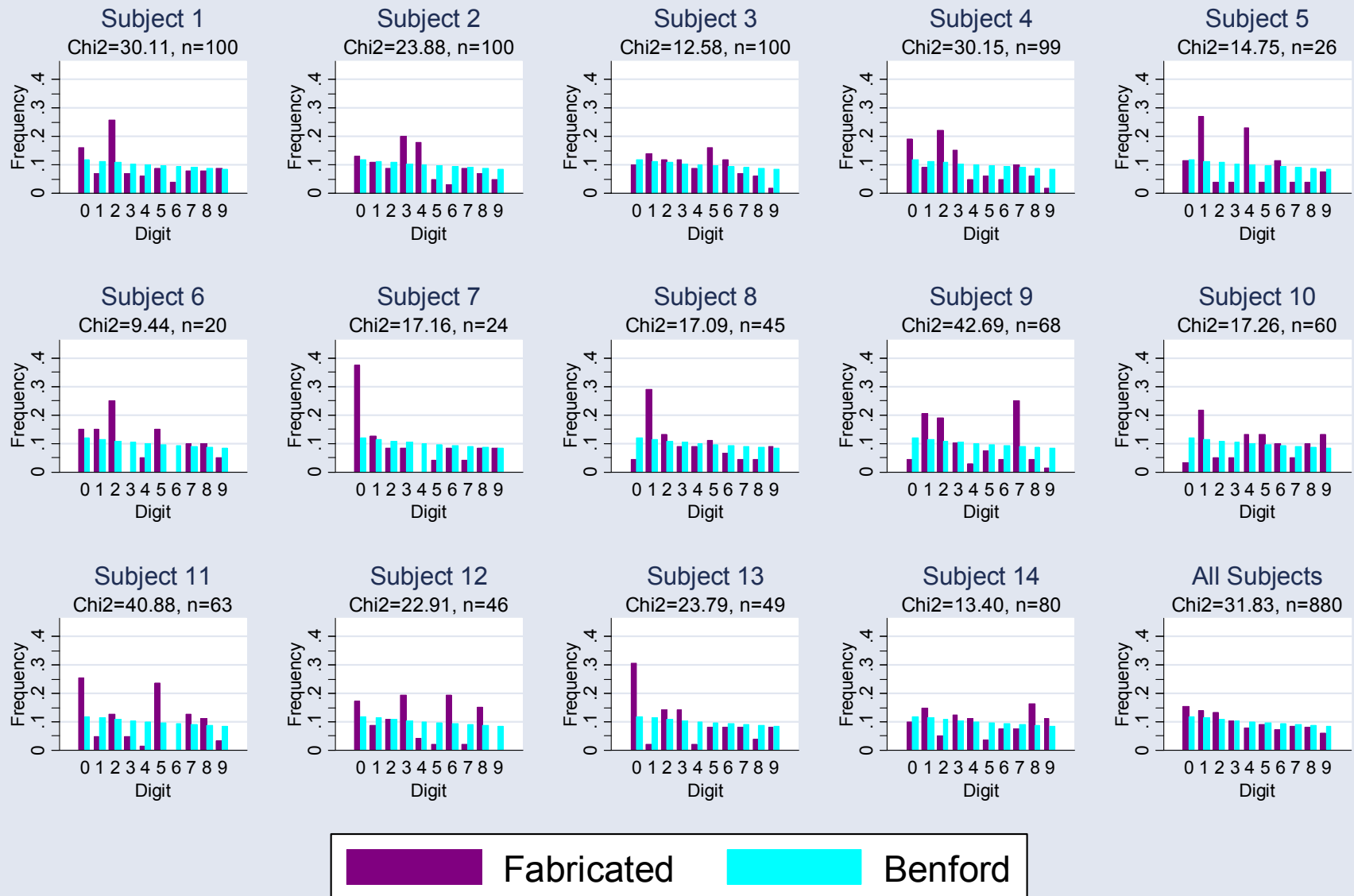
Rights / license:

