

СТАТИСТИЧЕСКИЙ АНАЛИЗ БАЗЫ ДАННЫХ И РЕЗУЛЬТАТОВ АНКЕТИРОВАНИЯ БЕЗРАБОТНЫХ

Н.Ф. Богаченко

к.ф.-м.н., доцент, e-mail: nfbogachenko@mail.ru

И.А. Герингер

студент, e-mail: ilya.tuponogov@gmail.com

Д.Н. Лавров

к.т.н., доцент, e-mail: dmitry.lavrov72@gmail.com

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. В работе анализируются данные о безработных, предоставленные службой занятости Омской области, а также данные по тестированию безработных. При помощи пакета MS Excel выявлен ряд закономерностей: получена статистика обращений по гендерному признаку и возрасту обратившихся, оценены услуги, предоставляемые центром занятости, причины снятия с учёта, характеристики работы центра занятости.

Ключевые слова: анализ данных, служба занятости, портрет безработного.

Введение

В данной работе в рамках выполнения научного проекта «Методология и инструменты оценки эффективности активной политики занятости на рынке труда» планируется проанализировать данные о безработных, выявить закономерности и понять, кто имеет наибольшие шансы быстро найти работу. Для анализа были получены два массива данных: сведения от службы занятости и данные по тестированию безработных.

В предоставленных данных содержится информация, в том числе о том, получает ли гражданин пособие по безработице, размер этого пособия, сведения о предыдущем месте работы, какие вакансии рассматривает гражданин и причины отказа, оценка услуг центра занятости, ранжирование характеристик центра занятости, динамика услуг, предоставляемых центром занятости, и сведения о гражданине: пол, возраст, семейное положение, образование.

В массиве данных от службы занятости 24035 записей, все данные типовые, анализ чистого текста не требуется. Требуется первоначальная обработка нескольких столбцов (например, в столбце «Вид услуги» перечислены услуги через запятую, а для анализа удобнее вынести каждую услугу в отдельный столбец и проставить для каждого кандидата 1, если услуга выбрана, или 0 в противном случае). Также требуется обработка дат (например вычислить возраст на момент обращения). Для выявления закономерностей необходимо

будет строить диаграммы и графики. Весь требуемый функционал есть в MS Excel [1], входящем в состав пакета Microsoft Office. Количество данных также не является препятствием для использования этой программы.

Анкетирование подразумевает несколько ответов в свободной форме. Качество данных невысокое. Например, человек отмечает «нет» на вопрос «3. Получаете ли Вы пособие по безработице?», но тут же указывает размер этого пособия. При ответе на вопрос о предыдущем месте работы «Если в предыдущем вопросе Вы выбрали вариант ответа “Другое”, то отметьте, что именно» были варианты ответов: «2 недели», «Июнь Июль», «От 3 до 5 месяцев». Или отвечают, что детей нет, но тут же указывают их возраст. Примерно в 3000 анкет перепутаны местами столбцы «Укажите, пожалуйста, Ваше имя и отчество», «Удобный для Вас способ связи (номер телефона / адрес электронной почты / имя в социальных сетях)» и «В каком Центре занятости населения Вы поставлены на учёт?» Это было исправлено. Хотя данные в свободной форме есть, но их количество небольшое и тематика заранее задана. Основной анализ также требует обработки табличных данных, сводных таблиц и построения графиков. Весь требуемый функционал также есть в MS Excel. Данных в этом наборе меньше (4674 записей), соответственно выбор был сделан именно в пользу пакета от Microsoft.

1. Предварительная обработка

Первые шаги анализа — определение проблемы и извлечение данных — не являются частью данной работы, поэтому не описываются. Очистка массива данных потребовалась в основном на этапе преобразования.

1.1. Преобразование данных в базе центра занятости

Для удобства анализа потребовался ряд преобразований. Столбцы «Особо учитываемая категория» и «Вид услуги» содержат несколько параметров через запятую. Была использована функция «текст по столбцам». Также были вычислены возраст на момент обращения в центр занятости, возраст, когда был признан безработным и возраст снятия с учёта. Данные были форматированы как таблицы.

