

Школа Инженерная школа ядерных технологий
 Направление подготовки 01.03.02 «Прикладная математика и информатика»
 Отделение школы (НОЦ) Отделение экспериментальной физики

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Кластеризация регионов России на основе социально-экономических показателей и ожидаемой продолжительности жизни

УДК 519.237:332.143(47+57)

Студент

Группа	ФИО	Подпись	Дата
0B51	Курникова Александра Олеговна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отделения экспериментальной физики	Шинкеев Михаил Леонидович	Доцент, кандидат ф-м. наук		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отделения социально – гуманитарных технологий	Конотопский Владимир Юрьевич	Доцент, кандидат э. наук		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель отделения общетехнических дисциплин	Исаева Елизавета Сергеевна			

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
	Крицкий Олег Леонидович	Доцент, кандидат ф-м. наук		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
0B51	Курниковой Александре Олеговне

Школа	ИЯТШ	Отделение школы (НОЦ)	ОЭФ
Уровень образования	Бакалавр	Направление/специальность	Прикладная математика и информатика

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:	
1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	
2. <i>Нормы и нормативы расходования ресурсов</i>	
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. <i>Организация и планирование комплекса работ</i>	
2. <i>Стоимостная оценка работы</i>	
3. <i>Оценка экономической эффективности исследования</i>	
Перечень графического материала (с точным указанием обязательных чертежей):	
1. <i>Календарный план – график выполнения работ</i>	

Дата выдачи задания для раздела по линейному графику	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отделения социально – гуманитарных технологий	Конотопский В.Ю.	Кандидат экономических наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0B51	Курникова Александра Олеговна		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
0B51	Курниковой Александре Олеговне

Школа	ИЯТШ	Отделение (НОЦ)	Экспериментальной физики
Уровень образования	Бакалавр	Направление/специальность	Прикладная математика и информатика

Тема ВКР:

Кластеризация регионов России на основе социально-экономических показателей и ожидаемой продолжительности жизни	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	<i>Целью данной работы является кластеризация регионов РФ по совокупности социально-экономических показателей и ОПЖ на основе данных за 2016 год</i>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	<i>Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 27.12.2018); ГОСТ 12.2.032-78 ССБТ; ГОСТ 21889-76; ГОСТ 22269-76; ГОСТ Р 50923-96; СанПиН 2.2.2/2.4.1340-03; СанПиН 4719-88; Федеральный закон от 22.08.1996 №125-ФЗ</i>
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<i>– приводится классификация пожароопасности помещений, указывается класс пожароопасности вашего помещения, перечисляются средства пожаробнаружения и принцип их работы, средства пожаротушения, принцип работы, назначение (какие пожары можно тушить, какие – нет), маркировка;</i>
3. Экологическая безопасность:	<i>– анализ воздействия при работе на ПЭВМ на атмосферу, гидросферу, литосферу;</i> <i>– наличие отходов (бумага, картриджи, компьютеры и т. д.);</i> <i>- методы утилизации отходов.</i>
4. Безопасность в чрезвычайных ситуациях:	<i>– выявление типичных аварийных ситуаций, причин их возникновения;</i> <i>– разрабатываются превентивные</i>

	<i>меры по предупреждению ЧС;</i> <i>– разработка мер по повышению устойчивости объекта к данной ЧС;</i> <i>– разработка действий в результате возникшей ЧС и мер по ликвидации её последствий;</i> <i>– ЧС природная – сильный мороз;</i> <i>ЧС техногенная – предполагаемая диверсия.</i>
--	---

Дата выдачи задания для раздела по линейному графику	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Исаева Елизавета Сергеевна			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0B51	Курникова Александра Олеговна		

Реферат

Выпускная квалификационная работа содержит 100 страниц, 29 таблиц, 8 рисунков, 20 источников литературы.

Ключевые слова: кластеризация, социально-экономические показатели, ожидаемая продолжительность жизни, метод главных компонент, дисперсионный анализ, кластерный анализ.

Объекты исследования: социально-экономические показатели, ожидаемая продолжительность жизни, методы кластеризации.

Цель работы: изучение влияния социально - экономических факторов и характеристик социальной среды на ОПЖ российских регионов статистическими методами.

Методы проведения исследования: теоретический (изучение литературы, методов факторизации и кластеризации) и практический: факторизация социально-экономических показателей, кластеризация регионов РФ на основе обобщенных факторов и ОПЖ.

Выпускная квалификационная работа представлена на USBflash носителе.

Оглавление

Введение	9
1. Теоретическая часть.....	12
1.1. Задача снижения размерности.....	12
1.2. Метод главных компонент	12
1.3. Критерии отбора главных компонент.....	17
1.4. Классификация факторов. Вращение факторов	18
1.5. Кластерный анализ.....	20
1.5.1. Меры близости объекта	21
1.5.2. Иерархические агломеративные методы кластерного анализа ...	23
1.5.2.1. Метод одиночной связи.....	25
1.5.2.2. Метод полной связи.....	26
1.5.2.3. Метод Варда.....	26
1.6. Дисперсионный анализ	26
1.6.1. Однофакторный дисперсионный анализ с фиксированными эффектами.....	27
Для сумм квадратов можно построить также следующую таблицу:	31
1.6.2. Непараметрический дисперсионный анализ	32
2. Практическая часть	33
3. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение ..	73
3.1. Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения	73
3.1.1 Продолжительность этапов работ	74
3.2 Расчет сметы затрат на выполнение проекта.....	79
3.2.1 Расчет затрат на материалы	80
3.2.2 Расчет заработной платы	81
3.2.3 Расчет затрат на социальный налог	81
3.2.4 Расчет затрат на электроэнергию	82
3.2.5 Расчет амортизационных расходов	83

3.2.6 Расчет прочих расходов	84
3.2.7 Расчет общей себестоимости разработки	84
3.2.8 Расчет прибыли	84
3.2.9 Расчет НДС	84
3.2.10 Цена разработки НИР	85
3.3 Оценка экономической эффективности проекта	85
4. Социальная ответственность	86
4.1. Правовые и организационные вопросы обеспечения безопасности.	86
4.1.1. Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства.	86
4.1.2. Организационные мероприятия при компоновке рабочей зоны...	87
4.2. Профессиональная социальная безопасность.	89
4.2.1. Анализ вредных и опасных факторов, которые может создать объект исследования.	89
4.2.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при проведении исследований.	90
4.2.3. Обоснование мероприятий по защите исследователя от действия опасных и вредных факторов.	91
4.3. Экологическая безопасность.	92
4.3.1. Анализ влияния объекта исследования на окружающую среду.	92
4.3.2. Анализ влияния процесса исследования на окружающую среду. ...	92
4.3.3. Обоснование мероприятий по защите окружающей среды.	92
4.4. Безопасность в чрезвычайных ситуациях.	93
4.4.1. Анализ вероятных ЧС, которые может инициировать объект исследований.	93
4.4.2. Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований.	93
4.4.3. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС.	94
4.5. Вывод и рекомендации	95

Заключение	97
Список публикаций студента	98
Список литературы	99

Введение

В самом простом случае средняя продолжительность жизни — это среднее арифметическое количества лет, прожитых определенной группой людей. Например, у вас есть пять человек, которые прожили до 50 лет, и пять человек, которые прожили до 100 лет. Средняя продолжительность жизни десяти этих человек — 75 лет.

Демографы такой показатель практически никогда не используют. Это неудобно: чтобы выяснить продолжительность жизни группы людей, рожденных в один год в конкретной стране или регионе (демографы называют такие группы когортами рождения), надо дожидаться, пока все ее представители умрут. Нам же интересно знать, что происходит и будет происходить с живущими сейчас людьми.

Эти ограничения вынуждают нас при ответе на простой, казалось бы, вопрос — сколько живут люди — прибегать к математическому моделированию, результатом которого оказывается показатель с обманчиво простым названием — ожидаемая продолжительность жизни. Именно кажущаяся простота показателя приводит к многочисленным ошибкам в интерпретации.

В основе расчетов ожидаемой продолжительности жизни лежит математическая модель — таблица смертности (в англоязычной традиции намного оптимистичнее — *lifetable*). Она моделирует процесс вымирания условного поколения.

Вместо того, чтобы терпеливо дожидаться, пока умрут все представители живущих сегодня поколений, мы можем представить, что рожденные в определенный год дети во всех возрастах на протяжении своей жизни будут умирать с интенсивностью, которую мы наблюдали в соответствующих возрастах в год их рождения.

Например, ожидаемая продолжительность жизни российских мужчин в 2016 году составила 66,5 года. Столько проживут новорожденные мальчики, если они будут умирать

в возрасте 1 года (в 2017 году) так же, как годовалые дети в 2016 году

в возрасте 5 лет (в 2021 году) — как пятилетние мальчики в 2016 году

в 65 лет (в 2081 году) — как 65-летние мужчины в 2016 году

То есть интерпретировать значение в 66,5 года для российских мужчин в 2016 году как среднюю продолжительность предстоящей жизни для рожденного поколения можно лишь при исполнении важнейшего условия — все возрастные коэффициенты смертности не изменятся с 2016 года. А это очень маловероятно.

В результате демографического перехода со второй половины XIX века в большинстве европейских стран смертность стремительно снижалась — сперва в ранних возрастах, затем во все более зрелых. На протяжении всего XX века и по сей день демографы чрезвычайно пессимистично оценивают перспективы увеличения ожидаемой продолжительности жизни. Однако в знаменитой статье 2002 года в журнале Science Джим Оеппен и Джим Вепель наглядно демонстрируют, как раз за разом успехи снижения смертности в разных странах опровергали сдержанные прогнозы демографов.

Получается, что, несмотря на привлекательное название, ожидаемая продолжительность жизни говорит вовсе не о будущем, а, скорее, представляет собой наиболее емкую характеристику уровня смертности на текущий момент. В условиях непрекращающегося снижения смертности реальная продолжительность жизни новорожденных когорт закономерно оказывается значительно больше оценок ожидаемой продолжительности жизни на момент их рождения.

Ожидаемая продолжительность жизни является один из ключевых показателей, входящих в расчет индекса развития человеческого потенциала - главного интегрального показателя для межстрановых сравнений уровня и качества жизни населения. Кроме того, ОПЖ населения страны и отдельных

ее регионов характеризует уровень жизни людей. Изменение ожидаемой продолжительности жизни может быть следствием проведения экономических, политических, социальных реформ, а также изменением идеологии общества.

Цель ВКР - изучение влияния социально - экономических факторов и характеристик социальной среды на ОПЖ российских регионов статистическими методами.

Для достижения поставленной цели необходимо решить след задачи:

1. Провести факторизацию пространства социально-экономических показателей
2. Выявить факторы, значимо влияющие на ОПЖ, оценить степень их влияния и, по возможности, дать интерпретацию факторов
3. Провести кластеризацию регионов РФ на основе значений обобщенных факторов и ОПЖ
4. Определить основные факторы, характеризующие каждый кластер и их влияние на ОПЖ.

1. Теоретическая часть

1.1. Задача снижения размерности

Обычно при исследовании некоторого явления или объекта рассматривается большое число параметров $\xi_1, \dots, \xi_k$, характеризующих данное явление или объект. Как правило, большинство этих параметров зависимы друг от друга, причем корреляционные связи между наблюдаемыми переменными могут определяться существованием меньшего числа реально ненаблюдаемых переменных или факторов. Кроме того, если число параметров k очень велико, то возникает проблема интерпретации результатов анализа всей совокупности. Поэтому у исследователя возникает естественное желание перейти к рассмотрению небольшого числа факторов (желательно, некоррелированных между собой), описывающих объект без существенной потери информативности. Основные цели, которые обычно преследуются в задачах снижения размерности можно разбить на две следующие группы:

- 1) построение обобщенных показателей для объекта исследования, характеризующих основную часть изменчивости исходных параметров и, соответственно, описание изучаемой системы значительно меньшим числом факторов;
- 2) поиск скрытых взаимосвязей между факторами, исследование внутренних причин формирующих специфику изучаемого явления, т.е. выявление скрытых обобщенных факторов которые стоят за соответствующими показателями (число которых опять так существенно меньше числа исходных параметров).

1.2. Метод главных компонент

Метод главных компонент осуществляет переход к новой совокупности некоррелированных признаков $\eta_1, \dots, \eta_k$, каждый из которых является линейной комбинацией исходных признаков $\xi_1, \xi_2, \dots, \xi_k$. При этом линейные комбинации выбираются таким образом, что среди всех возможных

линейных нормированных комбинаций исходных признаков первая главная компонента η_1 обладает наибольшей дисперсией. Вторая главная компонента имеет наибольшую дисперсию среди всех оставшихся линейных преобразований, некоррелированных с первой главной компонентой. Следующие главные компоненты определяются по аналогичной схеме. Геометрически это выглядит как ориентация новых координатных осей $\eta_1, \dots, \eta_k$ вдоль осей эллипсоида рассеивания исследуемой выборки в пространстве признаков $\xi_1, \xi_2, \dots, \xi_k$. При этом ось η_1 соответствует направлению наибольшей вытянутости эллипсоида рассеивания, ось η_2 интерпретируется как направление наибольшей вытянутости, перпендикулярное первой главной компоненте и так далее.

В дальнейшем из полученных величин можно оставить только m ($m < k$) наиболее значимых, вносящих максимальный вклад в суммарную дисперсию и использовать эти величины как некие интегральные факторы, характеризующие всю совокупность признаков.

Таким образом, от исходных признаков $\xi_1, \xi_2, \dots, \xi_k$, нам необходимо перейти к новым переменным $\eta_1, \dots, \eta_k$, которые попарно не коррелированы и упорядочены по убыванию дисперсии.

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_k)$ – центрированная многомерная случайная величина с матрицей ковариаций A . Положим

$\eta_i = \beta^{(i)T} \xi = \beta_1^{(i)} \xi_1 + \beta_2^{(i)} \xi_2 + \dots + \beta_k^{(i)} \xi_k$, где $\beta^{(i)}$ – векторы неизвестных коэффициентов преобразования, $i = \overline{1, m}$, ($m \leq k$). Будем называть величины η_i главными компонентами.

Так как величины ξ_i центрированы, то $M(\eta_i) = 0, i = \overline{1, m}$. Найдем дисперсию $D(\eta_i)$. Имеем:

$$D(\eta_i) = E(\eta_i \eta_i^T) = \beta^{(i)T} E(\xi \xi^T) \beta^{(i)} = \beta^{(i)T} A \beta^{(i)} \quad (1)$$

Определим коэффициенты преобразования для главных компонент η_i из условия максимума дисперсии $D(\eta_i)$. Поскольку данная задача (отыскания признаков с наибольшей дисперсией) может иметь решение только при ограничениях на коэффициенты преобразования $\beta^{(i)}$, необходимо наложить дополнительные условия, например на норму коэффициентов $\beta^{(i)}$. Будем искать решение задачи среди коэффициентов для которых

$$\|\beta^{(i)}\|^2 = \beta^{(i)T} \beta^{(i)} = 1. \quad (2)$$

Найдем первую компоненту η_1 из условия максимума дисперсии $D(\eta_1)$ при условии (2) с помощью метода множителей Лагранжа. Функция Лагранжа будет иметь вид:

$$L(\beta^{(1)}, \lambda) = \beta^{(1)T} A \beta^{(1)} - \lambda(\beta^{(1)T} \beta^{(1)} - 1). \quad (3)$$

Дифференцируя (4.3) по $\beta^{(1)}$ и приравнявая к нулю получаем:

$$\frac{\partial L}{\partial \beta^{(1)}} = 2A\beta^{(1)} - 2\lambda\beta^{(1)} = 0, \text{ или } (A - \lambda E)\beta^{(1)} = 0. \quad (4)$$

Последнее уравнение определяет задачу на собственные числа и собственные значения матрицы ковариаций A . Если многомерное распределение величины ξ не вырождено, то матрица ковариаций A положительно определена, и соответственно можно найти k положительных значений $\lambda_1, \lambda_2, \dots, \lambda_k$ и k линейно независимых векторов β . Выясним какое значение λ будет соответствовать наибольшей дисперсии $D(\eta_1)$. Для этого умножим (4) слева на $\beta^{(1)T}$: $\beta^{(1)T} A \beta^{(1)} - \lambda \beta^{(1)T} \beta^{(1)} = 0$, или $\lambda = D(\eta_1)$. Последнее означает, что для того, чтобы достичь максимума дисперсии, необходимо найти все собственные числа матрицы ковариаций A и выбрать максимальное.

Итак, если λ_1 - наибольшее собственное значение матрицы A , а $\beta^{(1)}$ - соответствующий этому значению собственный вектор, то

$$\eta_1 = \beta^{(1)T} \xi, \quad D(\eta_1) = \lambda_1$$

Для получения следующей компоненты η_2 потребуем некоррелированности η_2 с η_1 : $\text{cov}(\eta_1, \eta_2) = 0$. Найдем $\text{cov}(\eta_1, \eta_2)$, учитывая, что величины η_i центрированы:

$$\text{cov}(\eta_1, \eta_2) = E(\eta_1 \eta_2^T) = \beta^{(1)T} E(\xi \xi^T) \beta^{(2)} = \beta^{(1)T} A \beta^{(2)}.$$

Таким образом, ищем вектор $\beta^{(2)}$ из условия максимума дисперсии $D(\eta_2) = \beta^{(2)T} A \beta^{(2)}$, при условии нормировки на коэффициенты $\beta^{(2)T} \beta^{(2)} = 1$ и равенства нулю корреляции $\beta^{(1)T} A \beta^{(2)} = \beta^{(2)T} A \beta^{(1)} = 0$.

Как и ранее, решение осуществляем методом неопределенных множителей Лагранжа. Функция Лагранжа в этом случае будет иметь вид:

$$L(\beta^{(2)}, \lambda, \nu) = \beta^{(2)T} A \beta^{(2)} - \lambda(\beta^{(2)T} \beta^{(2)} - 1) - 2\nu \beta^{(2)T} A \beta^{(1)}.$$

Дифференцируя $L(\beta^{(2)}, \lambda, \nu)$ по $\beta^{(2)}$ и приравнявая к нулю, получаем:

$$A \beta^{(2)} - \lambda \beta^{(2)} - \nu A \beta^{(1)} = 0. \quad (5)$$

Умножая последнее равенство на $\beta^{(1)T}$ слева, получим:

$$\beta^{(1)T} A \beta^{(2)} - \lambda \beta^{(1)T} \beta^{(2)} - \nu \beta^{(1)T} A \beta^{(1)} = 0, \text{ или } \lambda \beta^{(1)T} \beta^{(2)} + \nu \beta^{(1)T} A \beta^{(1)} = 0 \quad (6)$$

Покажем, что $\beta^{(1)T} \beta^{(2)} = 0$. Действительно, по условию некоррелированности, $\beta^{(2)T} A \beta^{(1)} = 0$. Так как $\beta^{(1)}$ – собственный вектор матрицы ковариаций, соответствующий λ_1 , то $A \beta^{(1)} = \lambda_1 \beta^{(1)}$. Значит, $\beta^{(2)T} A \beta^{(1)} = \lambda_1 \beta^{(2)T} \beta^{(1)} = 0$, или $\beta^{(1)T} \beta^{(2)} = 0$, т.е. $\beta^{(1)}, \beta^{(2)}$ ортогональны друг другу.

Таким образом (6) принимает вид:

$$\nu \beta^{(1)T} A \beta^{(1)} = \nu D(\eta_1) = 0.$$

Последнее равенство возможно только тогда, когда параметр $\nu = 0$. Соответственно уравнение (5) примет вид

$$(A - \lambda E) \beta^{(2)} = 0.$$

То есть вектор $\beta^{(2)}$, удовлетворяющий заданным условиям, является также собственным вектором матрицы ковариаций, как и вектор $\beta^{(1)}$. Так как

$D(\beta^{(2)}) = \lambda$, то следует выбрать второе по величине собственное значение λ_2 (наибольшее λ_1 уже выбрано нами при нахождении $\beta^{(1)}$) с соответствующим ему собственным вектором $\beta^{(2)}$. Соответственно вторая компонента

$$\eta_2 = \beta^{(2)T} \xi, D(\eta_2) = \lambda_2.$$

Таким образом, получаем, что

$$\eta_i = \beta^{(i)T} \xi = \beta_1^{(i)} \xi_1 + \beta_2^{(i)} \xi_2 + \dots + \beta_k^{(i)} \xi_k, \quad i = \overline{1, k},$$

где $\beta^{(i)}$ есть собственный вектор матрицы ковариаций A , соответствующий собственному значению λ_i , причем $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$.

С точки зрения линейной алгебры переход к главным компонентам, по сути есть переход к новому базису, ортами которого служат собственные вектора матрицы ковариаций. Если ввести матрицу преобразования P , столбцами которой являются собственные векторы матрицы A - $\beta^{(i)}$, то формулы преобразования компонент при переходе от одного базиса к другому (учитывая ортогональность преобразования) примут вид: $\eta = P^T \xi$, $\xi = P \eta$. Из линейной алгебры известно, что $sp A$ является инвариантной величиной относительно ортогональных преобразований базиса. В исходном

$$\text{базисе: } sp A(\xi) = \sum_{i=1}^k D(\xi_i), \text{ в ортогональном базисе: } sp A(\eta) = \sum_{i=1}^k D(\eta_i) = \sum_{i=1}^k \lambda_i$$

$$\text{Следовательно, } \sum_{i=1}^k D(\xi_i) = \sum_{i=1}^k \lambda_i.$$

Оценка главных компонент на основе выборочных данных строится на основе выборочной матрицы ковариаций. Оценки собственных значений, являющиеся собственными числами выборочной матрицы ковариаций, в случае нормального распределения генеральной совокупности являются оценками максимального правдоподобия. Если единицы измерения исходных признаков различаются или их значения сильно различаются, то

лучше использовать при нахождении оценок главных компонент вместо выборочной матрицы ковариаций выборочную корреляционную матрицу.

1.3. Критерии отбора главных компонент

После нахождения всех компонент η_i задаемся вопросом, какие из них можно отбросить, чтобы уменьшить размерность вектора признаков η . Заметим, что математически строгих критериев отбора не существует. Обычно используют один из следующих эвристических методов. Во первых, зная $sp A = \sum_{i=1}^n \lambda_i$, можно выбрать те компоненты $\eta_i, i = \overline{1, m}$ из общего набора, которые бы объясняли не менее некоторой заданной доли q суммарной доли дисперсии признаков. Обычно q полагают не менее 0,7. Другим критерием отбора является критерий Кайзера, который предполагает использование для нахождения оценок собственных значений выборочной матрицы корреляций. Согласно данному критерию оставляют только те главные компоненты, дисперсия которых больше 1. По существу, это означает, что если фактор не выделяет дисперсию, эквивалентную, по крайней мере, дисперсии одной переменной, то он опускается.

Подход к оценке числа главных компонент по необходимой доле объяснённой дисперсии формально применим всегда, однако неявно он предполагает, что нет разделения на «сигнал» и «шум», и любая заранее заданная точность имеет смысл. Поэтому часто более продуктивна иная эвристика, основывающаяся на гипотезе о наличии «сигнала» (сравнительно малая размерность, относительно большая амплитуда) и «шума» (большая размерность, относительно малая амплитуда). С этой точки зрения метод главных компонент работает как фильтр: сигнал содержится, в основном, в проекции на первые главные компоненты, а в остальных компонентах пропорция шума намного выше. Вопрос, как оценить число необходимых главных компонент, если отношение «сигнал/шум» заранее неизвестно?

