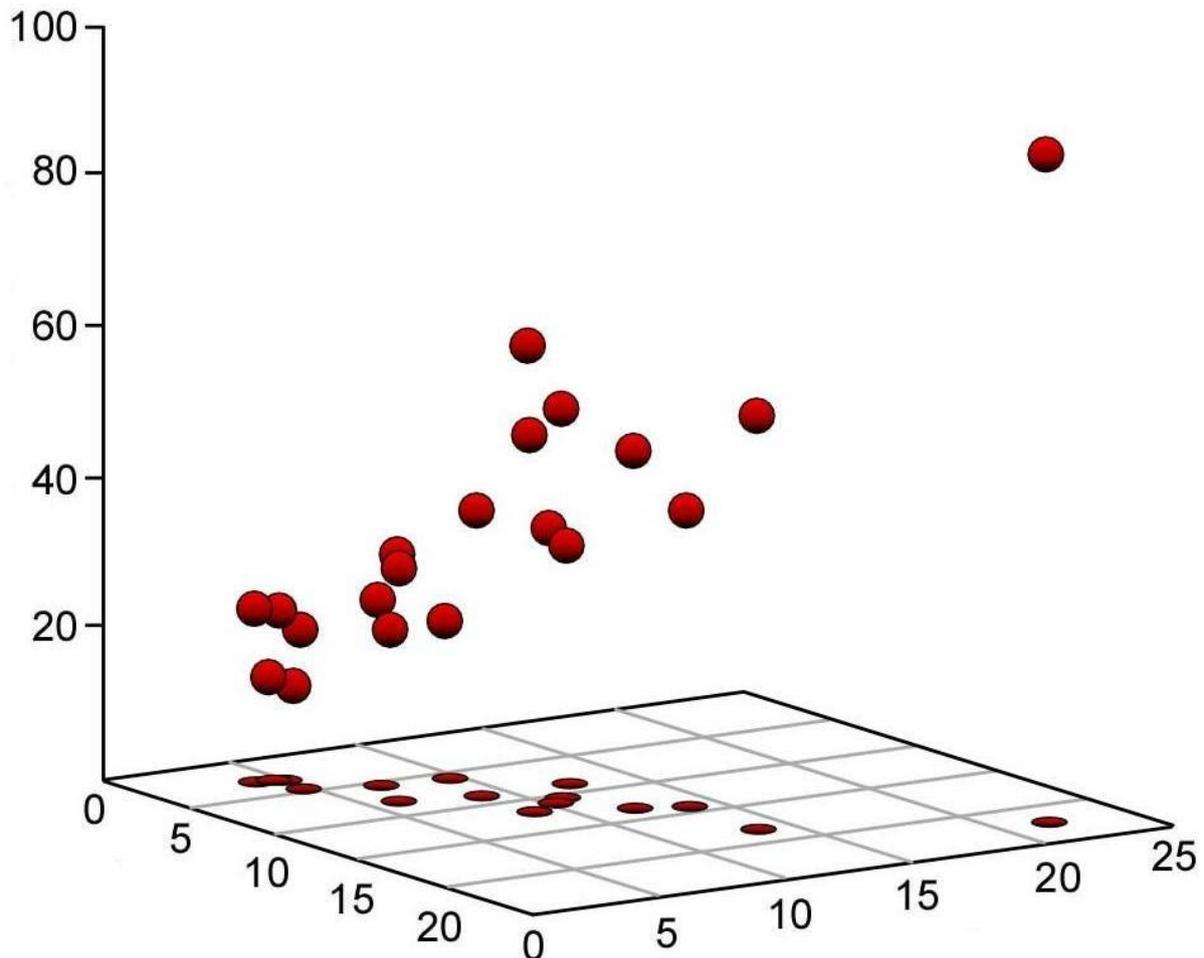


Р. И. Остапенко

Многомерный анализ данных для ПСИХОЛОГОВ



Остапенко Р. И. Многомерный анализ данных для психологов: учебно-методическое пособие / Р. И. Остапенко. – Воронеж.: ВГПУ, 2012. – 72 с.: ил.

В учебном пособии дается обзор основных понятий и методов применяемых при обработке данных. Процедуры применения методов многомерного анализа рассматриваются на конкретных примерах. Рассмотрены способы обработки с помощью MS Excel и русифицированного пакета SPSS 18.0. Пособие рассчитано на студентов, аспирантов, а также широкий круг специалистов, занимающихся научно-прикладными исследованиями в области гуманитарных и смежных наук.

СОДЕРЖАНИЕ

| | |
|---|----|
| ВВЕДЕНИЕ | 3 |
| ГЛАВА 1. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ | 4 |
| 1.1. Измерение. Измерительные шкалы | 4 |
| 1.2. Выборка | 6 |
| 1.3. Числовые характеристики распределений..... | 7 |
| 1.4. Степень свободы..... | 11 |
| 1.5. Нормальное распределение | 12 |
| 1.6. Статистические гипотезы. Уровень статистической значимости..... | 17 |
| ГЛАВА 2. МЕТОДЫ МНОГОМЕРНОГО АНАЛИЗА ДАННЫХ..... | 20 |
| 2.1. Корреляционный анализ | 20 |
| Коэффициент корреляции r -Спирмана | 24 |
| Коэффициент корреляции τ -Кендалла | 29 |
| Коэффициент корреляции r -Пирсона | 35 |
| Анализ корреляционных матриц | 37 |
| 2.2. Регрессионный анализ | 42 |
| Парный регрессионный анализ | 42 |
| Множественная регрессия | 46 |
| 2.3. Эксплораторный факторный анализ | 51 |
| ЗАКЛЮЧЕНИЕ..... | 66 |
| ЛИТЕРАТУРА | 67 |
| ПРИЛОЖЕНИЯ | 68 |

ВВЕДЕНИЕ

В педагогике, психологии и других смежных науках о человеке, подавляющее большинство изучаемых явлений не поддается прямому измерению. Умственные способности, компетентность, академическая успеваемость, личностные качества, толерантность, мобильность и другие абстракции есть понятия, выделяемые исследователями и позволяющие описывать отношения между наблюдаемыми переменными. Эти феномены могут быть описаны не только семантически, их существование может быть подтверждено эмпирически, а в частном случае и количественно.

В психолого-педагогических исследованиях системное описание явлений указывает на необходимость освоения и применения соответствующего математического аппарата. В исследованиях также используются математические методы многомерного анализа.

В первой главе учебника уделено внимание основным понятиям математической статистики: измерение, выборка, нормальное распределение и т.д. Во второй главе рассмотрены методы многомерной обработки количественных данных: корреляционный, регрессионный и эксплораторный факторный анализ. В конце пособия имеется заключение, список литературы и приложения.

Пособие рассчитано на студентов, магистрантов, аспирантов, преподавателей вузов, а также широкий круг специалистов, занимающихся научно-прикладными исследованиями в области гуманитарных и смежных наук.

Автор учебного пособия будет искренне признателен всем, кто сможет прислать свои критические замечания и предложения по поводу данной книги (ramiro@list.ru).

ГЛАВА 1. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1.1. Измерение. Измерительные шкалы

Главное отличие отраслей психологического знания использующих математические методы – их предмет может быть не только описан, но и измерен. Возможность измерения открывает доступ для применения количественного анализа.

Измерение – это приписывание чисел объектам или их свойствам по определенным правилам. Правила устанавливают соответствия между некоторыми свойствами рассматриваемых объектов и рядом чисел.

Любой вид измерения предполагает наличие единиц измерения. Психологические переменные не имеют собственных измерительных единиц. Значение психологического признака определяется при помощи специальных измерительных шкал. Выделяют четыре типа измерительных шкал: номинативной, порядковой, интервальной и отношений (см. Табл. 1).

Примером измерения в номинативной шкале является деление выборки по половому признаку на мужчин и женщин. Пример измерения в порядковой шкале – упорядоченная градация мнений по отношению к какому-либо утверждению: «полностью согласен», «согласен», «не знаю», «не согласен», «категорически не согласен». Примером измерения в интервальной шкале является расчет коэффициента интеллекта испытуемых с помощью теста Векслера. Пример измерения в шкале отношений – определение роста испытуемых в сантиметрах.

Математические преобразования с переменными находятся в тесной связи с измерительной шкалой, в которой измерено то или иное свойство. Так, например номинативная и порядковая шкалы являются неметрическими, а интервальная и отношений – метрические. Для неметрических шкал такие преобразования как сложение, умножение и другие операции не допустимы, либо не имеют смысла. Для метрических шкал допустимы операции сложения, умножения и т.д.

В психологии и педагогике большинство измерений относится к номинальному и порядковому уровням.

