МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ)

Институт Финансовых Технологий и Экономической Безопасности Кафедра Финансового мониторинга

**Лабораторная работа №4 по курсу**

**«Макростатический анализ и прогнозирование»**

|  |  |
| --- | --- |
| **Выполнил студент группы С21-703:** | Монастырский М.О. |
| **Проверил:** | Домашова Д.В. |

Москва, 2024

Оглавление

[1. Постановка задачи 3](#_Toc184234807)

[2. Снижение размерности признакового пространства методом главных факторов 4](#_Toc184234808)

[3. Метод Уорда (МГФ) 14](#_Toc184234809)

[К средних (МГФ) 17](#_Toc184234810)

1. Постановка задачи

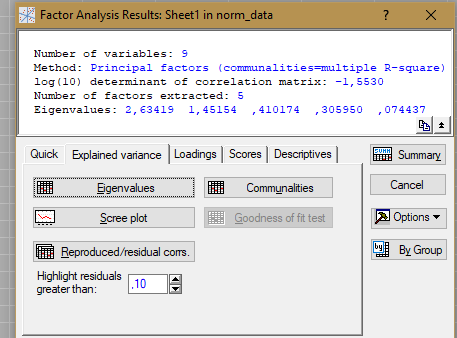
Субъекты РФ характеризуются социально-экономическими показателями, обозначение и наименование которых приведены в таблице 1. Ставится задача на основании статистических данных по показателям, соответствующим нужному варианту:

1. с помощью компонентного анализа и метода главных факторов снизить размерность признакового пространства, обеспечив уровень информативности не менее 70%;
2. при необходимости провести вращение пространства новых факторов;
3. дать экономическую интерпретацию факторам;
4. найти матрицу индивидуальных значений факторов.

|  |  |
| --- | --- |
|  | Число дорожно-транспортных происшествий и пострадавших в них на 100 000 человек населения |
|  | Смертность населения старше трудоспособного возраста, на 100 000 человек населения соответствующего возраста |
|  | Продажа сильно алкогольной продукции населению(тысяч декалитров)/**на тыс населения** |
|  | Средняя Стоимость минимального (условного) набора потребительских товаров и услуг |
|  | Число спортивных сооружений/ **на тыс населения** |
|  | Доходы консолидированных бюджетов субъектов Российской Федерации / **на тыс населения** |
|  | Предварительно расследовано преступлений, совершенных в состоянии алкогольного опьянения/ **на тыс населения** |
|  | Среднедушевые доходы населения (в месяц), руб. |
|  | Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на 10 000 человек населения, всего |

2. Снижение размерности признакового пространства методом главных факторов

Оценками общностей в данном алгоритме будут служить квадраты оценок множественных коэффициентов корреляции, т.е.: , .



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Value | Eigenvalues (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) | | | |
| |  | | --- | | Eigenvalue | | |  | | --- | | % Total variance | | |  | | --- | | Cumulative Eigenvalue | | |  | | --- | | Cumulative % | |
| |  | | --- | | 1 | | 2,634186 | 29,26873 | 2,634186 | 29,26873 |
| |  | | --- | | 2 | | 1,451542 | 16,12824 | 4,085728 | 45,39698 |
| |  | | --- | | 3 | | 0,410174 | 4,55749 | 4,495902 | 49,95447 |
| |  | | --- | | 4 | | 0,305950 | 3,39945 | 4,801852 | 53,35392 |
| |  | | --- | | 5 | | 0,074437 | 0,82708 | 4,876290 | 54,18099 |

Количество главных компонент можно посчитать с помощью Критерия Кайзера (отбираются факторы с собственными числами > 1, в данном примере их получается 2).