Одним из наиболее популярных подходов является правило сломанной трости (англ. Broken stick model). Набор нормированных собственных чисел ($\lambda_i^* / \text{tr} \bar{A}$, $i = \overline{1, k}$) сравнивается с распределением длин обломков трости единичной длины, сломанной в $k - 1$ случайно выбранной точке (точки разлома выбираются независимо и равномерно распределенными по длине трости). Пусть l_i ($i = \overline{1, k}$) - длины полученных кусков трости, занумерованные в порядке убывания длины: $l_1 \geq l_2 \geq \dots \geq l_k$. Тогда

математическое ожидание l_i : $L_i = M(l_i) = \frac{1}{k} \sum_{j=i}^k \frac{1}{j}$. По правилу сломанной

трости i -я компонента (в порядке убывания собственных чисел λ_i)

сохраняется в списке главных компонент, если $\frac{\lambda_1^*}{\text{tr} \bar{A}} > L_1, \frac{\lambda_2^*}{\text{tr} \bar{A}} > L_2, \dots, \frac{\lambda_i^*}{\text{tr} \bar{A}} > L_i$.

К подобным критериям относится также графический критерий каменистой осыпи Кэттелла. Критерий каменистой осыпи состоит в поиске точки, где убывание собственных значений замедляется наиболее сильно. Справа от этой точки должна находиться, по-видимому, только "факторная осыпь" ("осыпь" - это геологический термин для обломков, которые скапливаются в нижней части каменистого склона). Таким образом, число выделенных факторов не должно превышать количество факторов слева от этой точки (включая точку излома).

1.4. Классификация факторов. Вращение факторов

После определения факторов исследователю зачастую требуется оценить уровень информативности или вклад фактора в дисперсию каждого признака, а также выяснить, можно ли интерпретировать разумным образом полученные факторы и как это сделать.

Пусть имеется m - факторная модель. Пусть $f^{(p)}$ - некоторый фактор, $p \leq m$. Общностью фактора $f^{(p)}$ (или вкладом фактора в суммарную

дисперсию) признаков называется число $V^{(p)} = \sum_{j=1}^k [\alpha_j^{(p)}]^2$. Величину

$V_0 = \sum_{p=1}^m V^{(p)}$ называют суммарной общностью факторов $f^{(1)}, \dots, f^{(m)}$.

Отношение $V^{(p)} / V_0$ называется долей фактора $f^{(p)}$ в суммарной общности.

Пусть $\alpha_{s_1}^{(p)}, \alpha_{s_2}^{(p)}, \dots, \alpha_{s_l}^{(p)}$ факторные нагрузки фактора $f^{(p)}$ соответствующие признакам $\xi_{(s_1)}, \dots, \xi_{(s_l)}$. Тогда число

$$K_u = \frac{\sum_{j=1}^l (\alpha_{s_j}^{(p)})^2}{\|\alpha^{(p)}\|^2}$$

называется коэффициентом информативности признаков $\xi_{(s_1)}, \dots, \xi_{(s_l)}$. Данное число определяет, какой вклад в признаки $\xi_{(s_1)}, \dots, \xi_{(s_l)}$ вносит в фактор $f^{(p)}$. Если можно выделить какую либо группу признаков, коэффициент информативности которой для фактора $f^{(p)}$ существенно выше коэффициента информативности оставшихся признаков, то первая группа признаков и будет определять содержание фактора.

Если факторные нагрузки имеют более или менее равномерное распределение, то задача интерпретации фактора усложняется. В этом случае целесообразно прибегать к вращению факторов. Вспомним, что обобщенные факторы определяются с точностью до ортогонального преобразования. Следовательно, мы можем осуществлять ортогональное вращение факторного пространства, добиваясь такого расположения осей, при котором факторы допускают наиболее содержательную интерпретацию. Обычно используют следующие методы (стратегии) вращения, облегчающие интерпретацию факторов:

метод варимакс, при вращении максимизируют величину:

$$V = \sum_{j=1}^m \frac{1}{k} \sum_{i=1}^k \left((\alpha_i^{(j)})^2 - \frac{1}{k} \sum_{s=1}^k (\alpha_s^{(j)})^2 \right)^2,$$

которая характеризует различие столбцов матрицы факторных нагрузок;

метод квартимакс, при вращении максимизируют величину:

$$Q = \sum_{i=1}^k \frac{1}{m} \sum_{j=1}^m \left((\alpha_i^{(j)})^2 - \frac{1}{m} \sum_{s=1}^m (\alpha_i^{(s)})^2 \right)^2,$$

которая характеризует различие строк матрицы факторных нагрузок.

Учитывая, что метод варимакс основан на упрощении столбцов, а квартимакс на упрощении строк матрицы факторных нагрузок, можно построить обобщенный метод, основанный на максимизации величины $\alpha Q + \beta V$. В частности при $\alpha = \beta$ получим критерий, используемый в методе вращения биквартимакс.

Если C - матрица поворота, то в новой системе координат матрица факторных нагрузок определится как: $\alpha' = \alpha C$. Обычно в процедуре вращения осуществляют последовательное попарное ортогональное вращение осей. В этом случае полная матрица преобразования будет равна произведению матриц попарных поворотов. Например, для трехфакторной модели:

$$C = C_{12} C_{23} C_{31} = \begin{pmatrix} \cos \varphi_1 & -\sin \varphi_1 & 0 \\ \sin \varphi_1 & \cos \varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_2 & -\sin \varphi_2 \\ 0 & \sin \varphi_2 & \cos \varphi_2 \end{pmatrix} \cdot \begin{pmatrix} \sin \varphi_3 & 0 & \cos \varphi_3 \\ 0 & 1 & 0 \\ \cos \varphi_3 & 0 & -\sin \varphi_3 \end{pmatrix}$$

(в случае вращения против часовой стрелки).

Для того, чтобы избежать влияния на результаты вращения переменных с большой общностью применяют также нормировку факторных нагрузок, то есть, делят координаты векторов факторных нагрузок на соответствующие корни из общностей.

1.5. Кластерный анализ

Кластерный анализ — это анализ, позволяющий получить разбиение большого объема данных на классы или группы согласно некоторому

критерию или их совокупности. При этом объединяются в классы объекты близкие между собой в определенном смысле, причем предполагается отсутствие априорной информации о характеристиках объектов внутри классов.

Например, при сравнении стран, регионов, городов имеет смысл разбить всю совокупность изучаемых объектов на классы, близкие, например, по уровню жизни. При производстве товаров массового потребления имеет смысл разбить всех потенциальных покупателей на группы, близкие по потребительской способности и т.д.

В качестве основных задач, решаемых с помощью кластерного анализа, можно выделить:

- 1) классификацию объектов с учетом большого числа признаков, отражающих их сущность;
- 2) выявление скрытой структуры изучаемой совокупности объектов;
- 3) снижение размерности пространства изучаемых объектов.

1.5.1. Меры близости объекта

Пусть дана совокупность из n объектов, каждый из которых характеризуется k признаками $\xi_1, \xi_2, \dots, \xi_k$. Обозначим:

$X^{(i)} = \{X_1^{(i)}, X_2^{(i)}, \dots, X_k^{(i)}\}$ - вектор значений признаков для i -ого объекта, $X_j^{(i)}$ - значение j -го признака для i -го объекта. Определим меру близости (или меру однородности) объектов. В качестве меры близости будем использовать расстояние между объектами. Напомним, что расстоянием, или метрикой определенной на множестве M называется числовая функция $\rho(X, Y)$, $X, Y \in M$, удовлетворяющая следующим условиям:

- 1) $\rho(X, Y) = 0 \Leftrightarrow X = Y$;
- 2) $\rho(X, Y) = \rho(Y, X)$;
- 3) $\rho(X, Y) \leq \rho(X, Z) + \rho(Z, Y)$.

В теории кластерного анализа используются следующие метрики для измерения расстояния между отдельными объектами (точками в k-мерном пространстве):

1) Евклидово расстояние: $\rho(X^{(i)}, X^{(j)}) = \sqrt{\sum_{s=1}^k (X_s^{(i)} - X_s^{(j)})^2}$. Используется, если признаки однородны по физическому смыслу и их значения одинаково “весомы”.

2) Квадрат евклидова расстояния: $\rho(X^{(i)}, X^{(j)}) = \sum_{s=1}^k (X_s^{(i)} - X_s^{(j)})^2$. Применяется для придания большего веса более отдаленным друг от друга объектам.

3) Взвешенное евклидово расстояние: $\rho(X^{(i)}, X^{(j)}) = \sqrt{\sum_{s=1}^k w_s (X_s^{(i)} - X_s^{(j)})^2}$, $\sum_{s=1}^k w_s = 1$, где w_s — веса, пропорциональные важности признака в задаче классификации.

4) Хеммингово расстояние: $\rho(X^{(i)}, X^{(j)}) = \sum_{s=1}^k |X_s^{(i)} - X_s^{(j)}|$. Данную метрику также часто называют расстоянием городских кварталов (city-block). Используется для уменьшения влияния отдельных больших разностей (выбросов).

5) Расстояние Махаланобиса: $\rho(X^{(i)}, X^{(j)}) = \sqrt{(X^{(i)} - X^{(j)})^T A^{-1} (X^{(i)} - X^{(j)})}$, где A — матрица ковариаций признаков $\xi_1, \xi_2, \dots, \xi_k$ (или выборочная матрица ковариаций).

6) Расстояние Чебышева: $\rho(X^{(i)}, X^{(j)}) = \max_s |X_s^{(i)} - X_s^{(j)}|$.

7) Расстояние Минковского: $\rho(X^{(i)}, X^{(j)}) = \sqrt[p]{\sum_{s=1}^k |X_s^{(i)} - X_s^{(j)}|^p}$, $p \geq 1$.

8) Степенное расстояние или обобщенное расстояние Минковского:

$\rho(X^{(i)}, X^{(j)}) = \sqrt[r]{\sum_{s=1}^k |X_s^{(i)} - X_s^{(j)}|^p}$, $p \geq 1, r \geq 1$. Применяется в случае, когда необходимо увеличить или уменьшить вес, относящийся к размерности, для которой соответствующие объекты сильно отличаются. Параметр p

ответственен за постепенное взвешивание разностей по отдельным координатам, параметр r ответственен за прогрессивное взвешивание больших расстояний между объектами.

9) Процент несогласия $\rho(X^{(i)}, X^{(j)}) = \frac{1}{k} \sum_{s=1}^k I[X_s^{(i)} \neq X_s^{(j)}]$, где

$$I[X_s^{(i)} \neq X_s^{(j)}] = \begin{cases} 1, & X_s^{(i)} \neq X_s^{(j)} \\ 0, & X_s^{(i)} = X_s^{(j)} \end{cases}. \text{ Используется для категориальных данных.}$$

В качестве меры тесноты связи между объектами может выступать и величина коэффициента корреляции между объектами, рассматриваемыми как переменные с наблюдаемыми значениями, равными значениям соответствующих признаков:

$$r(X^{(i)}, X^{(j)}) = \frac{\sum_{s=1}^k (X_s^{(i)} - \bar{X}^{(i)})(X_s^{(j)} - \bar{X}^{(j)})}{\sqrt{\sum_{s=1}^k (X_s^{(i)} - \bar{X}^{(i)})^2 \sum_{s=1}^k (X_s^{(j)} - \bar{X}^{(j)})^2}}.$$

Чем ближе по абсолютной величине значение коэффициента корреляции к единице, тем более близки объекты между собой (обычно в качестве характеристики тесноты связи используют величину $1 - r$).

Оценка сходства между объектами сильно зависит от абсолютного значения признака и от степени его вариации в совокупности. Если абсолютные значения для различных признаков сильно различаются между собой, следует проводить нормировку исходных данных, преобразуя их, например, к стандартным величинам с выборочным средним равным нулю и выборочной дисперсией равной единице.

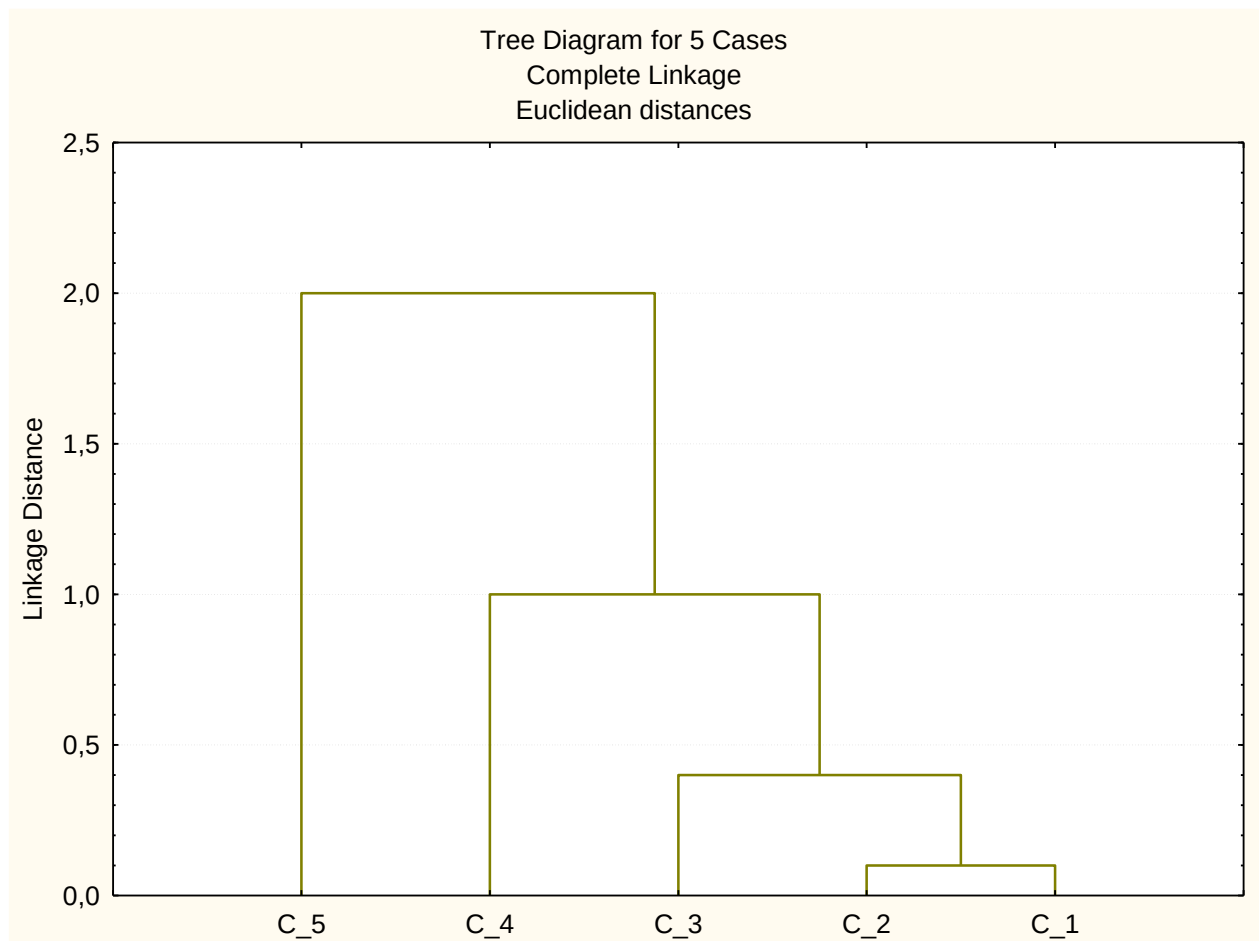
Выбор метрики осуществляется исследователем в зависимости от того, какой результат он хочет получить. Этот выбор неформализуем, так как зависит от многих факторов, в частности, от ожидаемого результата, от опыта исследователя, уровня его математической подготовки и т.д.

1.5.2. Иерархические агломеративные методы кластерного анализа

Иерархические (т.е. деревообразующие) агломеративные (т.е. объединяющие) методы являются наиболее простыми и

распространенными среди алгоритмов кластерного анализа. На первом шаге каждый из объектов $X^{(1)}, \dots, X^{(n)}$ исходных данных рассматривается как отдельный кластер. По вычисленной матрице расстояний $R = \| \rho(X^{(i)}, X^{(j)}) \| = \| \rho_{ij} \|$, ρ – некоторая метрика, на каждом шаге объединяются наиболее близкие друг к другу кластеры. При этом, каждый раз пересчитывается матрица расстояний (заметим, что при этом размерность матрицы на каждом шаге уменьшается на единицу). Очевидно, что процесс завершится через $(n-1)$ шаг, когда в результате все объекты будут объединены в один кластер. Процесс объединения может быть прерван на каком либо шаге, если изначально задать пороговое максимальное значение расстояния для объединения кластеров. Если расстояние между кластерами превосходит эту величину, то объединения кластеров не происходит.

Последовательность объединений можно представить в виде дендрограммы или дерева. На рисунке показано, что на первом шаге были объединены вектора $X^{(1)}, X^{(2)}$, так как расстояние между ними 0,1. На втором шаге к ним был присоединен вектор $X^{(3)}$, отстоящий от кластера $\{X^{(1)}, X^{(2)}\}$ на расстояние 0,4 и т.д. Очевидно, на последнем шаге объединяются все вектора в один кластер.



Иерархические методы являются наиболее простыми методами кластерного анализа и используются в первую очередь для анализа структуры множества объектов. В зависимости от того, как определяется расстояние между кластерами (как происходит объединение кластеров), выделяют методы одиночной и полной связи, методы взвешенных и невзвешенных попарных средних, взвешенный и невзвешенный центроидные методы, метод Варда.

1.5.2.1. Метод одиночной связи

В методе одиночной связи расстояние между кластерами определяется по методу “ближнего соседа”, то есть за расстояние между кластерами принимается расстояние между самыми близкими элементами двух кластеров: $\rho(K_i, K_j) = \min_{\substack{X_s \in K_i \\ X_k \in K_j}} \rho(X^{(s)}, X^{(k)})$.

К достоинствам данного метода следует отнести то, что он не чувствителен к монотонному преобразованию матрицы расстояний между исходными объектами. К недостаткам следует отнести то, что зачастую данный метод приводит к появлению длинных цепочек кластеров.

1.5.2.2. Метод полной связи

В методе полной связи расстояние между кластерами вычисляется по принципу «дальнего соседа», то есть за расстояние между кластерами принимается расстояние между самыми удаленными элементами двух кластеров: $\rho(K_i, K_j) = \max_{\substack{X_s \in K_i \\ X_k \in K_j}} \rho(X^{(s)}, X^{(k)})$.

1.5.2.3. Метод Варда

В методе Варда объединение кластеров происходит таким образом, что сумма квадратов расстояний объектов до центров кластеров получает наименьшее приращение. То есть, для каждого кластера K_s , состоящего из объектов $X^{(1)}, \dots, X^{(n_s)}$, в соответствии с выбранной метрикой, вычисляется величина $D_s = \sum_{i=1}^{n_s} \rho^2(X^{(i)}, \bar{X})$, где $\bar{X}$ - центр кластера K_s . Данная величина, по сути, характеризует внутрикластерную дисперсию для кластера K_s . Для совокупности кластеров $\{K_1, K_2, \dots, K_q\}$, соответственно, вычисляется суммарная сумма квадратов отклонений $V_q = \sum_{s=1}^q D_s$. На первом шаге все величины $D_s = 0, s = \overline{1, n}$, и на каждом шаге объединяются те кластеры, которые приводят к наименьшему изменению величины V_q после объединения (для этого потребуется рассмотреть все возможные объединения кластеров). Метод Варда на каждом шаге приводит к формированию кластеров с наименьшей внутрикластерной дисперсией.

1.6. Дисперсионный анализ

Дисперсионный анализ — статистический метод, позволяющий анализировать влияние различных факторов на результаты эксперимента,

путем исследования значимости различий в средних значениях. Суть дисперсионного анализа (analysis of variance — сокращенно ANOVA) заключается в разложении дисперсии измеряемого признака на независимые слагаемые, каждое из которых характеризует влияние того или иного фактора или их взаимодействия. Последующее сравнение таких слагаемых позволяет определить, какая часть вариации экспериментальных данных обусловлена влиянием факторов, а какая может быть отнесена на счет случайных отклонений и, соответственно, оценить значимость каждого изучаемого фактора, а также комбинаций факторов.

В терминах метода дисперсионного анализа та переменная, которая, как мы считаем, должна оказывать влияние на конечный результат, называется фактором. Например, если мы начнем строить модель объяснения различий в доходах различных опрошенных респондентов тем, что респонденты проживают в различных населенных пунктах, то переменная “место проживания респондента” будет выступать фактором. Конкретное значение фактора (например, определенный населенный пункт) называют уровнем фактора. Значение измеряемого признака (в нашем примере — величину дохода) называют откликом.

Если исследуется зависимость отклика только от одного фактора, то такой дисперсионный анализ называется однофакторным, если исследуется зависимость от двух и более факторов, то такой дисперсионный анализ называется многофакторным.

1.6.1. Однофакторный дисперсионный анализ с фиксированными эффектами

В однофакторной модели дисперсионного анализа исходят из следующей модели порождения данных:

$$y_{ij} = \mu_j + \varepsilon_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \quad i = \overline{1, n_j}, \quad j = \overline{1, k},$$

где: y_{ij} - i -ое наблюдаемое значение отклика в j -ой группе (для j -го уровня фактора);

μ - среднее значение отклика по всем уровням фактора (среднее по всей совокупности);

μ_j - среднее значение отклика для j -го уровня фактора;

$\alpha_j = \mu_j - \mu$ - дифференциальный эффект среднего, соответствующий j -му уровню фактора;

ε_{ij} - независимые случайные величины с математическим ожиданием равным нулю и одинаковой дисперсией σ^2 .

Очевидно, что при заданных величинах μ_j , величины α_j и μ определяются неоднозначно, поэтому необходимо наложить дополнительное условие, устанавливающее связь между этими величинами. Обычно используют одно из следующих условий: $\sum_{i=1}^k \alpha_i = 0$, $\sum_{i=1}^k n_i \alpha_i = 0$, $\alpha_k = 0$.

Выражение $y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$ можно представить в виде

$$y_{ij} = \mu + (\mu_j - \mu) + (y_{ij} - \mu_j),$$

или:

$$y_{ij} - \mu = (\mu_j - \mu) + (y_{ij} - \mu_j).$$

Данное соотношение говорит о том, что отклонение наблюдаемого значения отклика для j -ой группы складывается из суммы двух слагаемых: отклонения отклика от среднего значения j -ой группы: $(y_{ij} - \mu_j)$, и отклонения среднего значения j -ой группы от среднего значения всей совокупности: $(\mu_j - \mu)$. По сути это означает, что дисперсия отклика может быть представлена в виде суммы двух дисперсий, одна из которых характеризует внутригрупповую изменчивость, а вторая межгрупповую.

Разложение общей дисперсии на составляющие для выборочных данных обычно записывается в виде равенства сумм квадратов соответствующих отклонений:

$$SS_T = SS_B + SS_R,$$

где:

$SS_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$ – общая, или полная, сумма квадратов отклонений;

$SS_B = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$ – сумма квадратов отклонений групповых средних от общего среднего, или межгрупповая (межуровневая факторная) сумма квадратов отклонений, также называемая суммой квадратов эффекта фактора или просто эффектом фактора;

$SS_R = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ – сумма квадратов отклонений наблюдений от групповых средних, или внутригрупповая (остаточная) сумма квадратов отклонений, также называемая остаточным эффектом или эффектом ошибок;
 k – число уровней фактора,

n_j – число наблюдений для j -го уровня фактора,

$n = \sum_{j=1}^k n_j$ – общее число наблюдений.