Таблица 1

Классификация измерительных шкал

| Шкала | Характеристика | Примеры |
|--------------------|---|---|
| Номинативная шкала | Состоит в присваивании какому-либо свойству или признаку определенного обозначения или символа. При измерении в этой шкале осуществляется классификация или распределение на непересекающиеся классы. Символы не несут никакой информации, операции с ними не имеют смысла. | Типы темперамента; типы акцентуаций характера; варианты ответов испытуемых. Дихотомические (двоичные): «Да» и «нет»; «За» и «Против»; «Интроверт» и «Экстраверт»; «Полная семья» и «Неполная семья». |
| Порядковая шкала | Классифицирует совокупность измеренных признаков по принципу «больше-меньше», «выше-ниже», «сильнее-слабее». | Школьные оценки от 1 до 5; закодированные уровни от низкого до высокого; ранжируемые иерархии предпочтений или ценностей. |
| Интервальная шкала | Каждое из возможных значений измеренных величин отстоит от ближайшего на равном расстоянии. Нуль условен. При работе с этой шкалой измеряемому свойству или предмету присваивается число равное количеству единиц измерения, эквивалентное количеству измеряемого свойства. | Семантический дифференциал Ч.Осгуда; IQ Векслера; 16-ти факторный опросник Кеттела и другие тестовые шкалы, которые специально вводятся при обосновании их равноинтервальности.* |
| Шкала отношений | Обладает всеми свойствами интервальной шкалы и имеет твердо фиксированный нуль, который означает полное отсутствие свойства. Используется в химии, физике, психофизике, психофизиологии. | Рост; вес; число реакций; показатель силы; выносливости. |

В реальной ситуации с трудом можно найти шкалы, удовлетворяющие условиям интервальной шкалы и шкалы отношений. Психологические и педагогические измерения, а особенно клинические не поддаются какой-либо простой классификации, вроде порядковой или интервальной.

* Интервальные измерения часто есть порядковые, так как функциональные возможности человека меняются в зависимости от разных условий (усталость, мотивация)

1.2. Выборка

Генеральная совокупность – это любая совокупность объектов, относительно которой исследователь делает вывод. Теоретически генеральная совокупность неограниченна.

Выборка – любая подгруппа элементов (испытуемых, респондентов) выделенная из генеральной совокупности для проведения эксперимента.

Соотношение выборки и генеральной совокупности показано на рисунке 1:

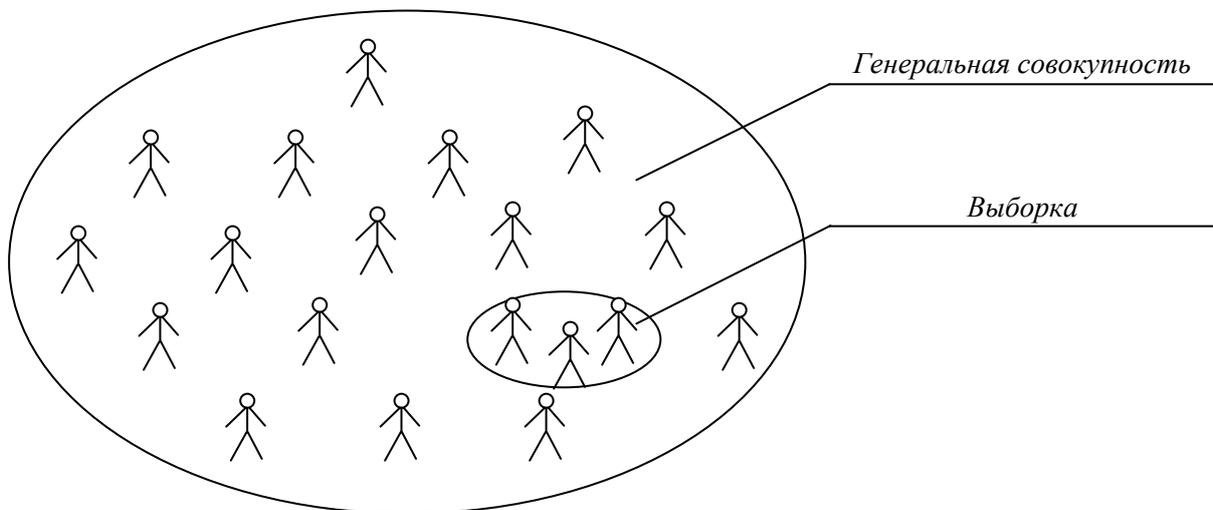


Рис. 1. Генеральная совокупность и выборка

Выборочное исследование – это исследование, при котором производится выбор ограниченного числа элементов из изучаемой генеральной совокупности.

Если процедура эксперимента и полученные результаты измерения некоторого свойства, проведенные на одной выборке, оказывают влияние на другую, то такие выборки называются *зависимыми (связными)*. Если процедура эксперимента и полученные результаты измерения некоторого свойства, проведенные на одной выборке, не оказывают влияние на другую, то такие выборки называются *независимыми (несвязными)*.

Требования к выборке:

1. Однородность. Выбор осуществляется на основаниях: возраст, уровень интеллекта, национальность, заболевания.

2. Репрезентативность. Качество выборки позволяющее распространять полученные на ней выводы на всю генеральную совокупность. Состав экспериментальной выборки это модель генеральной совокупности.

Следует отметить, что любая выборка может быть репрезентативной лишь в каких-то определенных, но не всех отношениях. Например, если выборка сделана по социально-образовательному признаку, это не значит, что она будет репрезентативна и для возрастной структуры населения или для разных типов семьи и т.д.

Рекомендуемый объем выборки: не менее 30-35 человек в изучаемой группе. При использовании методов многомерного анализа объем выборки должен быть существенно больше: от 100-150 испытуемых.

1.3. Числовые характеристики распределений

Исследователь, получая большое количество экспериментальных данных, сталкивается с необходимостью их сведения к «среднему». В психолого-педагогических исследованиях наиболее часто используется следующие характеристики: мода, медиана и среднее арифметическое.

Мода – это числовое значение, встречающееся в выборке наиболее часто. Обозначается через M_o .

Медиана – это значение, которое делит упорядоченный ряд пополам. Обозначается через M_e .

Среднее арифметическое – это сумма всех элементов, деленная на их количество. Обозначается через \bar{X} и определяется как:

$$\bar{X} = \frac{(X_1 + X_2 + \dots + X_n)}{n},$$

где X_1, X_2, \dots, X_n – элементы выборки, n – число элементов выборки.

В отличие от моды и медианы на величину среднего влияют значения всех результатов. Преимущество среднего арифметического заключается в его способности аккумулировать, уравнивать все индивидуальные отклонения. Например, рост испытуемых был бы распределен

следующим образом: 180 см, 175 см, 170 см, 165 см, 163 см (см. Рис.2).

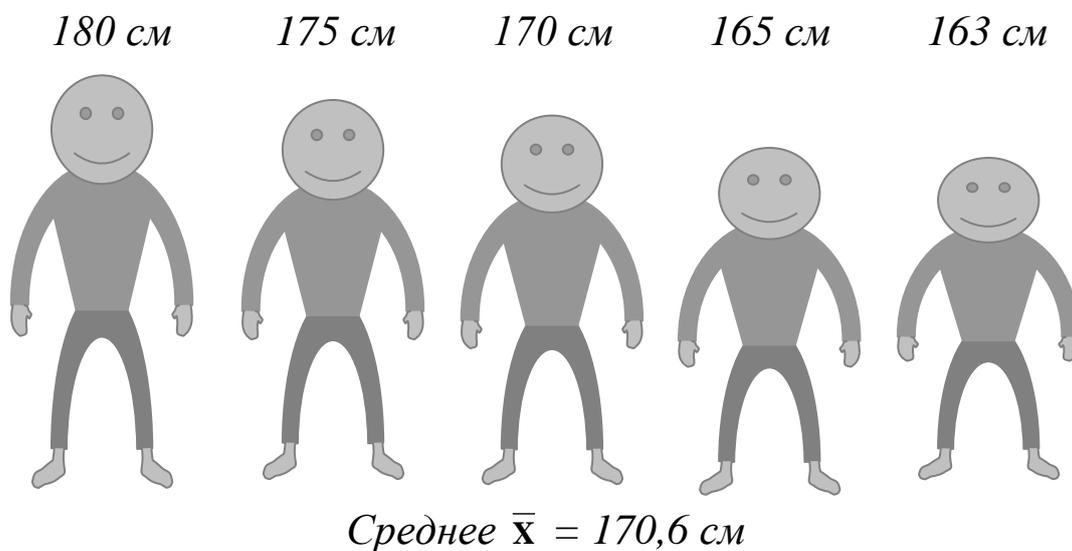


Рис. 2. Выборка испытуемых, упорядоченная по росту

Среднее арифметическое не всегда является объективной величиной заменяющей числовой ряд. Например, рост испытуемых был бы распределен следующим образом: 163 см, 163 см, 165 см, 180 см, 181 см (см. Рис.3).

