



# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES

## *Cheetah Basic Tutorial*



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 824091

# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES

## Outline

1	Introduction .....	2
2	Database description .....	2
3	Why study fast-growing firms.....	2
4	Basic user guide.....	3
4.1	Opening the file .....	3
4.2	Descriptive statistics.....	5
5	References .....	8

## 1 Introduction

The Cheetah dataset contains geographical, industry, accounting and ownership information on three cohorts of medium-sized firms that experienced sales or employment fast growth rates in the periods 2008-2011, 2009-2012 and 2010-2013. These fast growing medium-sized firms (FGMFs) are located in 30 European countries (EU-28, Norway and Switzerland) plus Israel. Cheetah covers an overall number of 42,369 firms.

Cheetah has been developed in the context of the WP20 of the project RISIS - Research Infrastructure for Research and Innovation Policy Studies, funded by the European Commission under the Seventh Framework Program. The aim of Cheetah is to cover the long-term economic performance of FGMFs, as one of the main pillars of the European industrial and technological system.

The legal name of the operating organization is POLITECNICO DI MILANO, Department of Management, Economics and Industrial Engineering, located in VIA LAMBRUSCHINI 4/B, MILANO, 20156, Italy represented by Alessandro Perego, Head of Department (or his authorized representative).

## 2 Database description

Firms included in the database must abide all of the following criteria:

1. Firms are established in Europe (28 EU countries, CH, NO) + Israel
2. Medium-sized criterion at the beginning of each observation period (2008, 2009, 2010):
  - 50 < employees < 5,000
  - turnover < €1.5 billion OR balance sheet total < €2 billion
3. Firms experienced fast-growth in at least one of the observation periods (2008-2011, 2009-2012, 2010-2013)

The database includes 42,369 FGMFs. Accounting and ownership information has been collected for each firm. For the first cohort of FGMFs (2008-2011) we identified 17,438 firms, for the second cohort (2009-2012) 24,993 firms, and for the third cohort (2010-2013) 16,085 firms. Indeed, the sum of firms in the three cohorts is higher than the number of firms in the database. This is because some firms may have experienced high growth in more than one observation period.

## 3 Why study fast-growing firms

Fast-growing firms has recently received considerable attention by academic scholars and policymakers (Coad et al. 2014).

Fast-growing firms, sometimes called in the literature Gazelles, have been studied for their role in employment, job-creation and economic growth.

In particular, a few rapidly growing firms generate a disproportionately large share of all new net jobs compared with non-high-growth firms (see (Henrekson and Johansson, 2010) for a review). On average, they are younger in age then other firms, although not necessarily smaller or larger.

Within the extant literature on fast growing firms, one category of firms received somewhat less attention: medium-sized firms. The aim of the Cheetah dataset is to map fast-growing mid-sized firms (FGMFs) in Europe.



## 4 Basic user guide

### 4.1 Opening the file

Cheetah DB is available in STATA format. The following example uses STATA 14 to illustrate the basic operations that can be run using Cheetah database.

Cheetah can be opened in Stata with the command *use*, specifying the file path. When using Stata, the commands can be saved in a Do-file, as shown in Figure 1. To view and edit data, Data editor can be accessed through the command *browse*, or by clicking the Data Editor icon in the manual menu as shown in Figure 2, or by selecting Data > Data Editor > Data Editor (Browse) from the menu as shown in Figure 3.

