

УДК 51.77:311.21

doi: 10.18101/2304-5728-2016-3-23-32

© *Нямхуу Батжаргал, А. С. Цыбиков, Т. Г. Дармаев, Э. Д. Дагбаев*

### **Корреспондентский анализ данных опроса молодежи Монголии<sup>1</sup>**

В области прикладного анализа данных, в частности в социологии, часто возникают проблемы с поиском закономерностей в данных неметрического типа. В данной статье рассматривается один современных математических инструментов анализа непараметрических данных – корреспондентский анализ. Описан пошаговый алгоритм его реализации. Суть метода заключается в представлении содержимого таблицы относительных частот (таблицы сопряженности) в виде расстояний между отдельными строками и/или столбцами таблицы в пространстве возможно более низкой размерности с метрикой  $\chi^2$ . Это позволяет визуализировать структуру взаимосвязей между переменными. Практическое применение данного метода с помощью пакета Statistica позволило получить новые данные о различиях социально-политических представлений молодежи Монголии, занимающихся разными видами деятельности. Полученная информация может быть полезна для стратегического планирования взаимоотношений России и Монголии.

**Ключевые слова:** корреспондентский анализ, молодежь Монголии, общественно-политическое сознание.

© *Nyamhu Batjargal, A. S. Tsybikov, T. G. Darmayev, E. D. Dagbayev*

### **Correspondence analysis of Mongolian youth survey data**

Troubles with finding patterns in the data of not the metric type often arise in the field of applied data analysis, particularly in sociology. This article discusses one of modern mathematical tools for analysis of non-parametric data – correspondent analysis. The algorithm of its realization is described step by step. The method consists in the presentation of the contents of the table of relative frequencies (contingency tables) as the distances between the rows and/or columns of the table in the space maybe of a lower dimension with metric  $\chi^2$ . It allows to visualize the structure of relationships between variables. The practical application of this method using Statistica package allowed us to obtain new data on the differences of social and political views of Mongolian young people involved in different activities. The information may be useful for strategic relations planning of Russia and Mongolia.

---

<sup>1</sup> Исследование выполнено при финансовой поддержке РГНФ. «Общественное сознание современной молодежи Монголии», проект № 15-23-03001

**Keywords:** correspondence analysis, young people of Mongolia, Mongolian youth, socio-political consciousness

### Введение

Как известно, в теории статистического анализа данных разработаны множество разведывательных параметрических и непараметрических методов, позволяющие искать и формализовать (моделировать) латентные закономерности в данных практически любой природы. В области анализа социологических данных, с большим количеством переменных (вопросов) и необходимым числом наблюдений (респондентов), одним из мощных современных непараметрических инструментов анализа является корреспондентский анализ (в некоторых источниках – «анализ соответствий») [1, 2, 3].

### Математическое описание метода

Процедура анализа соответствий предполагает применение описательных и разведочных методов анализа многомерных таблиц сопряженности. Непараметрические методы анализ многовходовых таблиц сопряженности позволяют определить (выявить) структуру взаимосвязи и обнаружить связи между категориальными переменными, образующих исходную таблицу данных. Назначение анализа соответствий — отображение таблицы относительных частот в форме расстояний между категориальными уровнями независимых переменных, т.е. отдельными строками и (или) столбцами таблицы сопряженности в пространстве относительно более низкой размерности. Стандартный корреспондентский анализ требует предварительной стандартизации относительных частот таблицы таким образом, чтобы сумма наблюдений равнялась единице во всех ячейках. Главным критерием качества отображения многовходовой таблицы в пространстве низкой размерности выступает значение статистики  $\chi^2$ .