Общий трудовой стаж был преобразован. Исходный формат вида «16 лет 11 мес 17 дн» был пересчитан к количеству месяцев. Использована формула:

=ЛЕВСИМВ([@[Общий трудовой стаж гражданина]];2) * 12 + ПРАВСИМВ(ЛЕВСИМВ([@[Общий трудовой стаж гражданина]];9);2) + ЕСЛИ(ЗНАЧЕН(ЛЕВСИМВ(ПРАВСИМВ([@[Общий трудовой стаж гражданина]];5);2))>=15;1;0)

Для столбца «Вид услуги» применена функция «текст по столбцам» с разделителями — запятыми. Это позволило получить список всех услуг, отмеченных соискателями. Список транспонирован и добавлен в таблицу в качестве столбцов, после этого для каждого столбца использована формула:

=НЕ(ЕОШИБКА(НАЙТИ(Безработные[#Заголовки];[Содействие в поиске работы]);[@[Вид услуги]])))

Эта формула позволила получить ИСТИНА либо ЛОЖЬ в качестве значений. Позднее для удобства расчётов ИСТИНА была заменена на 1, ЛОЖЬ – на 0.

Столбец «Итоговый статус» подвергся нормализации. Вместо предложенных причин снятия с учёта использована более общая классификация. Для этого заполнена таблица 1 и использована формула:

=ВПР([@[Причина снятия с учета]];КатегорииПричин;2;0)

При составлении категорий прежде всего требовалось понять, насколько центр занятости помог гражданину устроиться на работу. «Положительным» исходом можно считать трудоустройство и «обучение». «Условно положительным» — самостоятельное трудоустройство в том или ином виде. Статус «безработный» говорит о том, что работа с гражданином не закончена. Остальные статусы показывают, что центр занятости не смог решить вопрос с работой по тем или иным причинам.

Аналогичным образом сделана разбивка подразделений на город / область.

1.2. Преобразование данных анкетирования

Основная работа по предварительной обработке данных — чистка. Как описано ранее, есть много несоответствий в данных, например, профессия «Бухгалтер» в вопросе «35. По какой профессии Вы имеете образование?» была написана 38-ю способами (включая опечатки, перечисление нескольких профессий, уточнения и пр.).

Основные методы аналогичны описанным выше: скопировать столбец на отдельный лист, удалить дубликаты, поставить соответствие каждой строке и заменить значения в таблице с помощью функции ВПР().

Также для удобства расчётов текстовая система оценки («хорошо», «отлично», «скорее важно» и т. д.) была заменена на цифровую интерпретацию. Это позволило высчитать средний бал и проранжировать центры занятости.

2. Исследование и визуализация результатов

2.1. Данные центра занятости

Сравним общее количество обратившихся в центр занятости. В городе этот показатель, ожидаемо, выше. Однако если добавить к этим данным население каждого района (по данным Википедии) и рассчитать относительную величину, то результаты окажутся другими. Резко выделяются районы-«аутсайдеры» по количеству обратившихся (рис. 1). Столбчатой диаграммой обозначены обратившиеся граждане в разрезе категорий в абсолютных единицах, на графиках — те же граждане, в разрезе тех же категорий, но в относительных единицах. По графику видно, что хотя количество обратившихся в городе людей больше, работу найти проще. Так же мало людей обращается в службу занятости в

Таблица 1. Причины снятия с учёта

Причина снятия с учёта	Категория
Длительная (более месяца) неявка гражданина в органы СЗ без уважительных причин	Отказ
Трудоустройство самостоятельно	Самостоятельно
Назначение пенсии	Самостоятельно
Отказ от услуг СЗ	Отказ
Попытка получения или получение пособия обманным путём	Мошенничество
Трудоустройство по направлению службы занятости	Трудоустройство
Профессиональное обучение	Обучение
Прохождение очного курса обучения в образовательных учреждениях	Обучение
Трудоустройство на субсидируемое рабочее место	Трудоустройство
Временное трудоустройство испытывающих трудности в поиске работы	Трудоустройство
Предпринимательская деятельность без образования юридического лица	Трудоустройство
Трудоустройство на общественные работы	Трудоустройство
Призыв на военную службу	Отказ
Смерть	Отказ
Восстановление на прежней работе	Отказ
Переезд или переселение внутри субъекта РФ	Отказ
Осуждение	Отказ
Предпринимательская деятельность с образованием юридического лица	Самостоятельно
Переезд или переселение в другую местность	Отказ
Организация крестьянского (фермерского) хозяйства	Самостоятельно
Не задана	Безработный