В разложении дисперсии на составляющие заключена основная идея дисперсионного анализа: общая вариация переменной, порожденная влиянием фактора и измеренная суммой SS_T , складывается из двух компонент: SS_B и SS_R , характеризующих изменчивость этой переменной между уровнями фактора (SS_B) и внутри уровней фактора (SS_R).

В дисперсионном анализе анализируются не сами суммы квадратов отклонений, а так называемые средние квадраты, которые получаются делением сумм квадратов отклонений на соответствующее число степеней свободы. Число степеней свободы для суммы квадратов случайных величин определяется как общее число линейно независимых слагаемых.

Для полной суммы квадратов $SS_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$ число степеней свободы $\nu_T = n - 1$, так как при ее расчете используются n наблюдений,

связанных между собой одним уравнением для общего выборочного среднего всей совокупности.

Для суммы квадратов эффекта фактора $SS_B = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$ число степеней свободы $v_B = k - 1$, так как при ее расчете используются k групповых средних, связанных между собой также одним уравнением для общего выборочного среднего всей совокупности.

Для суммы квадратов ошибок $SS_R = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ число степеней свободы $v_R = n - k$, ибо при его расчете используются n наблюдений, связанных между собой k уравнениями для выборочных средних k групп.

Соответственно выражения для средних квадратов отклонений, которые являются оценками соответствующих дисперсий, имеют вид:

$$MS_T = \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \text{ - оценка общей дисперсии;}$$

$$MS_B = \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 \text{ - оценка межгрупповой дисперсии;}$$

$$MS_R = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \text{ - оценка остаточной дисперсии } \sigma^2.$$

При условии истинности $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$, статистики MS_B и MS_R , являются несмещенными оценками одной и той же дисперсии σ^2 . Поэтому проверка гипотезы H_0 сводится к проверке гипотезы о равенстве двух дисперсий на основе F - отношения:

$$F = \frac{MS_B}{MS_R} = \frac{n-k}{k-1} \frac{SS_B}{SS_R}$$

При истинности H_0 статистика F в случае нормального распределения величин ε_{ij} имеет распределение Фишера с $\nu_1 = k - 1$ и $\nu_2 = n - k$ числом степеней свободы. Если наблюдаемое значение статистики $F_{набл} \geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня α (или квантиль уровня $1 - \alpha$) с числом степеней свободы $\nu_1 = k - 1$ и $\nu_2 = n - k$, то нулевая гипотеза отклоняется и считается, что средние для различных уровней фактора значимо различаются.

Результаты дисперсионного анализа принято отображать в виде таблицы:

Источник дисперсии	Сумма квадратов	Степени свободы	Средние квадраты	F-отношение
Между группами	$SS_B = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$	$k - 1$	$MS_B = \frac{SS_B}{k - 1}$	$F = \frac{MS_B}{MS_R}$
Остаточная	$SS_R = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$n - k$	$MS_R = \frac{SS_R}{n - k}$	
Полная	$SS_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$n - 1$	$MS_T = \frac{SS_T}{n - 1}$	

Для сумм квадратов можно построить также следующую таблицу:

Сумма квадратов		Степени свободы	Средняя
Отклонений между группами	$SS_B = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$	$k - 1$	$MS_B = \frac{SS_B}{k - 1}$
Отклонений внутри групп	$SS_R = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$n - k$	$MS_R = \frac{SS_R}{n - k}$
Отклонений, полная	$SS_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$n - 1$	$MS_T = \frac{SS_T}{n - 1}$
Генерального среднего	$n\bar{y}^2$	1	$n\bar{y}^2$
Полная	$\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2$	n	$\frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2$

1.6.2. Непараметрический дисперсионный анализ

Для порядковых данных непараметрической альтернативой однофакторного дисперсионного анализа являются ранговый дисперсионный анализ Краскела–Уоллиса.

В основе метода дисперсионного анализа Краскела — Уоллиса лежит однофакторный дисперсионный анализ, в котором вместо значений переменных используется ранг переменных.

Если обозначить через R_{ij} ранг элемента x_{ij} , в общем вариационном ряду значений отклика, то величины $\bar{R}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} R_{ij}$ будут определять средние ранги для элементов j -ой группы, а величина $\bar{R} = \frac{1}{n} \sum_{i=1}^{n_j} \sum_{j=1}^k R_{ij} = \frac{n+1}{2}$ средний ранг всей совокупности. Соответственно, величина $\sum_{j=1}^k n_j (\bar{R}_j - \bar{R})^2$ будет характеризовать межгрупповой разброс рангов.

При условии истинности гипотезы H_0 равенства средних рангов групп, статистика

$$H = \frac{12}{n(n+1)} \sum_{j=1}^k n_j (\bar{R}_j - \bar{R})^2$$

будет иметь приближенно распределение Хи-квадрат с $k-1$ степенью свободы.

Если наблюдаемое значение статистики $H_{набл} \geq H_{кр}$, где $H_{кр}$ - критическая точка распределения Хи-квадрат с числом степеней свободы $k-1$ уровня α (или квантиль уровня $1-\alpha$), то нулевая гипотеза отклоняется и считается, что средние ранги для различных уровней фактора значимо различаются.

2. Практическая часть

Цель ВКР - изучение влияния социально - экономических факторов и характеристик социальной среды на ОПЖ российских регионов. Используя данные государственной статистики, мы осуществляем анализ связей между этими факторами и показателем продолжительности жизни в субъектах Российской Федерации.

Для выполнения задач исследования была использована информация о макро экономических и социальных характеристиках по 85-ти из 89-ти регионов РФ. Ограничение выборки было обусловлено неполнотой данных по ряду территориальных образований и исключением оставшихся регионов, оказывающих очень сильное влияние на результаты анализа, способное исказить закономерности, обнаруживаемые в большинстве субъектов Российской Федерации.

Выбрали следующие экономические параметры:

- Число разводов на 1000 браков;
- Численность занимавшихся в физкультурно-оздоровительных клубах, секциях и группах на 1000 населения;
- Продажа алкогольных напитков и пива населению по субъектам российской федерации на 1 жителя чистого спирта мл;
- Доля населения с доходами ниже прожиточного минимума (в процентах от общей численности населения);
- Коэффициент Джини;
- Соотношение среднедушевого дохода и прожиточного минимума;
- Число собственных легковых автомобилей на 1000 человек населения;
- Уровень безработицы (в процентах от численности экономически активного населения);

- Доля занятого населения с высшим образованием (в процентах);
- Среднедушевые месячные доходы (руб.);
- Инвестиции в основной капитал на душу населения, руб.;
- Общая площадь жилых помещений, приходящаяся в среднем на одного жителя (кв. м.);
- Удельный вес ветхого и аварийного жилищного фонда в общей площади всего жилищного фонда;
- Удельный вес общей площади, оборудованной водопроводом;
- Доля городского населения (в процентах);
- Число спортивных сооружений на 10 тыс. человек;
- Число зарегистрированных преступлений на 100 тыс. человек;
- Выбросы загрязняющих веществ в атмосферу (тонн на 1000 человек);
- Число дорожно-транспортных происшествий на 100 000 человек населения;
- Численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом на 1000 работающих;
- Количество посещений в поликлинику с профилактической целью на 1 чел.;
- Число больничных коек на 10 тыс. человек;
- Число посещений в смену на 10 тыс. человек;
- Число врачей на 10 тыс. человек;
- Число среднего медицинского персонала на 10 тыс. человек.

Описанные выше экономические и социальные показатели составляют набор переменных, используемых в дальнейшем при проведении анализа.

Таблица 1. Матрица корреляций исходных признаков и ОПЖ
(часть таблицы)

Variable	Correlations (Spreadsheet17)																
ОПЖ	ОПЖ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.000000	-0.418446	0.021969	-0.468599	-0.122376	0.129803	0.188521	-0.061483	0.253051	0.409298	-0.127516	-0.009238	-0.181722	-0.106681	0.420025	-0.209517	-0.307222
2	-0.418446	1.000000	0.088424	0.518831	-0.311290	0.041879	0.049789	0.484385	-0.551631	-0.136404	0.249008	0.229990	0.597797	-0.159186	-0.025026	0.352553	0.354568
3	0.021969	0.088424	1.000000	0.107824	-0.228018	0.355114	0.402540	0.184305	-0.202146	0.121172	0.239939	0.497810	0.083542	0.147582	0.048082	0.006858	0.186002
4	-0.468599	0.518831	0.107824	1.000000	-0.359307	0.191991	0.245458	0.379629	-0.548360	-0.017749	0.550166	0.308583	0.341140	0.090426	0.042871	0.669555	0.076586
5	-0.122376	-0.311290	-0.228018	-0.359307	1.000000	-0.532162	-0.818120	-0.332479	0.702583	-0.078423	-0.408071	-0.229125	-0.536443	0.312299	-0.472371	-0.503991	-0.003359
6	0.129803	0.041879	0.355114	0.191991	-0.532162	1.000000	0.797806	0.069094	-0.281970	0.158097	0.504471	0.394053	0.032192	0.022787	0.221267	0.198954	-0.012026
7	0.188521	0.049789	0.402540	0.245458	-0.818120	0.797806	1.000000	0.133560	-0.507827	0.236209	0.574828	0.335656	0.243439	-0.105313	0.415283	0.340745	-0.034415
8	-0.061483	0.484385	0.184305	0.379629	-0.332479	0.069094	0.133560	1.000000	-0.389136	0.003022	0.100584	-0.039479	0.558218	-0.200046	0.151191	0.333485	0.158884
9	0.253051	-0.551631	-0.202146	-0.548360	0.702583	-0.281970	-0.507827	-0.389136	1.000000	-0.158542	-0.314358	-0.101198	-0.583007	0.401637	-0.241621	-0.581068	-0.230637
10	0.409298	-0.136404	0.249008	0.229990	-0.017749	-0.078423	0.158097	0.236209	0.003022	-0.158542	1.000000	0.266501	0.108808	-0.141952	-0.176264	0.387571	0.184909
11	-0.127516	0.249008	0.239939	0.550166	-0.408071	0.504471	0.574828	0.100584	-0.314358	0.266501	1.000000	0.668484	0.009786	0.261812	0.304816	0.435850	-0.189012
12	-0.009238	0.229990	0.497810	0.308583	-0.229125	0.394053	0.335656	-0.039479	-0.101198	0.108808	0.668484	1.000000	-0.074211	0.366850	0.086097	0.178851	-0.036820
13	-0.181722	0.597797	0.083542	0.341140	-0.536443	0.332192	0.243439	0.558218	-0.583007	-0.141952	0.009786	-0.074211	1.000000	-0.377048	0.079265	0.322827	0.360832
14	-0.106681	-0.159186	0.147582	0.090426	0.312299	0.022787	-0.105313	-0.200046	0.401637	-0.176264	0.261812	0.366850	-0.377048	1.000000	-0.275115	-0.074102	-0.140638
15	0.420025	-0.025026	0.048082	0.042871	-0.472371	0.221267	0.415283	0.151191	-0.241621	0.387571	0.304816	0.086097	0.079265	-0.275115	1.000000	0.386093	-0.310521
16	-0.209517	0.352553	0.006858	0.669555	-0.503991	0.198954	0.340745	0.333485	-0.581068	0.184909	0.435850	0.178851	0.322827	-0.074102	0.386093	1.000000	-0.120244
17	-0.307222	0.354568	0.186002	0.076586	-0.003359	-0.012026	-0.034415	0.158884	-0.230637	-0.324990	-0.189012	-0.036820	0.360832	-0.140638	-0.310521	-0.120244	1.000000
18	-0.735126	0.273537	-0.014998	0.503079	0.243496	-0.049351	-0.213357	0.110423	-0.005774	-0.254765	0.150234	0.084778	-0.104459	0.267219	-0.430456	0.247944	0.173239
19	-0.132086	0.307361	0.310423	0.344928	-0.171735	0.347838	0.214045	-0.029288	-0.071873	-0.044344	0.580197	0.870868	-0.058696	0.341425	0.037287	0.234261	-0.026269
20	-0.452064	0.400844	-0.055712	0.337167	0.048481	-0.165179	-0.226416	0.328942	-0.327366	-0.099098	-0.165713	-0.135109	0.366675	-0.102184	-0.257292	0.233454	0.281492
21	-0.589919	0.513807	0.030346	0.504967	0.075387	-0.154486	-0.225332	0.273573	-0.172345	-0.274212	0.074906	-0.005016	0.216141	0.112093	-0.185467	0.326114	0.197512
22	-0.214980	0.334154	0.259133	0.263054	-0.023979	0.205836	0.147003	-0.019900	-0.136326	0.026286	0.513048	0.565979	-0.084394	0.377063	-0.003359	0.099450	0.062225
23	-0.703632	0.347570	0.005127	0.351059	0.133296	-0.092228	-0.093546	0.128666	-0.149122	-0.192022	0.339352	0.115396	0.125551	0.281912	-0.156025	0.253613	0.159037
24	-0.533277	0.452250	0.056090	0.444142	-0.058672	0.048207	0.072504	0.147701	-0.306379	-0.131163	0.347579	0.010813	0.299981	0.026132	-0.054180	0.336600	0.190744
25	-0.089310	0.118515	0.021782	0.213220	-0.075726	0.207130	0.209287	0.030966	-0.170174	0.350664	0.443062	0.160631	-0.041180	0.138885	0.227314	0.357997	-0.153284
25	-0.453889	0.387898	0.154739	0.331812	0.117858	0.042970	-0.024310	0.058623	-0.097910	-0.082395	0.368934	0.309532	0.034999	0.411197	-0.111632	0.204550	0.191489

Как видим, исходные признаки сильно коррелируют, как между собой, так и с ОПЖ, что не позволяет однозначно оценить их степень влияния на ОПЖ и затрудняет кластеризацию, поэтому проведем факторизацию пространства признаков методом главных компонент.

Получили следующие собственные значения и, соответственно, доли выделенной главными компонентами дисперсии (для всех факторов):

Таблица 2: Собственные значения и доли выделенной главными компонентами дисперсии

Value	Eigenvalues (диплом) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	6,270917	25,08367	6,27092	25,0837
2	4,163715	16,65486	10,43463	41,7385
3	3,572327	14,28931	14,00696	56,0278
4	1,945168	7,78067	15,95213	63,8085
5	1,232856	4,93142	17,18498	68,7399
6	1,071931	4,28772	18,25691	73,0277
7	0,975413	3,90165	19,23233	76,9293
8	0,791120	3,16448	20,02345	80,0938
9	0,665648	2,66259	20,68909	82,7564
10	0,561361	2,24545	21,25046	85,0018
11	0,531765	2,12706	21,78222	87,1289
12	0,500175	2,00070	22,28240	89,1296
13	0,447183	1,78873	22,72958	90,9183
14	0,392663	1,57065	23,12224	92,4890
15	0,351702	1,40681	23,47394	93,8958
16	0,297920	1,19168	23,77186	95,0875
17	0,275061	1,10024	24,04693	96,1877
18	0,203594	0,81438	24,25052	97,0021
19	0,200241	0,80097	24,45076	97,8030
20	0,169914	0,67966	24,62068	98,4827
21	0,143023	0,57209	24,76370	99,0548
22	0,105690	0,42276	24,86939	99,4776
23	0,060571	0,24229	24,92996	99,7198
24	0,037607	0,15043	24,96757	99,8703
25	0,032434	0,12973	25,00000	100,0000

Перейдем от пространства исходных признаков (показателей) к факторному пространству главных компонент, объясняющих не менее 70% вариации исходных признаков. Исходя из полученных данных таблицы 2 видим, что минимальное количество факторов, объясняющих не менее 70% дисперсии – 6.

Построим график собственных значений в зависимости от порядкового номера (метод каменистой осыпи):

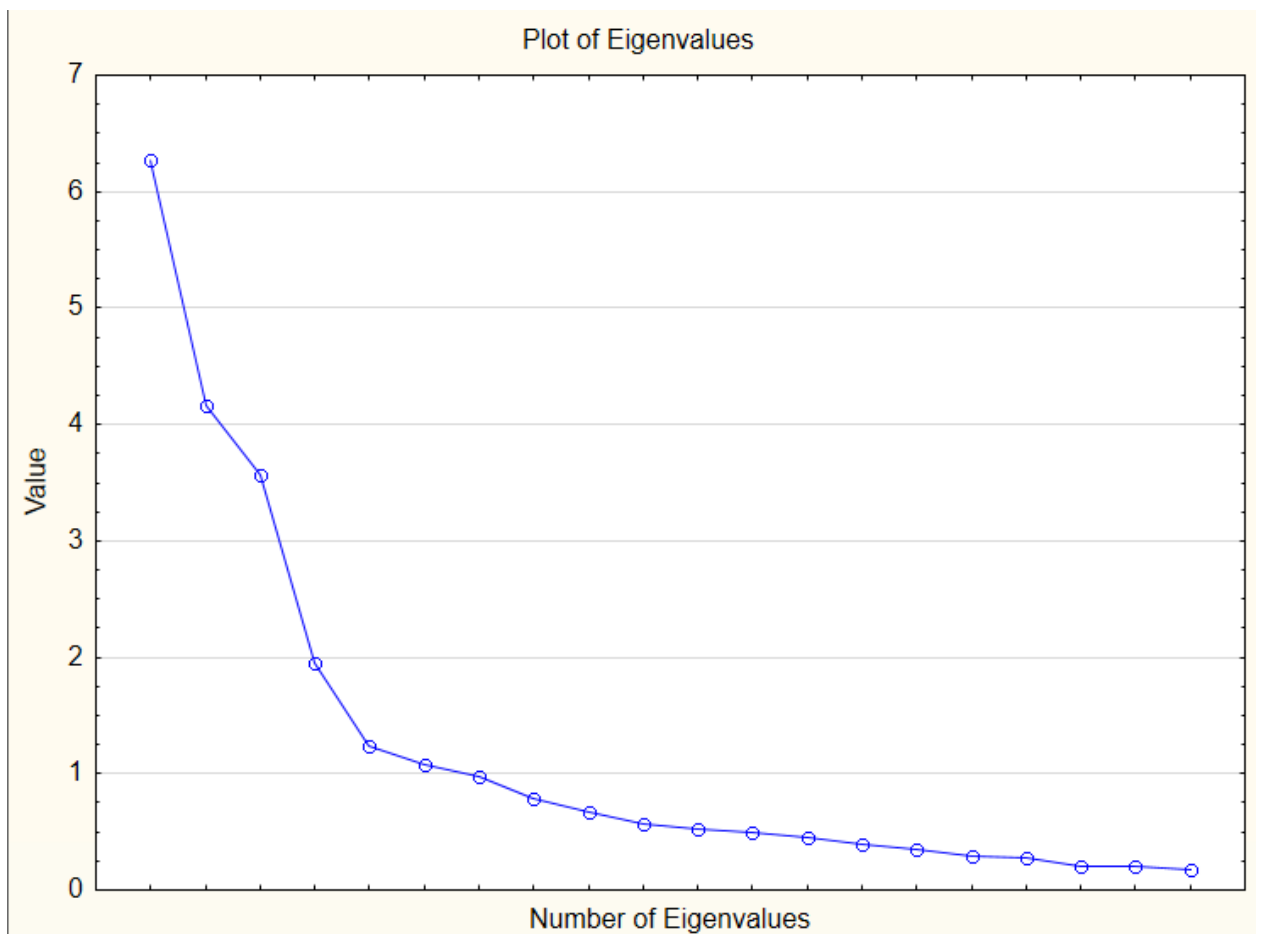


Рисунок 1. Метод каменистой осыпи

Согласно рисунку 1- ярко выражены 4 компоненты.

Итак, возможны варианты от шестифакторной до четырехфакторной модели, рассмотрим все варианты.

Рассмотрим шестифакторную модель.

При помощи пакета STATISTICA посчитаем факторные нагрузки:

Таблица 3. Факторные нагрузки (до вращения) шестифакторной модели

Variable	Factor Loadings (Unrotated) (дипном) Extraction: Principal components (Marked loadings are > .700000)					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Число разводов на 1000 браков	-0.682849	-0.250093	-0.350659	0.188774	-0.027496	0.307431
Численность занимавшихся в физкультурно-оздоровительных клубах, секциях и группах на 1000 населения	-0.345196	0.246158	0.191905	0.481748	-0.173872	0.104816
Продажа алкогольных напитков и пива населению по субъектам российской федерации на 1 жителя чистого спирта мл	-0.777161	-0.125701	-0.139494	-0.103357	0.401406	-0.085760
Доля населения с доходами ниже прожиточного минимума (в процентах от общей численности населения)	0.519147	-0.652533	0.366935	-0.120519	-0.076066	0.179325
Коэффициент Джини	-0.423033	0.566243	0.182737	0.199970	-0.066253	-0.423941
Соотношение среднедушевого дохода и прожиточного минимума	-0.483727	0.739051	-0.008189	0.115379	-0.119180	-0.377376
Число собственных легковых автомобилей на 1000 человек населения	-0.418275	-0.005721	-0.513957	0.062964	0.102442	0.243982
Уровень безработицы (в процентах от численности экономически активного населения)	0.650899	-0.262145	0.523312	-0.005200	0.078185	0.006310
Доля занятого населения с высшим образованием (в процентах)	-0.044155	0.483247	0.095445	-0.468410	-0.163590	0.377197
Среднедушевые месячные доходы (руб.)	-0.723277	0.315677	0.407953	-0.143649	0.080860	-0.052090
Инвестиции в основной капитал на душу населения, руб.	-0.533197	0.241118	0.575499	0.349152	0.194752	0.270109
Общая площадь жилых помещений, приходящаяся в среднем на одного жителя (кв. м.)	-0.432841	0.031701	-0.706631	0.213266	-0.117361	0.096709
Удельный вес ветхого и аварийного жилищного фонда в общей площади всего жилищного фонда	-0.104944	-0.286853	0.705020	0.076196	0.164408	-0.106639
Удельный вес общей площади, оборудованной водопроводом	-0.171850	0.642786	-0.095228	-0.389134	-0.024543	0.229182
Доля городского населения (в процентах)	-0.666583	0.144576	-0.230737	-0.406105	0.326794	-0.034248
Число спортивных сооружений на 10 тыс. человек	-0.176006	-0.326627	-0.307159	0.562600	-0.403779	-0.071520
Число зарегистрированных преступлений на 100 тыс. человек	-0.408092	-0.663316	0.139503	-0.087604	0.231014	-0.268680
Выбросы загрязняющих веществ в атмосферу (тонн на 1000 человек)	-0.557813	0.080671	0.545414	0.300711	0.252549	0.218783
Число дорожно-транспортных происшествий на 100 000 человек населения	-0.265946	-0.468199	-0.463991	-0.015131	0.071985	0.099680
Численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и бо	-0.481103	-0.608757	-0.154456	-0.056618	0.201166	-0.003974
Количество посещений в поликлинику с профилактической целью на 1 чел.	-0.560509	-0.076408	0.497991	0.149902	-0.241737	0.246138
Число больничных коек на 10 тыс. человек	-0.506280	-0.514848	0.142675	-0.214955	-0.197528	-0.127756
Число посещений в смену на 10 тыс. человек	-0.606870	-0.383440	-0.074853	-0.214319	-0.282573	-0.194301
Число врачей на 10 тыс. человек	-0.469626	0.015543	0.269903	-0.570885	-0.411612	-0.009725
Число среднего медицинского персонала на 10 тыс. человек	-0.590432	-0.439954	0.352587	-0.082420	-0.328276	0.054453
Expl.Var	6.270917	4.163715	3.723237	1.945168	1.232856	1.071931
Prp.Totl	0.250837	0.166549	0.142893	0.077807	0.049314	0.042877

Исходя из данных таблицы 3, можно заметить, что для четвертого, пятого и шестого факторов вообще сложно выделить существенно значимые координаты. Поэтому обратимся к процедуре вращения факторов.