Итак, для реализации алгоритма необходимо вычислить следующие матрицы:

- матрица  $P$  представляет исходную матрицу относительных частот, в которой каждый элемент вычисляется как абсолютная частота, деленная на сумму всех элементов таблицы ( $n_{ij} / n$ );
- матрица  $r$  представляет собой вектор, составленный из сумм элементов строк матрицы  $P$ ;
- матрица  $c$  представляет вектор, составленный из сумм элементов столбцов матрицы  $P$ ;
- матрица  $D_r$  представляет собой диагональную матрицу, соответствующие элементы главной диагонали  $D_r$  являются поэлементными суммами строк матрицы  $P$ ;

• матрица  $D_c$  представляет диагональную матрицу, соответствующие элементы главной диагонали  $D_r$  являются поэлементными суммами столбцов матрицы  $P$ .

Технология расчета координат точек-строк и точек-столбцов в пространстве низкой (меньшей) размерности основывается на обобщенном сингулярном разложении матрицы  $P$ . Матричное представление разложения, в данном случае, будет иметь следующий вид  $P = AD_u B$ , так что

$$AD_r^{-1}A = BD_c^{-1}B = I,$$

где  $A$  — матрица левосторонних обобщенных сингулярных векторов,  $B$  — матрица правосторонних обобщенных сингулярных векторов,  $D_u$  — диагональная матрица, диагональные элементы которой равны обобщенным сингулярным числам,  $I$  — единичная матрица.

Процедура расчета координат строк и столбцов зависит от применяемого способа стандартизации таблицы относительных частот. Расположение точек-строк и точек-столбцов относительно друг друга не зависит от выбора способа стандартизации. При канонической стандартизации координаты строк вычисляются по формуле

$$F = D_r^{-1}AD_u^s,$$

а координаты столбцов по формуле

$$G = D_c^{-1}BD_u^s.$$

Проведение анализа можно разбить на несколько шагов.

1. Вычисляются относительные частоты для таблицы данных рассматриваемых переменных, так что сумма всех элементов таблицы будет равна 1 (каждый элемент делится на общее число наблюдений). Полученная нормированная таблица показывает, как распределена единичная масса по ячейкам.

2. Вычисляется инерция. Инерция определяется как значение статистики хи-квадрат Пирсона для двухвходовой таблицы, деленное на общее количество наблюдений.

3. Вычисляются сингулярные значения, собственные значения, проценты объясненной инерции, кумулятивные проценты и вклады в статистику хи-квадрат каждого собственного значения.

4. Определяется оптимальная размерность по критерию каменистой осыпи.

5. Вычисляются координаты в пространстве новой размерности. Предварительно проводится стандартизация таблицы относительных частот. Используется метрика хи-квадрат.

6. Анализ качества модели (статистики «качество», «косинус2», «относительная инерция»).

7. Графическое представление результатов. Интерпретация результатов.

### Анализ рода деятельности и политической жизни молодежи Монголии

Вышеизложенный метод применен для поиска закономерностей в данных социологического опроса «Общественное сознание современной молодежи Монголии» с целью получения новых научных данных. Анкета опроса состоит из 75 общих, политических и социально-психологических вопросов [4]. В результате массового опроса с применением веб-технологий имеем таблицу данных из 75 переменных категориального типа и 367 наблюдений (опрошенных респондентов) [5,6].

Методика проведения анализа. Применение корреспондентского анализа для решения данной задачи осуществляется с помощью специализированного программного пакета Statistica 10 (модуль «Correspondence analysis») [1].

С помощью методики корреспондентского анализа исследована структура взаимосвязи рода деятельности молодежи Монголии с их гражданскими и политическими взглядами. В ходе анализа получены следующие результаты.

1. При анализе двумерной карты взаимосвязей вида деятельности респондентов с предпочтениями основных выпускаемых газет Монголии выявлено, что учащиеся дают предпочтение газете «Оноодор», рабочие отдают предпочтение газете «Одрийн сонин», безработные – «Онэн» (рис. 1). Два измерения дают 92,22% и 7,7% инерции соответственно, в сумме 99,92%. Это говорит о высокой точности воспроизведения относительных частот в метрике хи-квадрат и тем самым дает возможность по двумерному графику обоснованно интерпретировать полученный результат. Статистическая взаимосвязь между данными переменными отражает номинальная корреляция по  $\chi^2$ -Пирсону:  $\varphi=0,46$ ;  $p<0,05$ .