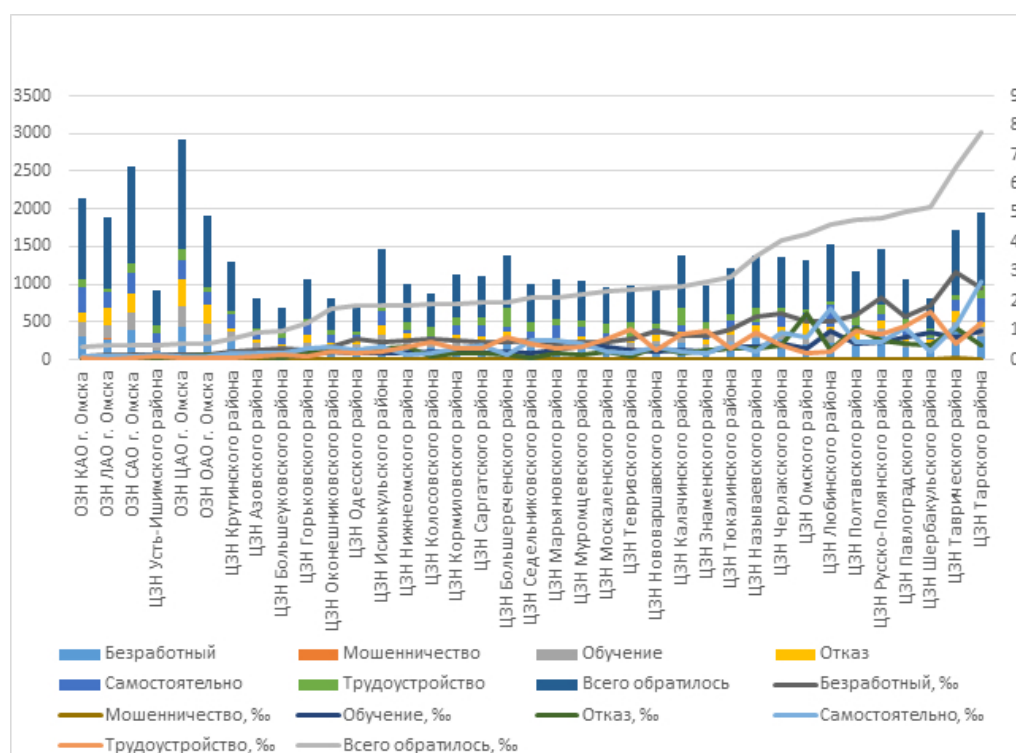


Рис. 1. Распределение результатов обращения граждан в центры занятости по районам. Ось X — районы Омской области, ось Y — количество безработных

Усть-Ишимском районе. Такое количество обращений может свидетельствовать о хорошем уровне занятости населения или о недоверии к центру занятости. Однако исследования вопроса качества работы подразделений выходят за рамки этой статьи. В Шербакульском районе, наоборот, в абсолютных цифрах количество обратившихся не велико (7896 человек, и это абсолютный минимум), но это 5,1 % населения, что является третьим результатом.

2.1.1. Гендерный признак

Если рассматривать обратившихся в центры занятости по гендерному признаку (рис. 2), то легко заметить, что в области в центры занятости обращаются чаще мужчины. В городе, наоборот, женщины.

В Центр занятости населения Кормиловского района обращаются равное количество мужчин и женщин.

Средние значения по области и городу также представлены на графике. Эти значения даны как сравнение и обозначены на графике менее ярким цветом.

2.1.2. Возраст

Для анализа возраста безработных потребовалось разбить граждан на возрастные группы [2]. Было решено использовать общепринятые возрастные группы (табл. 2).

