

Использование вторичных данных в социально-экономических исследованиях

*Потапова В.В. студентка 2 курса магистратуры
ФГБОУ ВО «Алтайский государственный университет»
e-mail: potapova-v-12@mail.ru
Россия, Барнаул*

Аннотация. В данной статье рассматривается понятие вторичных данных, выделены преимущества и недостатки использования таких данных в социально-экономических исследованиях. Также сообщается о том, какие условия необходимо соблюдать, чтобы получить наиболее достоверные данные из всех имеющихся. Приведено несколько источников вторичных данных.

Ключевые слова. Вторичные данные, социально-экономические исследования.

1. Введение

На сегодняшний день все актуальнее становится проведение различного рода исследований и анализа данных. Часто для проведения тех или иных маркетинговых исследований используются первичные и вторичные данные, более подробно о вторичных данных и их источниках рассмотрим в данной статье.

Вторичные данные - это данные, которые уже были собраны и доступны из других источников. Такие данные дешевле и более доступны, чем первичные данные, а также могут быть использованы, когда первичные данные не могут быть получены вообще.

2. Преимущества и недостатки вторичных данных

Выделим преимущества вторичных данных:

Самым большим преимуществом использования вторичных данных является экономия. Кто-то уже собрал данные, поэтому исследователю не нужно выделять деньги, время, энергию и ресурсы на этот этап исследований. Иногда необходимо приобретать вторичный набор данных, но стоимость почти всегда ниже, чем расходы на сбор аналогичного набора данных с нуля, что обычно влечет за собой заработную плату, проезд и транспорт, офисные помещения, оборудование и другие накладные расходы.

Кроме того, поскольку данные уже собраны и обычно очищаются и хранятся в электронном формате, исследователь может проводить большую часть своего времени, анализируя данные, вместо того, чтобы готовить данные для анализа.

Второе важное преимущество использования вторичных данных - широта доступных данных. Например, Федеральная служба государственной статистики проводит многочисленные исследования в крупном масштабе,

которые отдельные исследователи могли бы с трудом собрать. Многие из этих наборов данных были собраны для одного и того же региона в течение нескольких разных периодов времени. Это позволяет исследователям выявлять временные и пространственные закономерности, применяя различные методы анализа данных к имеющимся массивам данных [1, 2, 3].

Третьим важным преимуществом использования вторичных данных является то, что процесс сбора данных часто поддерживает уровень знаний и профессионализма, которые могут отсутствовать у отдельных исследователей или небольших исследовательских проектов. Например, сбор данных для многих федеральных наборов данных часто выполняется сотрудниками, которые специализируются на определенных задачах и имеют многолетний опыт работы в этой конкретной области и с этим конкретным обследованием. Многие небольшие исследовательские проекты не обладают таким уровнем знаний, так как многие данные собираются студентами, работающими неполный рабочий день.

Недостатки вторичных данных:

Основным недостатком использования вторичных данных является то, что они могут не соответствовать вопросам исследователя или не содержать информацию, которую он хотел бы получить. Также возможно, что данные были собраны не в том географическом регионе, или не в тот период, в котором заинтересован исследователь. Поскольку исследователь не собирал данные, он не контролирует то, что содержится в наборе данных. Часто это может ограничивать анализ или изменять исходные вопросы, на которые исследователь пытался ответить.

Связанная с этим проблема заключается в том, что переменные могут быть определены или классифицированы иначе, чем выбрал исследователь. Например, возраст может быть собран в категориях, а не как непрерывная переменная, или раса может быть определена как «белый» и «другой», а не содержать категории для каждой крупной расы.