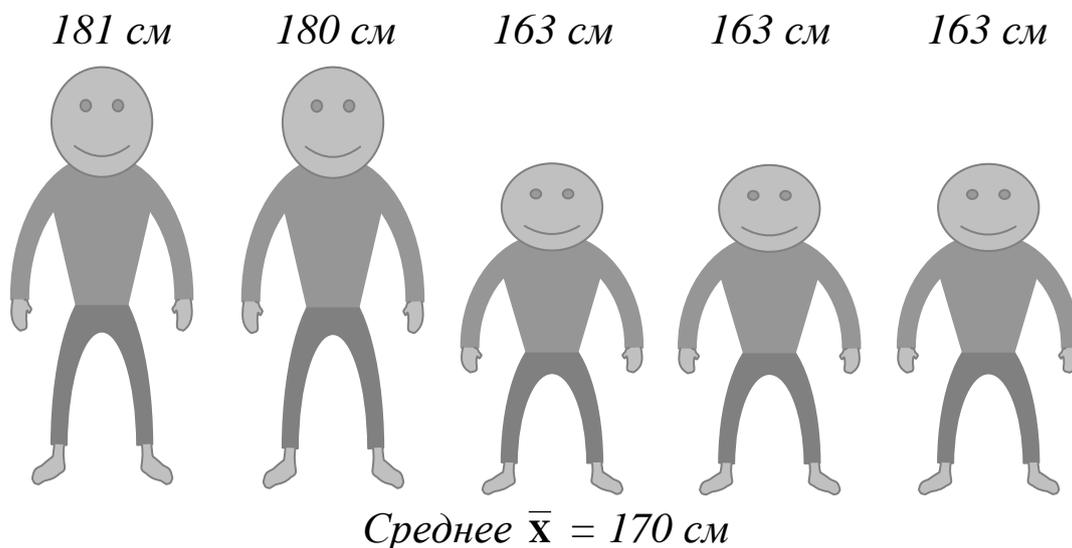


Рис. 3. Выборка испытуемых, упорядоченная по росту

В данном примере наиболее адекватным средним являются мода и медиана: $M_o = M_e = 163$ см. Использование средних для «замены» ряда данных часто оказывается недостаточным. Более важную роль играет постоянство значений, их «кучность». Например, на рисунках 4 и 5 представлены две группы испытуемых с одинаковым средним арифметическим. В первой выборке (см. Рис.4) изменчивость показателей роста выше, чем во второй выборке (см. Рис.5). В первой выборке рост колеблется от 163 до 180 см., а во второй – от 168 до 175 см.

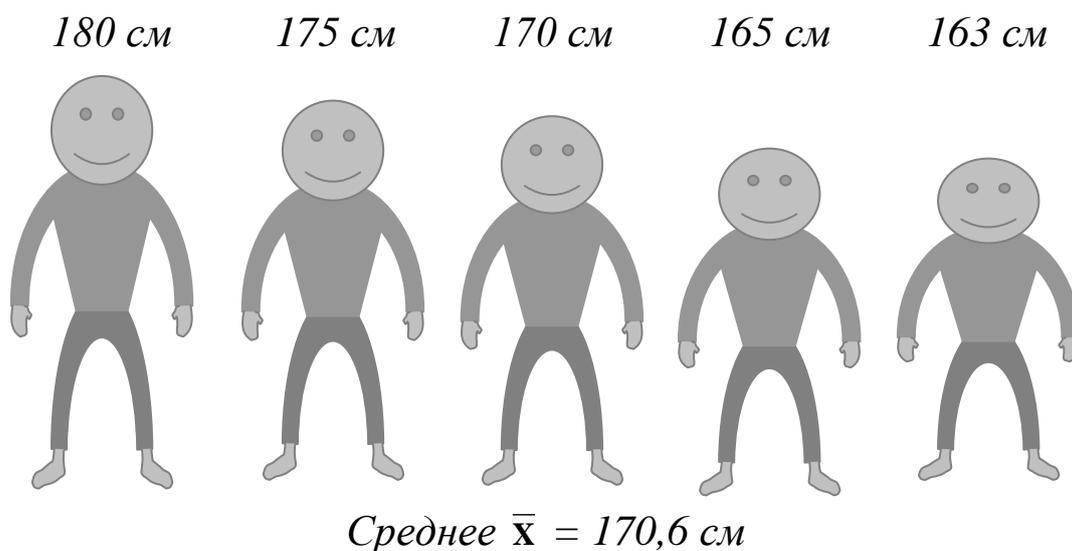


Рис. 4. Выборка испытуемых №1

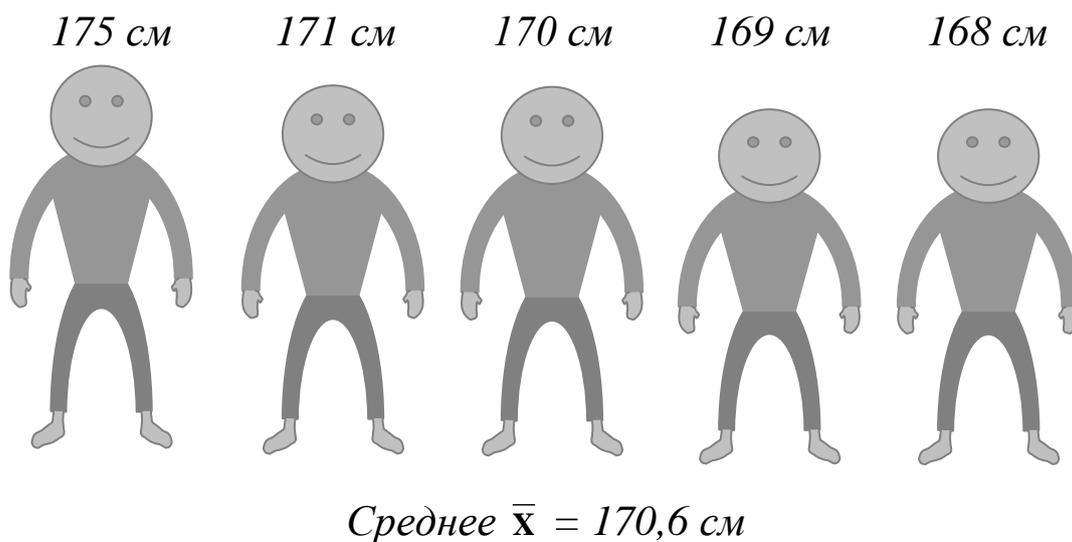


Рис. 5. Выборка испытуемых №2

Показатели роста испытуемых показаны на измерительной линейке:

Выборка №1

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 180 | 179 | 178 | 177 | 176 | 175 | 174 | 173 | 172 | 171 | 170 | 169 | 168 | 167 | 166 | 165 | 164 | 163 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Среднее $\bar{X} = 170,6$ см

Выборка №2

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 180 | 179 | 178 | 177 | 176 | 175 | 174 | 173 | 172 | 171 | 170 | 169 | 168 | 167 | 166 | 165 | 164 | 163 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Среднее $\bar{X} = 170,6$ см

Данный факт, указывает на необходимость использования при описании данных дополнительной числовой характеристики, показывающей степени рассеивания или разброса значений вокруг среднего.

Дисперсия – это мера разброса данных относительно среднего значения. Дисперсия обозначается через D и вычисляется как:

$$D = \frac{(x_1 - \bar{X})^2}{n-1} + \frac{(x_2 - \bar{X})^2}{n-1} + \dots + \frac{(x_n - \bar{X})^2}{n-1} = \frac{\sum (x_i - \bar{X})^2}{n-1}, \text{ где}$$

x_i – элемент ряда X, \bar{X} – среднее арифметическое элементов ряда X, n – число элементов в выборке.

Используя данные предыдущего примера, рассчитаем дисперсии для двух рядов данных имеющих одинаковые средние.

Выборка испытуемых №1: 180 см, 175 см, 170 см, 165 см, 163 см.

$$D_1 = \frac{(180-170,6)^2}{6-1} + \frac{(175-170,6)^2}{6-1} + \frac{(170-170,6)^2}{6-1} + \frac{(165-170,6)^2}{6-1} + \frac{(163-170,6)^2}{6-1} = 49,3 (\tilde{n}^2)$$

Выборка испытуемых №2: 175 см, 171 см, 170 см, 169 см, 168 см.

$$D_1 = \frac{(175-170,6)^2}{6-1} + \frac{(171-170,6)^2}{6-1} + \frac{(170-170,6)^2}{6-1} + \frac{(169-170,6)^2}{6-1} + \frac{(168-170,6)^2}{6-1} = 7,3 (\tilde{n}^2)$$

Таким образом, в выборке №1, где изменчивость показателей роста выше, дисперсия $D = 49,3 \text{ см}^2$, а в выборке №2, где изменчивость показателей роста ниже, дисперсия $D = 7,3 \text{ см}^2$.

Стандартное отклонение есть квадратный корень из дисперсии:

$$s = \sqrt{D} = \sqrt{\frac{\sum (x_i - \bar{X}_x)^2}{n-1}}$$

Стандартное отклонение является более удобным показателем в отличие от дисперсии. Извлекая из корня, мы избавляемся от квадратных единиц. Используя данные предыдущего примера, рассчитаем стандартные отклонения для двух выборок

$$\text{Выборка испытуемых №1: } s_1 = \sqrt{83,9} = 4,77 \tilde{n}i ;$$

$$\text{Выборка испытуемых №2: } s_2 = \sqrt{14,9} = 1,22 \tilde{n}i .$$

Для нормального распределения (см. пункт 1.5.) мы можем приблизительно узнать, какой процент данных лежит внутри одного, двух, трех и более стандартных отклонений от среднего.