Факторные нагрузки для шестифакторной модели после вращения:

Таблица 4. Факторные нагрузки (после вращения) шестифакторной модели

Variable	Factor Loadings (Varimax raw) (дипном) Extraction: Principal components (Marked loadings are > .700000)					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Число разводов на 1000 браков	0.241814	-0.036924	0.272354	0.085371	0.248782	0.761250
Численность занимавшихся в физкультурно-оздоровительных клубах, секциях и группах на 1000 населения	-0.265733	0.318862	0.512709	0.109795	0.022006	0.205389
Продажа алкогольных напитков и пива населению по субъектам российской федерации на 1 жителя чистого спирта мл	0.739592	0.211345	0.257673	0.047842	0.211765	0.336468
Доля населения с доходами ниже прожиточного минимума (в процентах от общей численности населения)	-0.175903	-0.787394	-0.058544	0.178374	0.166066	-0.410692
Коэффициент Джини	-0.006276	0.826081	0.254145	0.033647	0.051198	-0.079333
Соотношение среднедушевого дохода и прожиточного минимума	-0.009550	0.949220	0.151026	-0.139765	0.019875	0.079615
Число собственных легковых автомобилей на 1000 человек населения	0.252970	0.044323	0.017745	-0.079137	-0.042317	0.662418
Уровень безработицы (в процентах от численности экономически активного населения)	-0.233465	-0.478923	0.022117	0.144071	-0.129209	-0.671275
Доля занятого населения с высшим образованием (в процентах)	-0.118221	0.089473	0.049144	-0.776539	0.067901	-0.031829
Среднедушевые месячные доходы (руб.)	0.316512	0.479754	0.548048	-0.257728	0.344164	-0.059338
Инвестиции в основной капитал на душу населения, руб.	0.059888	0.211566	0.924405	-0.054736	0.017178	-0.016359
Общая площадь жилых помещений, приходящаяся в среднем на одного жителя (кв. м.)	0.083848	0.216869	-0.134252	0.069241	-0.001785	0.824263
Удельный вес ветхого и аварийного жилищного фонда в общей площади всего жилищного фонда	0.131570	-0.146926	0.483877	0.252578	0.233491	-0.493162
Удельный вес общей площади, оборудованной водопроводом	0.030230	0.340271	0.002598	-0.724788	-0.072854	0.094851
Доля городского населения (в процентах)	0.694148	0.300997	0.052770	-0.322851	0.176649	0.284077
Число спортивных сооружений на 10 тыс. человек	-0.315098	0.020600	-0.013510	0.564271	0.181318	0.515276
Число зарегистрированных преступлений на 100 тыс. человек	0.564976	-0.195930	0.111487	0.437643	0.442667	-0.041457
Выбросы загрязняющих веществ в атмосферу (тонн на 1000 человек)	0.181941	0.127301	0.872102	0.029094	0.078945	-0.016269
Число дорожно-транспортных происшествий на 100 000 человек населения	0.318510	-0.264160	-0.161489	0.194211	0.153744	0.511771
Численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и бо	0.535545	-0.257048	0.083016	0.292070	0.348734	0.320775
Количество посещений в поликлинику с профилактической целью на 1 чел.	-0.093318	0.043485	0.678080	-0.024515	0.482074	0.078157
Число больничных коек на 10 тыс. человек	0.274494	-0.102157	0.090607	0.176218	0.715237	0.081723
Число посещений в смену на 10 тыс. человек	0.267138	0.089003	-0.031973	0.138379	0.717764	0.264221
Число врачей на 10 тыс. человек	0.077976	0.156267	0.056821	-0.416642	0.758770	-0.077567
Число среднего медицинского персонала на 10 тыс. человек	0.081718	-0.100797	0.372724	0.115332	0.778583	0.093579
Expl.Var	2.458517	3.386085	3.324641	2.343205	3.192071	3.552394
Prp.Totl	0.098341	0.135443	0.132886	0.093728	0.127683	0.142096

Видим в таблице 4, что произошло, с одной стороны, перераспределение объясняемой факторами дисперсии между факторами (она более равномерно распределилась между факторами), а с другой

стороны, перераспределение дисперсии объясняемой каждой фактором по каждой компоненте (увеличилась разница между дисперсиями для отдельных компонент).

Из таблицы 4 видно, что:

- в первый фактор вносит основной вклад компонента 3 (продажа алкогольных напитков и пива населению по субъектам российской федерации на 1 жителя чистого спирта мл);
- во второй фактор вносят основной вклад компоненты 4 (доля населения с доходами ниже прожиточного минимума (в процентах от общей численности населения)), 5 (коэффициент Джини), 6 (соотношение среднедушевого дохода и прожиточного минимума);
- в третий фактор вносят основной вклад компоненты 11 (инвестиции в основной капитал на душу населения, руб.) и 18 (выбросы загрязняющих веществ в атмосферу (тонн на 1000 человек));
- в четвертый фактор вносят основной вклад компоненты 9 (доля занятого населения высшим образованием (в процентах)) и 14 (удельный вес общей площади, оборудованной водопроводом);
- в пятый фактор вносят основной вклад компоненты 22 (число больничных коек на 10 тыс. человек), 23 (число посещений в смену на 10 тыс. человек), 24 (число врачей на 10 тыс. человек) и 25 (число среднего медицинского персонала на 10 тыс. человек);
- в шестой фактор вносят основной вклад компоненты 1 (число разводов на 1000 браков) и 12 (общая площадь жилых помещений, приходящаяся в среднем на одного жителя (кв. м.)).

Следовательно, эти компоненты и определяют содержание данных шести факторов.

Значение факторов, которые будут нужны для дальнейшего анализа, содержатся в таблице 5:

Таблица 5. Значение факторов

Case	Factor Scores (диплом) Rotation: Varimax raw Extraction: Principal components					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	-1,81028	1,30509	0,08340	0,47321	-0,20526	1,37238
2	-0,53865	0,39573	-0,53727	0,29520	-0,10593	0,41598
3	0,89179	-0,22998	-0,64819	0,64052	-0,68080	0,76271
4	-0,90169	1,46622	-0,66486	0,67892	0,20687	0,78203
5	0,83083	-0,39718	-0,39733	0,09267	-0,58390	0,17365
6	0,58890	0,40601	-0,43593	0,56280	-0,75776	0,66614
7	0,21061	-0,11975	-0,10347	0,53630	-0,34549	0,34795
8	-0,93342	0,39856	-0,32315	-0,26741	0,02475	1,03830
9	-1,11833	1,09657	-0,38252	0,50588	0,15700	1,08226
10	1,20965	1,32168	-0,23874	-1,25232	-1,77897	0,50518
11	-0,62553	-0,31212	-0,60811	-0,17704	0,22107	0,89085
12	-0,77106	0,13028	-0,35875	-0,69669	-0,01031	1,02280
13	-0,39458	-0,00939	-0,75402	0,32605	0,35825	0,44163
14	-1,67167	0,93067	-0,23123	1,59040	-0,07498	1,38601

Рассмотрим пятифакторную модель.

При помощи пакета STATISTICA посчитаем факторные нагрузки:

Таблица 6. Факторные нагрузки (до вращения)
пятифакторной модели

Variable	Factor Loadings (Unrotated) (диплом) Extraction: Principal components (Marked loadings are >,700000)				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Число разводов на 1000 браков	-0,682849	-0,250093	-0,350659	0,188774	-0,027496
Численность занимавшихся в физкультурно-оздоровительных секциях населения	-0,345196	0,246158	0,191905	0,481748	-0,173872
Продажа алкогольных напитков и пива населению по оптовой цене	-0,777161	-0,125701	-0,139494	-0,103357	0,401406
Доля населения с доходами ниже прожиточного минимума	0,519147	-0,652533	0,366935	-0,120519	-0,076066
Коэффициент Джини	-0,423033	0,566243	0,182737	0,199970	-0,066253
Соотношение среднедушевого дохода и прожиточного минимума	-0,483727	0,739051	-0,008189	0,115379	-0,119180
Число собственных легковых автомобилей на 1000 человек	-0,418275	-0,005721	-0,513957	0,062964	0,102442
Уровень безработицы (в процентах от численности экономически активного населения)	0,650899	-0,262145	0,523312	-0,005200	0,078185
Доля занятого населения с высшим образованием (в процентах)	-0,044155	0,483247	0,095445	-0,468410	-0,163590
Среднедушевые месячные доходы (руб.)	-0,723277	0,315677	0,407953	-0,143649	0,080860
Инвестиции в основной капитал на душу населения, в среднем по годам	-0,533197	0,241118	0,575499	0,349152	0,194752
Общая площадь жилых помещений, приходящаяся в среднем на одного жителя	-0,432841	0,031701	-0,706631	0,213266	-0,117361
Удельный вес ветхого и аварийного жилищного фонда	-0,104944	-0,286853	0,705020	0,076196	0,164408
Удельный вес общей площади, оборудованной водопроводом	-0,171850	0,642786	-0,095228	-0,389134	-0,024543
Доля городского населения (в процентах)	-0,666583	0,144576	-0,230737	-0,406105	0,326794
Число спортивных сооружений на 10 тыс. человек	-0,176006	-0,326627	-0,307159	0,562600	-0,403779
Число зарегистрированных преступлений на 100 тыс. населения	-0,408092	-0,663316	0,139503	-0,087604	0,231014
Выбросы загрязняющих веществ в атмосферу (тонн на 100 тыс. человек)	-0,557813	0,080671	0,545414	0,300711	0,252549
Число дорожно-транспортных происшествий на 100 тыс. автомобилей	-0,265946	-0,468199	-0,463991	-0,015131	0,071985
Численность пострадавших при несчастных случаях	-0,481103	-0,608757	-0,154456	-0,056618	0,201166
Количество посещений в поликлинику с профилактической целью	-0,560509	-0,076408	0,497991	0,149902	-0,241737
Число больничных коек на 10 тыс. человек	-0,506280	-0,514848	0,142675	-0,214955	-0,197528
Число посещений в смену на 10 тыс. человек	-0,606870	-0,383440	-0,074853	-0,214319	-0,282573
Число врачей на 10 тыс. человек	-0,469626	0,015543	0,269903	-0,570885	-0,411612
Число среднего медицинского персонала на 10 тыс. человек	-0,590432	-0,439954	0,352587	-0,082420	-0,328276
Expl.Var	6,270917	4,163715	3,572327	1,945168	1,232856
Prp.Totl	0,250837	0,166549	0,142893	0,077807	0,049314

Исходя из данных таблицы 6, можно заметить, что для четвертого и пятого факторов вообще сложно выделить существенно значимые координаты. Поэтому обратимся к процедуре вращения факторов.

Факторные нагрузки для пятифакторной модели после вращения:

Таблица 7. Факторные нагрузки (после вращения) пятифакторной модели

Factor Loadings (Varimax raw) (диплом) Extraction: Principal components (Marked loadings are >.700000)					
Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Число разводов на 1000 браков	0,559779	0,394515	0,107061	-0,403094	0,212701
Численность занимавшихся в физкультурно-оздоровительных	-0,172027	0,262032	0,522824	-0,325129	0,039836
Продажа алкогольных напитков и пива населению по субъект	0,754742	0,354098	0,298398	0,101858	0,128882
Доля населения с доходами ниже прожиточного минимума (в	-0,030165	-0,887066	-0,211839	-0,058471	0,119057
Коэффициент Джини	-0,213606	0,494720	0,530046	0,053529	0,056298
Соотношение среднедушевого дохода и прожиточного миним	-0,244419	0,746256	0,417895	0,108237	0,069163
Число собственных легковых автомобилей на 1000 человек н	0,415002	0,488174	-0,105352	-0,174961	-0,036696
Уровень безработицы (в процентах от численности экономиче	-0,296315	-0,803570	0,003155	0,125823	-0,151708
Доля занятого населения с высшим образованием (в процент	-0,305292	0,285774	-0,039311	0,497729	0,258083
Среднедушевые месячные доходы (руб.)	0,151638	0,339916	0,663329	0,314434	0,372787
Инвестиции в основной капитал на душу населения, руб.	0,043342	0,101998	0,905145	0,018112	0,039664
Общая площадь жилых помещений, приходящаяся в средне	0,314134	0,653306	0,208593	-0,421688	-0,010921
Удельный вес ветхого и аварийного жилищного фонда в общ	0,102056	-0,535335	0,545765	0,075265	0,150541
Удельный вес общей площади, оборудованной водопроводс	-0,209882	0,547631	-0,009492	0,499874	0,099441
Доля городского населения (в процентах)	0,567768	0,521006	0,096408	0,392330	0,182113
Число спортивных сооружений на 10 тыс. человек	0,066896	0,107901	-0,051135	-0,827709	0,087927
Число зарегистрированных преступлений на 100 тыс. человек	0,692654	-0,319171	0,161272	-0,091818	0,266367
Выбросы загрязняющих веществ в атмосферу (тонн на 1000	0,194340	0,020795	0,852542	0,007114	0,065336
Число дорожно-транспортных происшествий на 100 000 чело	0,577829	0,085085	-0,288199	-0,284688	0,073514
Численность пострадавших при несчастных случаях на произ:	0,761523	-0,079727	0,011275	-0,189695	0,217232
Количество посещений в поликлинику с профилактической це	0,047175	-0,005064	0,612261	-0,166546	0,494023
Число больничных коек на 10 тыс. человек	0,440728	-0,133137	0,095540	-0,129506	0,623779
Число посещений в смену на 10 тыс. человек	0,432828	0,132743	0,004178	-0,180327	0,639691
Число врачей на 10 тыс. человек	0,014232	0,131807	0,086126	0,294118	0,823088
Число среднего медицинского персонала на 10 тыс. человек	0,294234	-0,140268	0,326567	-0,210228	0,723847
Expl.Var	3,778870	4,514572	3,833722	2,295568	2,762251
Prp.Totl	0,151155	0,180583	0,153349	0,091823	0,110490

В таблице 7 видим, что произошло, с одной стороны, перераспределение объясняемой факторами дисперсии между факторами (она более равномерно распределилась между факторами), а с другой стороны, перераспределение дисперсии объясняемой каждой фактором по каждой компоненте (увеличилась разница между дисперсиями для отдельных компонент).

Из таблицы видно, что:

- в первый фактор вносят основной вклад компоненты 3 (продажа алкогольных напитков и пива населению по субъектам российской федерации на 1 жителя чистого спирта мл) и 20 (численность пострадавших при несчастных случаях на производстве с утратой

трудоспособности на один рабочий день и более и со смертельным исходом на 1000 работающих);

- во второй фактор вносят основной вклад компоненты 4 (доля населения с доходами ниже прожиточного минимума (в процентах от общей численности населения)), 6 (соотношение среднедушевого дохода и прожиточного минимума) и 8 (уровень безработицы);
- в третий фактор вносят основной вклад компоненты 11 (инвестиции в основной капитал на душу населения, руб.) и 18 (выбросы загрязняющих веществ в атмосферу (тонн на 1000 человек));
- в четвертый фактор вносит основной вклад компонента 16 (Число спортивных сооружений на 10 тыс. человек);
- в пятый фактор вносят основной вклад компоненты 24 (число врачей на 10 тыс. человек) и 25 (число среднего медицинского персонала на 10 тыс. человек).

Следовательно, эти компоненты и определяют содержание данных пяти факторов.

Рассмотрим четырехфакторную модель.

При помощи пакета STATISTICA посчитаем факторные нагрузки:

*Таблица 8. Факторные нагрузки (до вращения)
четырёхфакторной модели*

Variable	Factor Loadings (Unrotated) (диплом) Extraction: Principal components (Marked loadings are >,700000)			
	Factor 1	Factor 2	Factor 3	Factor 4
Число разводов на 1000 браков	-0,682849	-0,250093	-0,350659	0,188774
Численность занимавшихся в физкультурно-с	-0,345196	0,246158	0,191905	0,481748
Продажа алкогольных напитков и пива насел	-0,777161	-0,125701	-0,139494	-0,103357
Доля населения с доходами ниже прожиточн	0,519147	-0,652533	0,366935	-0,120519
Коэффициент Джини	-0,423033	0,566243	0,182737	0,199970
Соотношение среднедушевого дохода и прож	-0,483727	0,739051	-0,008189	0,115379
Число собственных легковых автомобилей на	-0,418275	-0,005721	-0,513957	0,062964
Уровень безработицы (в процентах от числен	0,650899	-0,262145	0,523312	-0,005200
Доля занятого населения с высшим образова	-0,044155	0,483247	0,095445	-0,468410
Среднедушевые месячные доходы (руб.)	-0,723277	0,315677	0,407953	-0,143649
Инвестиции в основной капитал на душу нас	-0,533197	0,241118	0,575499	0,349152
Общая площадь жилых помещений, приходя	-0,432841	0,031701	-0,706631	0,213266
Удельный вес ветхого и аварийного жилищнс	-0,104944	-0,286853	0,705020	0,076196
Удельный вес общей площади, оборудованн	-0,171850	0,642786	-0,095228	-0,389134
Доля городского населения (в процентах)	-0,666583	0,144576	-0,230737	-0,406105
Число спортивных сооружений на 10 тыс. чел	-0,176006	-0,326627	-0,307159	0,562600
Число зарегистрированных преступлений на	-0,408092	-0,663316	0,139503	-0,087604
Выбросы загрязняющих веществ в атмосфер	-0,557813	0,080671	0,545414	0,300711
Число дорожно-транспортных происшествий	-0,265946	-0,468199	-0,463991	-0,015131
Численность пострадавших при несчастных с	-0,481103	-0,608757	-0,154456	-0,056618
Количество посещений в поликлинику с проф	-0,560509	-0,076408	0,497991	0,149902
Число больничных коек на 10 тыс. человек	-0,506280	-0,514848	0,142675	-0,214955
Число посещений в смену на 10 тыс. человек	-0,606870	-0,383440	-0,074853	-0,214319
Число врачей на 10 тыс. человек	-0,469626	0,015543	0,269903	-0,570885
Число среднего медицинского персонала на	-0,590432	-0,439954	0,352587	-0,082420
Expl.Var	6,270917	4,163715	3,572327	1,945168
Prp.Totl	0,250837	0,166549	0,142893	0,077807

Исходя из данных таблицы 8, можно заметить, что для четвертого факторов вообще сложно выделить существенно значимые координаты. Поэтому обратимся к процедуре вращения факторов.

Факторные нагрузки для четырехфакторной модели после вращения отображены в таблице 9:

*Таблица 9. Факторные нагрузки (после вращения)
четырёхфакторной модели*

Variable	Factor Loadings (Varimax raw) (диплом) Extraction: Principal components (Marked loadings are >.7000000)			
	Factor 1	Factor 2	Factor 3	Factor 4
Число разводов на 1000 браков	0,535591	0,563400	0,138652	-0,252834
Численность занимавшихся в физкультурно-оздоровительн	-0,107250	0,213505	0,589244	-0,210637
Продажа алкогольных напитков и пива населению по субъек	0,589419	0,478942	0,246401	0,111390
Доля населения с доходами ниже прожиточного минимума	0,214502	-0,826246	-0,284218	-0,187096
Коэффициент Джини	-0,217295	0,354844	0,604994	0,183902
Соотношение среднедушевого дохода и прожиточного мин	-0,289530	0,582898	0,527895	0,302157
Число собственных легковых автомобилей на 1000 человек	0,226028	0,610085	-0,091430	-0,107010
Уровень безработицы (в процентах от численности экономи	-0,184132	-0,846290	-0,085234	-0,094316
Доля занятого населения с высшим образованием (в проце	-0,195386	0,088416	0,038968	0,645355
Среднедушевые месячные доходы (руб.)	0,254622	0,224510	0,698520	0,454616
Инвестиции в основной капитал на душу населения, руб.	0,058008	0,027872	0,889564	0,008755
Общая площадь жилых помещений, приходящаяся в сред	0,159031	0,788988	-0,137967	-0,257563
Удельный вес ветхого и аварийного жилищного фонда в об	0,272631	-0,547066	0,470820	-0,029591
Удельный вес общей площади, оборудованной водопрово	-0,264226	0,369092	0,062386	0,627107
Доля городского населения (в процентах)	0,406327	0,550045	0,081832	0,457219
Число спортивных сооружений на 10 тыс. человек	0,179820	0,260436	0,025767	-0,669106
Число зарегистрированных преступлений на 100 тыс. чело	0,768031	-0,137830	0,076271	-0,137753
Выбросы загрязняющих веществ в атмосферу (тонн на 100	0,205156	-0,005207	0,814130	-0,025358
Число дорожно-транспортных происшествий на 100 000 чел	0,500985	0,297869	-0,313057	-0,260264
Численность пострадавших при несчастных случаях на про	0,755280	0,138769	-0,050493	-0,191971
Количество посещений в поликлинику с профилактической	0,365456	-0,049490	0,673759	0,022725
Число больничных коек на 10 тыс. человек	0,753764	-0,041948	0,123485	0,052734
Число посещений в смену на 10 тыс. человек	0,711282	0,221977	0,073156	0,079401
Число врачей на 10 тыс. человек	0,443061	0,010021	0,200949	0,618679
Число среднего медицинского персонала на 10 тыс. челове	0,716404	-0,098841	0,386489	0,030025
Expl.Var	4,751642	4,506916	4,134457	2,559112
Prp.Totl	0,190066	0,180277	0,165378	0,102364

В таблице 9 видим, что, с одной стороны, перераспределение объясняемой факторами дисперсии между факторами (она более равномерно распределилась между факторами), а с другой стороны, перераспределение дисперсии объясняемой каждой фактором по каждой компоненте (увеличилась разница между дисперсиями для отдельных компонент).

Из таблицы видно, что:

- в первый фактор вносят основной вклад компоненты 17 (число зарегистрированных преступлений на 100 тыс. человек), 20 (численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом на 1000 работающих), 22 (число больничных коек на 10 тыс. человек), 23 (число посещений в смену на 10 тыс. человек) и 25 (число среднего медицинского персонала на 10 тыс. человек);

- во второй фактор вносят основной вклад компоненты 4 (Доля населения с доходами ниже прожиточного минимума (в процентах от общей численности населения)), 8 (уровень безработицы) и 12 (общая площадь жилых помещений, приходящаяся в среднем на одного жителя (кв. м.));
- в третий фактор вносят основной вклад компоненты 11 (инвестиции в основной капитал на душу населения, руб.) и 18 (выбросы загрязняющих веществ в атмосферу (тонн на 1000 человек));
- четвертый фактор не вносит основной вклад в дисперсию.

Следовательно, эти компоненты и определяют содержание данных четырех факторов.

Оставляем к рассмотрению шестифакторную модель, так как она наиболее интерпретируема в отличие от других рассмотренных моделей (факторы не так ярко выражены). Можно также заметить, что второй фактор явно указывает на социальное неравенство, а пятый фактор говорит о развитии медицины в том или ином регионе страны.

