

# Kauffman Firm Survey (KFS)

## *Key Personnel*

Alicia Robb  
Principal Investigator  
Ewing Marion Kauffman Foundation  
[arobb@kauffman.org](mailto:arobb@kauffman.org)

## *Funding*

Ewing Marion Kauffman Foundation

## *URLs*

[www1.kauffman.org/kfs](http://www1.kauffman.org/kfs)

## *Type of Data*

Non-Governmental

## *Availability of Data Set*

Public-use download (free).  
Restricted access (sponsorship available).  
RDC version in Census.

## *Activity Measured*

2004-2011 Longitudinal

## *Data Collected*

2005-2012

## *Number of Observations*

4,928

## *Date of Overview Update*

7 April 2016

KAUFFMAN  
Compendium on  
Entrepreneurship  
and Innovation  
Data

## OVERVIEW

The Kauffman Firm Survey (KFS) is a panel study of 4,928 businesses founded in 2004 and tracked over their early years of operation. The survey focuses on the nature of new business formation activity; characteristics of the strategy, offerings, and employment patterns of new businesses; the nature of the financial and organizational arrangements of these businesses; and the characteristics of their founders.

For the Kauffman Firm Survey (KFS), the target population was all new businesses that were started in the 2004 calendar year in the United States (the 50 states plus the District of Columbia). This population excludes any branch or subsidiary owned by an existing business or a business inherited from someone else. The issue that arose immediately with this target definition is the meaning of *started*. For the study population, a business started in 2004 was defined as a new, independent business that was created by a single person or a team of people; the purchase of an existing business; or the purchase of a franchise.

There is rich detail about each firm and its owner(s). The information is collected each year so that changes can be tracked over time. These are some of the types of variables available on the KFS dataset:

- Firm characteristics: Industry, Legal Form, # of Owners, # of Employees (PT/FT), Types of Customers, Location
- Firm strategy and innovation: Product/Service Offerings, Intellectual Property, Licensing In and Licensing Out, R&D
- Detailed financial information: Equity & Debt Financing, Income Statement Info (Revenue, Expenses, Profits), Balance Sheet Info (Assets, Liabilities, Equity)
- Employees: Types of Benefits Offered, Task/Work Structure
- Owner characteristics and work behaviors (information on up to 10 owners): Education, Age, Race, Ethnicity, Gender, Citizenship, Immigrant Status, Hours Worked, Previous Years of Work Experience, Previous Start-up Experience (same/different industry as this firm)

## PROCEDURES FOR ACCESSING THE DATA SET

There are three different versions of the KFS dataset available to researchers: a public-use microdata set and then a more detailed confidential microdata set, which can either be accessed through the [NORC data enclave](#) or the [Census Research Data Centers \(RDCs\)](#). The main differences between the two types of files are that the confidential dataset has five-digit industry (NAICS) codes, geographical detail such as zip code, state, and metropolitan statistical area (MSA), and various continuous variables that are not on the public-use file. The public-use file has two-digit industry codes and Census region.

### **Public-use Microdata**

Available for free with registration on the web. Identifying features such as geographical location are omitted.

**Data Enclave Version**

Available via remote access and [by application only](#). More identifying variables available along with other matched variables such as credit scores. Teams of geographically distributed researchers can access and share code, results, and work in a collaborative environment. Other datasets can be linked to the KFS.

**RDC Version**

Available at the [Center for Economic Studies](#) or any of the [Census RDCs](#). Application must be submitted to the Center for Economic Studies. Data can be linked to other [Census datasets](#).

**METHODOLOGY**

The panel of businesses was created by using a random sample from [Dun & Bradstreet's \(D&B\)](#) database list of new businesses started in 2004, which totaled roughly two hundred fifty-thousand such businesses. The KFS oversampled these businesses based on the intensity of research and development employment in the businesses' primary industries. The KFS sought to create a panel that included new businesses founded by a person or team of people, purchases of existing businesses by a new ownership team, and purchases of franchises. To this end, the KFS excluded D&B records for businesses that were wholly owned subsidiaries of existing businesses, businesses inherited from someone else, and not-for-profit organizations. In response to the Kauffman Foundation's interest in understanding the dynamics of high-technology, medium-technology, and woman-owned businesses, the KFS is a stratified sample based on industrial technology level (High-Tech, Medium-Tech, and Non-Tech) and gender, which oversamples businesses in high- and medium-tech industries.

For more information, visit [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1024045](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1024045).

**SURVEY INSTRUMENTS**

The baseline annotated questionnaire, first, second, third, and fourth follow-up annotated questionnaires are available at <http://www1.kauffman.org/kfs/KFSWiki/Annotated-Questionnaires.aspx>.

**BIBLIOGRAPHY**

There is a database of research using KFS data, visit the [KFS bibliography page](#) or the [KFS eJournal on SSRN](#).