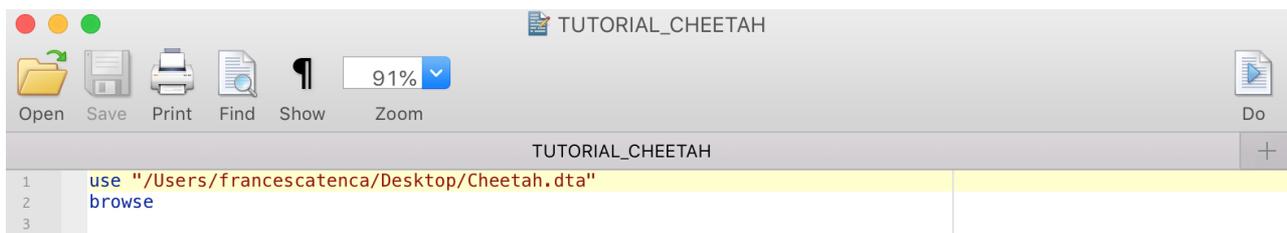


Figure 1: Use and browse commands

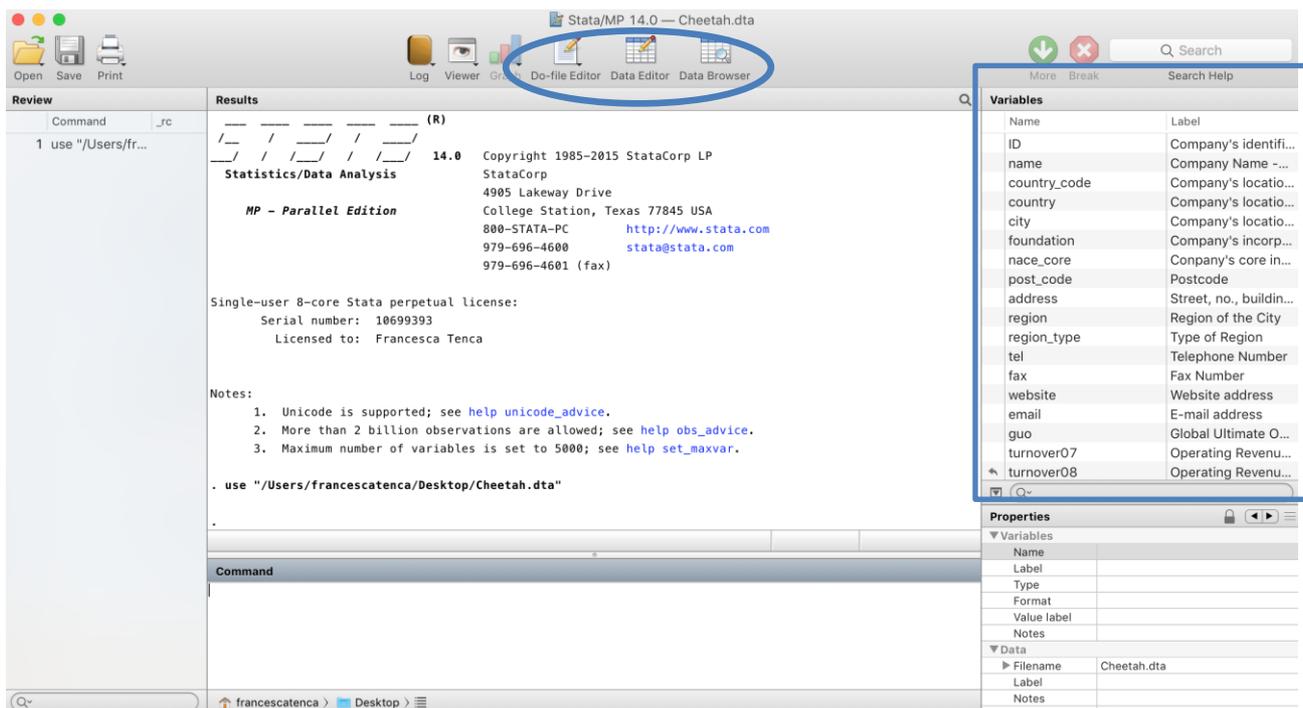


Figure 2: Stata menu and variables

# RISIS



RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES

ID	name
1	FGMFAT00001 ZUNO BANK /
2	FGMFAT00002 IMMORENT-BANK GME
3	FGMFAT00003 KOMMUNALKREDIT AUSTRIA /
4	FGMFAT00004 WIENER PRIVATBANK S
5	FGMFAT00005 POSOJILNICA-BANK BOROVLJE -CELOVEC, REGISTRIRANA ZADRUGA Z OMEJENIM JAMSTVO
6	FGMFAT00006 RAIFFEISEN REGIONALBANK MOEDLING EGE
7	FGMFAT00007 GENERALI HOLDING VIENNA /
8	FGMFAT00008 AUSTRIAN ANADI BANK /
9	FGMFAT00009 RAIFFEISENBANK REGION WALDVIERTEL MITTE EGE
10	FGMFAT00010 ERSTE BANK DER OESTERREICHISCHEN SPARKASSEN /
11	FGMFAT00011 WESTERN UNION INTERNATIONAL BANK GME
12	FGMFAT00012 RAIFFEISEN-BEZIRKSBAK KLAGENFURT, REG. GENOSSENSCHAFT M
13	FGMFAT00013 KATHREIN PRIVATBANK AKTIENGESELLSCHAFT
14	FGMFAT00014 RAIFFEISEN CENTROBANK /
15	FGMFAT00015 RAIFFEISENLANDESBANK VLBG. WAREN- U. REVISIONSVERBAND REG. GENOSSENSCHAFT ME
16	FGMFAT00016 OESTERREICHISCHE NATIONAL BANK /
17	FGMFAT00017 HYPO-BANK BURGENLAND AKTIENGESELLSCHAFT
18	FGMFAT00018 RAIFFEISENLANDESBANK KAERNTEN - RECHENZENTRUM UND REVISIONSVERBAND RGME
19	FGMFAT00019 POSOJILNICA BANK EGE
20	FGMFAT00020 VOLKSBANKEN HOLDING REGGENME
21	FGMFAT00021 VOLKSBANK KAERNTEN SUED E.GEN
22	FGMFAT00022 MUENZE OESTERREICH AKTIENGESELLSCHAFT
23	FGMFAT00023 RAIFFEISEN BANK INTERNATIONAL /
24	FGMFAT00024 RAIFFEISEN-LANDESBANK STEIERMARK /
25	FGMFAT00025 BKS BANK /
26	FGMFAT00026 HETA ASSET RESOLUTION /