Для оценки степени влияния факторов на ОЖП проведем регрессионный анализ на основе модели множественной линейной регрессии. В качестве зависимой переменной выступает ожидаемая продолжительность жизни.

Результаты регрессионного анализа приведены в таблице 10.

Таблица 10. Результаты регрессионного анализа

Regression Summary for Dependent Variable: Ожидаемая продолжительность жизни						
R= ,84107574 R²= ,70740840 Adjusted R²= ,68490135						
F(6,78)=31,431 p<,00000 Std. Error of estimate: 1,3540						
N=85	b*	Std.Err. of b*	b	Std.Err. of b	t(78)	p-value
Intercept			71,15412	0,146860	484,5022	0,000000
Ф1	-0,420292	0,061247	-1,01377	0,147732	-6,8623	0,000000
Ф2	0,135240	0,061247	0,32621	0,147732	2,2081	0,030174
Ф3	-0,007157	0,061247	-0,01726	0,147732	-0,1169	0,907269
Ф4	-0,497258	0,061247	-1,19942	0,147732	-8,1189	0,000000
Ф5	-0,474982	0,061247	-1,14569	0,147732	-7,7552	0,000000
Ф6	-0,198869	0,061247	-0,47969	0,147732	-3,2470	0,001720

Как видим, значимое влияние на ОПЖ оказывают все обобщенные факторы, кроме третьего, при этом связь между факторами и ОПЖ достаточно сильная (коэффициент детерминации модели $R^2 \approx 0,71$).

Проведем иерархическую кластеризацию регионов на основе построенных факторов (главных компонент), используя алгоритм Варда (рисунок 2).

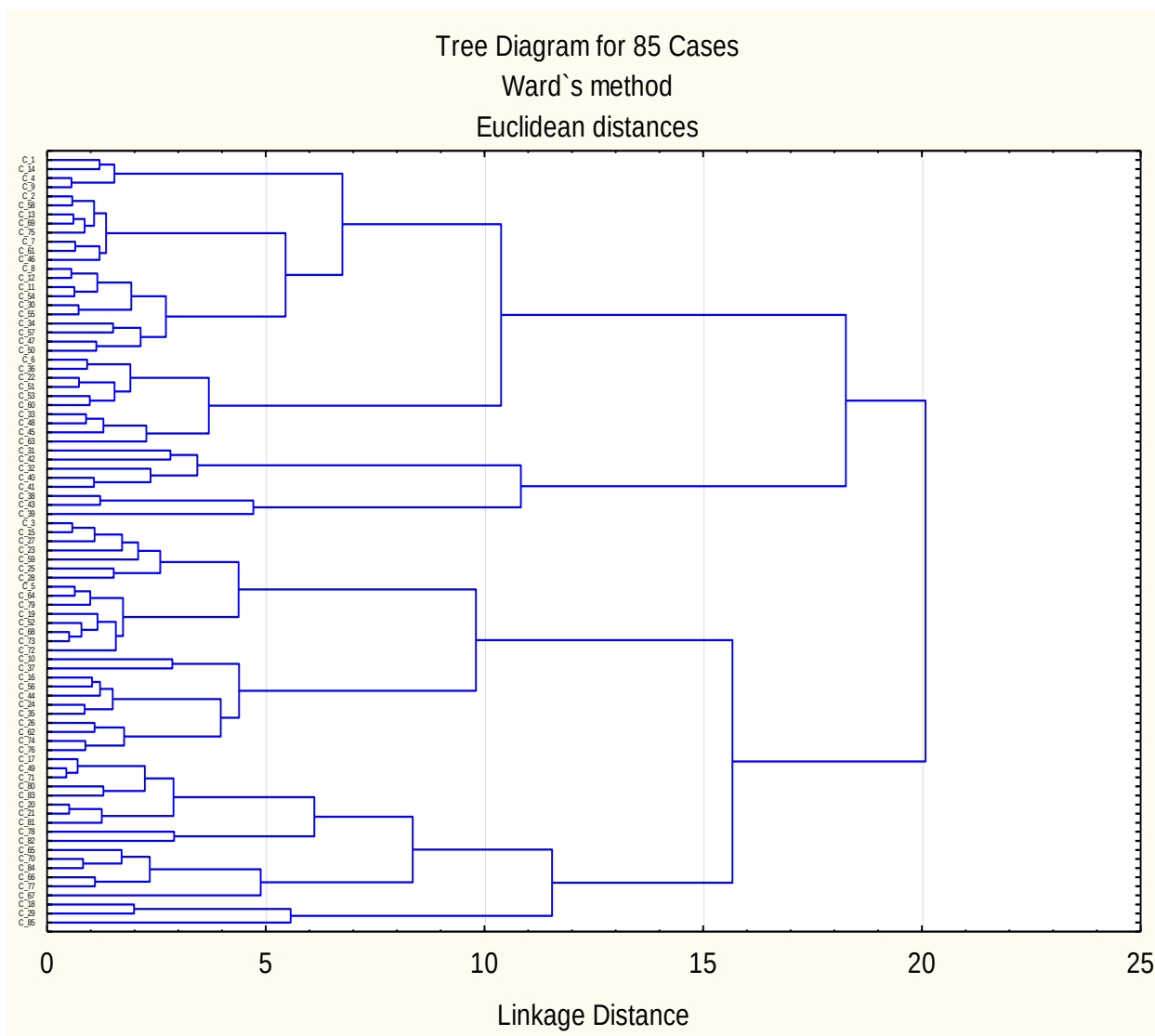


Рисунок 2. Иерархическая кластеризация регионов на основе главных компонент методом Варда.

На основе, полученной в результате кластеризации, дендрограммы выберем оптимальное разбиение совокупности объектов на кластеры. В качестве критерия оптимального разбиения используем следующий: выберем

множество кластеров, для которого: 1) по каждому фактору (главной компоненте) средние значения фактора по кластерам для всей совокупности кластеров значимо различаются; 2) для каждой пары различных кластеров существует значимое различие средних хотя бы для одного фактора (главной компоненты); 3) для любого другого разбиения с большим числом кластеров условия 1 или 2 не выполняются.

В результате получаем разбиение на 13 кластеров, оптимальное в смысле предложенного критерия.

Для установления значимости различий средних для совокупности кластеров используем методы дисперсионного анализа, а в качестве критерия сравнения средних используем критерий LSD Фишера (критерий наименьшей значимой разности).

В следующих таблицах (таблицы 11 – 16) приведены результаты дисперсионного анализа и множественного сравнения средних по кластерам для каждого фактора.

Дисперсионный анализ для тринадцати кластеров:

Таблица 11. Дисперсионный анализ для построенных 13 кластеров

Univariate Results for Each DV (23,04) Sigma-restricted parameterization Effective hypothesis decomposition													
Effect	Degr. of Freedom	Φ1 SS	Φ1 MS	Φ1 F	Φ1 p	Φ2 SS	Φ2 MS	Φ2 F	Φ2 p	Φ4 SS	Φ4 MS	Φ4 F	Φ4 p
Intercept	1	0.96817	0.968167	3.65882	0.059750	3.77612	3.776117	10.17776	0.002106	2.31928	2.319283	7.94764	0.006214
"CL1"	12	64.94797	5.412331	20.45387	0.000000	57.28681	4.773901	12.86709	0.000000	62.98893	5.249078	17.98736	0.000000
Error	72	19.05203	0.264612			26.71319	0.371016			21.01107	0.291820		
Total	84	84.00000				84.00000				84.00000			

	Univariate Results for Each DV (23,04) Sigma-restricted parameterization Effective hypothesis decomposition											
	Ф5 SS	Ф5 MS	Ф5 F	Ф5 p	Ф6 SS	Ф6 MS	Ф6 F	Ф6 p	Ожидаемая продолжительно сть жизни SS	Ожидаемая продолжительно сть жизни MS	Ожидаемая продолжительно сть жизни F	Ожидаемая продолжительно сть жизни p
Effect												
Intercept	9.01797	9.017971	24.47703	0.000005	3.25109	3.251089	11.62660	0.001069	0.01377	0.013774	0.26996	0.604955
"CL1"	57.47333	4.789444	12.99975	0.000000	63.86700	5.322250	19.03352	0.000000	10.76418	0.897015	17.58119	0.000000
Error	26.52667	0.368426			20.13300	0.279625			3.67353	0.051021		
Total	84.00000				84.00000				14.43771			

Проводим попарное сравнение средних для кластеров по каждому фактору (чтобы посмотреть различаются ли средние значения для каждого кластера в совокупности по каждому фактору).

Таблица 12. Множественное сравнение средних по кластерам для первого фактора (распространенность вредных привычек)

LSD test; variable Φ1 (23,04) Probabilities for Post Hoc Tests Error: Between MS = ,26461, df = 72,000														
Cell No.	CL1	{1} -1,375	{2} -2,492	{3} -1,090	{4} ,07175	{5} -1,613	{6} -1,412	{7} ,86033	{8} ,60232	{9} ,84442	{10} 1,3363	{11} ,01577	{12} ,77454	{13} -,6211
1	1		0,000629	0,351367	0,000010	0,493426	0,925315	0,000000	0,000000	0,000000	0,000000	0,000078	0,000008	0,193787
2	2	0,000629		0,000954	0,192520	0,000015	0,001329	0,000005	0,000656	0,000063	0,000215	0,343332	0,014048	0,497696
3	3	0,351367	0,000954		0,000003	0,067506	0,344199	0,000000	0,000000	0,000000	0,000000	0,000086	0,000013	0,387647
4	4	0,000010	0,192520	0,000003		0,000000	0,000039	0,000349	0,020957	0,002262	0,002215	0,833676	0,082014	0,203180
5	5	0,493426	0,000015	0,067506	0,000000		0,595031	0,000000	0,000000	0,000000	0,000000	0,000002	0,000000	0,082605
6	6	0,925315	0,001329	0,344199	0,000039	0,595031		0,000000	0,000000	0,000000	0,000000	0,000195	0,000014	0,186972
7	7	0,000000	0,000005	0,000000	0,000349	0,000000	0,000000		0,210477	0,943881	0,223007	0,001106	0,825291	0,006770
8	8	0,000000	0,000656	0,000000	0,020957	0,000000	0,000000	0,210477		0,314514	0,067520	0,027726	0,664486	0,025755
9	9	0,000000	0,000063	0,000000	0,002262	0,000000	0,000000	0,943881	0,314514		0,230416	0,003897	0,864045	0,008971
10	10	0,000000	0,000215	0,000000	0,002215	0,000000	0,000000	0,223007	0,067520	0,230416		0,002421	0,278443	0,002705
11	11	0,000078	0,343332	0,000086	0,833676	0,000002	0,000195	0,001106	0,027726	0,003897	0,002421		0,075012	0,255493
12	12	0,000008	0,014048	0,000013	0,082014	0,000000	0,000014	0,825291	0,664486	0,864045	0,278443	0,075012		0,029907
13	13	0,193787	0,497696	0,387647	0,203180	0,082605	0,186972	0,006770	0,025755	0,008971	0,002705	0,255493	0,029907	

Наибольшее среднее значение (1,3363) по первому фактору достигается в десятом кластере. В состав кластера входят следующие регионы: Камчатский край, Магаданская область со средней продолжительностью жизни 68,83 года.

Наименьшее среднее значение (-1,613) по первому фактору достигается в пятом кластере, соответствующие ему регионы России: Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания. Средняя продолжительность жизни в перечне республик 73,8 года.

Рассмотрим и промежуточные значения (в порядке возрастания).

Среднее значение по шестому кластеру -1,412 по первому фактору сопоставим со средней продолжительностью жизни 75,57 года в следующих регионах: Республика Дагестан, Республика Ингушетия, Чеченская Республика.

Среднее значение по первому кластеру -1,375 по первому фактору сопоставим со средней продолжительностью жизни 72,16 года в следующих регионах: Белгородская область, Воронежская область, Липецкая область, Тамбовская область.

Среднее значение по третьему кластеру -1,090 по первому фактору сопоставим со средней продолжительностью жизни 71,73 года в следующих регионах: Курская область, Орловская область, Рязанская область,

Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область.

Среднее значение по тринадцатому кластеру $-0,6211$ по первому фактору сопоставим со средней продолжительностью жизни 64,42 года в Чукотском автономном округе.

Среднее значение по второму кластеру $-0,2492$ по первому фактору сопоставим со средней продолжительностью жизни 70,71 года в следующих регионах: Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область.

Среднее значение по одиннадцатому кластеру $0,01577$ по первому фактору сопоставим со средней продолжительностью жизни 68,17 года в следующих регионах: Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область.

Среднее значение по четвертому кластеру $0,07175$ по первому фактору сопоставим со средней продолжительностью жизни 71,44 года в следующих регионах: Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область.

Среднее значение по восьмому кластеру $0,60232$ по первому фактору сопоставим со средней продолжительностью жизни 71,9 года в следующих регионах: Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты-Мансийский округ – Юрга, Новосибирская область, Томская область.

Среднее значение по двенадцатому кластеру $0,77454$ по первому фактору сопоставим со средней продолжительностью жизни 76,0 лет в городах Москва и Санкт-Петербург.

Среднее значение по девятому кластеру 0,8442 по первому фактору сопоставим со средней продолжительностью жизни 69,8 лет в следующих регионах: Ярославская область, Республика Коми, Архангельская область, Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область.

Среднее значение по седьмому кластеру 0,86033 по первому фактору сопоставим со средней продолжительностью жизни 65,32 года в следующих регионах: Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край.

Построим диаграмму рассеяния кластеров в системе координат «ОПЖ» и «среднее значение первого фактора».

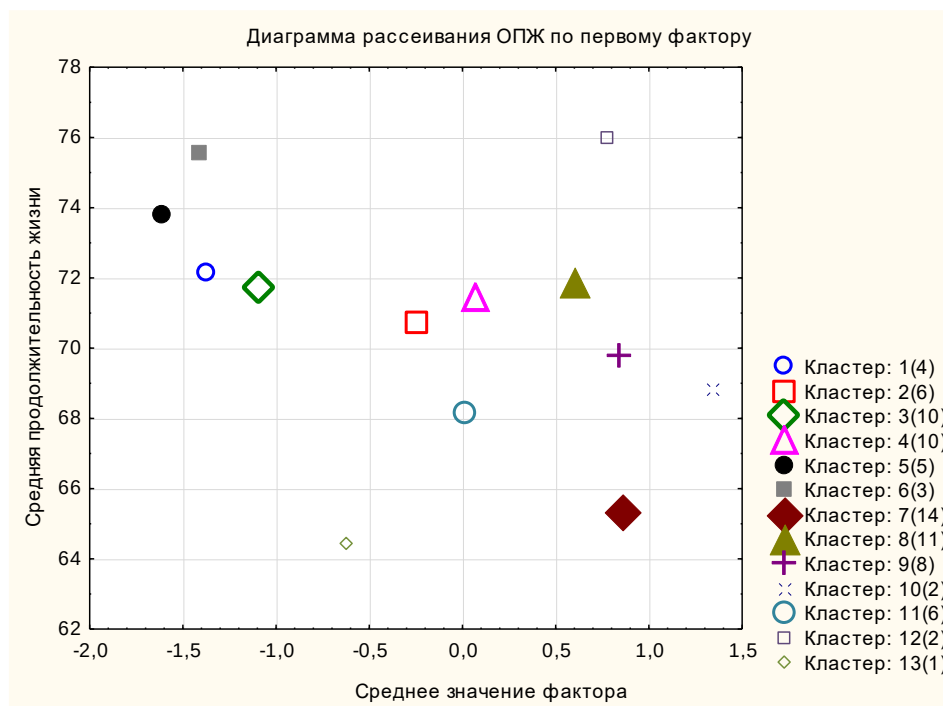


Рисунок 3. Диаграмма рассеяния кластеров в системе координат «ОПЖ» и «среднее значение первого кластера».

Из построенного графика видно, что сложно установить, как характеризуется средняя продолжительность жизни от первого фактора, который показывает распространенность вредных привычек (продажа

алкогольных напитков). Следовательно, среднее значение продолжительности жизни характеризуется совокупностью нескольких факторов.

Таблица 13. Множественное сравнение средних по кластерам для второго фактора (социальное неравенство)

LSD test; variable Φ2 (23,04) Probabilities for Post Hoc Tests Error: Between MS = ,37102, df = 72,000														
Cell No.	CL1	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}
		1,1996	-0,156	-2,199	1,1194	-1,437	,30036	-,5563	-,1708	,25907	-,8804	-,7769	2,3448	2,4991
1	1		0,001714	0,000187	0,824483	0,000000	0,057165	0,000002	0,000251	0,013895	0,000185	0,000004	0,033231	0,060366
2	2	0,001714		0,481797	0,000194	0,000109	0,446105	0,046285	0,585198	0,370183	0,076691	0,023509	0,000006	0,000220
3	3	0,000187	0,481797		0,000005	0,000495	0,198627	0,180292	0,854099	0,101742	0,165799	0,080820	0,000001	0,000062
4	4	0,824483	0,000194	0,000005		0,000000	0,044740	0,000000	0,000007	0,003954	0,000066	0,000000	0,011388	0,034135
5	5	0,000000	0,000109	0,000495	0,000000		0,000209	0,006537	0,000249	0,000006	0,278104	0,077539	0,000000	0,000000
6	6	0,057165	0,446105	0,198627	0,044740	0,000209		0,029303	0,238953	0,920526	0,037145	0,014657	0,000452	0,002555
7	7	0,000002	0,046285	0,180292	0,000000	0,006537	0,029303		0,115181	0,003130	0,481937	0,455927	0,000000	0,000007
8	8	0,000251	0,585198	0,854099	0,000007	0,000249	0,238953	0,115181		0,133233	0,133984	0,053783	0,000001	0,000076
9	9	0,013895	0,370183	0,101742	0,003954	0,000006	0,920526	0,003130	0,133233		0,020659	0,002384	0,000047	0,000891
10	10	0,000185	0,076691	0,165799	0,000066	0,278104	0,037145	0,481937	0,133984	0,020659		0,835654	0,000001	0,000023
11	11	0,000004	0,023509	0,080820	0,000000	0,077539	0,014657	0,455927	0,053783	0,002384	0,835654		0,000000	0,000004
12	12	0,033231	0,000006	0,000001	0,011388	0,000000	0,000452	0,000000	0,000001	0,000047	0,000001	0,000000		0,836791
13	13	0,060366	0,000220	0,000062	0,034135	0,000000	0,002555	0,000007	0,000076	0,000891	0,000023	0,000004	0,836791	

Наибольшее среднее значение (2,4991) по второму фактору достигается в тринадцатом кластере. В состав кластера входит Чукотский автономный округ со средней продолжительностью жизни 64,42 года.

Наименьшее среднее значение (-1,437) по второму фактору достигается в пятом кластере, соответствующие ему регионы России: Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания. Средняя продолжительность жизни в перечне республик 73,8 года.

Рассмотрим и промежуточные значения (в порядке возрастания).

Среднее значение по десятому кластеру -0,8804 по второму фактору сопоставим со средней продолжительностью жизни 68,83 года в следующих регионах: Камчатский край, Магаданская область.

Среднее значение по одиннадцатому кластеру -0,7769 по второму фактору сопоставим со средней продолжительностью жизни 68,17 года в следующих регионах: Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область.

Среднее значение по седьмому кластеру $-0,5563$ по второму фактору сопоставим со средней продолжительностью жизни $65,32$ года в следующих регионах: Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край.

Среднее значение по третьему кластеру $-0,2199$ по второму фактору сопоставим со средней продолжительностью жизни $71,73$ года в следующих регионах: Курская область, Орловская область, Рязанская область, Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область.

Среднее значение по восьмому кластеру $-0,1708$ по второму фактору сопоставим со средней продолжительностью жизни $71,9$ года в следующих регионах: Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты-Мансийский округ – Юрга, Новосибирская область, Томская область.

Среднее значение по второму кластеру $-0,0156$ по второму фактору сопоставим со средней продолжительностью жизни $70,71$ года в следующих регионах: Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область.

Среднее значение по девятому кластеру $0,25907$ по второму фактору сопоставим со средней продолжительностью жизни $69,8$ года в следующих регионах: Ярославская область, Республика Коми, Архангельская область, Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область.

Среднее значение по шестому кластеру $0,30036$ по второму фактору сопоставим со средней продолжительностью жизни $75,57$ года в следующих

регионах: Республика Дагестан, Республика Ингушетия, Чеченская Республика.

Среднее значение по четвертому кластеру 1,1194 по второму фактору сопоставим со средней продолжительностью жизни 71,44 года в следующих регионах: Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область.

Среднее значение по первому кластеру 1,1996 по второму фактору сопоставим со средней продолжительностью жизни 72,16 года в следующих регионах: Белгородская область, Воронежская область, Липецкая область, Тамбовская область.

Среднее значение по двенадцатому кластеру 2,3448 по второму фактору сопоставим со средней продолжительностью жизни 76,0 лет в Москве и Санкт-Петербурге.

Построим диаграмму рассеивания кластеров в системе координат «ОПЖ» и «среднее значение второго фактора».

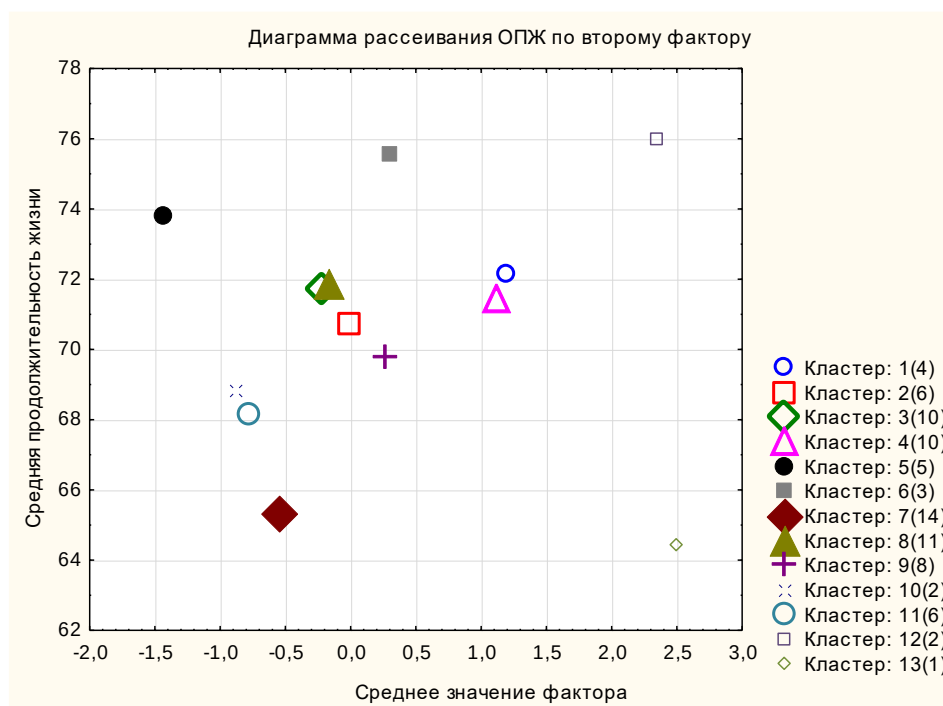


Рисунок 4. Диаграмма рассеивания кластеров в системе координат «ОПЖ» и «среднее значение второго фактора».

Из построенного графика видно, что есть возможность установить прямую зависимость средней продолжительности жизни от второго фактора, который показывает социальное неравенство (финансовое благополучие). Можно заметить, что в основном при увеличении среднего значения кластера идет увеличение ожидаемой продолжительности жизни.