2. Подобным образом, выявлена связь рода деятельности с предпочтениями фильмов, разбитых по странам производителя ( $\varphi=0,45$ ;  $p<0,05$ ). Учащиеся в явном большинстве предпочитают фильмы американского производства, рабочий класс склонен к выбору монгольских фильмов, служащие – российские (в соответствии с правилами интерпретации, угол образованный между двумя точками в начале координат острый). Значения инерции равны соответственно 98,88% и 1,12%, кумулятивная – 100% (рис. 2).

3. Рабочие придерживаются мнения, что мораль, культура и нравственность находятся в кризисе. Однако, учащиеся с этим не согласны. Безработные затрудняются ответить. Кумулятивная инерция – 100% (рис. 3). Статистическая взаимосвязь показателя о морали и нравственности и видом деятельности достоверна ( $p<0,05$ ) и равна  $\varphi=0,34$ .

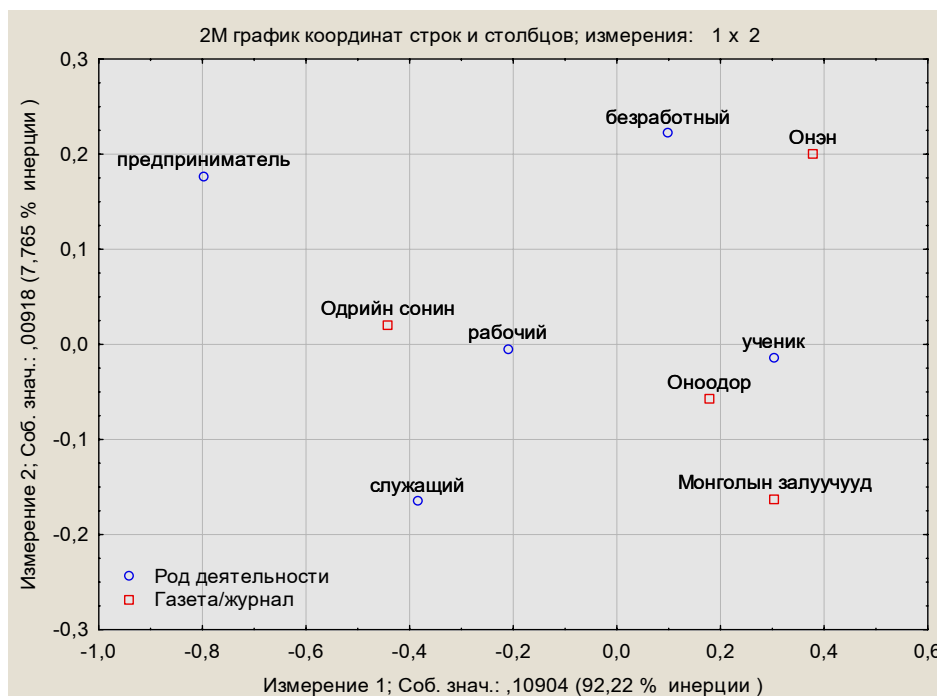


Рис. 1. Вид деятельности и выбор газеты

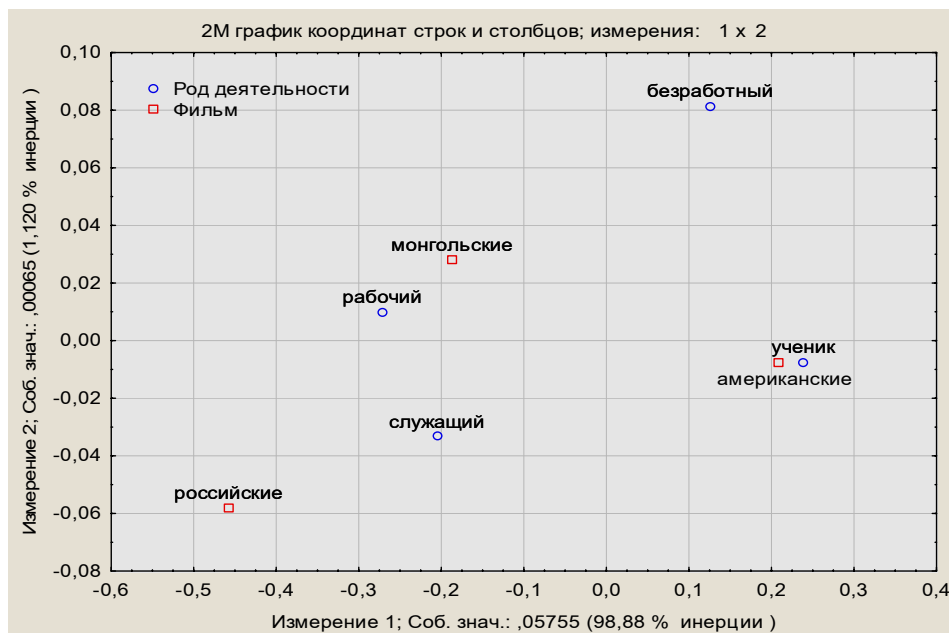


Рис. 2. Вид деятельности и предпочтение фильмов

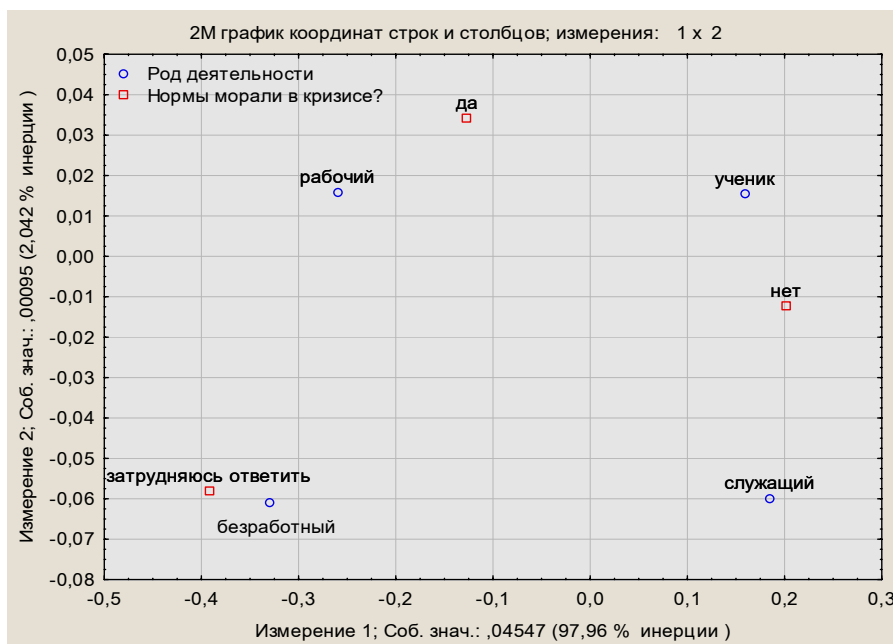


Рис. 3. Вид деятельности и мнение о морали, нравственности

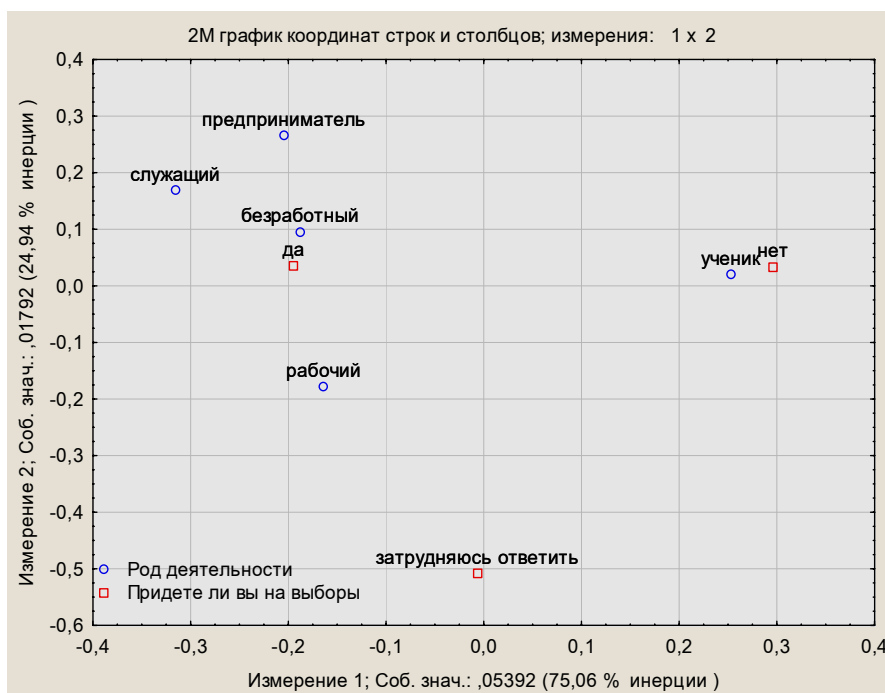


Рис. 4. Вид деятельности и посещение выборов

4. Безработные, служащие и предприниматели имеют самую высокую вероятность посещения выборов. Учащиеся (в данной выборке только студенты) не желают участвовать в выборах. Возможно, это объясняется тем, что они считают, что итоги выборов фальсифицируются, либо их голос ничего не значит. Статистическая взаимосвязь:  $\phi=0,32$ ;  $p<0,05$  (рис. 4).

5. Готовность участия в политических акциях, митингах и манифестах тесно связано с критериями отбора кандидатов на выборные должности ( $\phi=0,65$ ;  $p<0,01$ ). Так, например, респонденты-молодежь, готовые прийти на избирательный участок, основным критерием отбора кандидатов считают образование. Молодежь, ответившая отрицательно на вопрос об участии в политических акциях, чаще всего выбирала вариант ответа «коммуникабельность» и «деловые и нравственные качества» (рис. 5).