1.4. Степень свободы

Число степеней свободы – это число свободно варьирующих единиц в составе выборки. Если выборка состоит из n элементов и характеризуется средним \bar{X} , то любой элемент этого множества может быть получен как разность между $n - \bar{X}$, и суммой всех остальных элементов, кроме самого этого элемента.

Например, рассмотрим ряд чисел 4, 7, 15, 19, 25. Среднее арифметическое $\bar{X} = 14$. В ряду 5 чисел, т.е. $n = 5$. Выразим первый элемент ряда через другие элементы и среднее арифметическое:

$$4 = (5 - 14) + 25 - 19 + 7$$

Далее выразим второй элемент ряда через другие элементы и среднее арифметическое:

$$7 = (5 - 14) + 25 - 19 + 25 - 15 \text{ и т.д.}$$

Таким образом, один элемент выборки не имеет свободы вариации и всегда может быть выражен через другие элементы и среднее. В рассмотренном выше случае число степеней свободы k будет определяться как $k = n - 1$, где n – общее число элементов выборки.

При наличии нескольких ограничений свободы вариации, число степеней свободы, обозначаемое как ν («ню»), будет равно

$v = n - k$, где k – число ограничений свободы вариации.

В общем случае для таблицы экспериментальных данных число степеней свободы определяется по формуле:

$$v = (c - 1)(n - 1),$$

где c – число столбцов, а n – число строк.

Нахождение числа степеней свободы для каждого метода имеет свои специфические особенности.

1.5. Нормальное распределение

Понятие нормы в педагогике и психологии многозначно. Норма понимается как норматив, т.е. как эталон, на который необходимо равняться, оценивая по нему свое индивидуальное поведение (нормы питания, спортивные нормы и т.д.). Такие нормы (нормативы) являются условными и имеют значение только в определенной системе отсчета. Норма также понимается как функциональный оптимум, подразумевающий протекание всех процессов в системе с наиболее возможной слаженностью, эффективностью и экономичностью. Функциональная норма всегда индивидуальна, в ней лежит представление о неповторимости пути развития каждого человека, и ее нарушение определяется функциональными последствиями. Третьей системой отсчета является норма, понимаемая как статистически среднее, наиболее часто встречающееся, массовое в явлениях. «Нормальное» в статистическом смысле включает не только среднестатистическую величину, но и серию отклонений от нее в известном диапазоне. Ориентация на статистическую норму развития тех или иных психических качеств особенно важно на этапе первичной диагностики психического состояния испытуемых при определении характера выраженности какого-либо свойства.

Нормальный закон распределений лежит в основе измерений, разработки тестовых шкал и методов проверки гипотез. Нормальное распределение играет большую роль в математической статистике, так как многие статистические методы предполагают, что анализируемые данные распределены нормально.

Нормальное распределение часто встречается в природе. Нормальное распределение характеризует такие случайные величины, на которые воздействует большое количество разнообразных факторов. Например, если у испытуемых выбранных случайным образом измерять их рост, вес, интеллект, какие-либо свойства личности, а затем построить график частоты встречаемости показателей любой из этих величин, то мы получим распределение, у которого крайние значения встречаются редко, а от крайних значений к середине частота повышается. Таким образом, подавляющее большинство испытуемых будет иметь значения близкие к среднему.

График нормального распределения имеет вид симметричной, колоколообразной кривой (см. Рис. 6).

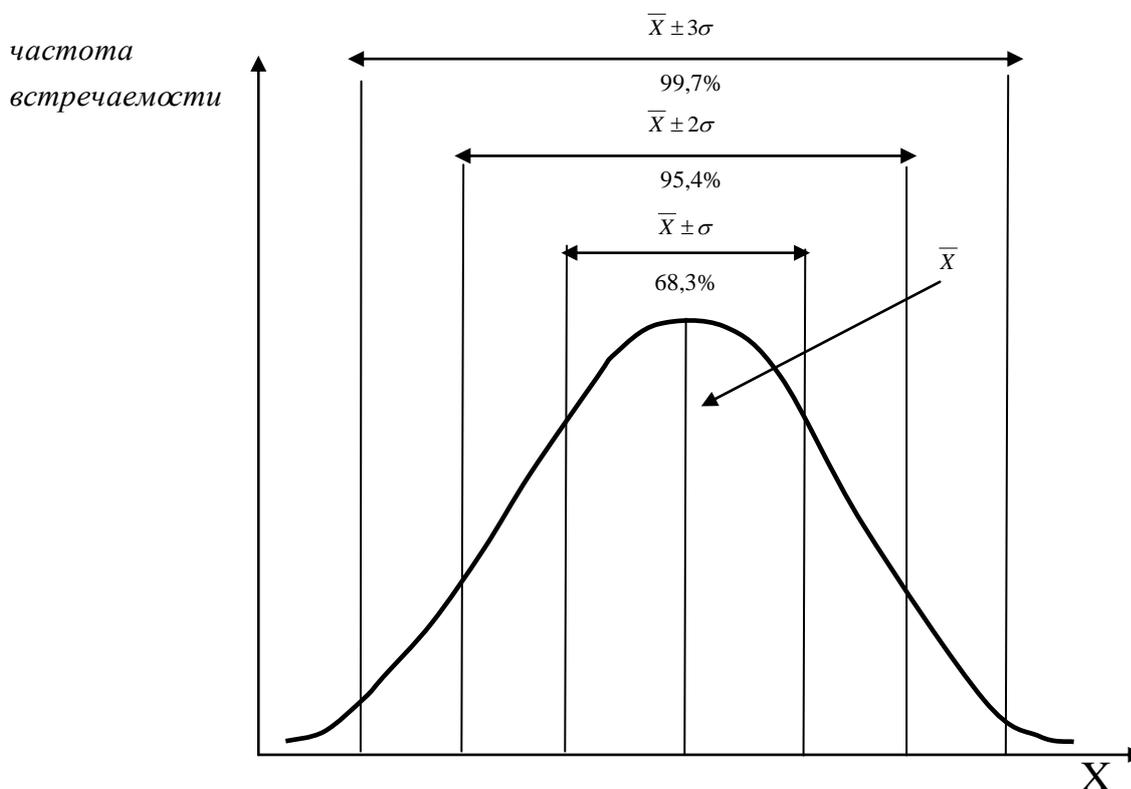


Рис. 6. График нормального распределения

Форма и положение графика нормального распределения определяется двумя параметрами: средним арифметическим (\bar{X}) и стандартным отклонением (σ). Среднее арифметическое задает положение кривой на числовой оси, а стандартное отклонение задает ширину этой кривой.

Любое нормальное распределение может быть сведено к одной кривой, если применить z-преобразование ко всем измерениям по формуле:

$$z_i = \frac{x_i - \bar{X}_x}{s_x},$$

x_i – элемент ряда X, \bar{X}_x – среднее арифметическое элементов ряда X, s_x – стандартное отклонение элементов ряда X.

В результате график нормального распределения будет иметь среднее арифметическое равное нулю, а стандартное отклонение равное единице. Это единичное нормальное распределение (см. Рис. 7).

Наиболее важным свойством единичного нормального распределения является, тот факт, что площадь между кривой и осью z равна 1. Площадь под кривой интерпретируется как вероятность, или относительная частота. Зная свойства единичного нормального распределения, мы можем узнать, какая доля генеральной совокупности имеет выраженность свойства от -1σ до $+1\sigma$ а также, какова вероятность того, что случайно выбранный представитель генеральной совокупности будет иметь выраженность свойства, на 2σ превышающую среднее значение?

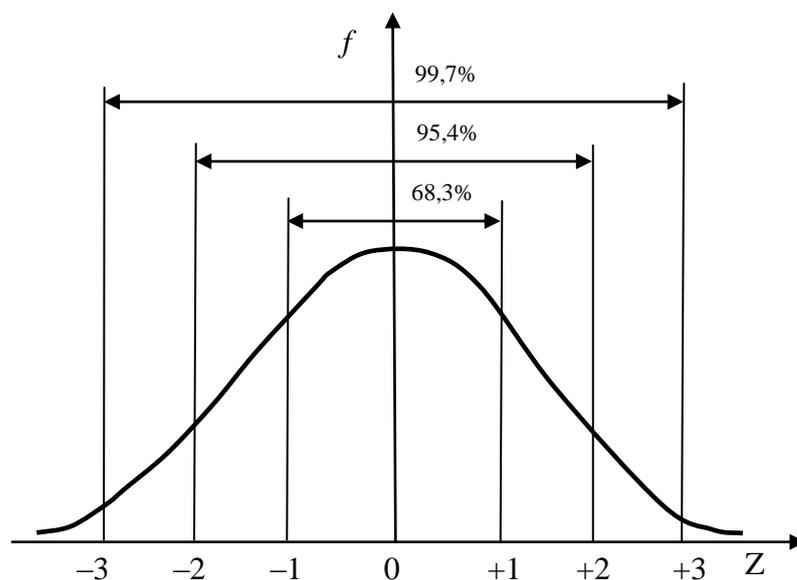


Рис. 7. Единичное нормальное распределение

Существует специальная таблица, позволяющая определить вероятность встречаемости значений признака из любого диапазона.