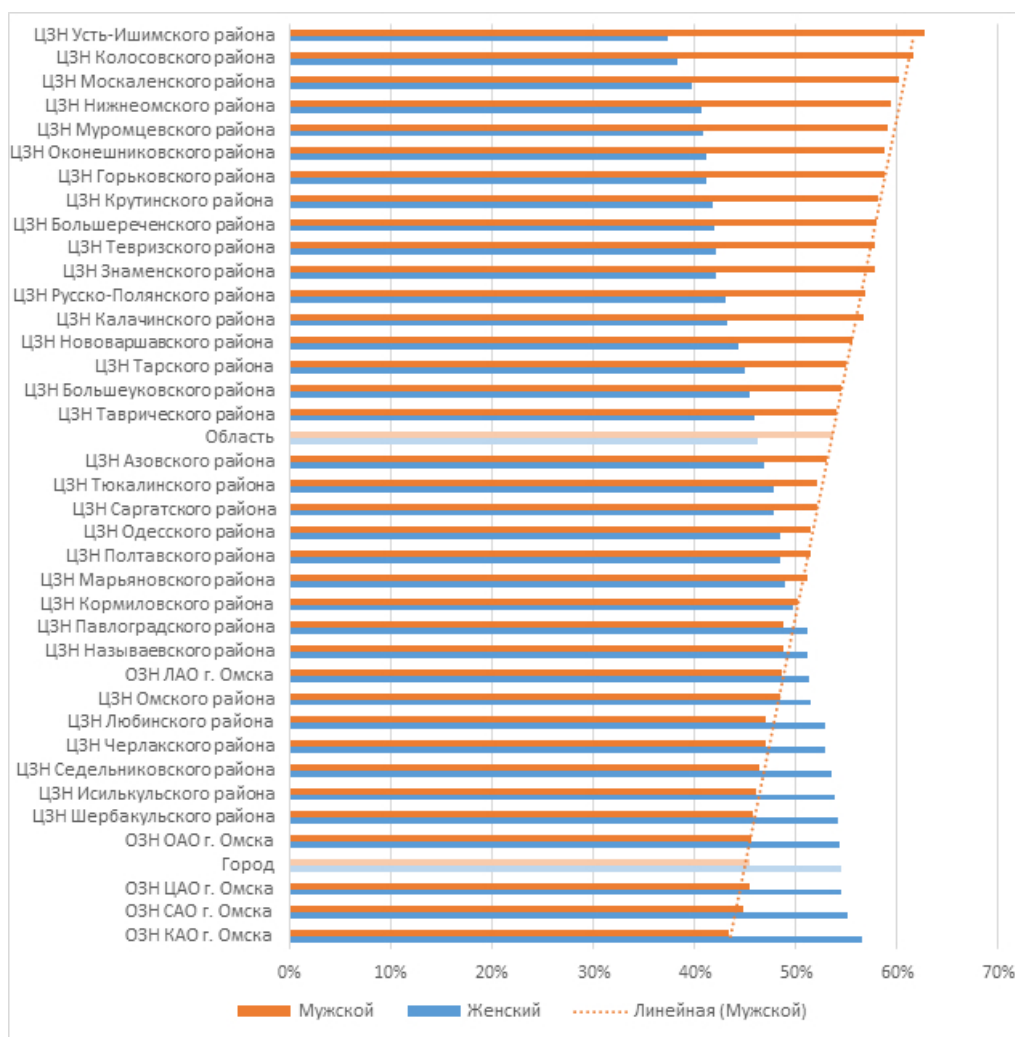


Рис. 2. Распределение обратившихся по гендерному признаку в разрезе географических районов. Ось X — процент мужчин / женщин, ось Y — центры занятости по географическому признаку

Таблица 2. Возрастные группы

Возрастная группа	Мужчины	Женщины
Юный возраст	до 21	До 20
Зрелый 1 период	От 21 до 35	От 20 до 35
Зрелый 2 период	От 35 до 60	От 35 до 55
Пожилой	Старше 60	Старше 55

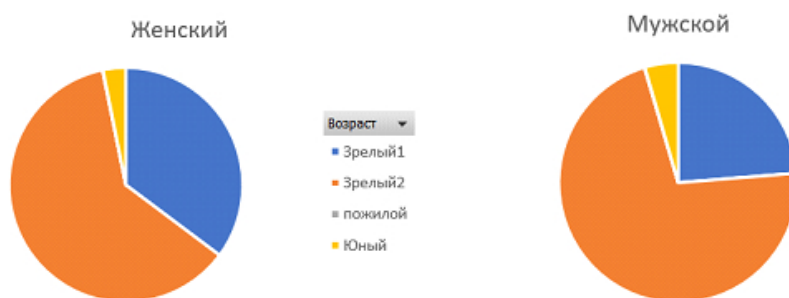


Рис. 3. Распределение возраста обратившихся в зависимости от гендерного признака

В ходе анализа было выяснено, что разбиение по возрасту не зависит от района области или города, но сильно зависит от гендерной принадлежности. Пример итоговых диаграммы для одного из центров занятости представлен на рисунке 3. Эта разбивка примерно одинаковая для всех районов (центров занятости).

2.1.3. Услуги центра занятости

В исходных данных услуги центра занятости были представлены в одной колонке через запятую. На этапе подготовки данных они были разнесены в разные колонки. Это позволило легко построить сводную таблицу и провести анализ.