Name	Label
<input checked="" type="checkbox"/> ID	Company's identification c...
<input checked="" type="checkbox"/> name	Company Name - updated...
<input checked="" type="checkbox"/> country_code	Company's location - coun...
<input checked="" type="checkbox"/> country	Company's location - coun...
<input checked="" type="checkbox"/> city	Company's location - city...
<input checked="" type="checkbox"/> foundation	Company's incorporation year
<input checked="" type="checkbox"/> nace_core	Company's core industry code
<input checked="" type="checkbox"/> post_code	Postcode
<input checked="" type="checkbox"/> address	Street, no., building etc, lin...
<input checked="" type="checkbox"/> region	Region of the City
<input checked="" type="checkbox"/> region_type	Type of Region
<input checked="" type="checkbox"/> tel	Telephone Number

Vars: 257 Order: Dataset Obs: 42.369 Length: 11 Filter: Off

Figure 2: Data Editor Browse mode

After opening the Cheetah database with STATA, we can see on the right side a panel (Figure 2 and Figure 3) with a summary of the variables available in the database. In particular, the variable have a Name, which can be used to refer to them in the command line, and a Label containing a short human-readable description of the variable.

To actual edit the variables in the DB, we have to open the Data Editor window in the Edit mode by clicking on Edit on the up-right corner (see Figure 3).

By clicking with the left mouse button on a variable, that variable will be selected, and additional information will be shown in the Properties panel. Figure 4 shows an example, where the variable `country_code` has been selected.

Variables	
Name	country_code
Label	Company's location - country code
Type	str16
Format	%16s
Value label	
Notes	

Data	
Filename	FGMF_official_geocoded.dta
Label	
Notes	
Variables	257
Observations	42,369
Size	109.90M
Memory	160M
Sorted by	ID

Figure 4: Properties panel

Additional information include the type of the variable (in this case, a string) and the format. Moreover, in the Data subpanel, additional information for the whole dataset is available, like the total number of variables, the number of observations, and the dimension of the database in memory.

## 4.2 Descriptive statistics

STATA has a user-friendly interface that allows us to run different types of descriptive statistics. To compute the mean of a variable, we can click on Statistics -> Summaries, tables and tests -> Summary and descriptive statistics -> Means

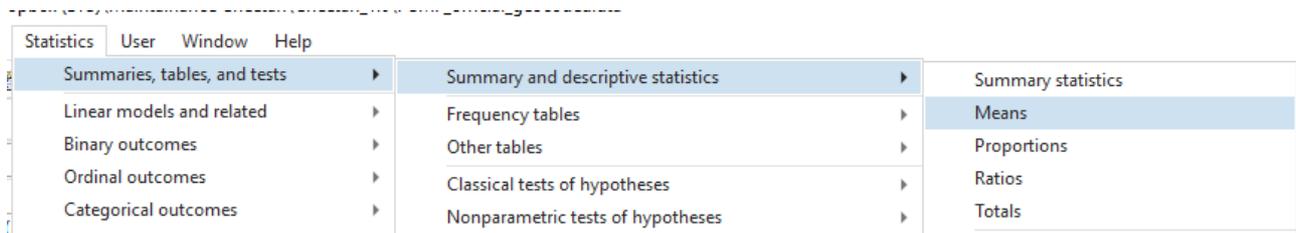


Figure 5: Means command

A user-friendly dialog-box opens up (Figure 6). Clicking on the arrow button will open a text box in which we can select the variables we want the mean of. Lets calculate the mean of the variables turnover07, turnover08 and turnover09.