В психологических исследованиях нормальное распределение используется при разработке и применении тестов интеллекта. Отклонения показателей интеллекта следуют закону нормального распределения. При среднем значении 100 для исследуемой выборки, стандартное отклонение будет равно 16. Используя таблицы стандартного нормального распределения, можно вычислить какая часть выборки имеет то или иное значение коэффициента интеллекта. Применительно к другим психологическим категориям и сферам (личностная, мотивационная) применение закона нормального распределения является дискуссионным.

Существует множество критериев проверки соответствия изучаемого распределения нормальному. Наиболее простой, неформальный критерий: если мода, медиана и среднее арифметическое приблизительно равны, то ряд имеет нормальное распределение.

Наиболее эффективным критерием при проверке нормальности распределения считается критерий Колмогорова-Смирнова. Критерий позволяет оценить вероятность того, что выборка принадлежит генеральной совокупности с нормальным распределением. Если вероятность $p < 0,05$, то данное эмпирическое распределение существенно отличается от нормального, а если $p > 0,05$, данное распределение приблизительно соответствует нормальному.

Проверка нормального распределения в SPSS.

1. Введем произвольные данные в столбец.
2. В верхнем меню выбираем *Анализ* → *Непараметрические критерии* → *Одновыборочный Колмогорова-Смирнова*.
3. В открывшемся окне переносим переменную в список проверяемых переменных (см. Рис. 8).

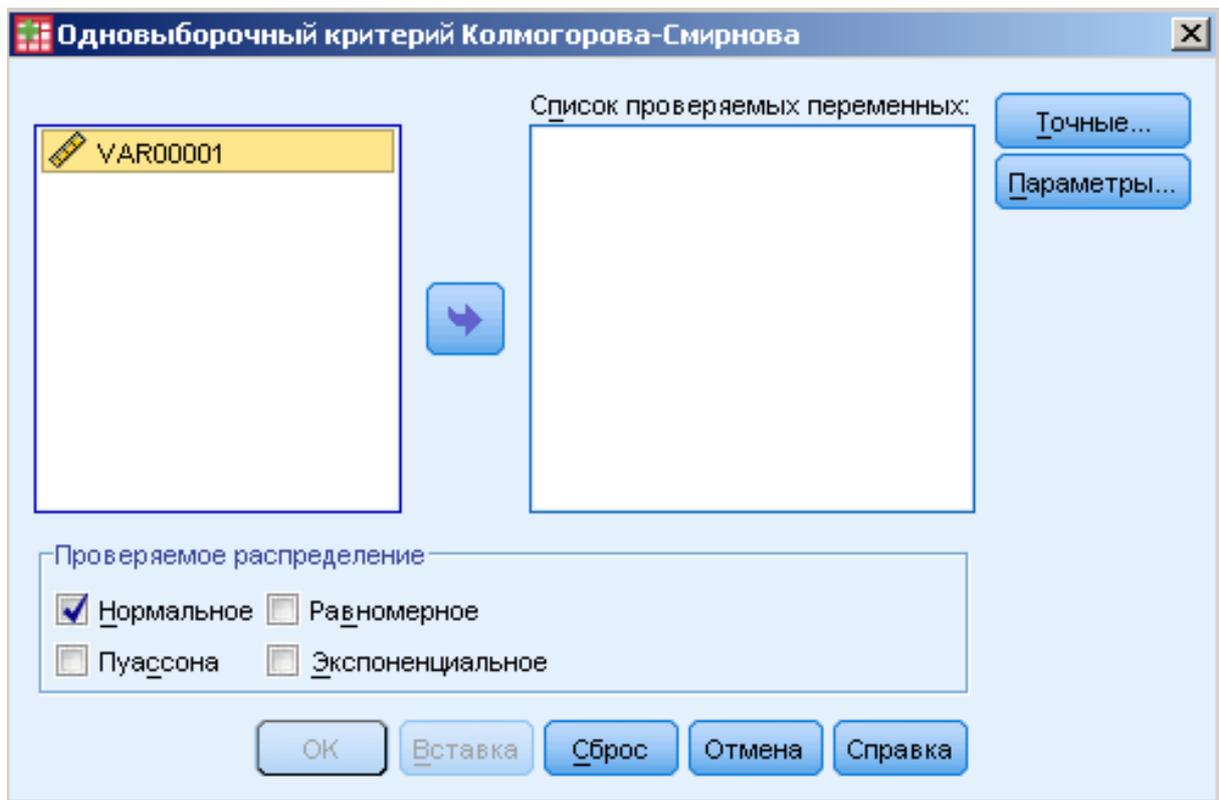


Рис.8. Окно программы SPSS

4. Нажимаем ОК и получаем следующий результат:

Таблица 2

Одновыборочный критерий Колмогорова-Смирнова

| | | VAR00001 |
|-------------------------------------|-----------------|----------|
| N | | 18 |
| Нормальные параметры ^{a,b} | Среднее | 51,1111 |
| | Стд. отклонение | 17,38116 |
| Разности экстремумов | Модуль | ,153 |
| | Положительные | ,137 |
| | Отрицательные | -,153 |
| Статистика Z Колмогорова-Смирнова | | ,649 |
| Асимпт. знч. (двухсторонняя) | | ,793 |

a. Сравнение с нормальным распределением.

b. Оценивается по данным.

Внизу таблицы находим значение **Статистика Z Колмогорова-Смирнова** равное 0,649 и вероятность **Асимпт. знч. (двухсторонняя)** равную 0,793.

Если *Асимпт. знч.* меньше или равно 0,05, то распределение существенно отличается от нормального. Если *Асимпт. знч.* больше 0,05, то отличий от нормальности не обнаружено.

В данном случае вероятность равна 0,793. Следовательно, ряд не имеет значимых отличий от нормального распределения.

В зависимости от соответствия/несоответствия ряда нормальному распределению, экспериментальные данные обрабатываются соответствующими математико-статистическими методами.

1.6. Статистические гипотезы. Уровень статистической значимости

Полученные в результате эксперимента данные служат основанием для суждения о генеральной совокупности. В силу действия случайных вероятностных причин оценка параметров генеральной совокупности, сделанная на основании экспериментальных данных всегда будет сопровождаться погрешностью. Подобного рода оценки должны рассматриваться как предположительные. Эти предположения называются статистическими гипотезами.

Статистическая гипотеза – это научная гипотеза, допускающая статистическую проверку. Пример: исследование интеллекта у подростков из полных и неполных семей. Можно ли сделать вывод о том, что неполная семья ведет к снижению интеллекта у подростков? Неработающие женщины имеют показатели самооценки ниже, чем работающие женщины. Можно ли утверждать, что трудовая занятость способствует повышению самооценки? Эти решения всегда вероятностны.

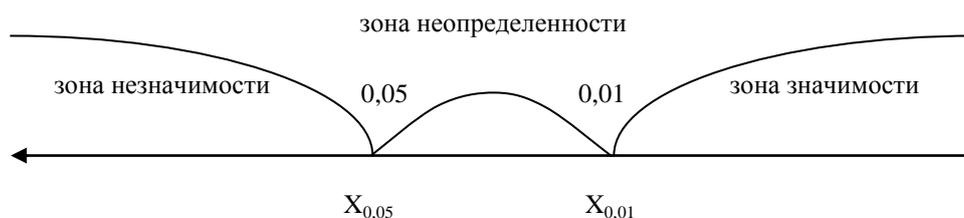
Выделяют нулевую и альтернативную гипотезы. Нулевая гипотеза (H_0) – гипотеза об отсутствии связи в генеральной совокупности. Альтернативная (H_1) – гипотеза о наличии связи.

При обосновании вопроса, где проходит линия между принятием и отвержением нулевой гипотезы в силу наличия в эксперименте случайных влияний эта граница не может быть проведена абсолютно точно. Она базируется на понятии уровня значимости. *Уровень значимости* – это вероятность ошибочного отклонения нулевой гипотезы. В психологии и педагогике

достаточный уровень значимости равен 0,05, высокий – 0,01, очень высокий – 0,001. Чем меньше уровень значимости, тем меньше вероятность того, что нами допущена ошибка и тем надежнее связь. Уровень $p = 0,05$ – означает, что вероятность допустить ошибку равна пяти ошибкам в выборке из ста элементов (случаев, испытуемых), или одной ошибке в выборке из 20 элементов. Таким образом, больше чем одна ошибка в выборке из двадцати элементов мы сделать не можем.

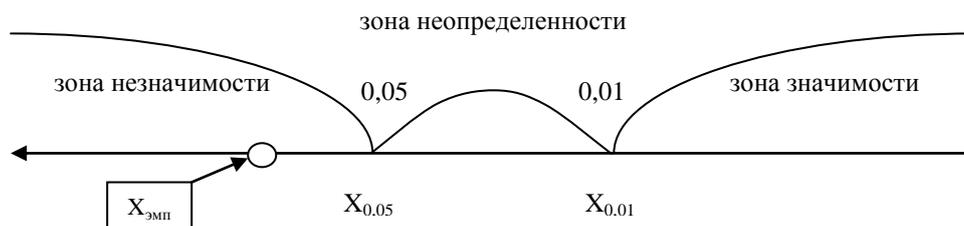
Правило принятия статистического решения: на основании полученных экспериментальных данных исследователь подсчитывает по выбранному им методу эмпирическое значение. Эмпирическое значение сравнивается с критическими, которые соответствуют уровню значимости для выбранного статистического метода. Критические величины находятся для данного статистического метода по соответствующим таблицам, приведенным в приложении к данному пособию. Сравнение полученного эмпирического значения с критическими (табличными) значениями удобно осуществлять с помощью «оси значимости».