Были определены самые популярные услуги. Для этого подсчитано количество раз, когда обратившийся гражданин пользовался каждой из услуг, предоставляемых центрами занятости. За исключением редких выбросов, распределение не зависит от подразделения, куда обратился гражданин, пола или возраста обратившегося. Данные в разбивке полов с проранжированными услугами (использована функция РАНГ.РВ, которая присваивает равный ранг при равенстве значений) представлены в таблицах 3, 4.

Из таблиц видно, что основные услуги — это «Содействие в поиске работы», «Информирование», «Другие услуги» и «Профориентация».

Все 15 услуг не оказано ни одному обратившемуся. Это связано в первую очередь с тем, что, например, «Содействие в поиске работы» и «Сопровождение при содействии занятости инвалидов» — взаимоисключающие пункты. Максимальное количество отмеченных пунктов для одного гражданина — 9. Распределение удобно посмотреть на графике (рис. 4).

Граждане с максимальным количеством запрошенных услуг — зрелого возраста. Двое мужчин второго периода зрелости, операторы котельной, обратились в одном районе в один день. Остальные данные по этим безработным тоже заполнены единообразно, что позволяет предположить, что они обращались вместе. Предположительно это бывшие коллеги. Третья — женщина 33 лет, парикмахер. Всем троим предложено профессиональное обучение.

Граждане, которым оказана единственная услуга, чаще всего выбирают

Таблица 3. Популярность услуг в зависимости от возраста обратившихся. Женщины

Услуга \ Возраст	Зрелый 1	ранг	Зрелый 2	ранг	Пожилой	ранг	Юный	ранг	Всего
Содействие в поиске работы	3920	1	6829	2	8	1	340	1	11097
Информирование	3886	2	6848	1	7	3	336	2	11077
Другие услуги	3624	3	6369	3	7	3	331	3	10331
Профориентация	3224	4	5815	4	8	1	297	4	9344
Психологическая поддержка	702	5	1353	5	0	6	55	6	2110
Профобучение	695	6	721	7	0	6	102	5	1518
Социальная адаптация	578	7	929	6	1	5	43	7	1551
Самозанятость	271	8	400	8	0	6	16	8	687
Общественные работы	164	9	343	9	0	6	13	9	520
Временное трудоустройство	78	10	148	10	0	6	8	10	234
Организация ярмарок вакансий	22	11	64	11	0	6	5	11	91
Содействие в переезде в другую местность	2	13	10	12	0	6	0	12	12
Сопровождение при содействии занятости инвалидов	4	12	1	14	0	6	0	12	5
Содействие в переселении в другую местность	0	15	2	13	0	6	0	12	2
Содействие в переселении для работы в сельской местности	1	14	0	15	0	6	0	12	1