Вклад двух главных факторов в суммарную дисперсию исходных признаков (в дисперсию процесса) составляет 45,40%.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Factor Loadings (Unrotated) (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) (Marked loadings are >,700000) | | | | |
| |  | | --- | | Factor 1 | | |  | | --- | | Factor 2 | | |  | | --- | | Factor 3 | | |  | | --- | | Factor 4 | | |  | | --- | | Factor 5 | |
| |  | | --- | | X1 | | -0,230074 | -0,598382 | 0,176050 | -0,111635 | -0,109549 |
| |  | | --- | | X2 | | -0,670846 | -0,455738 | -0,034723 | -0,060474 | 0,106714 |
| |  | | --- | | X3 | | -0,748329 | -0,043655 | -0,169933 | -0,322827 | -0,012205 |
| |  | | --- | | X4 | | 0,056665 | -0,025394 | -0,411551 | 0,092434 | 0,120805 |
| |  | | --- | | X5 | | -0,310735 | -0,266333 | 0,293893 | 0,094639 | 0,119476 |
| |  | | --- | | X6 | | -0,689085 | 0,519744 | 0,161186 | 0,224163 | -0,013744 |
| |  | | --- | | X7 | | -0,721658 | -0,269138 | -0,085050 | 0,261090 | -0,025184 |
| |  | | --- | | X8 | | -0,549012 | 0,684022 | 0,077839 | -0,192886 | 0,034901 |
| |  | | --- | | X9 | | 0,417625 | -0,042937 | 0,232545 | -0,111764 | 0,141386 |
| |  | | --- | | Expl.Var | | 2,634186 | 1,451542 | 0,410174 | 0,305950 | 0,074437 |
| |  | | --- | | Prp.Totl | | 0,292687 | 0,161282 | 0,045575 | 0,033994 | 0,008271 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable |  | Factor Loadings (Unrotated) (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) (Marked loadings are >,700000) | |
|  | |  | | --- | | Factor 1 | | |  | | --- | | Factor 2 | |
| Число дорожно-транспортных происшествий и пострадавших в них на 100 000 человек населения |  | -0,230074 | -0,598382 |
| Смертность населения старше трудоспособного возраста, на 100 000 человек населения соответствующего возраста |  | -0,670846 | -0,455738 |
| Продажа сильно алкогольной продукции населению(тысяч декалитров)/**на тыс населения** |  | -0,748329 | -0,043655 |
| Средняя Стоимость минимального (условного) набора потребительских товаров и услуг |  | 0,056665 | -0,025394 |
| Число спортивных сооружений/ **на тыс населения** |  | -0,310735 | -0,266333 |
| Доходы консолидированных бюджетов субъектов Российской Федерации / **на тыс населения** |  | -0,689085 | 0,519744 |
| Предварительно расследовано преступлений, совершенных в состоянии алкогольного опьянения/ **на тыс населения** |  | -0,721658 | -0,269138 |
| Среднедушевые доходы населения (в месяц), руб. |  | -0,549012 | 0,684022 |
| Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на 10 000 человек населения, всего |  | 0,417625 | -0,042937 |
| |  | | --- | | Expl.Var | |  | 2,634186 | 1,451542 |
| |  | | --- | | Prp.Totl | |  | 0,292687 | 0,161282 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Factor Loadings (Quartimax raw) (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) (Marked loadings are >,700000) | | | | |
| |  | | --- | | Factor 1 | | |  | | --- | | Factor 2 | | |  | | --- | | Factor 3 | | |  | | --- | | Factor 4 | | |  | | --- | | Factor 5 | |
| Число дорожно-транспортных происшествий и пострадавших в них на 100 000 человек населения | 0,**516797** | -0,308938 | 0,318535 | 0,038289 | 0,031609 |
| Смертность населения старше трудоспособного возраста, на 100 000 человек населения соответствующего возраста | 0,**812441** | 0,042198 | 0,032508 | 0,046819 | -0,094302 |
| Продажа сильно алкогольной продукции населению(тысяч декалитров)/**на тыс населения** | 0,635544 | 0,392742 | -0,050726 | 0,365763 | 0,025034 |
| Средняя Стоимость минимального (условного) набора потребительских товаров и услуг | 0,015592 | -0,109084 | -0,428584 | 0,023239 | 0,000724 |
| Число спортивных сооружений/ **на тыс населения** | 0,382542 | 0,019433 | 0,251506 | -0,195582 | -0,169910 |
| Доходы консолидированных бюджетов субъектов Российской Федерации / **на тыс населения** | 0,226247 | 0,846137 | 0,012817 | -0,227642 | 0,047654 |
| Предварительно расследовано преступлений, совершенных в состоянии алкогольного опьянения/ **на тыс населения** | 0,745297 | 0,211844 | -0,090168 | -0,223597 | 0,103871 |
| Среднедушевые доходы населения (в месяц), руб. | 0,022752 | 0,879269 | 0,007153 | 0,194977 | -0,045644 |
| Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на 10 000 человек населения, всего | -0,326133 | -0,252387 | 0,202756 | 0,017887 | -0,226541 |
| |  | | --- | | Expl.Var | | 2,191174 | 1,861387 | 0,401490 | 0,316385 | 0,105853 |
| |  | | --- | | Prp.Totl | | 0,243464 | 0,206821 | 0,044610 | 0,035154 | 0,011761 |



