

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ В СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Ключевые слова: корреляционный анализ, социологическое исследование.

Проведен корреляционный анализ результатов социологического исследования, посвященного проблеме отношения населения г.Казани к количеству выходных дней, отводимых на празднование Нового Года. Найдены статистические взаимосвязи между различными группами признаков. Выявлено, что практически во всех случаях величина коэффициента корреляции характеризует весьма слабую связь между переменными. Такие факторы, как гендерная принадлежность респондента, его образование и материальное положение в наименьшей степени коррелируют с переменными исследуемой проблемы, что позволяет говорить об отсутствии связи по данной группе признаков. Относительно выраженной взаимозависимостью обладает фактическое и желаемое количество выходных дней, а также отмечена корреляционная связь этих переменных с возрастом респондентов.

Keywords: correlation analysis, sociological research.

A correlation analysis of the results of a sociological study on the problem of the relation of the population to the number of Kazan weekend, bled to celebrate the New Year. Found statistical relationships between different groups of symptoms. Found that in almost all cases the value of the correlation coefficient characterizes a very weak relationship between the variables. Factors such as gender of the respondent, his education and financial situation of the least correlated with the variables studied problem, which suggests that there is no connection for this group of symptoms. Relative terms, interdependence has actual and desired number of days off, and marked correlation between these variables, with the age of respondents

Применение инструментов математического анализа позволяет дать численную оценку изучаемым явлениям, независимо от их природы [1, 2]. Интерпретация полученных результатов во многом зависит от того, насколько правильно были проведены расчёты, а также обоснованы применяемые методы. Одним из наиболее распространенных видов анализа статистической информации является корреляционный анализ. Его сущность заключается в поиске связи между двумя или более переменными. Причем наличие такой связи в первую очередь характеризуется тем, насколько сильно она выражена. Мерой тесноты двух коррелирующих величин является некоторый критерий, который получил название коэффициента корреляции. Этот коэффициент обозначается латинской буквой r , и может принимать значения от +1 до -1. Чем ближе модуль коэффициента корреляции к единице, тем более сильной является связь между измеряемыми величинами. Отсутствие связи характеризуется коэффициентом корреляции равным 0 или близким к нему значением. Существует следующая градация силы связи, представленная шкалой Чертока.

- $0 < |r| < 0,1$ - связь практически отсутствует,
- $0,1 < |r| < 0,3$ - слабая связь,
- $0,3 < |r| < 0,5$ - умеренная связь,
- $0,5 < |r| < 0,7$ - связь средней силы,
- $0,7 < |r| < 0,9$ - сильная связь,
- $0,9 < |r| < 1$ - очень сильная связь.

Есть мнение, что в социологических исследованиях значения коэффициентов корреляции выше 0,5 встречаются не очень часто, поэтому можно принимать во внимание те из них, которые равны или превышают 0,3 [3], т. е. характеризуют умеренную взаимосвязь признаков.

Если коэффициент корреляции отрицательный, это означает наличие

противоположной связи: чем выше значение одной переменной, тем ниже значение другой.

Широкое применение корреляционный анализ нашел в гуманитарных науках, в том числе и при изучении социологических проблем [4]. Объясняется это в первую очередь тем, что исследуемые социологической наукой вопросы могут включать в себя большое число влияющих факторов. То есть определенному значению одной переменной, соответствует целый комплекс значений другой, представляющий собой ряд распределения, причем при изменении данной величины меняется ряд распределения и его среднее. [5] Если каждую пару значений этих величин изобразить на плоскости в декартовой системе координат с помощью точек, то наносимые точки расположатся в виде «облака», называемое корреляционным полем.

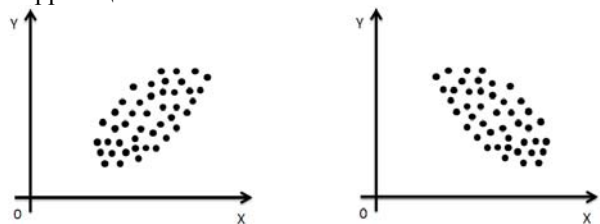


Рис 1. Корреляционное поле (слева – положительная корреляция, справа – отрицательная корреляция)

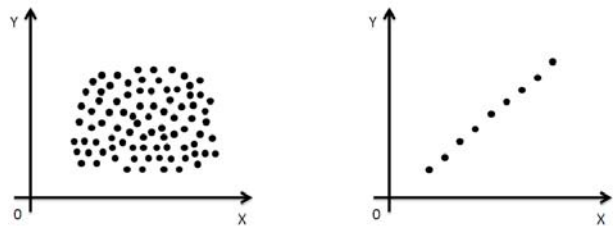


Рис 2. Корреляционное поле (слева – отсутствие связи, справа – функциональная связь)

Корреляционная зависимость не является абсолютно полной и точной. Она включает в себя множественность причин и следствий разных порядков. А любая социологическая проблема представляет собой явление, которое зависит от большого числа причин, действующих с разной силой. Изучая влияние одной переменной на другую, мы выделяем один фактор, но на зависимую величину оказывают влияние и многие другие, чем и обусловлен характер корреляционной зависимости.