Можно сделать предварительный вывод, что финансовое благополучие характеризует ОПЖ.

Таблица 13. Множественное сравнение средних по кластерам для четвертого фактора (качество жизни)

LSD test; variable Φ4 (23,04) Probabilities for Post Hoc Tests Error: Between MS = ,29182, df = 72,000														
Cell No.	CL1	{1} ,81212	{2} ,45343	{3} ,3111	{4} ,35102	{5} ,1,374	{6} ,06211	{7} ,50048	{8} ,1,092	{9} ,35364	{10} ,1,325	{11} 1,4989	{12} ,2,457	{13} ,3457
1	1		0,281852	0,000766	0,153422	0,000000	0,073253	0,308730	0,000000	0,170041	0,000020	0,052743	0,000000	0,059195
2	2	0,281852		0,003887	0,690610	0,000000	0,288199	0,842838	0,000000	0,712878	0,000086	0,000613	0,000000	0,167373
3	3	0,000766	0,003887		0,007725	0,000597	0,297446	0,000447	0,001464	0,011479	0,017945	0,000000	0,000002	0,951425
4	4	0,153422	0,690610	0,007725		0,000000	0,419210	0,500124	0,000000	0,991890	0,000149	0,000102	0,000000	0,222778
5	5	0,000000	0,000000	0,000597	0,000000		0,000510	0,000000	0,336873	0,000000	0,913668	0,000000	0,019183	0,086605
6	6	0,073253	0,288199	0,297446	0,419210	0,000510		0,203575	0,001599	0,427996	0,006338	0,000341	0,000003	0,515294
7	7	0,308730	0,842838	0,000447	0,500124	0,000000	0,203575		0,000000	0,536612	0,000027	0,000275	0,000000	0,133708
8	8	0,000000	0,000000	0,001464	0,000000	0,336873	0,001599	0,000000		0,000000	0,577252	0,000000	0,001572	0,190067
9	9	0,170041	0,712878	0,011479	0,991890	0,000000	0,427996	0,536612	0,000000		0,000193	0,000196	0,000000	0,226215
10	10	0,000020	0,000086	0,017945	0,000149	0,913668	0,006338	0,000027	0,577252	0,000193		0,000000	0,039641	0,143345
11	11	0,052743	0,000613	0,000000	0,000102	0,000000	0,000341	0,000275	0,000000	0,000196	0,000000		0,000000	0,002298
12	12	0,000000	0,000000	0,000002	0,000000	0,019183	0,000003	0,000000	0,001572	0,000000	0,039641	0,000000		0,002104
13	13	0,059195	0,167373	0,951425	0,222778	0,086605	0,515294	0,133708	0,190067	0,226215	0,143345	0,002298	0,002104	

Наибольшее среднее значение (1,4989) по четвертому фактору достигается в одиннадцатом кластере, соответствующие ему регионы России: Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область. Средняя продолжительность жизни в перечне республик 68,17 лет.

Наименьшее среднее значение (-2,457) по четвертому фактору достигается в двенадцатом кластере. В состав кластера входят города Москва и Санкт-Петербург со средней продолжительностью жизни 76,0 лет.

Рассмотрим и промежуточные значения (в порядке возрастания).

Среднее значение по пятому кластеру -1,374 по четвертому фактору сопоставим со средней продолжительностью жизни 73,8 года в следующих

регионах: Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания.

Среднее значение по десятому кластеру -1,325 по четвертому фактору сопоставим со средней продолжительностью жизни 68,83 года в следующих регионах: Камчатский край, Магаданская область.

Среднее значение по восьмому кластеру -1,092 по четвертому фактору сопоставим со средней продолжительностью жизни 71,9 года в следующих регионах: Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты-Мансийский округ – Юрга, Новосибирская область, Томская область.

Среднее значение по тринадцатому кластеру -0,3457 по четвертому фактору сопоставим со средней продолжительностью жизни 64,42 года в Чукотском автономном округе.

Среднее значение по третьему кластеру -0,3111 по четвертому фактору сопоставим со средней продолжительностью жизни 71,73 года в следующих регионах: Курская область, Орловская область, Рязанская область, Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область.

Среднее значение по шестому кластеру 0,06211 по четвертому фактору сопоставим со средней продолжительностью жизни 75,57 года в следующих регионах: Республика Дагестан, Республика Ингушетия, Чеченская Республика.

Среднее значение по четвертому кластеру 0,35102 по четвертому фактору сопоставим со средней продолжительностью жизни 71,44 года в следующих регионах: Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан,

Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область.

Среднее значение по девятому кластеру 0,35364 по четвертому фактору сопоставим со средней продолжительностью жизни 69,8 года в следующих регионах: Ярославская область, Республика Коми, Архангельская область, Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область.

Среднее значение по второму кластеру 0,45343 по четвертому фактору сопоставим со средней продолжительностью жизни 70,71 года в следующих регионах: Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область.

Среднее значение по седьмому кластеру 0,50048 по четвертому фактору сопоставим со средней продолжительностью жизни 65,32 года в следующих регионах: Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край.

Среднее значение по первому кластеру 0,81212 по четвертому фактору сопоставим со средней продолжительностью жизни 72,16 года в следующих регионах: Белгородская область, Воронежская область, Липецкая область, Тамбовская область.

Построим диаграмму рассеяния кластеров в системе координат «ОПЖ» и «среднее значение четвертого фактора».

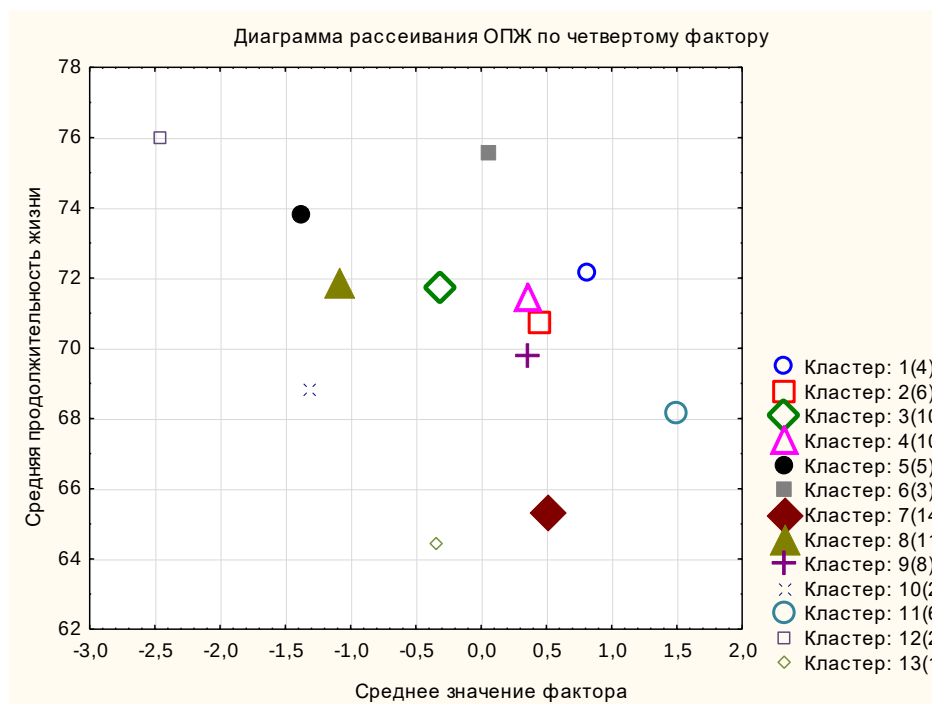


Рисунок 5. Диаграмма рассеивания кластеров в системе координат «ОПЖ» и «среднее значение четвертого фактора».

Из построенного графика видно, что есть возможность установить связь между средней продолжительности жизни от четвертого фактора, который показывает качество жизни. Можно заметить, что в основном при уменьшении среднего значения фактора идет увеличение ожидаемой продолжительности жизни.

Так, например, при самом высоком уровне жизни в Москве и Санкт-Петербурге наблюдается наибольшая продолжительность жизни.

Таблица 14. Множественное сравнение средних по кластерам для пятого фактора (уровень медицинского обслуживания)

LSD test; variable Φ5 (23,04) Probabilities for Post Hoc Tests Error: Between MS = ,36843, df = 72,000														
Cell No.	CL1	{1} _02091	{2} _04533	{3} _16712	{4} _5528	{5} _19450	{6} -1,837	{7} -5,422	{8} -5,007	{9} _73182	{10} 2,0235	{11} _99415	{12} _80414	{13} 4,1157
1	1		0,947802	0,685107	0,114496	0,671137	0,000147	0,103613	0,145417	0,059779	0,000290	0,015320	0,140594	0,000000
2	2	0,947802		0,673562	0,041326	0,667686	0,000019	0,030228	0,056786	0,026719	0,000099	0,005024	0,118189	0,000000
3	3	0,685107	0,673562		0,009832	0,934575	0,000004	0,005504	0,014022	0,053707	0,000181	0,010198	0,179686	0,000000
4	4	0,114496	0,041326	0,009832		0,027650	0,001954	0,965830	0,844845	0,000029	0,000001	0,000005	0,005144	0,000000
5	5	0,671137	0,667686	0,934575	0,027650		0,000019	0,021507	0,037142	0,124859	0,000578	0,032866	0,233895	0,000000
6	6	0,000147	0,000019	0,000004	0,001954	0,000019		0,001196	0,001170	0,000000	0,000000	0,000000	0,000009	0,000000
7	7	0,103613	0,030228	0,005504	0,965830	0,021507	0,001196		0,863940	0,000009	0,000000	0,000002	0,004328	0,000000
8	8	0,145417	0,056786	0,014022	0,844845	0,037142	0,001170	0,863940		0,000041	0,000001	0,000007	0,006619	0,000000
9	9	0,059779	0,026719	0,053707	0,000029	0,124859	0,000000	0,000009	0,000041		0,008832	0,426191	0,880618	0,000001
10	10	0,000290	0,000099	0,000181	0,000001	0,000578	0,000000	0,000000	0,000001	0,008832		0,041372	0,048300	0,006298
11	11	0,015320	0,005024	0,010198	0,000005	0,032866	0,000000	0,000002	0,000007	0,426191	0,041372		0,702560	0,000010
12	12	0,140594	0,118189	0,179686	0,005144	0,233895	0,000009	0,004328	0,006619	0,880618	0,048300	0,702560		0,000030
13	13	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000001	0,006298	0,000010	0,000030	

Из таблицы мы видим следующее: наибольшее среднее значение (4,1157) по пятому фактору достигается в тринадцатом кластере, соответствует Чукотскому автономному округу. Средняя продолжительность жизни в перечне республик 64,42 года.

Наименьшее среднее значение (-1,837) по пятому фактору достигается в шестом кластере. В состав кластера входят следующие регионы: Республика Дагестан, Республика Ингушетия, Чеченская Республика со средней продолжительностью жизни 75,57 года.

Рассмотрим и промежуточные значения (в порядке возрастания).

Среднее значение по четвертому кластеру -0,5528 по пятому фактору сопоставим со средней продолжительностью жизни 71,44 года в следующих регионах: Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область.

Среднее значение по седьмому кластеру -0,5422 по пятому фактору сопоставим со средней продолжительностью жизни 65,32 года в следующих регионах: Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская

область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край.

Среднее значение по восьмому кластеру -0,5007 по пятому фактору сопоставим со средней продолжительностью жизни 71,9 года в следующих регионах: Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты-Мансийский округ – Юрга, Новосибирская область, Томская область.

Среднее значение по первому кластеру 0,02091 по пятому фактору сопоставим со средней продолжительностью жизни 72,16 года в следующих регионах: Белгородская область, Воронежская область, Липецкая область, Тамбовская область.

Среднее значение по второму кластеру 0,04533 по пятому фактору сопоставим со средней продолжительностью жизни 70,71 года в следующих регионах: Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область.

Среднее значение по третьему кластеру 0,16712 по пятому фактору сопоставим со средней продолжительностью жизни 71,73 года в следующих регионах: Курская область, Орловская область, Рязанская область, Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область.

Среднее значение по пятому кластеру 0,19450 по пятому фактору сопоставим со средней продолжительностью жизни 73,8 года в следующих регионах: Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания.

Среднее значение по девятому кластеру 0,73182 по пятому фактору сопоставим со средней продолжительностью жизни 69,8 года в следующих регионах: Ярославская область, Республика Коми, Архангельская область,

Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область.

Среднее значение по двенадцатому кластеру 0,80414 по пятому фактору сопоставим со средней продолжительностью жизни 76,0 года в городах Москва и Санкт-Петербург.

Среднее значение по одиннадцатому кластеру 0,99415 по пятому фактору сопоставим со средней продолжительностью жизни 68,17 года в следующих регионах: Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область.

Среднее значение по десятому кластеру 2,0235 по пятому фактору сопоставим со средней продолжительностью жизни 68,83 года в следующих регионах: Камчатский край, Магаданская область.

Построим диаграмму рассеивания кластеров в системе координат «ОПЖ» и «среднее значение пятого фактора».

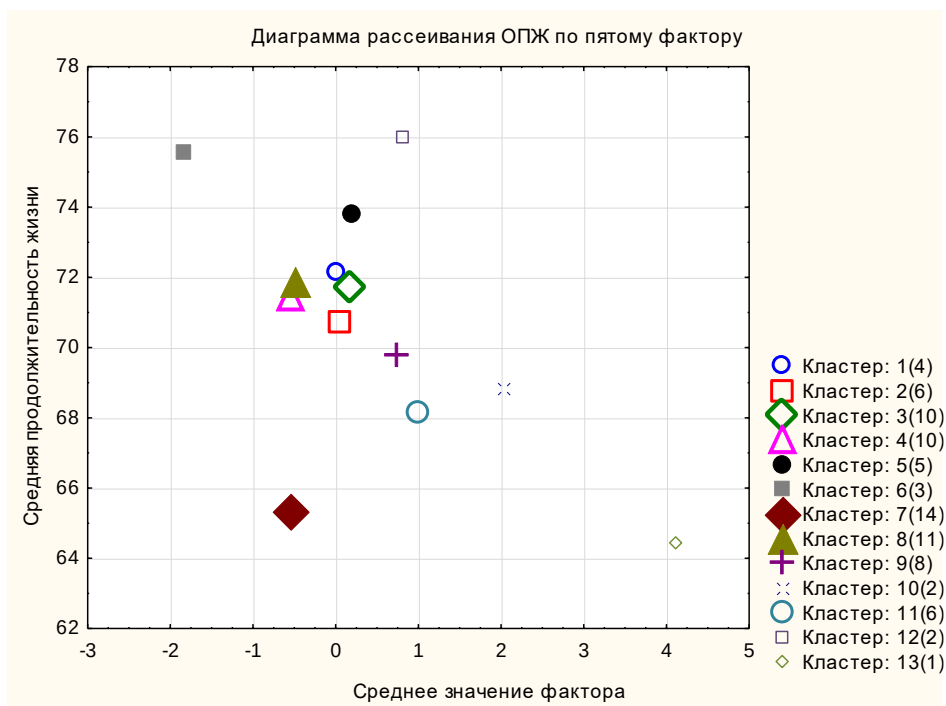


Рисунок 6. Диаграмма рассеивания кластеров в системе координат «ОПЖ» и «среднее значение пятого фактора».

Из построенного графика видно, что сложно установить прямую связь средней продолжительности жизни от пятого фактора, который показывает уровень медицинского обслуживания. Следовательно, среднее значение продолжительности жизни характеризуется совокупностью нескольких факторов.

Таблица 15. Множественное сравнение средних по кластерам для шестого фактора (плотность населения)

LSD test; variable Φ6 (23,04) Probabilities for Post Hoc Tests Error: Between MS = ,27963, df = 72,000														
Cell No.	CL1	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}
		1,1557	,38280	,76142	,01409	-,3569	-3,306	,41351	,00527	-,1746	1,0716	-1,289	-,9332	-1,146
1	1		0,019629	0,211655	0,000495	0,000060	0,000000	0,014926	0,000384	0,000104	0,854832	0,000000	0,000020	0,000219
2	2	0,019629		0,135562	0,145923	0,016553	0,000000	0,894846	0,128795	0,038484	0,103793	0,000000	0,002393	0,008064
3	3	0,211655	0,135562		0,002307	0,000244	0,000000	0,111431	0,001638	0,000377	0,451383	0,000000	0,000094	0,000976
4	4	0,000495	0,145923	0,002307		0,204290	0,000000	0,068387	0,969646	0,454276	0,011865	0,000009	0,023600	0,040049
5	5	0,000060	0,016553	0,000244	0,204290		0,000000	0,006174	0,208184	0,547235	0,001873	0,004812	0,196894	0,177584
6	6	0,000000	0,000000	0,000000	0,000000	0,000000		0,000000	0,000000	0,000000	0,000000	0,000001	0,000005	0,000709
7	7	0,014926	0,894846	0,111431	0,068387	0,006174	0,000000		0,055698	0,013226	0,102648	0,000000	0,001162	0,005620
8	8	0,000384	0,128795	0,001638	0,969646	0,208184	0,000000	0,055698		0,466455	0,010625	0,000008	0,023833	0,040726
9	9	0,000104	0,038484	0,000377	0,454276	0,547235	0,000000	0,013226	0,466455		0,003917	0,000213	0,073759	0,087689
10	10	0,854832	0,103793	0,451383	0,011865	0,001873	0,000000	0,102648	0,010625	0,003917		0,000001	0,000309	0,001023
11	11	0,000000	0,000000	0,000000	0,000009	0,004812	0,000001	0,000000	0,000008	0,000213	0,000001		0,413180	0,803113
12	12	0,000020	0,002393	0,000094	0,023600	0,196894	0,000005	0,001162	0,023833	0,073759	0,000309	0,413180		0,743843
13	13	0,000219	0,008064	0,000976	0,040049	0,177584	0,000709	0,005620	0,040726	0,087689	0,001023	0,803113	0,743843	

Из таблицы мы видим следующее: наибольшее среднее значение (1,1557) по шестому фактору достигается в первом кластере, соответствует следующим регионам: Белгородская область, Воронежская область, Липецкая область, Тамбовская область. Средняя продолжительность жизни в перечне республик 72,16 года.

Наименьшее среднее значение (-3,306) по шестому фактору достигается в шестом кластере. В состав кластера входят следующие регионы: Республика Дагестан, Республика Ингушетия, Чеченская Республика со средней продолжительностью жизни 75,57 года.

Рассмотрим и промежуточные значения (в порядке возрастания).

Среднее значение по одиннадцатому кластеру -1,289 по шестому фактору сопоставим со средней продолжительностью жизни 68,17 года в следующих регионах: Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область.

Среднее значение по тринадцатому кластеру -1,146 по шестому фактору сопоставим со средней продолжительностью жизни 64,42 года в Чукотском автономном округе.

Среднее значение по двенадцатому кластеру -0,9332 по шестому фактору сопоставим со средней продолжительностью жизни 76,0 года в Москве и Санкт-Петербурге.

Среднее значение по пятому кластеру -0,3569 по шестому фактору сопоставим со средней продолжительностью жизни 73,8 года в следующих регионах: Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания.

Среднее значение по девятому кластеру -0,1746 по шестому фактору сопоставим со средней продолжительностью жизни 69,8 года в следующих регионах: Ярославская область, Республика Коми, Архангельская область, Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область.

Среднее значение по восьмому кластеру 0,00527 по шестому фактору сопоставим со средней продолжительностью жизни 71,9 года в следующих регионах: Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты-Мансийский округ – Юрга, Новосибирская область, Томская область.

Среднее значение по четвертому кластеру 0,01409 по шестому фактору сопоставим со средней продолжительностью жизни 71,44 года в следующих регионах: Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область.

Среднее значение по второму кластеру 0,38280 по шестому фактору сопоставим со средней продолжительностью жизни 70,71 года в следующих

регионах: Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область.

Среднее значение по седьмому кластеру 0,41351 по шестому фактору сопоставим со средней продолжительностью жизни 65,32 года в следующих регионах: Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край.

Среднее значение по третьему кластеру 0,76142 по шестому фактору сопоставим со средней продолжительностью жизни 71,73 года в следующих регионах: Курская область, Орловская область, Рязанская область, Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область.

Среднее значение по десятому кластеру 1,0716 по шестому фактору сопоставим со средней продолжительностью жизни 68,83 года в следующих регионах: Камчатский край, Магаданская область.

Построим диаграмму рассеяния кластеров в системе координат «ОПЖ» и «среднее значение пятого фактора».

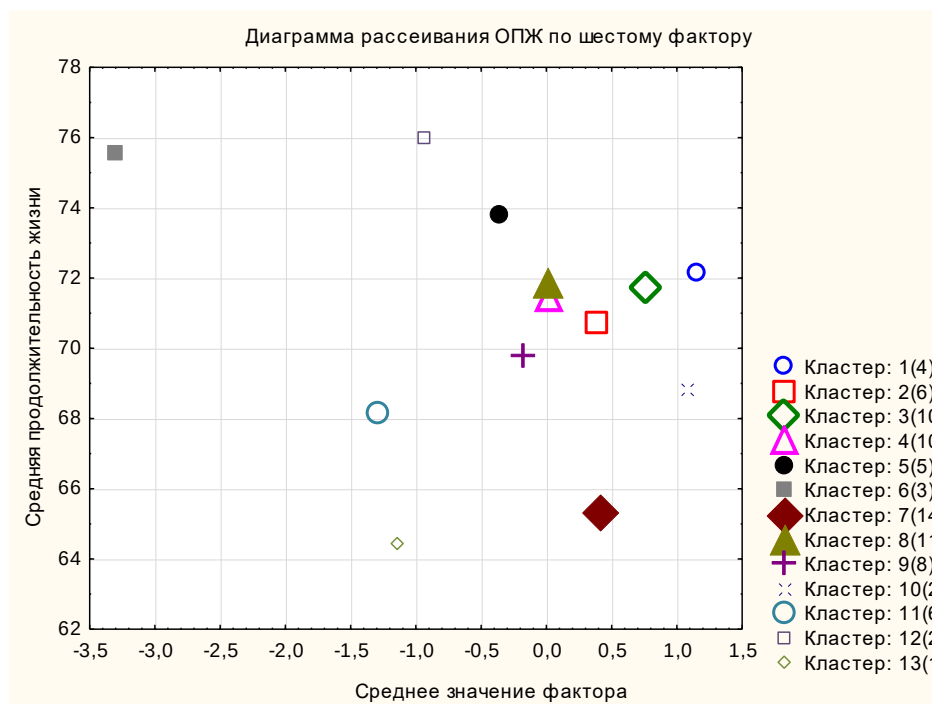


Рисунок 7. Диаграмма рассеивания кластеров в системе координат «ОПЖ» и «среднее значение шестого фактора».

Из построенного графика видно, что сложно установить прямую связь средней продолжительности жизни от шестого фактора, который показывает плотность населения. Следовательно, среднее значение продолжительности жизни характеризуется совокупностью нескольких факторов.