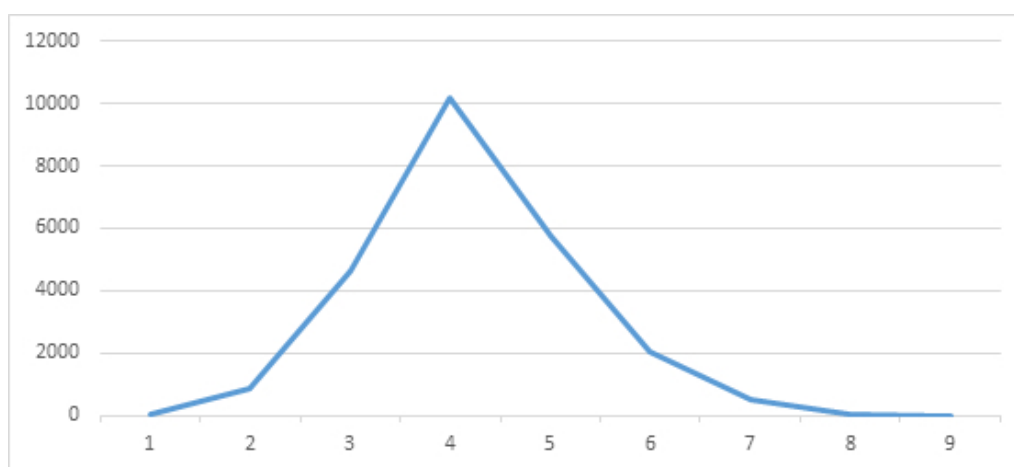
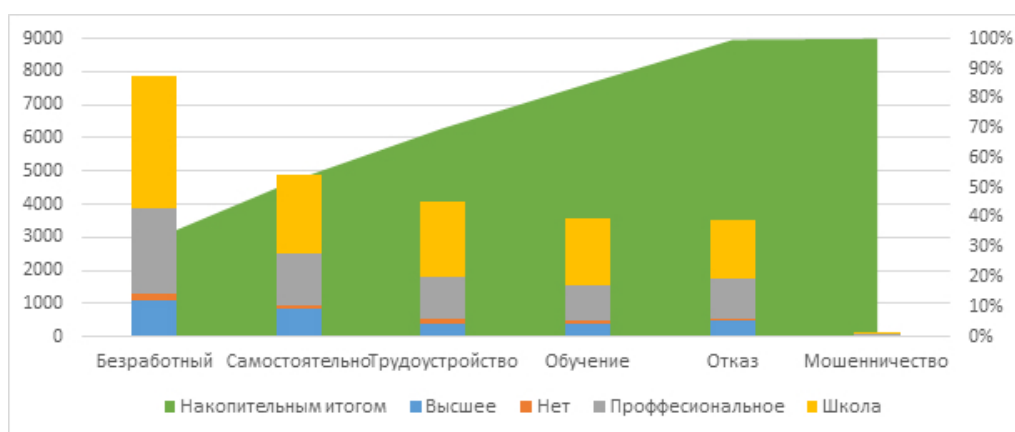
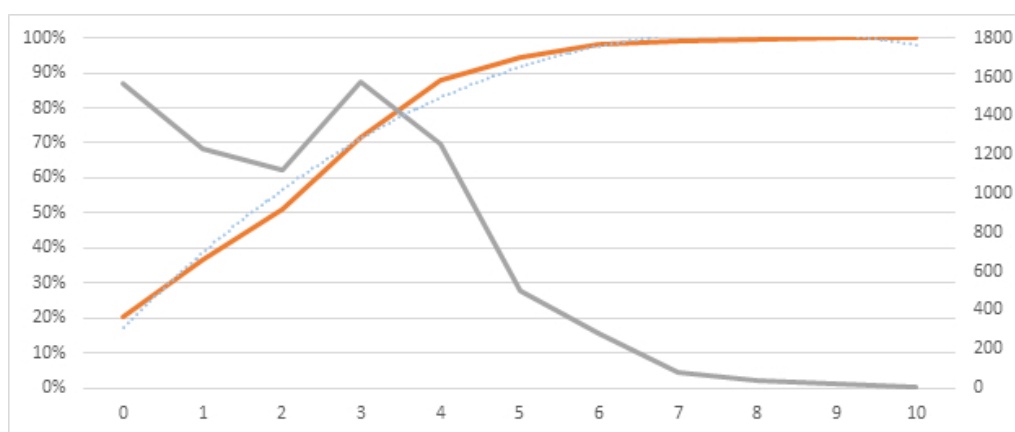
Рис. 4. Количество услуг, оказанных одному гражданину. Ось X — количество выбранных услуг, ось Y — количество граждан

Таблица 4. Популярность услуг в зависимости от возраста обратившихся. Мужчины

Услуга \ Возраст	Зрелый 1	ранг	Зрелый 2	ранг	Пожилой	ранг	Юный	ранг	Всего
Содействие в поиске работы	2836	1	8492	2	7	3	528	1	11863
Информирование	2804	2	8494	1	8	1	524	2	11830
Другие услуги	2644	3	7912	3	7	3	496	3	11059
Профориентация	2311	4	7239	4	8	1	448	4	10006
Психологическая поддержка	381	7	1606	5	2	5	71	6	2060
Профобучение	569	5	1382	6	0	8	155	5	2106
Социальная адаптация	421	6	1306	7	0	8	69	7	1796
Самозанятость	188	8	443	9	1	6	18	9	650
Общественные работы	134	9	473	8	1	6	29	8	637
Временное трудоустройство	57	10	193	10	0	8	15	10	265
Организация ярмарок вакансий	27	11	133	11	0	8	3	11	163
Содействие в переезде в другую местность	5	12	13	12	0	8	0	13	18
Сопровождение при содействии занятости инвалидов	5	12	6	13	0	8	1	12	12
Содействие в переселении в другую местность	0	14	1	14	0	8	0	13	1
Содействие в переселении для работы в сельской местности	0	14	0	15	0	8	0	13	0

Рис. 5. Результаты обращений. Ось X — результаты обращений, ось Y — количество анкетРис. 6. Устроившиеся на работу граждане. Ось X — количество месяцев с момента обращения, ось Y — количество граждан с завершившимся периодом безработицы накопительным итогом и в абсолютных величинах

«Информирование» (32 человека), «Профорientацию» (17), «Содействие в поиске работы» (10) и «Другие услуги» (1).