Другой причиной удобства использования корреляционного анализа в социологии является применимость его для переменных, относящихся к различным шкалам. От вида шкалы зависит метод вычисления коэффициента корреляции:

- переменные с интервальной и с номинальной шкалой: коэффициент корреляции Пирсона (корреляция моментов произведений).

- по меньшей мере, одна из двух переменных имеет порядковую шкалу либо не является нормально распределённой: ранговая корреляция по Спирману.

- одна из двух переменных является дихотомической: точечная двухрядная корреляция.

- обе переменные являются дихотомическими: четырёхполевая корреляция.

Вместе с тем важным аспектом применения корреляционного анализа в социологических исследованиях является корректная интерпретация полученных данных. Так, значительная корреляция между двумя случайными величинами всегда является свидетельством существования некоторой статистической связи в данной выборке, но эта связь не обязательно должна наблюдаться для другой выборки и иметь причинно-следственный характер. Коэффициенты корреляции устанавливают лишь статистические взаимосвязи и свидетельствуют о том, что либо одно из двух выделяемых явлений является причиной другого, либо оба явления – следствие общих причин. В литературе приводится множество примеров вышесказанного. Так, рассматривая пожары в конкретном городе, можно выявить весьма высокую корреляцию между ущербом, который нанес пожар, и количеством пожарных, участвовавших в ликвидации пожара, причём эта корреляция будет положительной. Из этого, однако, не следует вывод «большее количество пожарных приводит к большему ущербу», и тем более не имеет смысла попытка минимизировать ущерб от пожаров путем ликвидации пожарных бригад [6].

Однако, несмотря на всю простоту корреляционного анализа, существует ряд ограничений, когда его применение невозможно. Во-первых, необходимо выполнение условия о наличии достаточного количества наблюдений для изучения. На практике считается, что число наблюдений должно не менее, чем в 5-6 раз превышать число факторов (также встречается рекомендация использовать пропорцию не менее, чем в 10 раз превышающую количество факторов) [6]. Во-вторых, необходимо, чтобы совокупность

значений всех факторных и результативного признаков подчинялась многомерному нормальному распределению. В случае, если объём совокупности недостаточен для проведения формального тестирования на нормальность распределения, то закон распределения определяется визуально на основе корреляционного поля. Если в расположении точек на этом поле наблюдается линейная тенденция, то можно предположить, что совокупность исходных данных подчиняется нормальному закону распределения. Если связь, к примеру, U-образная (неоднозначная), то коэффициент корреляции непригоден для использования в качестве меры силы связи: его значение стремится к нулю [7]. В-третьих, Исходная совокупность значений должна быть качественно однородной [6].

С целью иллюстрации применения корреляционного анализа в социологии, была проведена обработка данных, полученных в результате изучения отношения жителей г. Казани к количеству выходных дней, отводимых на празднование Нового Года. Исследование проводилось в населённых пунктах РТ в феврале 2012 года. Размер выборки составил 584 респондента. Выборка репрезентативно представляет генеральную совокупность по полу, возрасту и образованию. Метод опроса – интервью *facetoface*. В качестве статистических методов анализа использованы частотные распределения, таблицы сопряженности и корреляционный анализ.

Предложенная респондентам анкета состояла из двух блоков: содержательного и социально-демографического. В первом, респондентам предлагалось ответить на вопросы о том, сколько дней они отдыхали на Новогодние праздники в 2012 году, а также, сколько выходных дней государству необходимо выделять именно на новогодний период, а сколько на майские праздники. Во втором блоке анкеты была получена информация о гендерной принадлежности респондента, его возрасте, образовании, занятости и о материальном положении. Обработка данных проводилась с использованием пакета SPSSv.13, обладающим широкими возможностями в области статистического анализа социологической информации [8].

