1. **О проекте и базе данных «Российские предприятия в глобальной экономике» (RUFIGE)**

База данных создана в рамках проекта «Российские предприятия в глобальной экономике» (<http://iims.hse.ru/rfge/>). Проект финансировался Программой фундаментальных исследований НИУ ВШЭ, выполнен Институтом анализа предприятий и рынков НИУ ВШЭ в 2014 г. Руководители проекта Б.Кузнецов и А.Яковлев.

Обследовано **2092** предприятия обрабатывающей промышленности России. Опрос проводился компанией ГФК-Русь методом личных интервью по стандартизованной анкете, состоящей из 111 вопросов (<http://iims.hse.ru/rfge/question>). Объекты обследования – предприятия обрабатывающей промышленности с числом занятых свыше 10 человек. Случайная стратифицированная выборка фирм репрезентативна в разрезе секторов и размерных групп предприятий, но не репрезентативна в разрезе регионов (подробнее о выборке см. «Методологические комментарии»).

Эмпирическая база данных опроса позволяет решать разнообразные исследовательские и образовательные задачи. Вопросы анкеты сфокусированы, прежде всего, на оценке конкурентных преимущества российских фирм обрабатывающей промышленности на внутреннем и глобальной рынках. Более 60% вопросов анкеты RUFIGE идентичны или сопоставимы с вопросами анкеты аналогичного европейского опроса предприятий обрабатывающей промышленности в 7 странах ЕС (проект European Firms in a Global Economy - EFIGE, [www.efige.org](http://www.efige.org)), что обеспечивает возможность проведения комплексных сравнительных исследований конкурентоспособности и поведения предприятий. Данные позволяют также проводить исследования в области финансов фирмы, ее организации и управления, инноваций, внешнеторговой активности, а также изучать инвестиционную среду и ее влияние на деятельность предприятий.

Часть вопросов анкеты обследования идентична предыдущим мониторинговым обследованиям крупных и средних предприятий обрабатывающей промышленности России, осуществленным ИАПР НИУ ВШЭ в 2005 и 2009 гг. Следует подчеркнуть, что панельная часть всех трех раундов составляет менее 100 предприятий, поэтому при сравнении результатов опросов 2005, 2009 и 2014 годов следует подходить с осторожностью, учитывать различия в методологии и ограниченную сопоставимость вопросов.

В период с апреля 2015 по апрель 2016 доступ к эмпирической базе опроса RUFIGE получают только исследовательские группы НИУ ВШЭ, представленные в Координационном совете проекта (см. <http://iims.hse.ru/rfge/kord>), через своих представителей в Координационном совете. База предоставляется по запросу и при условии подписания соглашения о конфиденциальности (форма соглашения содержится в файле Confidentiality\_agreement в разделе «Документы»).

**Методологические комментарии**

* 1. Методология формирования выборки

Информационная база представляет собою результаты опроса российских предприятий обрабатывающей промышленности (ОКВЭД коды 15–35) с численностью занятых не менее 10 человек. Опрос проводился по сформированной адресной базе предприятий, включенных в выборочную совокупность, методом прямого интервьюирования.

Для целей обследования была сформирован типическая выборка, полученная путем стратификация Генеральной совокупности (ГС) по состоянию на начало 2013 года, сформированной на основе информационных ресурсов (баз данных AMADEUS и CПАРК) по двум квалификационным признакам объектов наблюдения: отрасль и размер (численность занятых).

Для определения необходимого числа обследуемых единиц из каждой страты ГС применен метод оптимального размещения (нейманово размещение). Оптимальное размещение учитывает не только численность групп, но и степень вариации в них количественного признака.