«Ось значимости» – это прямая, имеющая три выделенные зоны: зона незначимости, зона неопределенности, зона значимости. Границами трех зон являются критические значения $X_{0,05}$ и $X_{0,01}$ для уровней $p = 0,05$ и $p = 0,01$ соответственно.



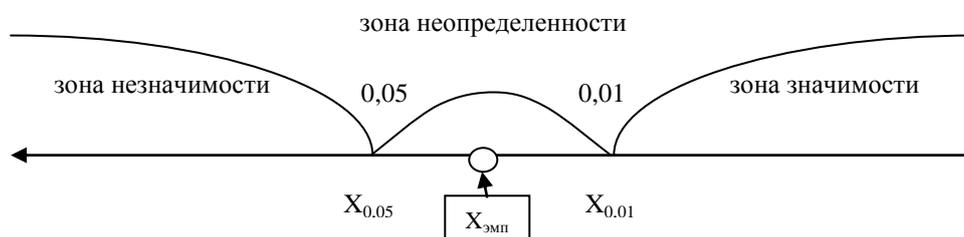
Эмпирическое значение, подсчитанное по какому либо методу должно обязательно попасть в одну из трех зон.

1 случай. Эмпирическое значение попало в зону незначимости:



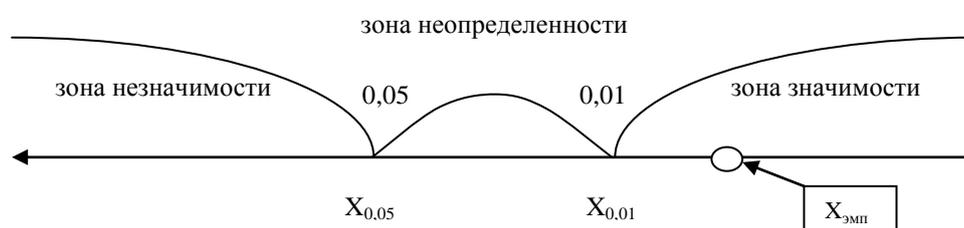
В этом случае принимается нулевая гипотеза H_0 об отсутствии различий (связи).

2 случай. Эмпирическое значение попало в зону неопределенности:



В этом случае принимается альтернативная гипотеза H_1 о наличии различий (связи) на уровне значимости $p < 0,05$.

3 случай. Эмпирическое значение попало в зону значимости:



В этом случае принимается альтернативная гипотеза H_1 о наличии различий (связи) на уровне значимости $p < 0,01$.

При совпадении эмпирического значения с теоретическим принимается альтернативная гипотеза H_1 на уровне значимости $p = 0,05$ или $p = 0,01$.

ГЛАВА 2. МЕТОДЫ МНОГОМЕРНОГО АНАЛИЗА ДАННЫХ

Развитие методов многомерного анализа данных обусловлено объективным многообразием изучаемых явлений и их сложностью. Многомерные методы эффективны при измерении и отражении результатов, в тех случаях, где расчеты средних, дисперсий и других одномерных характеристик оказываются недостаточными.

Предметом наблюдения и изучения в многомерном анализе выступают все виды и формы связей, недоступных для простого наблюдения и изучения. Изучением латентных, ненаблюдаемых явлений традиционная статистика мало интересовалась и не занималась в силу сложности самого предмета.

Бурное развитие программных средств обработки информации, систематизация знаний позволяет осуществить более действенное знакомство с многомерными методами анализа данных. Это позволит усилить методологическое и диагностическое значение проводимых исследований, актуализирует развитие исследовательской логики, расширит познавательные способности в изучении научной картины мира, действительности.

2.1. Корреляционный анализ

Корреляция – это согласованное изменение признаков. Если при изменении одной (или нескольких) величин изменяются другая (другие), то между показателями этих явлений будет наблюдаться корреляция. Наличие корреляции двух переменных ничего не говорит о причинно-следственных зависимостях между ними, однако дает возможность выдвинуть такую гипотезу.

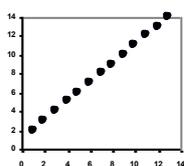
Корреляция является отрицательной, если увеличение одной переменной связано с уменьшением другой (чем выше способности к манипуляции людьми в межличностном общении, тем, ниже степень сопереживания к окружающим).

Корреляция является положительной, если увеличение одной переменной связано с увеличением другой переменной (чем выше интеллект, тем выше успеваемость учащегося).

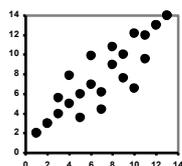
Корреляция является нулевой, при которой отсутствуют связи между переменными (связь между ростом учеников и их агрессивностью).

Корреляции также могут быть линейными и нелинейными. Если с увеличением или уменьшением одной переменной вторая переменная в среднем также либо растет, либо убывает, то связь линейна. Если при увеличении одной величины характер изменения другой величины нелинейен, а описывается другими законами, то связь нелинейна. В психологических и педагогических исследованиях сильная линейная корреляционная связь встречается достаточно редко.

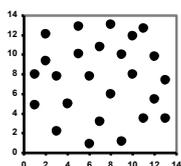
Линейную корреляцию можно количественно измерить. Степень связи между признаками выражается величиной, называемой коэффициентом корреляции. Обозначается r . Значения данного коэффициента могут находиться в диапазоне от -1 до $+1$. Возможные варианты связей, соответствующие им коэффициенты корреляции и их интерпретации изобразим на диаграммах рассеивания:



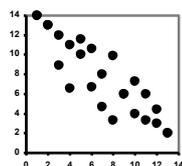
$r = +1$
сильная
прямая
связь



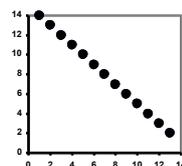
$r = +0,5$
слабая
прямая
связь



$r = 0$
нет
связи



$r = -0,5$
слабая
обратная
связь



$r = -1$
сильная
обратная
связь

Коэффициенты корреляции характеризуются не только силой, но и значимостью. Сильная корреляция может оказаться случайной при малом объеме выборки, а слабая корреляция может оказаться высокозначимой при большом объеме выборки.

Пример. В таблице 1 представлены результаты роста и веса 11 испытуемых. Видно, что связь статистическая, тесная и прямая, т.е. испытуемые, имеющие высокий рост как правило, имеют и больший вес, а испытуемые имеющих низкий рост – меньший вес (см. Рис.1).