Таблица 16. Множественное сравнение средних по кластерам для ОПЖ

LSD test; variable Ожидаемая продолжительность жизни (23,04) Probabilities for Post Hoc Tests Error: Between MS = ,05102, df = 72,000														
Cell No.	CL1	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}	{12}	{13}
		,17461	-,0686	,09847	,05206	,45408	1,0764	-,2217	,12804	-,2327	-,3995	-,5135	,83118	-,1157
1	1		0,082952	0,570595	0,362172	0,069237	0,000002	0,002616	0,725076	0,004347	0,004478	0,000011	0,001264	0,000001
2	2	0,082952		0,123320	0,263849	0,000124	0,000000	0,125855	0,065049	0,150465	0,068006	0,000499	0,000003	0,000022
3	3	0,570595	0,123320		0,647330	0,005319	0,000000	0,000876	0,765269	0,002834	0,005765	0,000001	0,000079	0,000001
4	4	0,362172	0,263849	0,647330		0,001759	0,000000	0,004054	0,443870	0,009675	0,011900	0,000007	0,000030	0,000003
5	5	0,069237	0,000124	0,005319	0,001759		0,000329	0,000000	0,009215	0,000001	0,000024	0,000000	0,049783	0,000000
6	6	0,000002	0,000000	0,000000	0,000000	0,000329		0,000000	0,000000	0,000000	0,000000	0,000000	0,238266	0,000000
7	7	0,002616	0,125855	0,000876	0,004054	0,000000	0,000000		0,000213	0,911738	0,299426	0,009275	0,000000	0,000146
8	8	0,725076	0,065049	0,765269	0,443870	0,009215	0,000000	0,000213		0,000979	0,003316	0,000000	0,000128	0,000001
9	9	0,004347	0,150465	0,002834	0,009675	0,000001	0,000000	0,911738	0,000979		0,353615	0,024271	0,000000	0,000245
10	10	0,004478	0,068006	0,005765	0,011900	0,000024	0,000000	0,299426	0,003316	0,353615		0,538402	0,000001	0,007743
11	11	0,000011	0,000499	0,000001	0,000007	0,000000	0,000000	0,009275	0,000000	0,024271	0,538402		0,000000	0,010173
12	12	0,001264	0,000003	0,000079	0,000030	0,049783	0,238266	0,000000	0,000128	0,000000	0,000001	0,000000		0,000000
13	13	0,000001	0,000022	0,000001	0,000003	0,000000	0,000000	0,000146	0,000001	0,000245	0,007743	0,010173	0,000000	

Из таблицы мы видим следующее: наибольшее среднее значение (1,0764) по ОПЖ достигается в шестом кластере, соответствует следующим регионам: Республика Дагестан, Республика Ингушетия, Чеченская

Республика. Средняя продолжительность жизни в перечне республик 75,57 года.

Наименьшее среднее значение (-1,157) по ОПЖ достигается в тринадцатом кластере. В состав кластера входит Чукотский автономный округ со средней продолжительностью жизни 64,42 года.

Рассмотрим и промежуточные значения (в порядке возрастания).

Среднее значение по одиннадцатому кластеру -0,5135 по ОПЖ сопоставим со средней продолжительностью жизни 68,17 года в следующих регионах: Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область.

Среднее значение по десятому кластеру -0,3995 по ОПЖ сопоставим со средней продолжительностью жизни 69,8 года в следующих регионах: Ярославская область, Республика Коми, Архангельская область, Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область.

Среднее значение по девятому кластеру -0,2327 по ОПЖ сопоставим со средней продолжительностью жизни 68,83 года в следующих регионах: Камчатский край, Магаданская область.

Среднее значение по седьмому кластеру -0,2217 по ОПЖ сопоставим со средней продолжительностью жизни 65,32 года в следующих регионах: Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край.

Среднее значение по второму кластеру -0,0686 по ОПЖ сопоставим со средней продолжительностью жизни 70,71 года в следующих регионах: Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область.

Среднее значение по четвертому кластеру 0,05206 по ОПЖ сопоставим со средней продолжительностью жизни 71,44 года в следующих регионах: Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область.

Среднее значение по третьему кластеру 0,09847 по ОПЖ сопоставим со средней продолжительностью жизни 71,73 года в следующих регионах: Курская область, Орловская область, Рязанская область, Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область.

Среднее значение по восьмому кластеру 0,12804 по ОПЖ сопоставим со средней продолжительностью жизни 71,9 года в следующих регионах: Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты-Мансийский округ – Юрга, Новосибирская область, Томская область.

Среднее значение по первому кластеру 0,12804 по ОПЖ сопоставим со средней продолжительностью жизни 72,16 года в следующих регионах: Белгородская область, Воронежская область, Липецкая область, Тамбовская область.

Среднее значение по пятому кластеру 0,45408 по ОПЖ сопоставим со средней продолжительностью жизни 73,8 года в следующих регионах: Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания.

Среднее значение по двенадцатому кластеру 0,83118 по ОПЖ сопоставим со средней продолжительностью жизни 76,0 года в г. Москва, г. Санкт-Петербург.

Теперь рассмотрим совокупность всех факторов и кластеров, пользуясь таблицей 17.

Таблица 17. Кластеры

Номер кластера	Субъекты страны	Средняя продолжительность жизни (СПЖ)	Регионы
1	Юго-западная часть центрального региона РФ	72,16	Белгородская область, Воронежская область, Липецкая область, Тамбовская область
2	Западные часть центрального, приволжского и сибирского регионов РФ	70,71	Смоленская область, Республика Марий Эл, Ульяновская область, Тюменская область, Алтайский край, Омская область
3	Центральная часть северо-западного и южная часть приволжского регионов РФ	71,73	Курская область, Орловская область, Рязанская область, Республика Адыгея, Астраханская область, Республика Мордовия, Чувашская Республика, Оренбургская область, Пензенская область, Саратовская область
4	Приволжский и северная часть уральского регионов РФ	71,44	Калужская область, Ненецкий автономный округ, Краснодарский край, Ростовская область, Республика Башкортостан, Республика Татарстан, Пермский край, Нижегородская область, Челябинская область, Свердловская область
5	Южный регион РФ	73,8	Республика Калмыкия, Республика Крым, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия-Алания
6	Юг приволжского региона РФ	75,57	Республика Дагестан, Республика Ингушетия, Чеченская Республика
7	Западная часть северо-западного и южная часть	65,32	Владимирская область, Ивановская область, Тверская область, Республика Карелия, Вологодская область

	сибирского регионов РФ		область, Ленинградская область, Новгородская область, Псковская область, Кировская область, Курганская область, Республика Хакасия, Иркутская область, Кемеровская область, Приморский край
8	Юг Уральского региона РФ	71,9	Московская область, Тульская область, Калининградская область, Мурманская область, Волгоградская область, г. Севастополь, Ставропольский край, Самарская область, Ханты- Мансийский округ – Юрга, Новосибирская область, Томская область
9	Большая часть северо-западный, центральная часть сибирский и южная часть Дальневосточного регионы РФ	69,8	Ярославская область, Республика Коми, Архангельская область, Удмуртская Республика, Красноярский край, Хабаровский край, Амурская область, Сахалинская область
10	Северо-восточная часть дальневосточного региона РФ	68,83	Камчатский край, Магаданская область
11	Дальневосточный и южная часть сибирского региона РФ	68,17	Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Республика Саха (Якутия), Еврейская автономная область
12	Крупнейшие города центрального и северо-западного региона РФ	76,0	г. Москва, г. Санкт-Петербург
13	Северо-восток дальневосточного региона РФ	64,42	Чукотский автономный округ

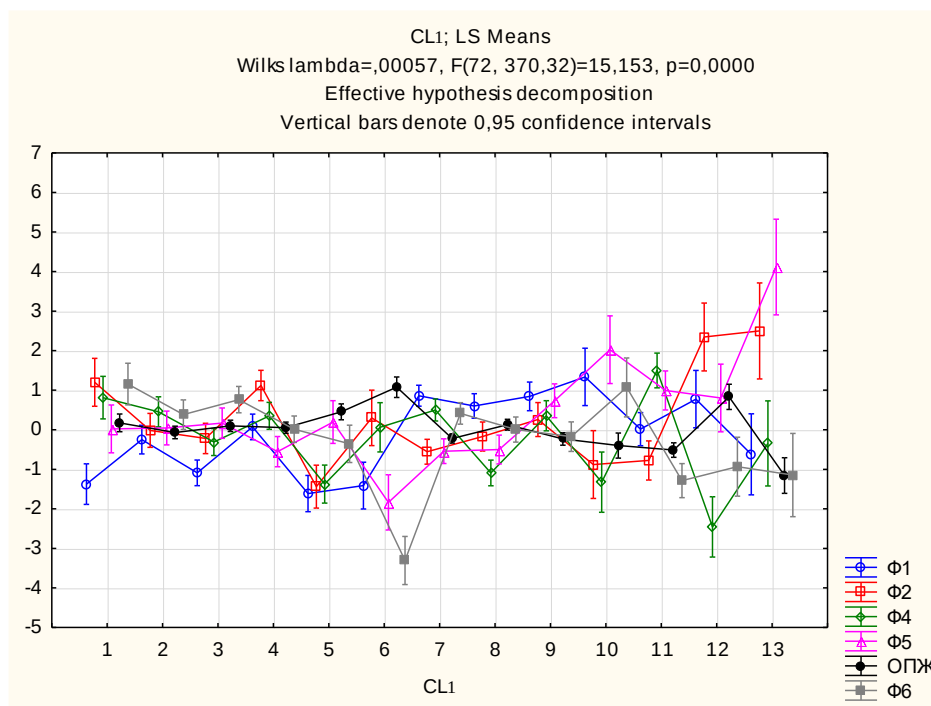


Рисунок 8. График зависимости кластеров и средних значений по каждому фактору в системе координат «среднее значение по каждому фактору» и «номер кластера»

Итак, подведем итоги по первому кластеру: средняя продолжительность жизни – 72,16 года, что соответствует средней ПЖ по РФ в 2016 году, где это величина составляет 72,1 года. Наблюдается самое низкое значение по первому фактору (распространение вредных привычек), одно из наиболее высоких показателей по второму фактору (финансовое благополучие), что скорее всего имеет прямую зависимость. Сравнительно высокие показатели плотность населения (шестой фактор) и качество жизни (четвертый фактор), что не удивительно для центрального региона РФ. Средний уровень медицинского обслуживания (пятый фактор).

Во втором кластере средняя продолжительность жизни также находится в пределах средней ПЖ по России – 70,71. Видно, что показатели по всем факторам достаточно близки между собой, то есть примерно на уровне находятся значения первого, второго и пятого факторов, которые чуть ниже нормы, и также близки значения четвертого и шестого факторов, которые чуть выше нормы. Это объясняется «разбросом» регионов по стране, входящих в кластер.

В третьем кластере заметно уменьшение (по отношению к первому кластеру) среднего значения первого и четвертого фактора, увеличение среднего значения по шестому фактору и почти не меняется среднее значение по второму фактору (по сравнению со вторым кластером), что немного отражается на средней продолжительности жизни данного кластера в лучшую сторону. Значит можно сделать вывод о том, что плотность населения (шестой фактор) в небольшой степени характеризует на СПЖ в данных регионах.

Наблюдаются средние значения по ряду факторов для четвертого кластера: распространенность вредных привычек (первый фактор), качество жизни (четвертый фактор) и плотность населения (шестой фактор). Высокое значение второго фактора, отвечающего за социальное неравенство. Низкий уровень медицинского обслуживания (пятый фактор). В совокупности эти факторы влияют на СПЖ в этом кластере, значение которой составляет 71,44, что тоже находится в пределах СПЖ по России, что не удивительно, так как области «равномерно раскиданы» по западной части РФ.

В пятом кластере можно заметить, что значение ОПЖ принимает значение, равное 73,8 года, это можно связать с тем, что среднее значение по пятому фактору (уровень медицинского обслуживания) усредняется и существенно изменяется среднее значение по первому, второму и четвертому факторам, что положительно влияет на СПЖ (так как идет обратная зависимость). Так можно судить о том, что значение по шестому фактору (плотность населения) не оказывает существенное влияние на продолжительность жизни.

В шестом кластере СПЖ принимает почти максимальное значение (75,57). Довольно странно, что среднее значение первого фактора (распространенность вредных привычек) одно из наиболее высоких (возможно это результат погрешности вычислений). А вот значение шестого фактора (плотность населения) самая низкая относительно всего исследования, но если обратить внимание на другие кластеры (а точнее

взаимосвязи шестого фактора с СПЖ, то можно заметить, что в целом продолжительность жизни не зависит от плотности населения). Также достаточно низкий пятый фактор, что может объясняться тем, что идет существенная нехватка медицинского персонала (что подтверждает статистика и данные на официальных сайтах). Средние значения по второму и четвертому факторам, что тоже является нормой для этих регионов РФ.

В седьмом факторе замечаем, что все показатели снова «группируются», то есть находятся примерно на одном и том же уровне относительно друг друга. Средняя продолжительность жизни вновь на уровне 65,32, что существенно меньше СПЖ по России. Первый фактор принимает почти максимальное значение (смотря на весь график в целом). Возможно, именно первый фактор влияет на резкое уменьшение значения СПЖ в регионах этого кластера.

В восьмом кластере можно наблюдать небольшое отличие значения СПЖ по сравнению со значением по седьмому кластеру. И изменение среднего значения первого фактора, остальные факторы находятся в пределах нормы. Значит можно говорить о том, что видна зависимость продолжительности жизни и первого фактора.

Поведение СПЖ и первого фактора в девятом кластере подтверждает наблюдение: увеличение среднего значения первого фактора приводит к уменьшению продолжительности жизни в этих регионах. Так же можно заметить увеличение значений по остальным факторам.

Десятый кластер, СПЖ которого равно 68,83 года. Снова можно заметить, что есть взаимосвязь первого фактора и продолжительности жизни. Видна связь шестого фактора и его связь с пятым фактором (можно сделать вывод, что от плотности населения зависит уровень медицинского обслуживания).

В одиннадцатом кластере начинается значимый и заметный разброс по всем средним значениям факторов. Уменьшается первый, пятый и шестой факторы, что не вписывается в «модель реакции», так как ранее замечалось,

что при уменьшении плотности населения увеличивается СПЖ. Но подтверждает прямую взаимосвязь шестого и пятого факторов. Довольно сильно увеличивается второй и четвертый факторы.

По двенадцатому кластеру нельзя ничего сказать однозначно, так как идет резкое изменение значений по всем факторам в сравнении с предыдущим кластером. Можно заметить, что четвертый фактор принимает наименьшее значение из всех. Но увеличивается среднее значение продолжительности жизни до 76,0 года (максимальное значение СПЖ) и второй фактор (социальное неравенство). Уменьшается плотность населения (шестой фактор). Пятый фактор практически не меняет своего значения. В целом заметна обратная зависимость четвертого фактора и СПЖ.

В тринадцатом кластере видим максимальные значения по второму и пятому факторам. Резко сокращается СПЖ и принимает самое маленькое среднее значение (64,42 года).

Итак, СПЖ не зависит от шестого фактора (плотности населения). Есть обратная связь между первым, четвертым и пятым факторами и СПЖ. Прямая взаимосвязь со вторым фактором, исключениями являются пятый и тринадцатый кластеры.

3. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Целью данного раздела работы является экономическое обоснование дипломного проекта. Необходимо рассмотреть следующие вопросы:

- организация и планирование комплекса работ на создание проекта;
- стоимостная оценка разработки;
- оценка уровня научной новизны.

3.1. Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

Трудоемкость выполнения ВКР оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов.

Для выполнения работы должны быть задействованы исполнители. Ими могут быть:

- руководитель проекта (Р);
- исполнитель (И).

Разделим выполнение дипломной работы на этапы, представленные в таблице 18.

Таблица 18. Комплекс работ по разработке проекта

№ раб.	Наименование работ	Исполнители
Подготовительный		
1	Составление и утверждение ТЗ: • Выбор темы и описание задания.	И, Р
2	Подбор и изучение материалов по теме: • Нахождение статей по теме; • Изучение книг по статистическому анализу.	И, Р

Исследование и анализ предметной области		
3	Сбор исходных данных.	И
4	Выбор методов исследования. Выбраны: математические методы	И, Р
5	Календарное планирование работ по теме: Выбраны сроки выполнения отдельных частей исследования и работы в целом.	И, Р
Теоретические и экспериментальные исследования		
6	Применение выбранных методов к данным, т.е. применение полученной модели к исходным данным.	И
Обобщение и оценка результатов		
7	Анализ результатов работы, который отражен в выводах и результатах, а также в заключении.	И
8	Составление отчета по работе.	И

3.1.1 Продолжительность этапов работ

Для определения вероятных (ожидаемых) значений продолжительности работ $t_{ож}$ применяется формула:

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5},$$

где t_{min} – минимальная продолжительность работы, дн.;

t_{max} – максимальная продолжительность работы, дн.;

Для построения линейного графика необходимо рассчитать длительность этапов в рабочих днях, а затем перевести ее в календарные дни. Расчет продолжительности выполнения каждого этапа в рабочих днях ($T_{РД}$) ведется по формуле:

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д}$$

где $t_{ож}$ – продолжительность работы, дн.;

$K_{\text{вн}}$ – коэффициент выполнения работ, учитывающий влияние внешних факторов на соблюдение предварительно определенных длительностей, в частности, возможно $K_{\text{вн}} = 1$;

$K_{\text{д}}$ – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ ($K_{\text{д}} = 1,2$)

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{\text{КД}} = T_{\text{РД}} \cdot T_{\text{К}},$$

где $T_{\text{КД}}$ – продолжительность выполнения этапа в календарных днях;

$T_{\text{К}}$ – коэффициент календарности, позволяющий перейти от длительности работ в рабочих днях к их аналогам в календарных днях, и рассчитываемый по формуле:

$$T_{\text{К}} = \frac{T_{\text{КАЛ}}}{T_{\text{КАЛ}} - T_{\text{ВД}} - T_{\text{ПД}}}$$

где $T_{\text{КАЛ}}$ – календарные дни ($T_{\text{КАЛ}} = 365$);

$T_{\text{ВД}}$ – выходные дни ($T_{\text{ВД}} = 52$);

$T_{\text{ПД}}$ – праздничные дни ($T_{\text{ПД}} = 10$).

$$T_{\text{К}} = \frac{365}{365 - 52 - 10} = 1,2$$

В таблице 19 определена продолжительность этапов работ и их трудоемкость по исполнителям, занятым на каждом этапе. В столбцах (3–5) реализован экспертный способ по формуле $t_{\text{ож}} = \frac{3 \cdot t_{\text{min}} + 2 \cdot t_{\text{max}}}{5}$. Столбцы 6 и 7

содержат величины трудоемкости этапа для каждого из двух участников проекта (научный руководитель и исполнитель) с учетом коэффициента $K_{\text{д}} = 1,2$. Каждое из них в отдельности не может превышать соответствующее значение $t_{\text{ож}} \cdot K_{\text{д}}$. Столбцы 8 и 9 содержат те же трудоемкости, выраженные в календарных днях путем дополнительного умножения на $T_{\text{К}}$ (здесь оно равно 1,2). Итог по столбцу 5 дает общую ожидаемую продолжительность работы над проектом в рабочих днях, итоги по столбцам 8 и 9 – общие трудоемкости

для каждого из участников проекта. Две последних величины далее будут использованы для определения затрат на оплату труда участников и прочие затраты. Величины трудоемкости этапов по исполнителям $T_{\text{КД}}$ (данные столбцов 8 и 9 кроме итогов) позволяют построить линейный график осуществления проекта (Таблица 19).

Таблица 19. Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.- дн.			
					$T_{РД}$		$T_{КД}$	
		t_{min}	t_{max}	$t_{ож}$	Р	И	Р	И
1	2	3	4	5	6	7	8	9
Постановка задачи	Р	2	4	2,8	3,36	-	4,032	-
Разработка и утверждение технического задания (ТЗ)	И, Р	1	4	2,2	2,64	2,64	3,168	3,168
Подбор и изучение материалов по тематике	И, Р	9	20	13,4	16,08	16,08	19,296	19,296
Обсуждение литературы	И, Р	5	8	6,2	7,44	7,44	8,928	8,928
Выбор объектов исследования	И	2	6	3,6	-	4,32	-	5,184
Выбор методов исследования	И, Р	2	5	3,2	3,84	3,84	4,608	4,608
Разработка календарного плана	И, Р	1	5	2,6	3,12	3,12	3,744	3,744
Сбор исходных данных	И	15	20	17	-	20,4	-	24,48
Расчет дополнительных параметров	И	5	6	5,4	-	6,48	-	7,776
Обработка данных	И	7	9	7,8	-	9,36	-	11,232
Оформление расчетно-пояснительной записки	И	6	9	7,2	-	8,64	-	10,368
Оформление графического материала	И	5	6	5,4	-	6,48	-	7,776

Подведение итогов	И, Р	5	8	6,2	7,44	7,44	8,928	8,928
Итого:				71,4	36,48	82,32	43,776	98,784

Таблица 20. Линейный график работ

№	Р	И	Февраль			Март			Апрель			Май		
			10	20	30	40	50	60	70	80	90	100	110	120
1	4,032	-												
2	3,168	3,168												
3	19,296	19,296												
4	8,928	8,928												
5	-	5,184												
6	4,608	4,608												
7	3,744	3,744												
8	-	24,48												
9	-	7,776												
10	-	11,232												
11	-	10,368												
12	-	7,776												
13	8,928	8,928												

Р –



И –



3.2 Расчет сметы затрат на выполнение проекта

Определение затрат на выполнение ВКР производится путем составления калькуляции по отдельным статьям затрат всех видов необходимых ресурсов. Калькуляция является основным документом, на основании которого осуществляется планирование и учет затрат на научные исследования.

3.2.1 Расчет затрат на материалы

Данная статья отражает стоимость всех материалов, используемых при разработке проекта, включая расходы на их приобретение и, при необходимости доставку. Транспортные расходы (если таковые имеются) принимаются в пределах 3-5% от стоимости материалов.

Расчет затрат на материалы производится в таблице.

Таблица 21 – Затраты на материалы

Наименование материалов и покупных изделий	Единица измерения	Кол-во	Цена за ед., руб.	Сумма, руб.
Бумага, формат А4	Пачка	1	218	218
Картридж HP122, черный	Штук	1	630	630
Flashcard, 4 Гб	Штук	1	500	500
Ручка	Штук	1	15	15
Блокнот	Штук	1	50	50
Ноутбук	Штук	1	45000	45000
Итого				46413

Допустим, что ТЗР составляют 5 % от отпускной цены материалов, тогда расходы на материалы с учетом ТЗР равны $C_{\text{мат}} = 46413 * 1,05 = 48733,65$ руб.

3.2.2 Расчет заработной платы

В этой статье расходов планируется и учитывается основная заработная плата исполнителей, непосредственно участвующих в проектировании разработки. Среднедневная тарифная заработная плата ($ЗП_{\text{дн-т}}$) рассчитывается по формуле:

$$ЗП_{\text{дн-т}} = МО/25,083$$

учитывающей, что в году 301 рабочий день и, следовательно, в месяце в среднем 25,083 рабочих дня (при шестидневной рабочей неделе).

Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы 2. Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов: $K_{\text{пр}} = 1,1$; $K_{\text{доп.ЗП}} = 1,188$; $K_{\text{р}} = 1,3$. Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент $K_{\text{и}} = 1,1 * 1,188 * 1,3 = 1,699$.

Таблица 22. Затраты на заработную плату

Исполнитель	Оклад, руб./мес	Среднедневная ставка, руб./раб.день	Затраты времени, раб.дни	Коэффициент	Фонд з/платы, руб.
НР	33 664	1342,09	37	1,699	84 367,8
И	15 470	616,75	83	1,62	82 928,2
Итого:					167 296

3.2.3 Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30 % от полной заработной платы по проекту, т.е. $C_{\text{соц.}} = C_{\text{зп}} * 0,3$. Итак, в нашем случае $C_{\text{соц.}} = 167\,296 * 0,3 = 50\,188,8$ руб.