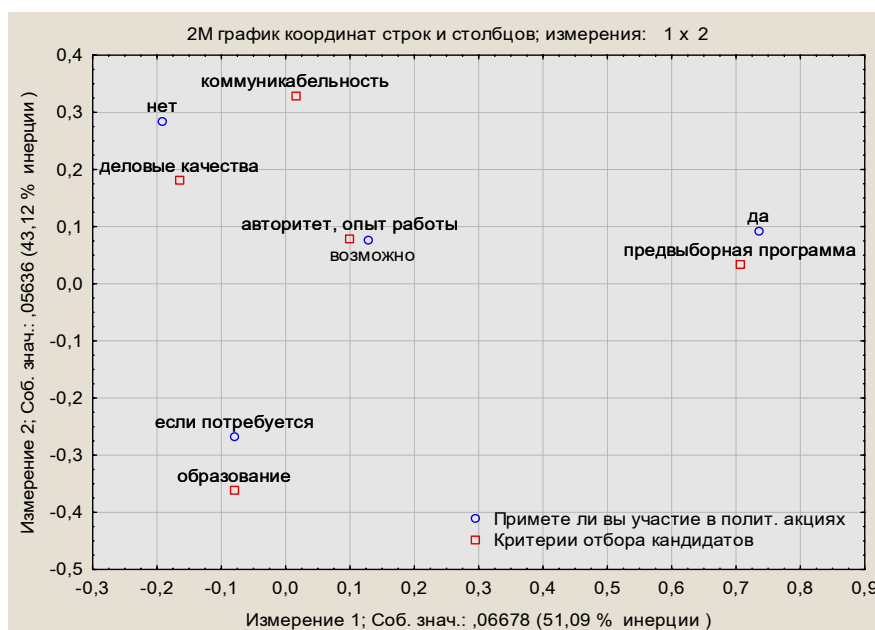


Рис. 5. Участие в политических акциях и критерии отбора кандидатов

6. Значение слова «самостоятельность» монголы разного рода деятельности понимают по-разному ( $\phi=0,45$ ;  $p<0,05$ ). Рабочие и безработные склонны считать, что самостоятельность – это быть хозяином самому себе. Для учащихся синонимом данного слова является готовность к любым изменениям или свобода. Для служащего – независимость (рис. 6).

7. Определена достоверная связь вида деятельности молодежи со взглядом на основные проблемы в Улан-Батаре ( $\phi=0,44$ ;  $p<0,05$ ). Рабочие считают, что основная проблема – смог, учащиеся (студенты) – низкий уровень проживания (рис. 7). Общая инерция – 99,82%.

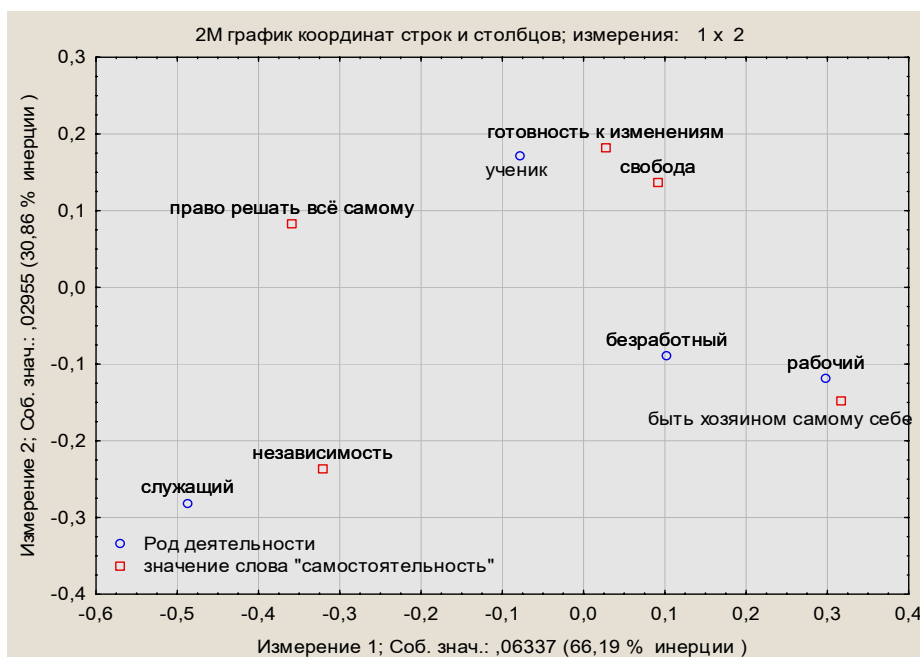


Рис. 6. Род деятельности и понимание слова «самостоятельность»