Другим существенным недостатком использования вторичных данных является то, что исследователь не знает точно, как был выполнен процесс сбора данных и насколько хорошо он выполнялся. Исследователь обычно не связан с информацией о том, насколько серьезно на данные влияют такие проблемы, как низкий уровень ответа или непонимание респондентами конкретных заданий опроса. Иногда эта информация легкодоступна, как в случае со многими федеральными наборами данных. Однако многие другие вторичные наборы данных не сопровождаются данным типом информации, и аналитик должен научиться читать между строк и учитывать, какие проблемы могут сопровождать процесс сбора данных.

Помимо вышеперечисленных проблем, данные также могут быть устаревшими [11].

Из-за вышеупомянутых недостатков вторичных данных приходится проводить оценку вторичных данных. Оценка означает, что должны соблюдаться следующие четыре требования:

- 1) **Доступность.** Необходимо видеть, что нужный тип данных вам доступен. Если он недоступен, вам необходимо перейти на первичные данные.
- 2) **Релевантность** - она должна соответствовать требованиям проблемы. Для этого у нас есть два критерия:
Единицы измерения должны быть одинаковыми.
Используемые понятия должны быть одинаковыми, а валюта данных не должна устаревать.
- 3) **Точность.** Чтобы определить, насколько точны данные, необходимо учитывать следующие моменты:
Используемая спецификация и методология.
Необходимо проанализировать маржу ошибки.
Необходимо убедиться в надежности источника.
- 4) **Достаточность.** Должны быть доступны адекватные данные.

Другими словами, есть некоторые важные вещи, которые необходимо сделать, прежде чем использовать вторичные данные в анализе. Поскольку исследователь не собирал данные, для него важно ознакомиться с набором данных: как были собраны данные, каковы категории ответов для каждого вопроса, нужно ли применять веса во время анализа, не нужно ли учитывать кластеры или стратификацию, кто был изучаемым населением, и многое другое.

Вторичный анализ - практика использования вторичных данных в исследованиях. Как метод исследования, он экономит время и деньги и позволяет избежать ненужного дублирования исследовательских усилий. Вторичный анализ обычно противопоставляется первичному анализу, который представляет собой анализ первичных данных, независимо собранных исследователем.

3. Примеры источников вторичных данных

Рассмотрим несколько источников вторичных данных:

- 1) Единая межведомственная информационно-статистическая система (<https://www.fedstat.ru/>) - ЕМИСС представляет собой государственный информационный ресурс, объединяющий официальные государственные информационные статистические ресурсы, формируемые субъектами официального статистического учета в рамках реализации федерального плана статистических работ [4].

Доступ к официальной статистической информации, включенной в состав статистических ресурсов, входящих в межведомственную систему, осуществляется на безвозмездной и недискриминационной основе.

- 2) Федеральная служба государственной статистики (<http://www.gks.ru/>) – один из самых полных источников вторичных данных, содержит большое количество данных разных категорий, все данные находятся в свободном доступе [6].

- 3) Мировой банк (<http://www.worldbank.org/>) – есть возможность осуществить поиск, просмотр и отображение на карте по более чем 10 000 проектов начиная с 1947 года и по настоящее время. Графики, карты и

сравнения более 1000 временных рядов показателей, включенных в публикацию «Показатели мирового развития». Ресурс не переведен на русский язык. Для формирования таблицы данных нужно выбрать страну(ы), показатели и желаемый период. После заполнения этих обязательных полей, можно получить таблицу данных, диаграмму или карту. Данные можно скачать в различных форматах [10].

4) Мультистат (<http://www.multistat.ru/>) – многофункциональный статистический портал, большинство данных доступно только зарегистрированным пользователям, ресурс платный.

5) Центральный банк РФ (<http://www.cbr.ru/>) – большой объем макроэкономических показателей и банковской статистики, обзоры инфляции. Чтобы получить статистические данные, нужно перейти в раздел «статистика» и далее выбрать интересующую категорию. Данные таблиц можно скачать [7].