In Copyright - Non-Commercial Use Permitted

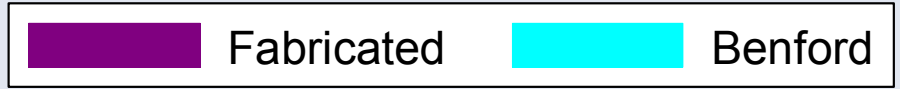
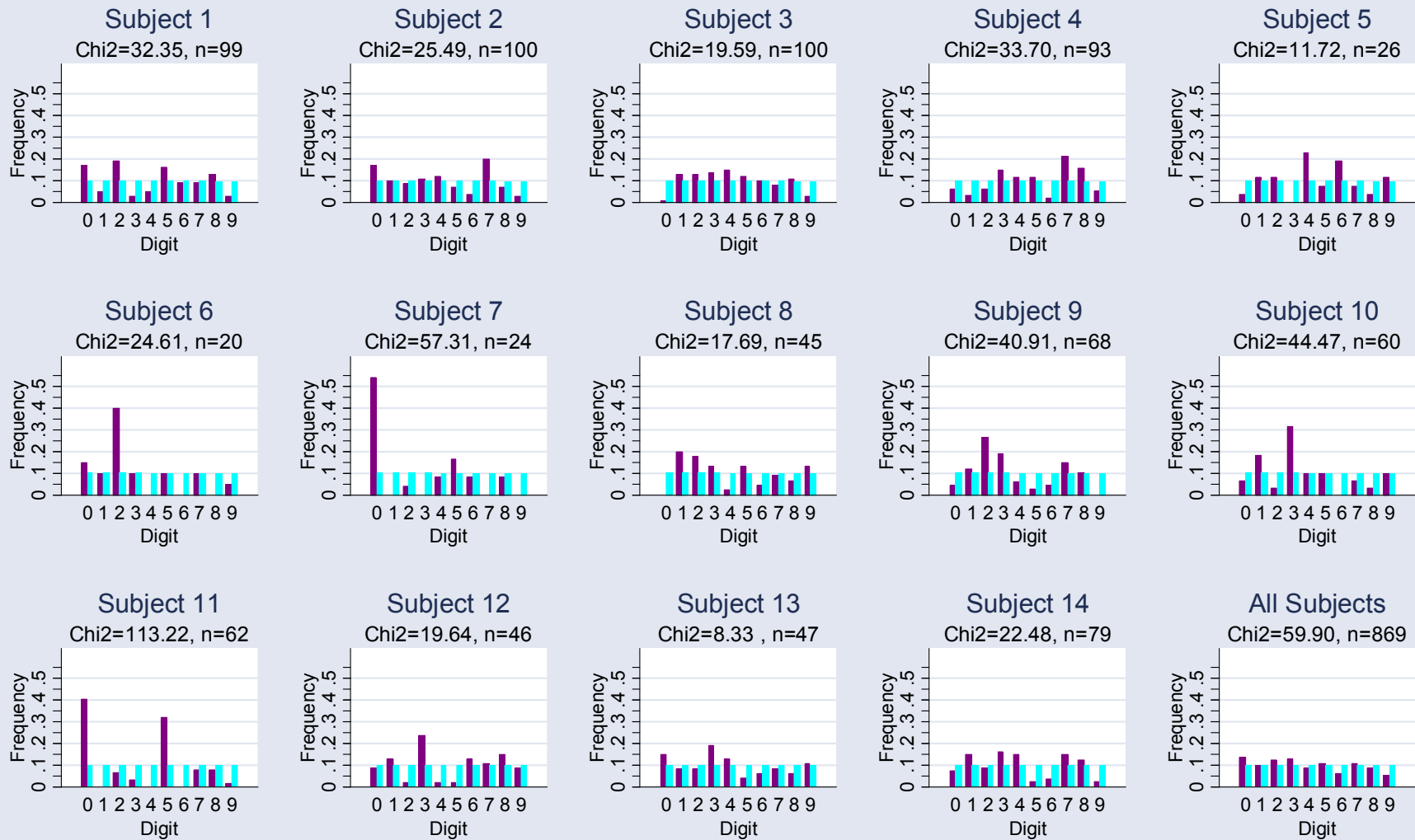
Falsified Regression Coefficients: First Digit