Первый фактор тесно связан с Алкоголизмом, ДТП и смертностью, его можно охарактеризовать как показатель маргинального образа жизни

Второй фактор связан с доходами бюджетов и населения, его можно назвать – «Доходы»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Communalities (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) Rotation: Quartimax raw | | | | | |
| |  | | --- | | From 1 Factor | | |  | | --- | | From 2 Factors | | |  | | --- | | From 3 Factors | | |  | | --- | | From 4 Factors | | |  | | --- | | From 5 Factors | | |  | | --- | | Multiple R-Square | |
| |  | | --- | | X1 | | 0,267079 | 0,362522 | 0,463987 | 0,465453 | 0,466452 | 0,388837 |
| |  | | --- | | X2 | | 0,660060 | 0,661841 | 0,662898 | 0,665090 | 0,673982 | 0,590637 |
| |  | | --- | | X3 | | 0,403917 | 0,558163 | 0,560736 | 0,694519 | 0,695146 | 0,627354 |
| |  | | --- | | X4 | | 0,000243 | 0,012143 | 0,195827 | 0,196367 | 0,196368 | 0,129942 |
| |  | | --- | | X5 | | 0,146338 | 0,146716 | 0,209971 | 0,248224 | 0,277093 | 0,216776 |
| |  | | --- | | X6 | | 0,051188 | 0,767136 | 0,767300 | 0,819121 | 0,821392 | 0,746782 |
| |  | | --- | | X7 | | 0,555468 | 0,600346 | 0,608477 | 0,658472 | 0,669261 | 0,604932 |
| |  | | --- | | X8 | | 0,000518 | 0,773632 | 0,773683 | 0,811699 | 0,813783 | 0,746572 |
| |  | | --- | | X9 | | 0,106363 | 0,170062 | 0,211172 | 0,211492 | 0,262813 | 0,219266 |

Далее представлена таблица вкладов факторов в дисперсию исходных признаков. Элементы первого столбца таблицы равны квадратам соответствующих элементов первого столбца матрицы В, элементы второго столбца – сумме квадратов соответствующих элементов первого и второго столбцов матрицы В, элементы третьего столбца – сумме квадратов элементов первого, второго и третьего столбцов матрицы В. Таким образом, в третьем столбце представлены оценки общностей, рассчитанные по матрице В, а в четвертом столбце – оценки общностей, рассчитанные по формуле:

, 

Т.к. исходные признаки процентированы и пронормированы, а главные факторы не коррелированы между собой, то оценки характерностей можно рассчитать следующим образом:

Центрировано-нормированные исходные признаки связаны с главными и характерными факторами следующими выражениями: ():

