

УДК 004

РАБОТА С ФАКТОРАМИ ПРИ СТАТИСТИЧЕСКОМ АНАЛИЗЕ СОЦИОЛОГИЧЕСКОЙ И ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИЙ С ПОМОЩЬЮ ПРОГРАММЫ SPSS

Луцик А.П. Щелочкова А.С.,

к.т.н. Кулаков С.В.

Европейский университет Донецкий филиал,
кафедра математики и компьютерных дисциплин

E-mail: Kosmoskonti@rambler.ru

Аннотация

Луцик А.П. Щелочкова А.С., к.т.н. Кулаков С.В. Работа с факторами при статистическом анализе социологической и экономической информации с помощью программы SPSS.

В данной работе описана технология работы с вычисленными значениями факторов с помощью программы SPSS. Она основана на преобразовании факторной переменной к ранговой с приданием смыслового значения уровням ранговой переменной. Дальнейшие корреляционный и регрессионный анализы выполняются с новыми ранговыми переменными. В качестве примера рассмотрена работа с факторами для базы данных группы банковских сотрудников сети российских банков.

Общая постановка проблемы

В работе описан процесс выявления, вращения и интерпретации факторов для группы банковских сотрудников сети российских банков, выполненный с помощью программы SPSS. В результате статистического анализа выявлены три фактора: квалификация, опыт работы, опыт работы в данном банке. В результате расчета сгенерированы три новые переменные, названные fac1_1, fac2_1 и fac3_1, которые содержат вычисленные значения факторов. Эти переменные имеют нормализованные значения без определенного смысла.

Исследования

SPSS Statistics (аббревиатура англ. «Statistical Package for the Social Sciences» — «статистический пакет для социальных наук») — компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках. Этот пакет разрабатывается компанией «SPSS: An IBM Company». Для исследований выбран пакет SPSS 13.01 релиз от 12 декабря 2004, так как он наиболее распространён в странах СНГ.

Факторный анализ - метод многомерного статистического анализа, позволяющий на основе экспериментального наблюдения признаков объекта выделить группу переменных, определяющих корреляционную взаимосвязь между признаками. Суть факторного анализа - на основе исследования корреляционных взаимосвязей признаков находить причины, определяющие эти взаимосвязи. В общем случае моделью описываемой взаимосвязи является набор линейных уравнений. Коэффициентами этих уравнений являются так называемые нагрузки, которые показывают "вес" каждого из факторов для данного признака (статистической переменной). В матричном виде эта система уравнений может быть записана как $X = S * F + E$, где X - матрица признаков (или переменных), S - матрица нагрузок, F - матрица новых переменных-факторов, E - матрица остатков. Это уравнение, по сути, описывает переход от первичных переменных (признаков) к новым переменным (факторам).

Можно выделить следующие *основные этапы факторного анализа* :

Вычисление корреляционной матрицы для всех переменных, участвующих в анализе;
Извлечение факторов;
Вращение факторов для создания упрощенной структуры;
Интерпретация факторов.

Далее можно выполнять численную оценку выделенных факторов, отбор факторов для анализа того или иного показателя и находить зависимость между отдельными факторами. При этом обычно исходят из принципа: чем больший комплекс признаков исследуется, тем точнее будут результаты анализа. Вместе с тем необходимо иметь в виду, что если этот комплекс признаков рассматривается как механическая сумма, без учета их взаимодействия, без выделения главных, определяющих, то выводы могут быть ошибочными. В анализе хозяйственной деятельности (АХД) и при социологическом анализе взаимосвязанное исследование влияния факторов на величину результативных показателей достигается с помощью их систематизации.

Важным методологическим вопросом в факторном анализе является *определение формы зависимости* между факторами и результативными показателями: функциональная она или стохастическая, прямая или обратная, прямолинейная или криволинейная. Здесь используется теоретический и практический опыт, а также способы сравнения параллельных и динамических рядов, аналитических группировок исходной информации, графический и др.

В соответствии с работой [2,3] для базы данных группы банковских сотрудников были сгенерированы три новые переменные, названные *fac1_1*, *fac2_1* и *fac3_1*, которые содержат вычисленные значения факторов. Если просмотреть текущий файл после поведения факторного анализа, то можно увидеть имеющие нормализованные значения факторов. Факторные значения безразмерны и, как правило, лежат в пределах от -3 до +3.