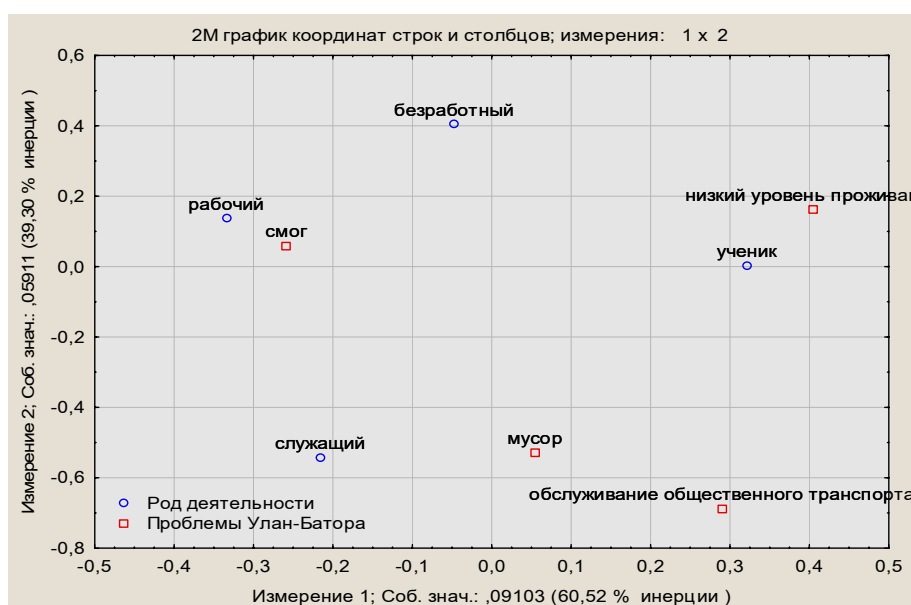


Рис. 7. Рода деятельности и взгляд на проблемы Улан-Батора

### Заключение

Итак, с помощью математического аппарата «анализ соответствий» и программного пакета Statistica получены данные, позволяющие яснее и глубже понять общественно-политическое сознание молодежи Монголии.



Так, например, если обобщить полученные данные, то можно увидеть кардинальные расхождения представлений студентов и рабочих (рис. 8).