2.1.4. Результат обращения

Большинство (33 %) обратившихся имеют незавершенный период безработицы, самостоятельно нашли работу 20 %, ещё 17 % были трудоустроены, 15 % направлены на обучение. Если сложить успехи центра занятости (трудоустройство + обучение), получим 32 % (рис. 5).

Представляет интерес зависимость трудоустройства от времени ожидания накопительным итогом (рис. 6, оранжевый график). Зависимость достаточно точно описывается полиномом 3-ей степени (синяя пунктирная линия). Самые большие шансы устроиться на работу — в первый и третий месяцы после обращения (рис. 6, серый график). Можно сделать вывод, что если после 4-х месяцев работу не получилось найти, то стоит задуматься о смене профессии.

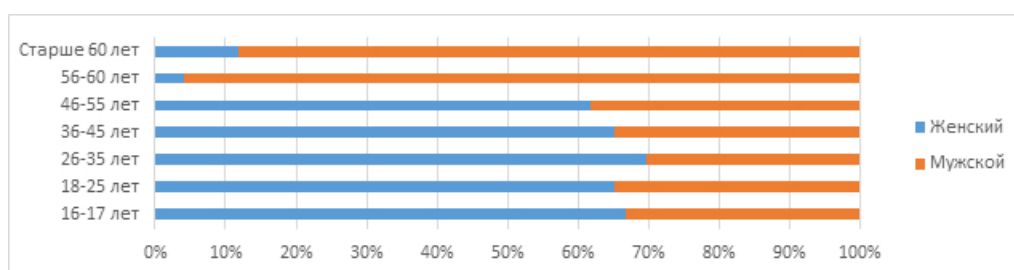


Рис. 7. Гендерная принадлежность обратившихся граждан. Ось X — гендерная принадлежность граждан, ось Y — возраст обратившихся

2.2. Анкеты безработных

2.2.1. Сведения о людях

По понятным причинам переданные для анализа анкеты почти не содержат персональных данных.

По имеющимся данным можно проанализировать предпочитаемый способ связи. Электронную почту оставили 706 человек (15 %), номер мобильного телефона — 3767 человек (80 %). 95 человек (2 %) оставили оба контакта, а 297 (6 %) человек не оставили контактных данных вообще.

Анализ гендерной принадлежности в разрезе возраста можно посмотреть на диаграмме (рис. 7). Из этой диаграммы можно сделать вывод о том, что в Омской области женщины обращаются в службу занятости чаще мужчин. Учитывая, что пенсионный возраст для женщин в 2020 году был установлен 55,5 лет, логично выглядит смещение 56-60-летнего возраста в пользу мужчин. Однако в возрасте старше 60-ти лет начинают обращаться чаще женщины.

2.2.2. Оценки подразделений

572 человека (12 %) отметили, что они не воспользовались ни одной из услуг центра занятости. 852 человека воспользовались хотя бы одной услугой, но не поставили оценок. Усреднённые оценки оказанных услуг представлены в таблице 5. Усреднённые оценки работы центра занятости представлены в таблице 6.

Тот факт, что самую высокую оценку получила скорость предоставления услуг, свидетельствует о том, что посещение центра занятости не является самоцелью. Предположительно, большинство граждан обращаются туда для скорейшего получения работы.

2.2.3. Развитие услуг центра занятости

В анкете присутствует вопрос о развитии услуг, предоставляемых центром занятости. То или иное изменение заметили почти все опрошенные. Большинство — 2070 человек обратили внимание на возможность получать услуги через интернет. Опрошенные выбирали один из 3-х вариантов — «улучшилось», что

Таблица 5. Средняя оценка в разрезе услуг

Услуга	Средняя оценка	Количество оценок
[Содействие в поиске подходящей работы]	3,32	3139
[Психологическая поддержка]	3,06	1026
[Профессиональное обучение]	2,73	814
[Профориентация]	3,14	1510
[Социальная адаптация на рынке труда]	2,85	827
[Содействие в самозанятости]	2,38	625
[Общественные работы]	2,31	586
[Другое]	1,86	112
Итого	3,22	8639