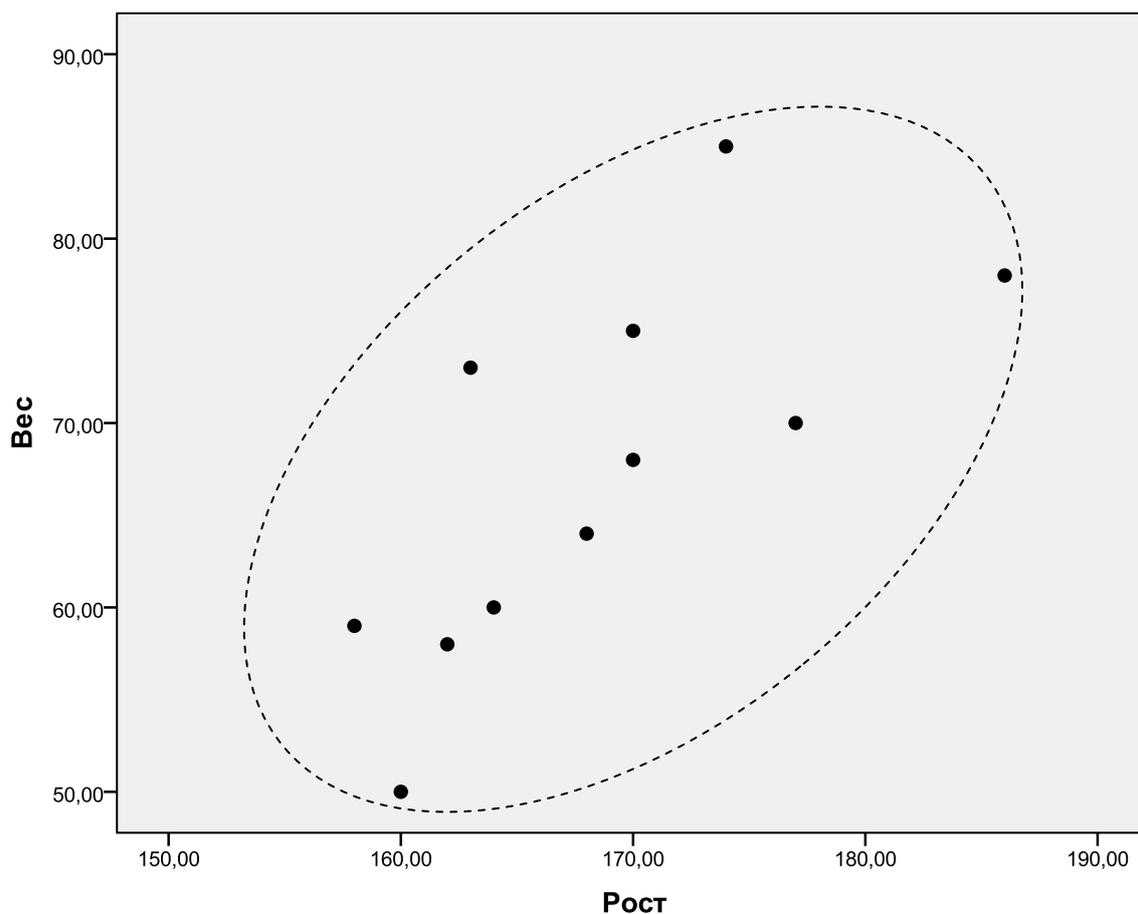


Рис. 1. Связь веса и возраста испытуемых

Таблица 1

Данные роста и веса испытуемых

| № | Рост, см | Вес, кг |
|----|----------|---------|
| 1 | 162 | 58 |
| 2 | 158 | 59 |
| 3 | 160 | 50 |
| 4 | 163 | 73 |
| 5 | 164 | 60 |
| 6 | 174 | 85 |
| 7 | 177 | 70 |
| 8 | 170 | 68 |
| 9 | 170 | 75 |
| 10 | 168 | 64 |
| 11 | 186 | 78 |

Наиболее простой вариант расчета коэффициента корреляции: с помощью программы MS Excel, вызвав статистическую функцию КОРРЕЛ (см. Рис.2).

| | В | С | Д | Е | Ф | Г | Н |
|--|----|------|-----|---|------------------------|---|---|
| | | | | | | | |
| | № | рост | вес | | | | |
| | 1 | 162 | 58 | | =КОРРЕЛ(C3:C13;D3:D13) | | |
| | 2 | 158 | 59 | | | | |
| | 3 | 160 | 50 | | | | |
| | 4 | 163 | 73 | | | | |
| | 5 | 164 | 60 | | | | |
| | 6 | 174 | 85 | | | | |
| | 7 | 177 | 70 | | | | |
| | 8 | 170 | 68 | | | | |
| | 9 | 170 | 75 | | | | |
| | 10 | 168 | 64 | | | | |
| | 11 | 186 | 78 | | | | |

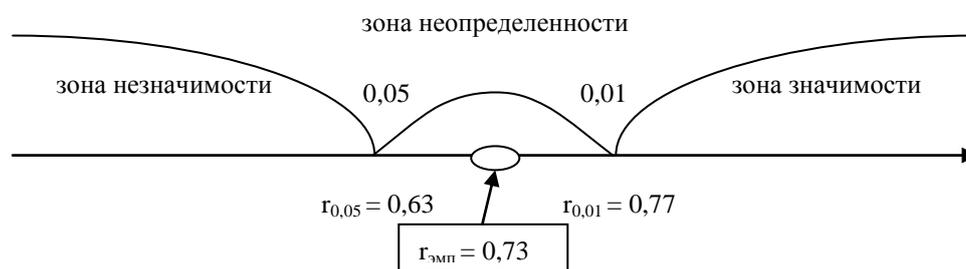
Рис.2. Рабочее окно MS Excel

В результате выполнения программы, коэффициент $r = 0,728$ или приблизительно $r = 0,73$. Возможно данный коэффициент при малой выборке испытуемых, является случайным, поэтому его необходимо проверить на значимость. Существует множество способов проверки коэффициента корреляции на значимость. Наиболее простой это сопоставление эмпирического значения коэффициента с критическими по таблице.

Критические значения для коэффициента корреляции находим по таблице 1 (см. Приложение). Поиск критических величин ведется по числу степеней свободы $k = n - 2$. В нашем примере $n = 10$, поэтому $k = 8$ и часть статистической таблицы выглядит следующим образом:

| k | p | |
|-----|------|------|
| | 0,05 | 0,01 |
| 8 | 0,63 | 0,77 |

Построим «ось значимости», на которой расположим критические значения $r_{0,05} = 0,63$, $r_{0,01} = 0,77$ и эмпирическое значение $r_{\text{эмп}} = 0,73$.



Полученная величина $r_{\text{эмп}}$ попала в зону неопределенности. Принимается гипотеза H_1 о том, что существуют статистически достоверная корреляционная связь между показателями роста и веса на уровне значимости $p < 0,05$.

Коэффициент корреляции r-Спирмана

Назначение. Вычисление ранговой корреляции позволяет определить силу и направление корреляционной связи между двумя признаками, измеренными в ранговой шкале или между двумя иерархиями признаков.

Если исходные данные были измерены в интервальной шкале или шкале отношений, то они переводятся в ранговую, т.е. меньшему значению по шкале присваивается ранг 1, следующему большему по величине значению присваивается ранг 2 и т.д.:

| | | | | | | |
|----------|----|----|----|----|---|----|
| Значение | 24 | 34 | 18 | 14 | 8 | 12 |
| Ранг | 5 | 6 | 4 | 3 | 1 | 2 |

Если в исходных данных присутствуют одинаковые значения, то между их рангами находится среднее арифметическое:

| | | | | | | |
|----------|----|----------|----------|----|---|----|
| Значение | 24 | 14 | 14 | 10 | 8 | 12 |
| Ранг | 6 | <u>4</u> | <u>5</u> | 2 | 1 | 3 |
| Ранг | 6 | 4,5 | 4,5 | 2 | 1 | 3 |

В указанном случае ранги 4,5 принимают два значения 14.
Коэффициент ранговой корреляции r -Спирмана вычисляется по формуле:

$$r = 1 - \frac{6 \cdot \sum d^2}{(n-1) \cdot n \cdot (n+1)}$$

где d – разность рангов, n – число пар объектов.

Пример. Исследователь протестировал по методике «Удовлетворены ли вы браком?» 10 пар супругов. Его цель выяснить, существует ли связь между показателями удовлетворенности, полученными между супругами?

Таблица 2

Результаты исследования по методике «Удовлетворены ли вы браком?»

| Пара | Жена | Муж |
|------|------|-----|
| 1 | 24 | 26 |
| 2 | 20 | 24 |
| 3 | 20 | 26 |
| 4 | 25 | 35 |
| 5 | 40 | 36 |
| 6 | 32 | 19 |
| 7 | 26 | 33 |
| 8 | 19 | 20 |
| 9 | 24 | 26 |
| 10 | 20 | 24 |

Решение. Построим дополнительные столбцы необходимые для дальнейшей работы с методом (см. Таблицу 3).

В четвертом и пятом столбцах таблицы 3 проранжированы показатели жен и мужей соответственно. В шестом столбце таблицы представлены величины разности рангов со знаками между данными пятого и шестого столбца. В последнем столбце эти величины возведены в квадрат.

Сумма разностей рангов равна нулю. Это является показателем правильности ранжирования и подсчета разностей.

Сформулируем статистические гипотезы:

H_0 – отсутствует корреляционная связь между группами.

H_1 – существует корреляционная связь между группами.

Таблица 3

Результаты исследования по методике
«Удовлетворены ли вы браком?»

| Пара | Жена | Муж | Ранг (Жена) | Ранг (Муж) | Разность рангов | Квадрат разности рангов |
|-------|------|-----|----------------|---------------|--------------------|-------------------------------|
| 1 | 24 | 26 | 5,5 | 6 | -0,5 | 0,25 |
| 2 | 20 | 24 | 3 | 3,5 | -0,5 | 0,25 |
| 3 | 20 | 26 | 3 | 6 | -3 | 9 |
| 4 | 25 | 35 | 7 | 9 | -2 | 4 |
| 5 | 40 | 36 | 10 | 10 | 0 | 0 |
| 6 | 32 | 19 | 9 | 1 | 8 | 64 |
| 7 | 26 | 33 | 8 | 8 | 0 | 0 |
| 8 | 19 | 20 | 1 | 2 | -1 | 1 |
| 9 | 24 | 26 | 5,5 | 6 | -0,5 | 0,25 |
| 10 | 20 | 24 | 3 | 3,5 | -0,5 | 0,25 |
| СУММА | | | 55 | 55 | 0 | 79 |