Оценка редуцированной матрицы парных коэффициентов корреляции, рассчитанная по матрице нагрузок В, и оценка остаточной матрицы парных коэффициентов корреляции представлены ниже:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Residual Correlations (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) (Marked residuals are > ,100000) | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | |
| |  | | --- | | X1 | | 0,53 | 0,01 | 0,03 | -0,07 | -0,00 | -0,03 | 0,04 | -0,02 | 0,01 |
| |  | | --- | | X2 | | 0,01 | 0,33 | 0,05 | 0,03 | 0,05 | -0,01 | 0,02 | -0,01 | 0,01 |
| |  | | --- | | X3 | | 0,03 | 0,05 | 0,30 | 0,01 | -0,02 | -0,02 | -0,01 | 0,05 | -0,03 |
| |  | | --- | | X4 | | -0,07 | 0,03 | 0,01 | 0,80 | -0,02 | -0,03 | 0,04 | -0,01 | 0,01 |
| |  | | --- | | X5 | | -0,00 | 0,05 | -0,02 | -0,02 | 0,72 | 0,02 | -0,00 | -0,01 | 0,02 |
| |  | | --- | | X6 | | -0,03 | -0,01 | -0,02 | -0,03 | 0,02 | 0,18 | 0,05 | 0,05 | -0,02 |
| |  | | --- | | X7 | | 0,04 | 0,02 | -0,01 | 0,04 | -0,00 | 0,05 | 0,33 | -0,02 | -0,02 |
| |  | | --- | | X8 | | -0,02 | -0,01 | 0,05 | -0,01 | -0,01 | 0,05 | -0,02 | 0,19 | 0,03 |
| |  | | --- | | X9 | | 0,01 | 0,01 | -0,03 | 0,01 | 0,02 | -0,02 | -0,02 | 0,03 | 0,74 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Reproduced Correlations (Sheet1 in norm\_data) Extraction: Principal factors (comm.=multiple R-square) | | | | | | | | |
| |  | | --- | | X1 | | |  | | --- | | X2 | | |  | | --- | | X3 | | |  | | --- | | X4 | | |  | | --- | | X5 | | |  | | --- | | X6 | | |  | | --- | | X7 | | |  | | --- | | X8 | | |  | | --- | | X9 | |
| |  | | --- | | X1 | | 0,47 | 0,42 | 0,21 | -0,09 | 0,26 | -0,15 | 0,29 | -0,25 | -0,03 |
| |  | | --- | | X2 | | 0,42 | 0,67 | 0,55 | -0,00 | 0,33 | 0,20 | 0,59 | 0,07 | -0,25 |
| |  | | --- | | X3 | | 0,21 | 0,55 | 0,70 | -0,00 | 0,16 | 0,39 | 0,48 | 0,43 | -0,32 |
| |  | | --- | | X4 | | -0,09 | -0,00 | -0,00 | 0,20 | -0,11 | -0,10 | 0,02 | -0,09 | -0,06 |
| |  | | --- | | X5 | | 0,26 | 0,33 | 0,16 | -0,11 | 0,28 | 0,14 | 0,29 | -0,00 | -0,04 |
| |  | | --- | | X6 | | -0,15 | 0,20 | 0,39 | -0,10 | 0,14 | 0,82 | 0,40 | 0,70 | -0,30 |
| |  | | --- | | X7 | | 0,29 | 0,59 | 0,48 | 0,02 | 0,29 | 0,40 | 0,67 | 0,15 | -0,34 |
| |  | | --- | | X8 | | -0,25 | 0,07 | 0,43 | -0,09 | -0,00 | 0,70 | 0,15 | 0,81 | -0,21 |
| |  | | --- | | X9 | | -0,03 | -0,25 | -0,32 | -0,06 | -0,04 | -0,30 | -0,34 | -0,21 | 0,26 |

3. Метод Уорда (МГФ)

Методом Уорда при пороговом значении расстояния ρпор=6,920 все субъекты РФ разбиваются на 5 классов S1= {S11, S12, S13, S14, S15}, состав которых приведен в таблице 6.