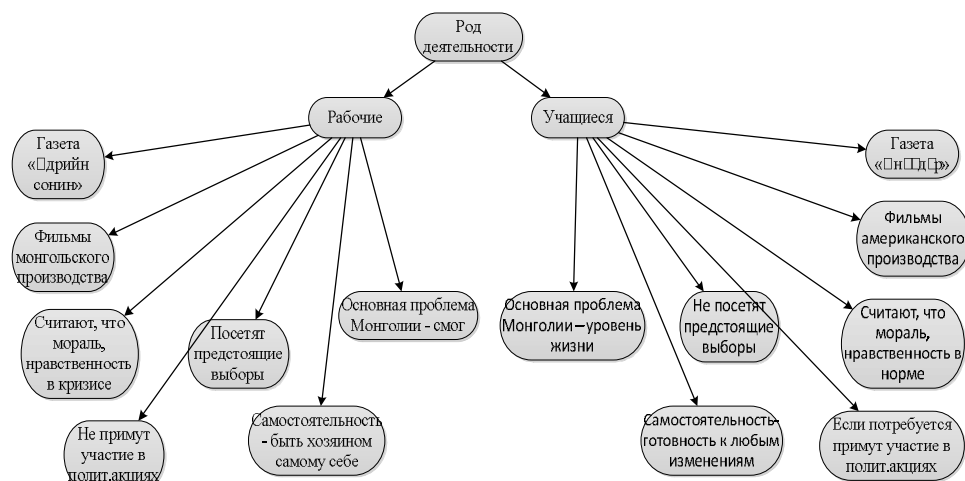


Рис. 8. Представления учащихся и рабочих

### Литература

1. StatSoft, Inc. (2012). Электронный учебник по статистике. Москва, StatSoft. WEB: <http://www.statsoft.ru/home/textbook/default.htm>.
2. Blasius J. Correspondence Analysis in Social Science Research / Correspondence Analysis in the Social Sciences (pp.23-52). San Diego, CA: Academic Press. 1994.
3. Клишина Ю. Н. Применение анализа соответствий в обработке нечисловой информации // Социология: 4М. — 1991. — №2. — С.105 – 119.
4. Нямхуу Б., Актамов И. Г., Халбашкеев А. В., Дагбаев Э. Д., Дармаев Т. Г. Методологические вопросы изучения общественного сознания современной монгольской молодежи // Вестник Бурятского государственного университета. — 2015. — 2015/14. — С. 79 – 83.
5. Цыбиков А.С., Батжаргал Нямхуу, Дагбаев Э. Д., Дармаев Т. Г. Анализ объема и структуры выборки социологического опроса молодежи Монголии // Вестник Бурятского государственного университета. — 2015. — №1. — С. 25 – 32.
6. Батжаргал Нямхуу, Хабитуев Б. В., Дармаев Т. Г. Информационные технологии в социологических исследованиях // Вестник Бурятского государственного университета. Математика, информатика. — 2015. — № 3. — С. 33 – 38.

### References

1. StatSoft, Inc. (2012). Jelektronnyj uchebnyj po statistike. Moskva, StatSoft. WEB: <http://www.statsoft.ru/home/textbook/default.htm>.

2. Blasius J. Correspondence Analysis in Social Science Research / Correspondence Analysis in the Social Sciences (pp.23-52). San Diego, CA: Academic Press. 1994.

3. Klishina Ju. N. Primenenie analiza sootvetstvij v obrabotke nechislovoj informacii // Sociologija: 4M. — 1991. — №2. — S.105 – 119.

4. Njamhuu B., Aktamov I. G., Halbashkeev A. V., Dagbaev Je. D., Darmaev T. G. Metodologicheskie voprosy izuchenija obshhestvennogo soznaniya sovremennoj mongol'skoj molodezhi // Vestnik Burjatskogo gosudarstvennogo universiteta. — 2015. — 2015/14. — S. 79 – 83.

5. Cybikov A.S., Batzhargal Njamhuu, Dagbaev Je. D., Darmaev T. G. Analiz ob#ema i struktury vyborki sociologicheskogo oprosa molodezhi Mongolii // Vestnik Burjatskogo gosudarstvennogo universiteta. — 2015. — №1. — S. 25 – 32.

6. Batzhargal Njamhuu, Habituev B. V., Darmaev T. G. Informacionnye tehnologii v sociologicheskikh issledovanijah // Vestnik Burjatskogo gosudarstvennogo universiteta. Matematika, informatika. — 2015. — № 3. — S. 33 – 38.

*Нямхуу Батжаргал*, профессор Института иностранного языка Монгольского университета науки и технологий, e-mail: intel\_bat@yahoo.com.

*Цыбиков Анатолий Сергеевич*, кандидат педагогических наук, старший научный сотрудник Научно-образовательного и инновационного центра системных исследований и автоматизации Бурятского государственного университета, e-mail: cas313@gambler.ru.

*Дармаев Тумэн Гомбоцыренович*, кандидат физико-математических наук, доцент, заведующий лабораторией вычислительных и геоинформационных систем Бурятского государственного университета, e-mail: dtumen@mail.ru.

*Дажбаев Эрдэм Данзанович*, доктор социологических наук, профессор кафедры политологии и социологии Бурятского государственного университета, e-mail: edagbaev@mail.ru.

*Nyamhu Batjargal*, Professor of Mongolian University of Science and Technology.

*Tsybikov Anatoly Sergeevich*, senior researcher, scientific and educational innovation center of system research and automation, Institute of Mathematics and Computer Science, Buryat State University.

*Darmaev Tumen Gombotsyrenovich*, candidate of physical and mathematical sciences, associate professor, head of laboratory of calculation and geoinformational technologies, scientific and educational innovation center of system research and automation, Institute of Mathematics and Computer Science, Buryat State University.

*Dagbaev Erdem Danzanovich*, Doctor of social sciences, professor of Political Science and Sociology of the Buryat State University.