Если рассмотрим факторную переменную *fac1_1* (квалификация), то можно отметить, что она включает в себя несколько элементарных переменных. Далее желательно выбрать другое имя и выражение для метки этой факторной переменной. Большое положительное значение фактора означает одобрение элементарных переменных, то есть положений, входящих в этот фактор.

В файле исходных имеются другие переменные. Эти переменные можно использовать для того, чтобы устанавливать связи для факторных значений. Самым распространённым методом для этого является разбиение факторных значений на 4-6 групп процентилей. Покажем это на примере первого факторного значения (переменная *fac1_1*):

- Выберем в главном меню SPSS **Transform** (Трансформировать) - **Rank**
- **Cases** (Создать иерархию наблюдений). Откроется диалоговое окно **Rank Cases** (Создать иерархию наблюдений);
- Переменную *fac1_1* перенесите в список тестируемых переменных;
- Щёлкнем на выключателе **Rank Types...** (Типы иерархии), деактивируем
- установленную по умолчанию опцию **Rank** (Ранг) и активируем опцию **Fractional**
- **rank as %** (Дробный ранг как процентиля). Оставим установленное по умолчанию количество групп равное 5;
- Подтвердите свой выбор нажатием на **Continue** (Далее) и затем на ОК.

Будет создана переменная *nfac1_1*, которая содержит значения 1 до 5 с примерно равномерной частотой. Наряду с этим создается переменная *rfac1_1*, дающая значение в %.

Перейдем в редактор данных и изменим имя переменной *nfac1_1* на более удобное имя – **Квалификация_сотрудника**, а в поле метки значениям присвоим например следующие метки: 1 = отсутствует, 2 = слабое, 3 = среднее, 4 = сильное, 5 = очень сильное. Для факторных переменных *fac2_1* и *fac3_1* следует сделать тоже самое.

После такого перехода к ранговым переменным можно исследовать связи между факторными переменными и оставшимися переменными, строить таблицы сопряженности и уравнения регрессии.

В файле исходных данных находятся ещё несколько дополнительных переменных, а именно:

Таблица 1-показывает дополнительные переменные.

jobcat	Категория сотрудника (0=неизвестна; 1=канцелярский работник; 2=практикант в офисе; 3=сотрудник службы безопасности;
Minority	Национальное меньшинство.
salgrp	Группа по заработной плате.
salchange	Изменения зарплаты.

Из этих переменных первые две имеют номинальную шкалу, третья – ранговую, четвертая – количественную.

Вычислим ранговый коэффициент корреляции Спирмена, определяющий степень связи между переменной **Изменение зарплаты** (salchange) и фактором **Квалификация**. Для этого введем команду **Analyze>Correlate>Bivariate**.

Результаты приведены в таблице 2. Значение коэффициента корреляции 0.657 показывает среднюю тесноту связи.

Для построения регрессионной кривой используем команду **Analyze>Regression>Curve Estimation**.

В выводимой программой SPSS таблице 3 приведены коэффициенты регрессионного уравнения, которое в данном случае имеет следующий вид:

$$\text{Salchange}^T = 90,03 - 40,99 * \text{nfac1} + 11,04 * \text{nfac1}^2$$

Таблица 2- ранговый коэффициент корреляции Спирмена.

Correlations

		Изменение зарплаты	Квалификация сотрудника
Изменение зарплаты	Pearson Correlation	1	,657**
	Sig. (2-tailed)		,000
	N	464	440
Квалификация сотрудника	Pearson Correlation	,657**	1
	Sig. (2-tailed)	,000	
	N	440	440

** . Correlation is significant at the 0.01 level (2-tailed).

Таблица 3-показывает коэффициенты регрессивного уравнения.

Model Summary and Parameter Estimates

Dependent Variable: Изменение зарплаты

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Quadratic	,547	264,247	2	437	,000	90,025	-40,994	11,039

The independent variable is Квалификация сотрудника.

Диаграмма рассеивания (рис.1) показывает, что связь будет явно нелинейной и применение квадратичной зависимости оправдано. На рис. 1 показана квадратная парабола на диаграмме рассеивания.