Таблица 6. Средняя оценка в разрезе работы центра занятости

Характеристика работы центра занятости	Количество оценок	Средняя оценка
[Внимательность, отзывчивость и вежливость обслуживающего персонала]	4522	3,39
[Скорость предоставления услуги]	4557	3,50
[График работы центра занятости]	4476	3,13
[Квалификация персонала]	4478	3,34
[Понятность правил получения услуг]	4455	3,09
[Наличие необходимой информации о предоставлении услуг]	4497	3,23
[Территориальное удобство расположения здания центра занятости]	4491	3,47
[Комфорт за пределами и внутри помещения (наличие парковок, удобных мест для сиденья и т. д.)]	4489	3,48
[Другое]	1703	1,92

Таблица 7. Развитие услуг центров занятости

	Количество ответов	Стало лучше (+1)	Не изменилось (0)	Стало хуже (-1)	Итого
[Возможность получать услугу полностью или частично через интернет]	2070	1674	314	82	1592
[Комфорт за пределами и внутри помещения центра занятости]	1971	923	951	97	826
[Очереди на обслуживание в центре занятости]	1963	1148	728	87	1061
[Наличие информации об услугах]	1931	1015	888	28	987
[Вежливость и отзывчивость персонала]	1964	1024	912	28	996
[График работы центра занятости]	1938	793	1121	24	769

давало +1 балл, «не изменилось» — 0 или «ухудшилось» — минус 1 балл. Сводные данные представлены в таблице 7.

Заключение

В ходе работы была проанализирована база данных обращений в центр занятости населения и анкеты безработных. Получена статистика обращений по гендерному признаку и возрасту обратившихся. Портреты различаются для города и области. В городе это женщина 39 лет с высшим или средним профессиональным образованием, в области — мужчина 41 год, со средним образованием.

Были также проанализированы услуги, предоставляемые центрами занятости. В целом они положительно оцениваются населением и показывают положительную динамику.

В предоставленных данных 33 % граждан — с незавершённым периодом безработицы. Трудоустройство либо обучение от центра занятости происходит в 32 % случаев обращений (17 % и 15 % соответственно). Ещё 20 % обратившихся находят работу самостоятельно, что позволяет сказать: большая часть граждан (52 %) получают положительный результат после посещения центра.

Среди безработных большинство (73 %) обращается в первый раз за последние 3 года.

Дальнейшие исследования будут направлены на предсказание времени поиска работы гражданином.

Благодарности

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 20-010-00489.

ЛИТЕРАТУРА

1. Функции Excel (по категориям). URL: <https://support.microsoft.com/ru-ru/office/функции-excel-по-категориям-5f91f4e9-7b42-46d2-9bd1-63f26a86c0eb> (дата обращения: 12.09.2021).
2. Возрастная периодизация. URL: https://ru.wikipedia.org/wiki/Возрастная_периодизация (дата обращения: 12.09.2021).

STATISTICAL ANALYSIS OF THE DATABASE AND THE RESULTS OF A SURVEY OF THE UNEMPLOYED

N.F. Bogachenko

Ph.D.(Phys.-Math.), Associate Professor, e-mail: nfbogachenko@mail.ru

D.N. Lavrov

Ph.D.(Eng.), Associate Professor, e-mail: dmitry.lavrov72@gmail.com

I.A. Geringer

Student, e-mail: ilya.tuponogov@gmail.com

Dostoevsky Omsk State University, Omsk, Russia

Abstract. In the paper the data on the unemployed provided by the employment service of the Omsk region as well as data on testing the unemployed are analyzed. Using the MS Excel package, a number of regularities were revealed: statistics of requests by gender and age of applicants were obtained, the services provided by the employment center, reasons for deregistration, characteristics of the work of the employment center were assessed.

Keywords: data analysis, employment service, portrait of the unemployed.

REFERENCES

1. Funktsii Excel (po kategoriyam). URL: <https://support.microsoft.com/ru-ru/office/funktsii-excel-po-kategoriyam-5f91f4e9-7b42-46d2-9bd1-63f26a86c0eb> (12.09.2021). (in Russian)
2. Vozrastnaya periodizatsiya. URL: https://ru.wikipedia.org/wiki/_ (12.09.2021). (in Russian)

Дата поступления в редакцию: 18.09.2021