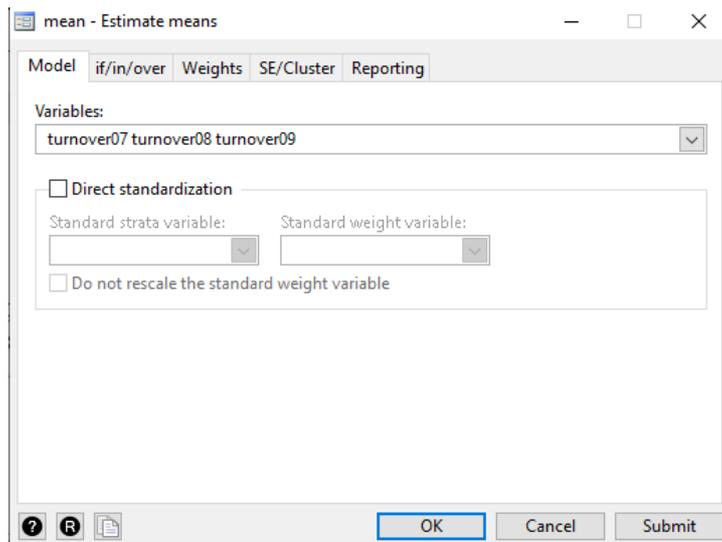


Figure 6: Means dialog-box

By clicking on OK button, results are shown in the main windows (Figure 7).

```
. mean turnover07 turnover08 turnover09
```

	Mean	Std. Err.	[95% Conf. Interval]	
turnover07	54694.61	1418.491	51914.29	57474.93
turnover08	57655.06	1647.71	54425.46	60884.66
turnover09	53755.35	1343.785	51121.46	56389.25

Figure 7: Means for different turnover variables

STATA allow us to plot variables too. To plot an histogram, we can click on Graphics -> Histogram. A dialog-box pop up. In the Variable text box, we can type num\_emp08 to obtain an histogram for the number of employees in 2008. We can define the number of bins in the histogram by checking the check box and defining, for example, 10 as the number of bins (Figure 8).