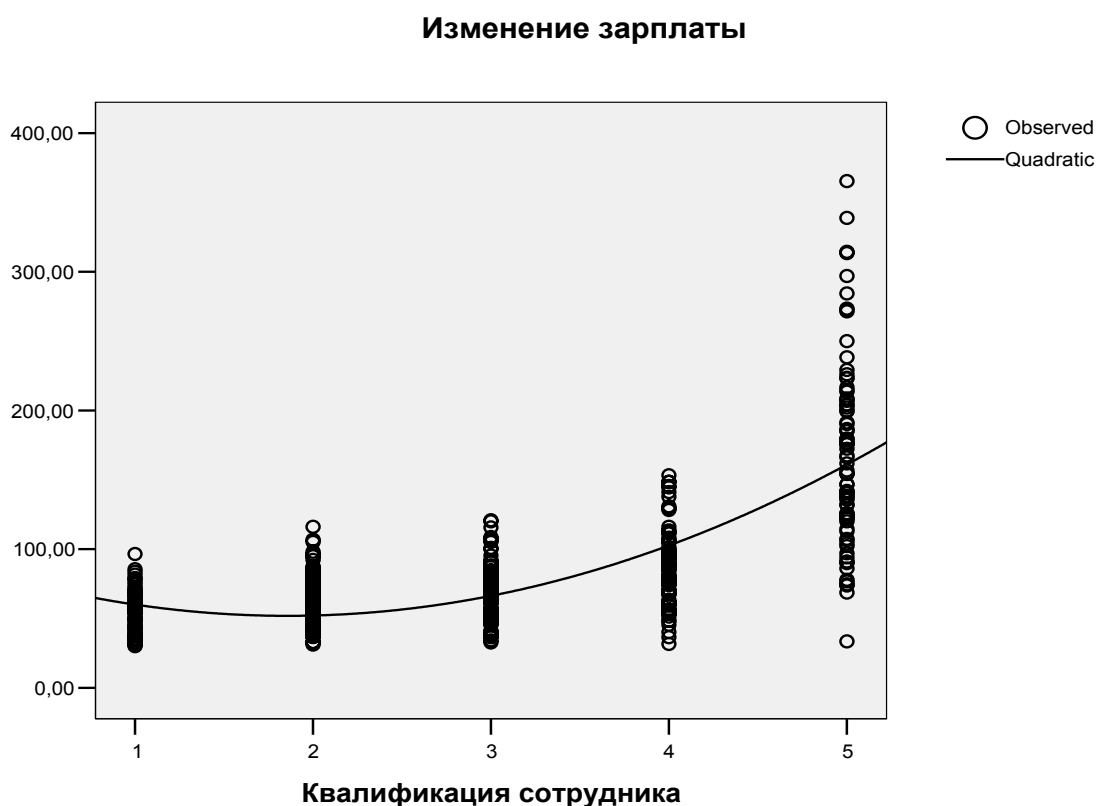


Рисунок 1 – Диаграмма рассеивания с регрессионной кривой.

Выводы

Программа Statistical Package for the Social Sciences (SPSS) является специализированным продуктом для социологических и экономических исследований. Кроме вычисления факторных значений можно выполнять работу с факторами путем преобразования их к ранговым переменным.

В работе проанализированы связи фактора Квалификация сотрудника банка с переменными, не входящими в факторный анализ.

Список литературы

1. Гмурман В.К. Теория вероятностей и математическая статистика. Учебное пособие для вузов. – 8-е изд. стер. - М.: Высш. шк., 2002. -479 с.
2. Кулаков С.В., Балко Е.В., Лифарь Д.В. Факторный анализ данных для группы банковских сотрудников, выполненный с помощью программы SPSS// Вісник Макіївського економіко-гуманітарного інституту, №4(17). –Макіївка, 2010. – с. 164-170.
3. Кулаков С.В. Использование программы SPSS для проведения факторного анализа экономической и социологической информации. – Проблемы современности: наука, практика, методология (коллективная монография), Донецк, 2010. – 584 с.
4. Наследов А.Д. Компьютерный анализ данных в психологии и социальных науках. – СПб: Питер, 2005. – 418 с.