В силу доминирования в ГС (по числу) малых предприятий и высокой вариации численности работников на них, пропорциональная стратам ГС выборка при ограничениях на общее число опрашиваемых фирм не позволяла получить достаточное для надежного статистического анализа количество наблюдений в группах средних и крупных предприятий (КСП). Поэтому структура выборки была смещена в пользу страт КСП, которые были «перепредставлены» примерно вдвое. Аналогичная процедура реализована и в европейском проекте EFIGE (http://www.bruegel.org/publications/publication-detail/publication/753-the-eu-efigebruegel-unicredit-dataset/) . Для приведения структуры выборки в соответствие со структурой ГС при анализе данных опроса необходимо использовать процедуру «перевзвешивания» (подробнее см. ниже в подразделе «Веса: корректировка результатов выборочного наблюдения с учетом отклонения структуры фактической выборки от планируемой»).

* 1. Результаты опроса

Опрос по заказу Института анализа предприятий и рынков НИУ ВШЭ был проведен Международным институтом маркетинговых и социальных исследований «ГФК-Русь» в период с мая по октябрь 2014 года в 60 регионах (субъектах федерации) России методом личного интервьюирования руководителей предприятий. Респондентами выступали следующие категории руководителей:

* генеральный директор,
* исполнительный директор,
* заместитель директора по экономике/финансам,
* директор по экономике,
* финансовый директор (но не являющийся при этом главным бухгалтером),
* коммерческий директор.

Замена топ-менеджеров на менеджеров среднего звена (например, на руководителей отделов) или иные категории сотрудников компании не допускалась.

Всего в обследовании приняло участие **2092** предприятия, которые и вошли в базу данных.

Следует отметить, что особенностью опросов, реализуемых без участия официальных органов статистики, является добровольность участия в них. Другими словами, гарантировать опрос именно тех предприятий, которые были выбраны для вхождения в выборку, невозможно – неизбежно высокое число отказов в проведении интервью. Поэтому, помимо основного списка, был сформирован список предприятий, максимально близких по основным параметрам на замену в случае невозможности опроса предприятий из основного списка. В ряде случаев даже этих списков «запасных» предприятий оказалось недостаточно, и замены проводились региональными менеджерами «ГФК-Русь» на основе информации из региональных баз данных предприятий, т.е. часть из опрошенных предприятий изначально не входила в адресную базу и была включена в список респондентов непосредственно в ходе полевых работ. Проверка показала, что «добранные» таким образом предприятия не всегда соответствовали квалификационным признакам и, хотя эти наблюдения могут в ряде случаев использоваться в анализе, они не могут быть учтены при оценке репрезентативности выборки. В связи с этим описание выборки ниже базируется на данных опроса **1950** предприятий (основная база), которые отвечают критериям формирования выборки и не вызывают сомнения в части их соответствия параметрам ГС. Для возможности «фильтрации» данных в базе содержится переменная «Flag», принимающая значение 1, если предприятие входит в основную базу и 0 в противном случае.

Основная выборка с вероятностью P = 0,954 обеспечила представительность данных как в целом по российской обрабатывающей промышленности, так и в разрезе большинства ee подотраслей. Исключение составили лишь две подотрасли: производство транспортных средств и оборудования и деревообработка и целлюлозно-бумажная промышленность, где из-за малого объема фактически обследованного числа предприятий относительная ошибка оказалась достаточно высокой.

**Подчеркнем, что задача обеспечения репрезентативности по отдельным субъектам Российской Федерации в данном обследовании не ставилась и в региональном разрезе выборка не является репрезентативной.**

* 1. Структура выборки

Фактически сложившееся распределение предприятий, включенных в выборочную совокупность, по двум ключевым стратифицирующим признакам, таким как вид экономической деятельности и размерная группа представлено на Рисунках 1 и 2 и в Таблицах 1 и 2.