|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Кол-во объектов** | **Состав класса** |
| Кластер 1 {S11} | 27 | Алтайский край  Владимирская область  Вологодская область  Ивановская область  Иркутская область  Кемеровская область — Кузбасс  Кировская область  Костромская область  Курганская область  Ленинградская область  Нижегородская область  Новгородская область  Пензенская область  Пермский край  Приморский край  Псковская область  Республика Башкортостан  Республика Бурятия  Республика Марий Эл  Республика Хакасия  Смоленская область  Тверская область  Тюменская область без автономных округов  Удмуртская Республика  Челябинская область  Чувашская Республика — Чувашия  Ярославская область |
| Кластер 2 {S12} | 7 | Кабардино-Балкарская Республика  Карачаево-Черкесская Республика  Республика Дагестан  Республика Ингушетия  Республика Северная Осетия — Алания  Ставропольский край  Чеченская Республика |
| Кластер 3 {S13} | **11** | Амурская область  Архангельская область без автономного ок  Еврейская автономная область  Забайкальский край  Мурманская область  Республика Алтай  Республика Карелия  Республика Коми  Республика Тыва  Сахалинская область  Хабаровский край |
| Кластер 4 {S14} | 6 | г. Москва  Камчатский край  Магаданская область  Ненецкий автономный округ  Чукотский автономный округ  Ямало-Ненецкий автономный округ |
| Кластер 5 {S15} | 34 | Астраханская область  Белгородская область  Брянская область  Волгоградская область  Воронежская область  Калининградская область  Калужская область  Краснодарский край  Красноярский край  Курская область  Липецкая область  Московская область  Новосибирская область  Омская область  Оренбургская область  Орловская область  Республика Адыгея (Адыгея)  Республика Калмыкия  Республика Крым  Республика Мордовия  Республика Саха (Якутия)  Республика Татарстан (Татарстан)  Ростовская область  Рязанская область  Самарская область  Санкт-Петербург  Саратовская область  Свердловская область  Севастополь  Тамбовская область  Томская область  Тульская область  Ульяновская область  Ханты-Мансийский автономный округ — Югра |

К средних (МГФ)

|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Кол-во объектов** | **Состав класса** |
| Кластер 1 {S11} | 3 | Ненецкий автономный округ  Чукотский автономный округ  Ямало-Ненецкий автономный округ |
| Кластер 2 {S12} | 34 | Астраханская область  Белгородская область  Брянская область  Волгоградская область  Воронежская область  Калининградская область  Калужская область  Краснодарский край  Красноярский край  Курская область  Ленинградская область  Липецкая область  Нижегородская область  Новосибирская область  Омская область  Оренбургская область  Орловская область  Республика Калмыкия  Республика Крым  Республика Марий Эл  Республика Мордовия  Республика Саха (Якутия)  Республика Татарстан (Татарстан)  Рязанская область  Самарская область  Саратовская область  Свердловская область  Севастополь  Тамбовская область  Томская область  Тульская область  Тюменская область без автономных округов  Ульяновская область  Челябинская область. |
| Кластер 3 {S13} | 33 | Алтайский край  Амурская область  Архангельская область без автономного ок  Владимирская область  Вологодская область  Еврейская автономная область  Забайкальский край  Ивановская область  Иркутская область  Кемеровская область — Кузбасс  Кировская область  Костромская область  Курганская область  Мурманская область  Новгородская область  Пензенская область  Пермский край  Приморский край  Псковская область  Республика Алтай  Республика Башкортостан  Республика Бурятия  Республика Карелия  Республика Коми  Республика Тыва  Республика Хакасия  Сахалинская область  Смоленская область  Тверская область  Удмуртская Республика  Хабаровский край  Чувашская Республика — Чувашия  Ярославская область |
| Кластер 4 {S14} | 2 | Камчатский край  Магаданская область |
| Кластер 5 {S15} | 13 | г. Москва  Кабардино-Балкарская Республика  Карачаево-Черкесская Республика  Московская область  Республика Адыгея (Адыгея)  Республика Дагестан  Республика Ингушетия  Республика Северная Осетия — Алания  Ростовская область  Санкт-Петербург  Ставропольский край  Ханты-Мансийский автономный округ — Югра  Чеченская Республика |