3.2.4 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{\text{эл.об.}} = P_{\text{об}} \cdot t_{\text{об}} \cdot \text{Ц}_{\text{э}}$$

где $P_{\text{об}}$ – мощность, потребляемая оборудованием, кВт;

$\text{Ц}_{\text{э}}$ – тариф на 1 кВт·час;

$t_{\text{об}}$ – время работы оборудования, час.

Для ТПУ $\text{Ц}_{\text{э}} = 5,748$ руб./кВт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы для инженера ($T_{\text{рд}}$) из расчета, что продолжительность рабочего дня равна 8 часов.

$$t_{\text{об}} = T_{\text{рд}} \cdot K_t$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{\text{рд}}$, определяется исполнителем самостоятельно. В ряде случаев возможно определение $t_{\text{об}}$ путем прямого учета, особенно при ограниченном использовании соответствующего оборудования.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{\text{об}} = P_{\text{ном.}} \cdot K_c$$

где $P_{\text{ном.}}$ – номинальная мощность оборудования, кВт;

$K_c \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности $K_c = 1$.

Расчет затрат на электроэнергию для технологических целей приведен в таблице 23.

Таблица 23. Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования $t_{об}, \text{час}$	Потребляемая мощность $P_{об},$ кВт	Затраты $\Sigma_{об},$ руб.
Персональный компьютер	664*0,85	0,3	973,25
Струйный принтер	30	0,1	17,24
Итого:			990,49

3.2.5 Расчет амортизационных расходов

В статье «Амортизационные отчисления» рассчитывается амортизация используемого оборудования за время выполнения проекта.

Используется формула

$$C_{AM} = \frac{H_A * C_{об} * t_{рф} * n}{F_D},$$

где H_A – годовая норма амортизации единицы оборудования;

$C_{об}$ – балансовая стоимость единицы оборудования с учетом ТЗР.

F_D – действительный годовой фонд времени работы соответствующего оборудования, берется из специальных справочников или фактического режима его использования в текущем календарном году.

$t_{рф}$ – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

n – число задействованных однотипных единиц оборудования.

Стоимость ПК 45000 руб., время использования 664 часов, тогда для него $C_{AM}(ПК) = (0,4*45000*664*1)/2408 = 4\,963,46$ руб. Стоимость принтера 12000 руб., его $F_D = 500$ час.; $H_A = 0,5$; $t_{рф} = 2$ час., тогда его $C_{AM}(Пр) = (0,5*12000*2*1)/500 = 360$ руб. Итого начислено амортизации 5 323,46 руб.

3.2.6 Расчет прочих расходов

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях, их следует принять равными 10% от суммы всех предыдущих расходов, т.е.

$$C_{\text{проч.}} = (C_{\text{мат}} + C_{\text{зп}} + C_{\text{соц}} + C_{\text{эл.об.}} + C_{\text{ам}}) \cdot 0,1$$

Для нашего случая это

$$C_{\text{проч.}} = (48733,65 + 167\,296 + 50\,188,8 + 990,49 + 5\,323,46) \cdot 0,1 = 27\,253,24 \text{руб.}$$

3.2.7 Расчет общей себестоимости разработки

Проведя расчет по всем статьям сметы затрат на разработку, можно определить общую себестоимость проекта «Макет демонстрационной модели принципов КТ».

Таблица 24. Смета затрат на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	48733,65
Основная заработная плата	$C_{\text{зп}}$	167 296
Отчисления в социальные фонды	$C_{\text{соц}}$	50 188,8
Расходы на электроэнергию	$C_{\text{эл.}}$	990,49
Амортизационные отчисления	$C_{\text{ам}}$	5 323,46
Прочие расходы	$C_{\text{проч}}$	27 253,24
Итого:		299 785,64

Таким образом, затраты на разработку составили $C = 299\,785,64$ руб.

3.2.8 Расчет прибыли

Размер прибыли возьмем как 7 % от полной себестоимости проекта. В итоге она составит 20 984,9 руб.

3.2.9 Расчет НДС

НДС составляет 20% от суммы затрат на разработку и прибыли. В нашем случае это $(299\,785,64 + 20\,984,9) \cdot 0,2 = 320\,770,54 \cdot 0,2 = 64\,154,11$ руб.

3.2.10 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС, в нашем случае

$$C_{\text{НИР(кр)}} = 299\,785,64 + 20\,984,9 + 64\,154,11 = 386\,925,08 \text{ руб.}$$

3.3 Оценка экономической эффективности проекта

Исследование имеет фундаментально-исследовательский характер. В общем случае экономическая оценка невозможна, так как требуются дополнительные исследования и дополнительная целевая информация о характере исследования полученных результатов.

4. Социальная ответственность

Цель ВКР - изучение влияния социально - экономических факторов и характеристик социальной среды на ОПЖ российских регионов. Используя данные государственной статистики, мы осуществляем анализ связей между этими факторами и показателем продолжительности жизни в субъектах Российской Федерации.

В современной жизни компьютер широко применяется в жизни человека: дома, в офисе, в магазине, в производстве и даже в бытовой технике. Другими словами, компьютеры прочно вошли в повседневную жизнь людей и их использование постоянно увеличивается.

Несоблюдение требований безопасности приводит к тому, что при работе за компьютером сотрудник может ощущать дискомфорт: возникают головные боли и резь в глазах, появляются усталость и раздражительность. У некоторых людей нарушается сон, аппетит, ухудшается зрение, начинают болеть руки, шея, поясница и тому подобное. При ненормированной работе вполне возможно нервное истощение.

Внедрение компьютерных технологий принципиально изменило характер труда и требования к организации и охране труда. Работники используют персональные ЭВМ (ПЭВМ) практически во всех отраслях деятельности как в производстве, так и в научно-исследовательских работах. Одновременно с этим, работая с компьютером, работник подвергается вредному воздействию, что может привести к различным профессиональным заболеваниям.

4.1. Правовые и организационные вопросы обеспечения безопасности.

4.1.1. Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства.

На основании статьи 107 ТК РФ одним из видов времени отдыха признаются перерывы в процессе рабочего дня.

Согласно статье 109 ТК РФ некоторые виды труда подразумевают необходимость предоставления работникам возможности отдохнуть в течение рабочего дня, что вызвано спецификой технологии, производства и условий

работы. Конкретный перечень таких работ и порядок предоставления соответствующих перерывов должны предусматриваться правилами внутреннего распорядка.

В статье 27 Закона о санитарно-эпидемиологическом благополучии от 30 марта 1999 года отмечается, что работа с машинами и различными устройствами не должна вредно влиять на здоровье.

Порядок организации работы для лиц, труд которых связан с компьютерной техникой, регулируют, в частности, СанПиН 2.2.2/2.4.1340-03 (утверждены 30 мая 2003 года).

В зависимости от вида работ и степени нагрузки Приложение 7 к СанПиН 2.2.2/2.4.1340-03 устанавливает, что время отдыха при рассматриваемом виде работ должно составлять в течение рабочего дня от 50 до 140 минут. И это с учетом того, что данные перерывы длительность работы увеличивать не должны.

За нарушение действующих санитарных правил статьей 6.3 КоАП РФ предусмотрена административная ответственность.

4.1.2. Организационные мероприятия при компоновке рабочей зоны

Под рабочим местом будем понимать часть рабочей зоны, где работник пребывает временно или же постоянно в процессе своей трудовой деятельности. Рабочее место должно гарантировать возможность удобно выполнять работу, принимая во внимание её технологические особенности и физическую тяжесть, и кроме того должно в себя включать пространство, которое требуется для передвижений в ней работника.

Согласно СанПиНу 2.2.2/2.4.1340-03, где прописаны общие требования к организации и оборудованию рабочих мест с ПЭВМ, во время работы инженера за своим рабочим местом (столом), конструкции стола и стула должны гарантировать для работающего оптимальное положение тела. Нормы для рабочего места при работе с ПЭВМ приведены в таблице 25.

Таблица 25 . Параметры рабочего места при работе

Параметры	Значение параметра	Реальные значения
Высота рабочей поверхности стола	От 600 до 800, мм	770 мм
Высота клавиатуры	600-700, мм	630 мм
Удаленность клавиатуры	Не менее 80, мм	85 мм
Удаленность экрана монитора	500-700, мм	650 мм
Высота сидения	400-500, мм	470 мм
Угол наклона монитора	0-30, град.	10 мм
Наклон подставки ног	0-20, град.	0 мм

Исходя из данных таблицы 25, параметры рабочего стола удовлетворяют нормативным требованиям.

Требование параметров для монитора при работе с ПЭВМ отмечены в таблице 26.

Таблица 26. Допустимые визуальные параметры устройств отображения информации

Параметры	Допустимые значения
Яркость белого поля	Не менее 35 кд/м ²
Неравномерность яркости рабочего поля	Не более $\pm 20\%$
Контрастность (для монохромного режима)	Не менее 3:1
Временная нестабильность изображения (непреднамеренное изменение во времени яркости изображения на экране дисплея)	Не должна фиксироваться

Исходя из данных представленных в таблице и зная характер работ, стоит отдать предпочтение малококонтрастным оттенкам, которые бы не способствовали рассеиванию внимания в рабочей зоне. Это в первую очередь связано с тем, что работа требует концентрации и собранности.

В связи с тем, что при работе с ПЭВМ на работника оказывают влияние вредные и опасные производственные факторы. К тому же может накапливаться утомляемость, что довольно негативно может сказаться на

самочувствии работника. Поэтому продолжительность непрерывной работы с компьютером без регламентированного перерыва не должна превышать 2-х часов.

4.2. Профессиональная социальная безопасность.

4.2.1. Анализ вредных и опасных факторов, которые может создать объект исследования.

Основные вредные и опасные факторы, которые могут возникнуть при работе за компьютером, представлены в таблице 27.

Таблица 27. Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015СБТ)	Этапы работ			Нормативные документы
	Разраб отка	Изгото вление	Экспл уатац ия	
Отклонение показателей микроклимата	+	+	+	ГОСТ 12.1.005–88
Недостаточная освещенность рабочей зоны	+	+	+	СП 52.13330.2016 СанПиН 2.2.1/2.1.1.1278-03
Воздействие электромагнитных полей	+	+	+	СанПиН 2.2.2/2.4.1340-03
Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека	+	+	+	ГОСТ 12.1.002–84

4.2.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при проведении исследований.

1. Работа, которую выполняет математик-экономист, производится в основном в положении сидя и зачастую сопровождается незначительным физическим напряжением. Следовательно, можно сделать вывод о том, что этот вид работ относится к категории Ia – работа с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт). Для данной категории допустимые нормы микроклимата помещения представлены в таблице 28.

Таблица 28. Допустимые нормы микроклимата в рабочей зоне производственных помещений

Сезон года	Категория тяжести выполняемых работ	Температура, С°		Относительная влажность, %		Скорость движения воздуха, м/сек	
		Фактическое значение	Допустимое значение	Фактическое значение	Допустимое значение	Фактическое значение	Допустимое значение
Холодный	Ia	(21÷23)	(20÷25)	55	(15÷75)	0,1	0,1
Теплый	Ia	(22÷24)	(21÷28)	55	(15÷75)	0,1	0,1

Исходя из данных таблицы 28, можно сделать вывод о том, что нормы соответствующие микроклимату для производственного помещения соблюдаются.

2.В производственном помещении используют как искусственное (люминесцентные лампы типа ЛД), так и естественное освещение.

Таблица 29. Допустимые нормы освещенности в рабочей зоне производственных помещений

Освещенность, лк	
Фактическое значение	Допустимое значение
256	Не менее 300

Исходя из данных таблицы 29, можно говорить о том, фактическое значение параметра освещенности в помещении соответствует требуемым нормам.

3. Воздействие ЭПМ происходит по время непосредственной работы и только от монитора ЭВМ. Проверив паспортные данные компьютера, можно судить о том, что они соответствуют нормам ТСО-99, ТСО-03

4. Помещение, где выполнялась ВКР, - помещение относится к категории без повышенной опасности по степени вероятности поражения электрическим током, поэтому к оборудованию предъявляются следующие требования:

- применение приэкранных фильтров, специальных экранов;
- экран монитора должен находиться на расстоянии не менее 50 см от пользователя (расстояния от источника).

4.2.3. Обоснование мероприятий по защите исследователя от действия опасных и вредных факторов.

1. В случае если параметры отклоняются от заявленной нормы, то необходимо обеспечить обогревателем, вентилятором или увлажнителем воздуха (в зависимости от требуемых условий работы).

2. В случае если параметры отклоняются от заявленной нормы, то необходимо обеспечить дополнительными источниками света, такими как точечные светильники, настольные или подвесные лампы и т.д.

3. Бакалаврская работа велась за компьютером, который снабжен монитором с жидкокристаллическим дисплеем, именно поэтому, чтобы снизить вредное воздействие, необходимо ограничить время пребывания на рабочем месте.

4. Чтобы обезопасить себя от поражения электрическим током необходимо перед тем, как начать работать убедиться в том, что:

- отсутствуют свешивающиеся со стола или висящие под столом провода электропитания;
- вилки и провода электропитания находятся в целостности;
- отсутствуют видимые повреждения аппаратуры и рабочей мебели;

- отсутствуют повреждения;
- имеется наличие заземления приэкранного фильтра.

4.3. Экологическая безопасность.

4.3.1. Анализ влияния объекта исследования на окружающую среду.

Объект исследования является теоретическим и не оказывает влияния на окружающую среду.

4.3.2. Анализ влияния процесса исследования на окружающую среду.

В процессе написания ВКР мною были использованы следующие ресурсы:

- бумага;
- люминесцентные лампы;
- электроэнергия для работы компьютера.

Можно заметить, что с точки зрения потребления ресурсов компьютер потребляет достаточно небольшое количество электроэнергии, что благоприятно сказывается на общей экономии потребления электроэнергии в целом.

В ходе выполнения работы в атмосферу, почву и водные источники вредных выбросов не производилось, не наблюдались чрезвычайные ситуации, поэтому не оказывались существенные воздействия на окружающую среду. Можно сделать вывод о том, что не было нанесено никакого ущерба окружающей среде.

4.3.3. Обоснование мероприятий по защите окружающей среды.

Так как до сих пор большая часть информации все еще содержится на бумажных носителях, то можно утверждать, что утилизация бумаги играет одну из важных ролей.

Среди основных методов уничтожения следует отметить следующие:

- шредирование;
- сжигание документов;

- химическая обработка;
- закапывание.

Переработка оргтехники включает в себя несколько этапов:

Первый этап – удаление всех опасных компонентов.

Второй этап – удаление всех крупных пластиковых частей. В большинстве случаев эта операция также осуществляется вручную. Оставшиеся после разборки части отправляют в большой измельчитель, и все дальнейшие операции автоматизированы.

Третий этап – измельченные в гранулы остатки компьютеров подвергаются сортировке. Сначала с помощью магнитов извлекаются все железные части. Затем приступают к выделению цветных металлов, которых в ПК значительно больше.

Перегоревшие люминесцентные лампы можно отнести в свой районный ДЕЗ или РЭУ, где установлены специальные контейнеры. Там их должны бесплатно принять.

4.4. Безопасность в чрезвычайных ситуациях.

4.4.1. Анализ вероятных ЧС, которые может инициировать объект исследований.

В течение выполнения ВКР возможно возникновение следующих чрезвычайных ситуаций: мороз, диверсия, пожар.

4.4.2. Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований.

В Томской области преобладают континентально-циклонический климат. Такие опасные природные явления как наводнения, землетрясения, засухи, ураганы отсутствуют. Возможными чрезвычайными ситуациями, которые могут возникнуть в городе Томске, являются пожар, сильные морозы, диверсия.

Чтобы гарантировать безопасность людей и сохранение их материальных ценностей, создана пожарная безопасность. Основная система, которой состоит

из системы предотвращения пожара и противопожарной защиты, куда могут быть включены организационно-технические мероприятия.

Для областей, которые относятся к сибирскому региону, в холодное время года характерны сильные морозы. Достижение критически низких температур может повлечь за собой не только аварии в системах теплоснабжения и жизнеобеспечения, но и приостановке работы, а также обморожениям и даже жертвам среди населения.

ЧС, которые могут возникнуть в результате диверсий, к огромному сожалению, возникают все чаще в наше время. Довольно часто такие угрозы являются ложными. Но иногда в действительности случаются и взрывы.

4.4.3. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС.

1. В случае возникновения пожара на территории предприятия действия всех работников должны быть направлены на немедленное сообщение о нем в пожарную охрану, обеспечение безопасности людей и их эвакуации, а также тушение возникшего пожара. Для оповещения людей о пожаре должны использоваться тревожные или звуковые сигналы.

Каждый работник, обнаруживший пожар обязан:

- немедленно сообщить об этом по телефону «01» или в пожарную охрану (при этом необходимо назвать адрес, место возникновения пожара, а также сообщить свою фамилию и должность);
- сообщить руководителю или его заместителю о пожаре;
- принять меры по организации эвакуации людей (эвакуацию начинать из помещения, где возник пожар, а также из помещений, которым угрожает опасность распространения огня и дыма);
- одновременно с эвакуацией людей приступить к тушению пожара своими силами и имеющимися средствами пожаротушения (огнетушители, вода, песок и т.д.).

2. Чтобы обезопасить себя от переморозки труб необходимо наличие запасных обогревателей. Количества и мощности которых должно хватить для того, чтобы работа на производстве не прекращалась.

3. Для того чтобы предотвратить вероятность осуществление диверсии предприятие настоятельно рекомендуется оборудовать круглосуточной охраной, пропускной системой, системой видеонаблюдения, надежной системой связи, а также исключить распространения информации о системе охраны объекта, расположении помещений и оборудования в помещениях, системах охраны, сигнализаторах, их местах установки и количестве.

4.5. Вывод и рекомендации

Проведя анализ рабочего места и условий труда, можно говорить о том, что помещение, где была написана ВКР, удовлетворяет нормативным требованиям.

Также можно упомянуть, что при соблюдении правил пользования компьютером и техники безопасности работа в рассмотренном помещении не сможет привести к ухудшению здоровья работника.

Рабочее место сотрудника, как и помещение, в котором происходило написание бакалаврской работы, соответствуют всем нормативным требованиям, поэтому воздействие вредных и опасных факторов сведено к минимуму.

В ходе выполнения работы в атмосферу, почву и водные источники вредных выбросов не производилось, не наблюдались чрезвычайные ситуации, поэтому не оказывались существенные воздействия на окружающую среду. Можно сделать вывод о том, что не было нанесено никакого ущерба окружающей среде.

Бакалаврская работа велась за компьютером, который снабжен монитором с жидкокристаллическим дисплеем, служащий источником ЭПМ – вредного фактора, который негативно сказывается на здоровье работника при продолжительной непрерывной работе, что может привести к снижению его работоспособности. Именно поэтому, чтобы снизить вредное воздействие,

необходимо ограничить время пребывания на рабочем месте, а так же проводить специализированные комплексы упражнений для глаз.

Заключение

1. Проведена факторизация пространства социально-экономических показателей, в результате которой из 25 исходных показателей было сформировано 6 обобщенных факторов (главных компонент), из которых 5 высоко значимо коррелируют с ОПЖ.
2. На основе матрицы значений обобщенных факторов и ОПЖ была проведена иерархическая кластеризация регионов РФ методом Варда.
3. Используя предложенный критерий кластеризации, было выбрано оптимальное разбиение (на основе данного критерия), состоящее из 13 кластеров.
4. Определены основные факторы, характеризующие каждый кластер.
5. Оценена связь социально-экономических факторов и ОПЖ.

Список публикаций студента

1. Чжэн Кэли. Дефицит квалифицированных кадров в России / Чжэн Кэли, А. О. Курникова; науч. рук. Е. С. Киселева // Научная инициатива иностранных студентов и аспирантов российских вузов: сборник докладов VIII Всероссийской научно-практической конференции, Томск, 16-18 мая 2018 г. — Томск : Изд-во ТПУ, 2018. — [С. 246-251].

Список литературы

1. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. М.: ЮНИТИ-ДАНА, т.1, 2001. С. 283
2. Алексеев А.И., Николина В.В. География: население и хозяйство России. 2004 г.
3. «Анализ тенденций развития регионов России». Программа ЕС «Тасис». Москва. 2006 г.
4. Балдин С.В., Быстров О.Ф., Соколов М.М. Эконометрика Москва. 2004г/ Учеб.пособие для вузов. — М.: ЮНИТИ-ДАНА, 2015. С. 52
5. Безопасность жизнедеятельности: Учебник для вузов/ С.В. Белов, А.В. Ильницкая, А.Ф. Козьяков и др.; Под общ.ред. С.В. Белова.7-е изд., стер. — М.: Высш.шк., 2007. — 616 с
6. Бочаров В. В. Инвестиции : учебник для вузов / В. В. Бочаров. — 2-е изд. — СПб: Питер, 2009. — 381 с. — Учебник для вузов.
7. «Введение в экономическую географию». Учебник для ВУЗов, А. Ю. Скопин - М, 2006 г.
8. Вишневский А.Г. // Демографические процессы в социальном контексте -- Демографическая модернизация, частная жизнь и идентичность в России, Тезисы докладов. М. 2007. С. 5.
9. Голосовский С. И. Эффективность научных исследований в промышленности / С. И. Голосовский. — Москва: Экономика, 1986. — 159 с.
- 10.Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы //Учебное пособие. — М: Финансы и статистика, 2003. С. 2557.
- 11.Зубченко Л. Обзор зарубежных периодических изданий по статистике и демографии // Вопросы статистики. М, 2003 г.
- 12.Ким О. Дж., Мьюллер Ч.У., Клетка У.Р., и др. Факторный, дискриминантный и кластерный анализ — М.: Финансы и статистика, 1989. С. 113

13. Кнышова Е. Н. Экономика организации: учебник / Е. Н. Кнышова, Е. Е. Панфилова. – Москва: Форум Инфра-М, 2012. – 334 с.: ил. – Профессиональное образование.
14. Морозова Т.Г., Шишов С.С. «Экономическая география России». Москва. 2007 г.
15. Назаренко О.Б. Безопасность жизнедеятельности: учебное пособие / О.Б. Назаренко; Национальный исследовательский Томский политехнический университет. – 2-е изд., перераб. и доп. – Томск: Изд-во Томского политехнического университета, 2010. – 144 с.
16. Население России 2000. М. 2001. С. 154.
17. Несветаев Ю. А. Экономическая оценка инвестиций: учебное пособие / Ю. А. Несветаев; Московский Государственный индустриальный университет; Институт дистанционного образования. – 3-е изд., стер. – Москва: Изд-во МГИУ, 2006. – 162 с.
18. Сошникова Л.А., Тамашевич В.Н. Многомерный статистический анализ в экономике/ Учеб.пособие для вузов. — М.: ЮНИТИ-ДАНА, 1999. С. 507
19. Староверова Г. С. Экономическая оценка инвестиций : учебное пособие / Г. С. Староверова, А. Ю. Медведев, И. В. Сорокина. – 2-е изд., стер. – Москва: КноРус, 2009. – 312 .
20. Учебно-методическое пособие по многомерным статистическим методам для студентов специальности 080116 «Математические методы в экономике». Томск: Изд. ТПУ, 2007. С. 50