**Рисунок 1. Распределение предприятий выборки по видам экономической деятельности, % (объем выборки 1 950 предприятий)**



**Таблица 1. Распределение предприятий выборки по видам экономической деятельности**

| Вид экономической деятельности | Число опрошенных компаний | |
| --- | --- | --- |
| ед. | % к итогу |
| *А* | *1* | *2* |
| Производство пищевых продуктов, включая производство напитков и табака (DA) | 444 | 22.8 |
| Текстильное и швейное производство + Производство кожи, изделий из кожи и производство обуви (DB+DC) | 175 | 9.0 |
| Обработка древесины и производство изделий из дерева + Целлюлозно-бумажное производство; Издательская деятельность (DD+DE) | 212 | 10.9 |
| Производство кокса, нефтепродуктов и ядерных материалов + Химическое производство (DF+DG) | 103 | 5.3 |
| Производство резиновых и пластмассовых изделий (DH) | 108 | 5.5 |
| Производство прочих неметаллических минеральных продуктов (DI) | 166 | 8.5 |
| Металлургическое производство и производство готовых металлических изделий (DJ) | 237 | 12.2 |
| Производство машин и оборудования (DK) | 265 | 13.6 |
| Производство электрооборудования, электронного и оптического оборудования (DL) | 134 | 6.9 |
| Производство транспортных средств и оборудования (DM) | 106 | 5.4 |
| **ИТОГО** | **1 950** | **100** |

**Рисунок 2. Структура выборки по размерным группам предприятий, % (N= 1 950)**



**Таблица 2. Структура выборки по размерным группам предприятий (N= 1 950)**

| Число занятых, чел. | Число опрошенных компаний | |
| --- | --- | --- |
| ед. | % к итогу |
| *А* | *1* | *2* |
| от 10 до 19 | 434 | 22.3 |
| от 20 до 49 | 491 | 25.2 |
| от 50 до 100 | 202 | 10.4 |
| от 101 до 249 | 303 | 15.5 |
| от 250 до 499 | 139 | 7.1 |
| свыше 500 | 381 | 19.5 |
| **ИТОГО** | **1950** | **100.0** |

* 1. Веса:

Поскольку структура выборки по размерным группам не совпадает со структурой генеральной совокупности (крупные фирмы перепредставлены), также как и доли отдельных отраслей/секторов в выборки отличаются от соответствующих пропорций ГС, то во многих случаях результаты расчетов по выборочной совокупности не могут быть непосредственно распространены на всю обрабатывающую промышленность. Для приведения структуры выборки к структуре ГС в данные добавлены две весовые переменные – относительный вес *rel\_weight и* абсолютный вес *abs\_weight*. Выборку можно представить себе разбитой на 54 ячейки (9 секторов и 6 размерных классов) , по которым и производилась стратификация.

Относительный вес *rel\_weight* равен



- где *Pfirmsks* – кол-во фирм сектора *k* и размерной группы *s* в генеральной совокупности, а *Sfirmsks* - кол-во фирм сектора k и размерной группы s в выборке. *Pfirms* и *Sfirmsn* –общее кол-во фирм в генеральной совокупности и выборке соответственно. Сумма весов по всей выборке равна числу фирм в выборке. С помощью этой весовой переменной можно корректировать пропорции в числе фирм по ячейкам.

Абсолютный вес *abs\_weight* равен



Теперь сумма всех весов равна кол-ву фирм в генеральной совокупности. Все фирмы, принадлежащие одной ячейке сектор/размер, имеют одинаковые значения весов.

Примечание: Пользователям предоставляется база данных с оригинальными ответами на вопросы анкеты, после процедуры двойного ввода и проверки, проведенной «ГФК-Русь». В таком массиве данных возможно наличие ошибок и противоречий в ответах респондента. Любые исправления и преобразования оригинальных данных остаются на усмотрение пользователя. Дополнительно уточнены показатели выручки за 2013 год (вопрос В14), с исключением (заменой на пропущенные значения) заведомо неправдоподобных результатов и исправлением ошибок размерности. В базе сохранен оригинальный ответ на вопрос В14 и добавлена переменная В14\_corr со скорректированной выручкой.