6) Мировой атлас данных (<http://knoema.ru/>) - Мировая и региональная статистика. Есть разбивка по странам, регионам, множество различных показателей, данные выводятся в таблицу, графики. Также можно выполнить сравнение двух стран по какому-либо показателю, посмотреть рейтинг страны. Но не все данные можно скачать бесплатно, т.к. требуется регистрация и дальнейшие покупки [9].

7) Мировая экономика (<http://www.ereport.ru/>) - сайт содержит раздел «новости» с ежедневно-обновляемой информацией в области экономики, раздел «Обзоры», включающий обзоры специалистов по развитию экономики страны, сравнению экономик нескольких стран и т.д. В разделе «Котировки» указаны актуальные значения курсов валют, цен на биржевые товары, цены на металлы LME. В раздел «Статьи» собраны публикации по экономике, маркетингу и финансам. Статистические данные по странам мира можно найти в разделе «Статистика» [8].

4. Выводы

Следует помнить: чтобы провести значимый вторичный анализ, исследователи должны потратить немало времени на чтение и изучение истоков наборов данных.

Благодаря тщательной проверке исследователи могут определить:

- Цель, для которой материал был собран или создан.
- Конкретные методы его сбора.
- Пределы набора данных (какая информация не запрашивалась, не собиралась или не представлялась).
- Исторические и / или политические обстоятельства, связанные с созданием или сбором материала.

Кроме того, перед использованием вторичных данных исследователь должен учитывать, как данные кодируются или классифицируются и как это может повлиять на результаты вторичного анализа данных. Он также должен рассмотреть вопрос о том, должны ли данные быть адаптированы или скорректированы каким-либо образом до проведения его собственного анализа.

Качественные данные обычно создаются в известных обстоятельствах по названным физическим лицам для определенной цели. Это позволяет относительно легко анализировать данные с пониманием предубеждений, пробелов, социального контекста и других проблем.

Однако количественные данные могут потребовать более критического анализа. Не всегда ясно, как были собраны данные, почему определенные типы данных были собраны, в то время как другие не были, или было ли какое-либо смещение, связанное с созданием инструментов, используемых для сбора данных. Опросы, опросные листы и интервью могут быть разработаны для достижения заранее определенных результатов.

Хотя предвзятые данные могут быть чрезвычайно полезными, абсолютно важно, чтобы исследователь знал о предвзятости, ее цели и ее степени [11].

5. Список литературы

1. Мальцева А.В., Шилкина Н.Е., Махныткина О.В. Data Minig в социологии: опыт и перспективы проведения исследования//Социологические исследования. 2016. № 3 (383). С. 35-44
2. Мальцева А.В., Махныткина О.В., Шилкина Н.Е. Многомерные классификационные модели в сравнительном анализе социально-структурной специфики регионов россии//Интернет-журнал Науковедение. 2015. Т. 7. № 6 (31). С. 64.
3. Чудова О.В., Славский В.В. Нейронные сети с самоорганизацией на основе конкуренции в экономике, бизнесе и финансах//Известия Алтайского государственного университета. 2005. № 1 (45). С. 080-082.
4. ЕМИИС [Электронный ресурс]. – режим доступа : <https://www.fedstat.ru/>
5. Мультистат: многофункциональный статистический портал [Электронный ресурс]. – режим доступа : <http://www.multistat.ru/>
6. Федеральная служба государственной статистики [Электронный ресурс]. – режим доступа : <http://www.gks.ru/>
7. Центральный банк Российской Федерации [Электронный ресурс]. – режим доступа : <http://www.cbr.ru/>
8. EREPORT.RU мировая экономика [Электронный ресурс]. – режим доступа : <http://www.ereport.ru/>
9. Кноема [Электронный ресурс]. – режим доступа : <https://knoema.ru/>
10. The world bank [Электронный ресурс]. – режим доступа : <http://www.worldbank.org/>
11. Understanding Secondary Data and How to Use It in Research [Электронный ресурс]. – режим доступа : <https://www.thoughtco.com/secondary-analysis-3026573>

Контактный телефон +79237489669