Как правило, первым этапом анализа статистических данных является частотный анализ [6]. В работе использовался статистический подход к анализу одномерных частотных распределений, с помощью которого решаются задачи выявления закона распределения вероятностей случайной величины и одномерных статистик данного распределения: моды, медианы, среднего арифметического, дисперсии, асимметрии и эксцесса. Приведем результаты одномерного частотного распределения по содержательному блоку анкеты. Так, на вопрос о том, сколько дней фактически отдыхали респонденты на новогодние праздники в 2012 году были получены следующие ответы:

- от 1 до 3 дней – 18,5%;

- от 4 до 7 дней – 14,4%;
- от 8 до 10 дней – 61,7%;
- 11 дней и больше – 0,7%;

Не удалось отдохнуть в период празднования Нового Года 28 респондентам, что составило 4,8% от общего массива опрошенных.

Премьер-министр России Владимир Путин, победивший на президентских выборах, в феврале 2012 года выступил с инициативой о сокращении новогодних каникул за счет увеличения количества выходных в майские праздники. В связи с этим одной из целей проводимого исследования было выяснить отношение граждан к этому предложению. Респондентам было предложено указать количество дней, которое они считают необходимым и достаточным для отдыха в новогодние и майские праздники. Интересен тот факт, что распределение ответов оказалось практически одинаковым в разрезе по дням для зимних и весенних «каникул». Результаты представлены в таблице 1.

Таблица 1- Распределение ответов респондентов на вопросы о необходимом и достаточном количестве выходных дней в период новогодних и майских праздников

Количество дней	Процент ответивших (%)	
	Новогодние праздники	Майские праздники
1-3	28,9	28,9
4-7	43,4	43,1
8-10	25,3	26,5
11 и больше	2,4	1,6
Итого	100	100

Также были получены таблицы распределения по второму блоку анкеты, что позволило получить социально-демографический портрет респондентов.

Полученные распределения позволили сформулировать гипотезу о наличии взаимосвязи между рассматриваемыми вопросами. С целью подтверждения или опровержения выдвинутой гипотезы были построены таблицы сопряженности, которые позволяют получить представление совместного распределения двух переменных и предназначены для исследования связи между ними. Для этой цели в SPSS реализован тест χ^2 , при котором проверяется, есть ли значимое различие между наблюдаемыми и ожидаемыми частотами. Несмотря на кажущуюся зависимость при одномерном распределении между вопросами о желаемом количестве выходных дней в новогодние и майские праздники, данные таблиц сопряженности по этим переменным не позволили подтвердить её однозначное наличие. Также был проведен двумерный анализ по другим признакам. Во всех рассмотренных случаях результаты позволяют судить о некоторой тенденции или её отсутствии, не характеризуя это конкретным числовым критерием. Таким критерием, выявляющим статистическую взаимосвязь между переменными, как было сказано выше, является

коэффициент корреляции. В круг задач исследования входил поиск возможных закономерностей и связей, как между вопросами содержательного блока, так и возможное влияние социально-демографических характеристик на распределение ответов о количестве выходных дней. В связи с этим были рассчитаны соответствующие коэффициенты корреляции Пирсона и Спирмэна, в зависимости от типа шкал, к которым принадлежат переменные. Результаты представлены в сводной таблице 2.

Таблица 2 - Сводная таблица коэффициентов корреляции

	Вопрос № 6	Вопрос № 7	Вопрос № 8
Пол респондента	0,058	-0,011	0,076
Возраст респондента	0,121	-0,325	0,108
Образование респондента	-0,038	0,021	-0,072
Занятость респондента (определение группы работников)	0,174	0,014	0,039
Материальное положение респондента	0,048	-0,091	0,090
Вопрос № 6	1,000	0,150	0,230
Вопрос № 7	0,150	1,000	0,114
Вопрос № 8	0,230	0,114	1,000

Здесь:

- «вопрос 6» - Количество дней, которые респондент отдыхал в новогодние праздники 2012 г.;
- «вопрос 7» - Количество дней, которое респондент считает необходимым и достаточным для отдыха в последующие новогодние праздники;
- «вопрос 8» - Количество дней, которое респондент считает необходимым и достаточным для отдыха в период майских праздников.