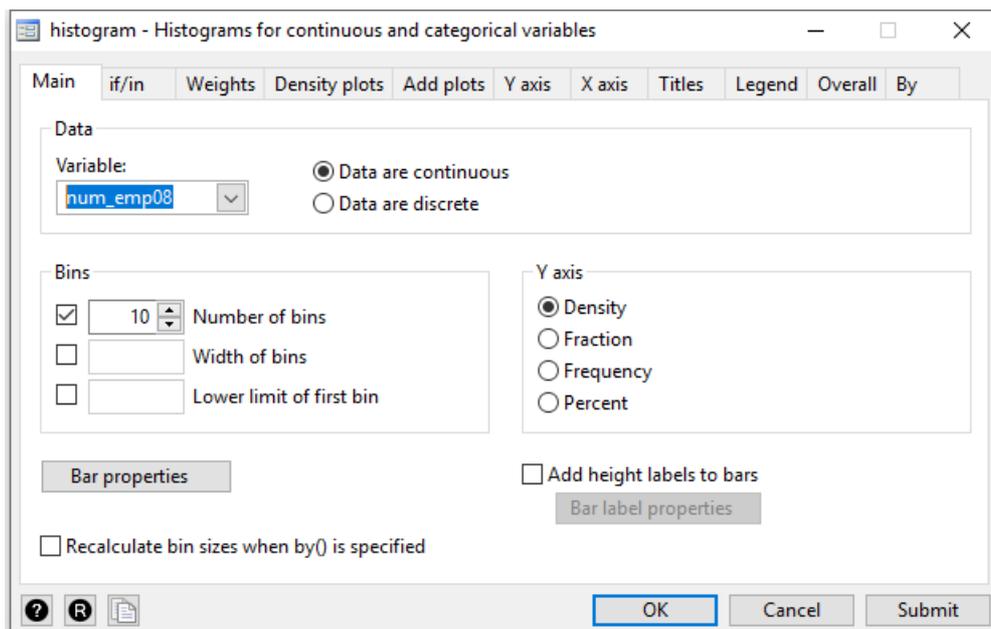
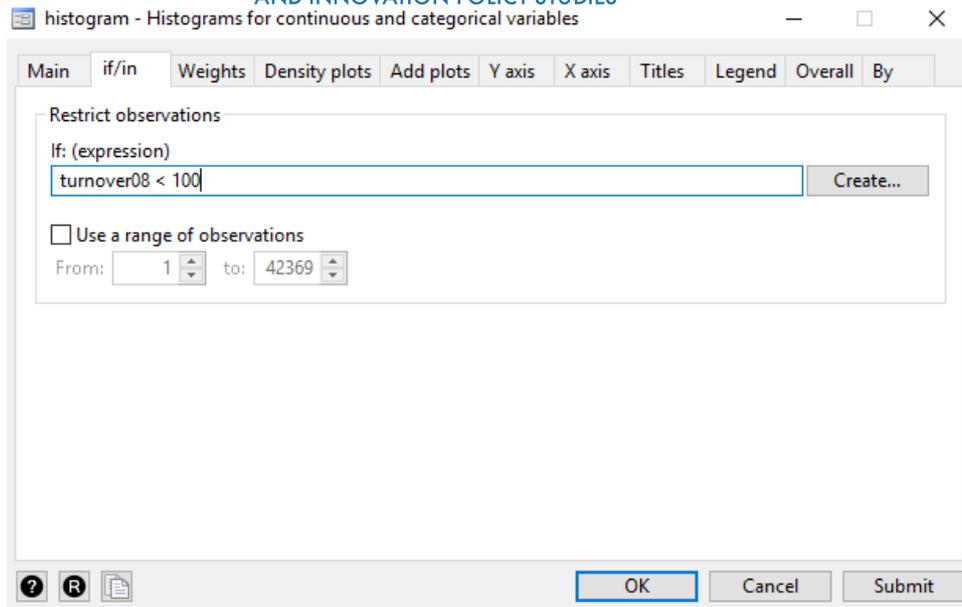


Figure 8: Histogram dialog-box

We can restrict the observations to consider for the plot by opening the if/in tab and specify the restrictions under which we want to plot the observations. For example, to obtain an histogram only for the firms with a turnover less than 100k, we can type turnover08 < 100 in the if: (expression) text box (Figure 9).



By clicking on OK we obtain the requested histogram, as shown in Figure 10.

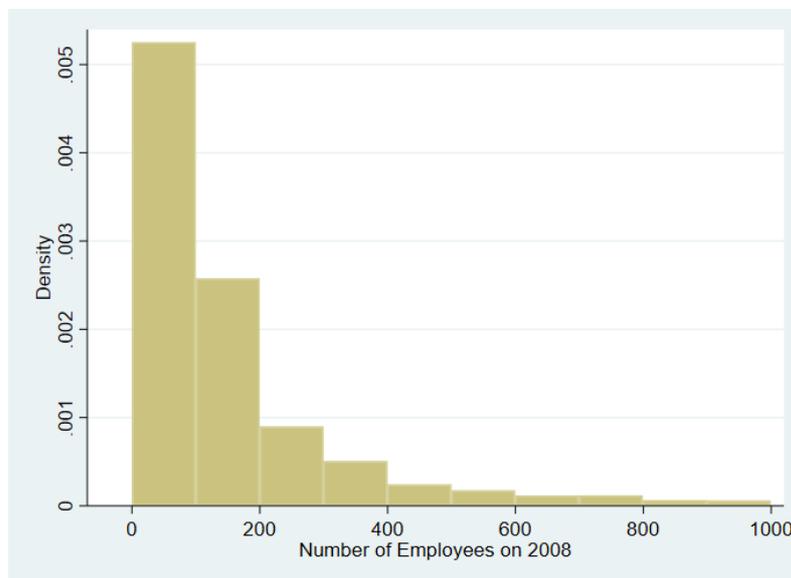


Figure 10: Histogram of number of employees in 2008

## 5 References

- Coad, A., S. O. Daunfeldt, W. Holzl, D. Johansson, and P. Nightingale. 2014. High-Growth Firms: Introduction to the Special Section. *Industrial and Corporate Change* 23(1):91112.
- Henrekson, Magnus and Dan Johansson. 2010. Gazelles as Job Creators: A Survey and Interpretation of the Evidence. *Small Business Economics* 35(2):22744.