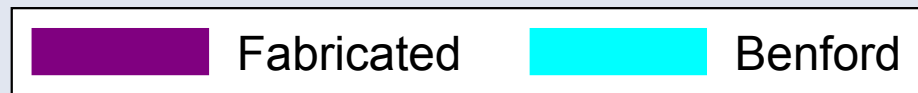
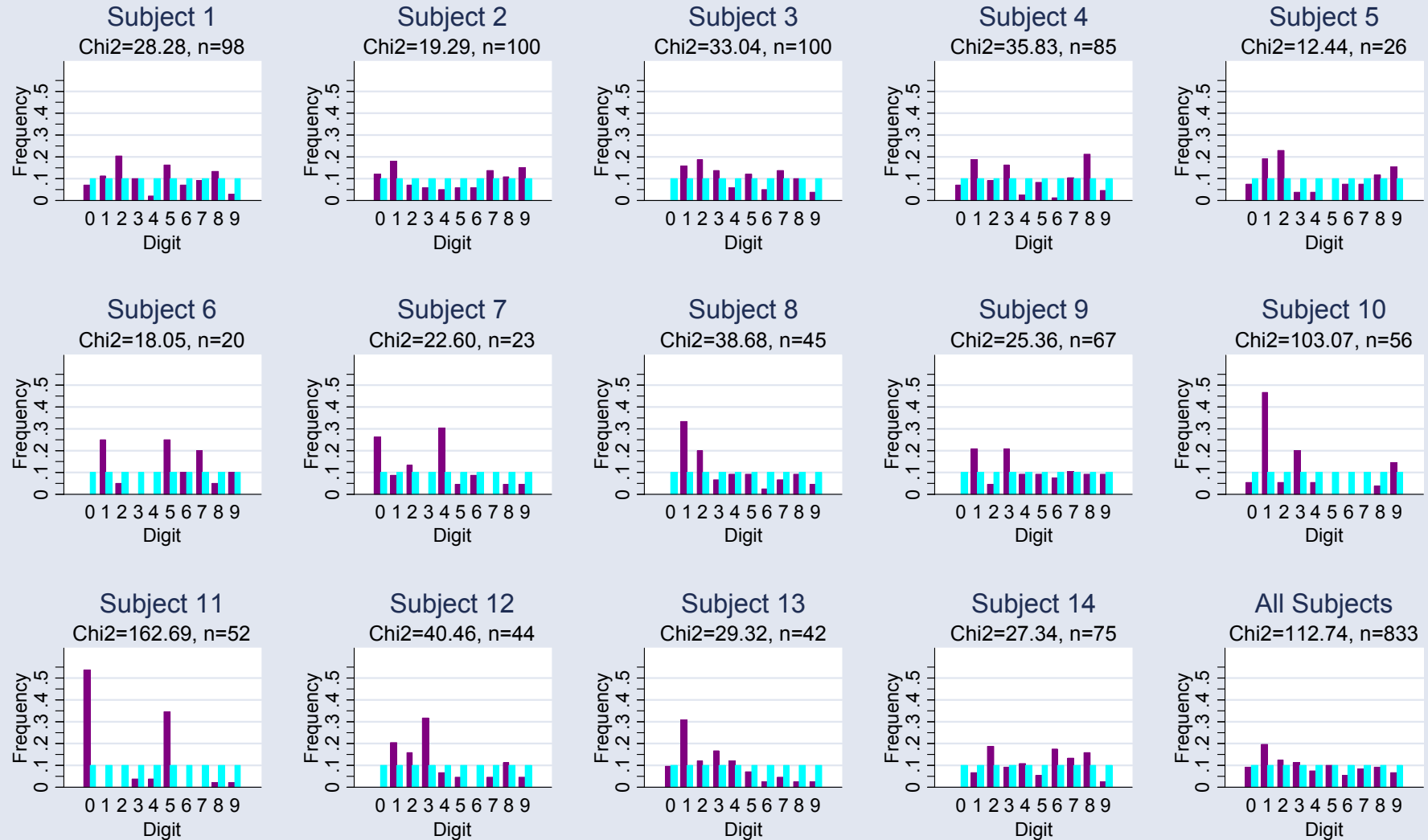
Falsified Regression Coefficients: Second Digit



Falsified Regression Coefficients: Third Digit



Falsified Regression Coefficients: Fourth Digit



Appendix 2: Questionnaire for the Fabrication Experiments 1 and 2*

Name:

Major:

Semester:

Your task is to construct a table of (unstandardized) regression coefficients (for a multiple linear regression) that support the following hypothesis:

**“The higher the unemployment benefits, the longer unemployment will last.”
The values should be plausible and they should seem to you to have been produced by actual data analysis.**

A few more things to consider:

1. Keep in mind that a coefficient can be meaningfully interpreted only for a certain scale. If, for example, unemployment benefits are measured in Swiss francs, then you will have to select different coefficients depending on whether one unit of the unemployment benefits variable is equal to 100 francs or 1,000 francs. You should take the units of all the other variables into account in a similar way. First select a scale (by placing an x next to the option you choose) and then fill in the table with coefficients that you think would produce realistic results.
2. Be sure to put down a standard error as well as a coefficient. As you know, a coefficient with a probability of error of $\alpha = .05$ is significant if the value of the coefficient is more than twice as large as the value of the standard error. **Please denote significant coefficients with an asterix.**
3. As you also know, the regression coefficient for a dichotomous- 0/1 coded- variable denotes the amount by which the dependent variable changes when the independent variable is equal to 1 versus when it is equal to 0. For example, the coefficient for a variable that takes on the values of 1 for a city and 2 for a town or a rural area might be - 3.642. If the length of the unemployment spell is measured in weeks, then the length of the unemployment spell in a city is 3.642 weeks shorter in a city than in a town or a rural area.
4. Be sure to note the coefficients and standard errors to **four digits, not including the zeroes before the first digit.** For example, the numbers 0.001438 or 91.24 would both fulfill this condition.

*A slightly modified version of this questionnaire was used in experiment 3.

So, let's get started:

First, select a scale for the length of the unemployment spell:

Days:

Weeks:

Months:

Table: Determinants of the length of unemployment: Estimates from a multiple regression (standard errors in parentheses)

Independent Variables	Regression Coefficients (Standard Errors)
Unemployment benefits In units of CHF 1 (.....)
CHF 100	
CHF 1000	
Years of education (.....)
Years of job experience (.....)
Mother's years of education (.....)
Father's years of education (.....)
Sex (Female = 1) (.....)
Marital status (married = 1 , otherwise 0) (.....)
Last position was in the service sector (service sector = 1, otherwise 0) (.....)

Monthly income for the last job held, in units of CHF 1 CHF 100 CHF 1000 (.....)
Distance between residence and place of business in units of: 1 km 10 km (.....)
Adjusted multiple R-squared
Number of cases (N)