Очевидно, что далеко не все переменные коррелируют между собой, тем не менее, согласно результатам расчетов можно выделить некоторые значения, которые в большей степени позволяют утверждать о наличии статистической связи в ряде случаев. Так, $r=-0,375$ свидетельствует о заметной связи возраста респондента и желаемом количестве выходных дней, которое должно отводиться на новогодние праздники. Построение таблицы сопряженности подтвердило этот вывод. Более сложной задачей является интерпретация полученного результата. Кажущаяся очевидность причинно следственной связи может быть обманчива. Разумеется, количество выходных дней, которое респондент выбрал, абсолютно не оказывает никакого влияния на его возраст, а скорее наоборот. Но вероятнее всего, что оба этих фактора являются следствием другой общей причины. В данном расчете получен отрицательный коэффициент корреляции. Это свидетельствует о том, что более молодые респонденты предпочитают

отдыхать в новогодние праздники большее количество дней, чем люди старшего возраста. Объяснением может служить тот факт, что динамичный темп жизни современной молодежи допускает необходимого количества выходных дней и новогодние праздники становятся своего рода «передышкой» перед новым календарным стартом. Большинство людей среднего возраста ведут более размеренный ритм жизни, вероятно, это является одной из причин того, что они готовы отдыхать меньшее количество дней по сравнению с молодым поколением. Подкрепляет полученные выводы анализ ответов старшей возрастной страты, в которую входит большее число неработающих, чем в другие и, следовательно, не нуждающихся в дополнительных выходных днях. Вместе с тем, стоит обратить внимание, что коэффициент корреляции в данном случае составляет всего -0,375, а это значит, что взаимозависимость рассматриваемых переменных составляет всего 37,5%. Полученные выводы косвенно подтверждает коэффициент корреляции переменных возраста респондентов и количества дней, которые они фактически отдыхали в новогодние праздники 2012 года. Здесь $r=0,121$, корреляция положительная, но достаточно слабая. Тем не менее, это свидетельствует от прямо пропорциональной зависимости переменных, а как следствие о том, что более молодые респонденты отдыхали меньшее количество дней, чем люди старшего поколения.

Согласно полученным результатам некоторые социально-демографические характеристики слабо коррелируют с вопросами изучаемой проблемы, некоторые совсем не коррелируют (пол респондента, образование, материальное положение) – коэффициент корреляции по модулю меньше 0,1. А коэффициенты, полученные по такому признаку как «занятость» формируют иную картину. Наблюдается некоторая зависимость между этим параметром и фактическим количеством выходных дней по вопросу №6. Так как группы работников в вопросе о занятости были расположены так, что первые позиции занимали категории рабочих, специалистов, предпринимателей и т.д., а последние – пенсионеры, безработные и домохозяйки, то вполне объяснима положительная корреляция между степенью занятости респондента и количеством дней, которое он отдыхал в новогодние праздники 2012 года.

Однако нет практически никакой статистической связи между группой работника и вопросами о желаемом количестве выходных дней на новогодние и майские праздники. Такое различие можно объяснить тем, что официальное количество выходных дней непосредственно зависит от того, работает респондент или нет, но интенсивность труда в каждой группе может сильно отличаться и поэтому респонденты, принадлежащие к одной и той же категории занятости могут нуждаться в разном количестве выходных, как в новогодние, так и в майские праздники.

Таким образом, полученные коэффициенты корреляции позволяют говорить о наличии статистической взаимосвязи между исследуемыми переменными, однако поиск причинно-следственной связи и интерпретация этих значений требует не столько применение аппарата математической статистики, сколько законов и правил социологии. Таким образом, корреляционный анализ не является заменой специальных знаний, но может сделать теоретическое мышление социолога более эффективным, так как позволяет отбросить несуществующие связи и очертить круг поисков.

Литература

1. С.Г. Смердова, Н.Н. Умарова, Л.В. Петухова, Г.И. Сибгатуллина, Вестник Казанского технологического университета, 5, 259-266 (2010).
2. И.А. Абдуллин, О.И. Белобородова, Н.И. Лаптев, Е.Л. Москвичева, А.Д. Горяйнов, Вестник Казанского технологического университета, 12, 477-482(2010)
3. Г.П. Артемов, Политическая социология. Учебное пособие. Логос, Москва, 2002, С.51.
4. А.Н. Ершов А.Н., А.П. Кулапин, Вестник Казанского технологического университета, 2, 118-121 (2012).
5. В.И. Паниотто, В.С. Максименко, Количественные методы в социологических исследованиях. Киев, 2003, С.44.
6. И.И. Елисеева, М.М. Юзбашев. Общая теория статистики: Учебник. 4-е издание, переработанное и дополненное. Финансы и Статистика, Москва, 2002, С.228-230.
7. Общая теория статистики: Учебник / Под ред. Р.А. Шмойловой. 3-е издание, переработанное. Финансы и Статистика, Москва, 2002, С.275.
8. В.В. Пациорковский, В.В. Пациорковская. SPSS для социологов. Учебное пособие. ИСЭПН РАН, Москва, 2005, С.9