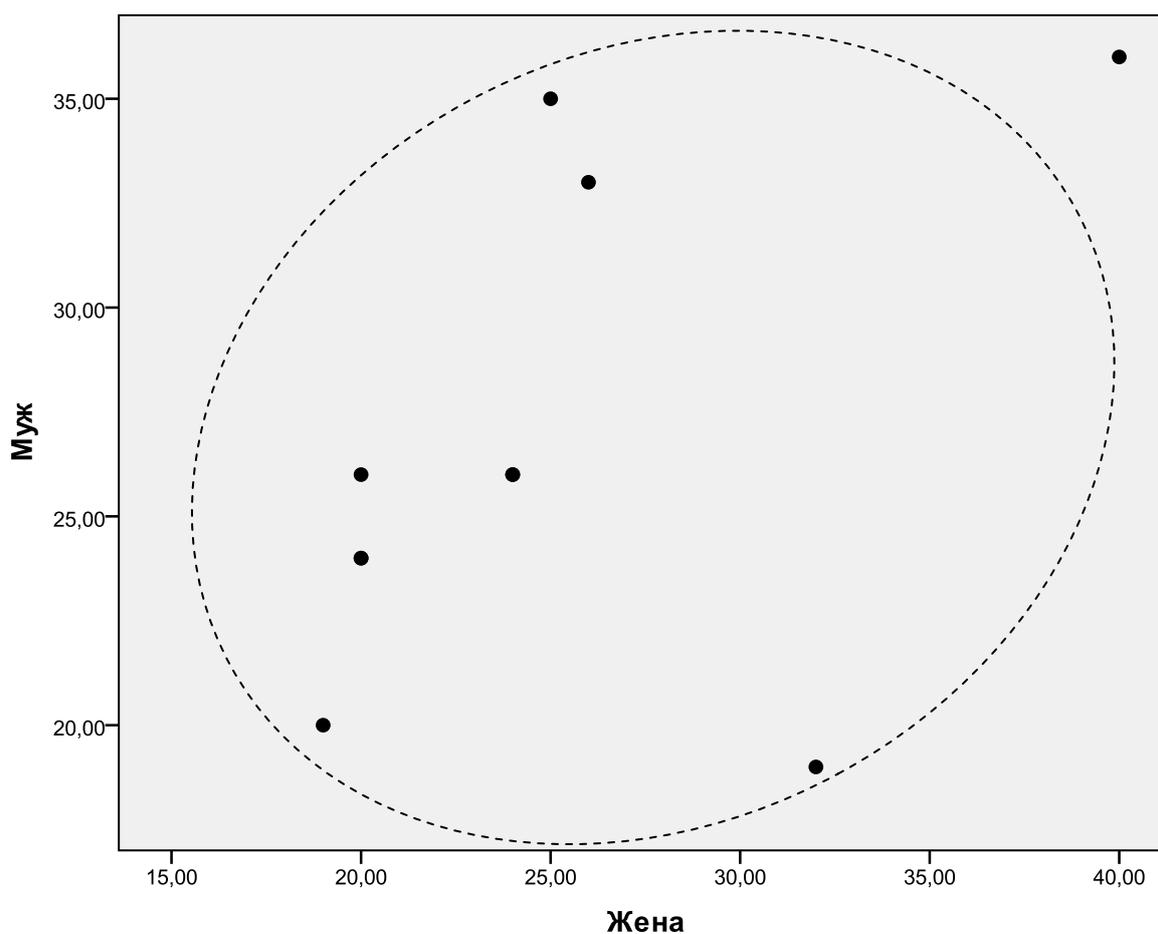


Рис. 3. График связи между показателями удовлетворенности

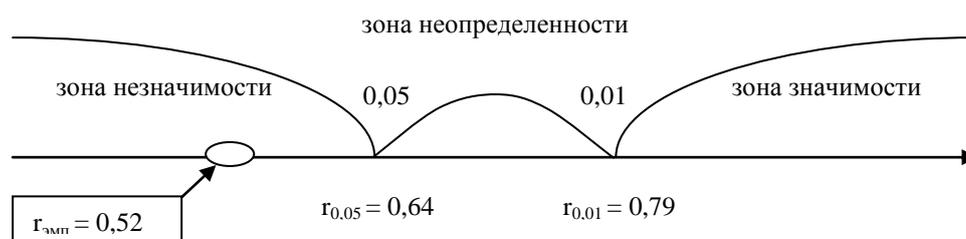
Вычислим коэффициент корреляции r -Спирмана:

$$r_{\text{эмп}} = 1 - \frac{6 \cdot \sum d^2}{(n-1) \cdot n \cdot (n+1)} = 1 - \frac{6 \cdot 79}{(10-1) \cdot 10 \cdot (10+1)} = 1 - \frac{474}{990} = 0,52$$

Критические значения для коэффициента корреляции r -Спирмена находим по таблице 2 (см. Приложение). Поиск критических величин ведется по числу пар испытуемых. В нашем примере $n = 10$, поэтому часть статистической таблицы выглядит следующим образом:

| n | p | |
|-----|------|------|
| | 0,05 | 0,01 |
| 10 | 0,64 | 0,79 |

Построим «ось значимости», на которой расположим критические значения $r_{0,05} = 0,64$, $r_{0,01} = 0,79$ и эмпирическое значение $r_{\text{эмп}} = 0,52$.



Полученная величина $r_{\text{эмп}}$ попала в зону незначимости. Принимается гипотеза H_0 о том, что отсутствуют статистически достоверная корреляционная связь между показателями двух групп. Следовательно, наблюдается рассогласованность в степени удовлетворенности браком между супругами.

Коэффициент корреляции r -Спирмана: обработка в SPSS. Решим предыдущую задачу с помощью компьютерной программы SPSS.

1. Введем данные в таблицу в два столбца.
2. В верхнем меню выбираем *Анализ* → *Корреляции* → *Парные*.
3. В открывшемся окне переносим два столбца в правую часть для переменных. В этом же окне в разделе *Коэффициенты корреляции* снимаем флажок с коэффициента *Пирсона*, который стоит по умолчанию, и ставим флажок на коэффициенте *Спирмана* (см. Рис.4).

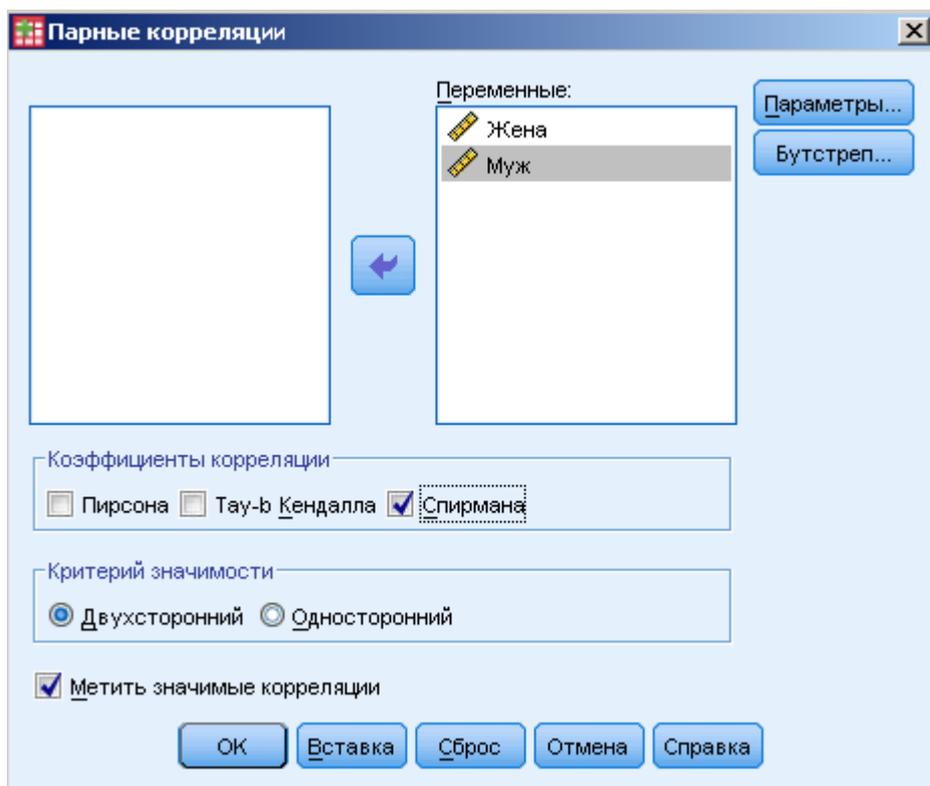


Рисунок 4. Окно программы SPSS

4. Нажимаем **OK** и получаем следующий результат:

Таблица 4

Корреляции

| | | | Жена | Муж |
|-------------|------|------------------------|-------|-------|
| ро Спирмена | Жена | Кoeffициент корреляции | 1,000 | ,506 |
| | | Знч. (2-сторон) | . | ,135 |
| | N | | 10 | 10 |
| | Муж | Кoeffициент корреляции | ,506 | 1,000 |
| | | Знч. (2-сторон) | ,135 | . |
| | N | | 10 | 10 |

В результате получаем симметричную матрицу, в которой содержится коэффициент корреляции равный 0,506 и р-уровень значимости (**Знч. (2-сторон)**). В данном случае он равен 0,135. Уровень значимости $0,135 > 0,05$, следовательно, принимается гипотеза H_0 об отсутствии статистически достоверной корреляционной связи.

Коэффициент корреляции τ -Кендалла

Назначение. Коэффициент корреляции τ «тау» Кендалла является непараметрическим. Коэффициент предназначен для обработки данных полученных в ранговой шкале. Он основан на вычислении суммы инверсий и совпадений, где под инверсией понимается нарушение порядка в ранжированных числовых данных, а в совпадениях – отсутствие нарушений.

Коэффициент ранговой корреляции τ -Кендалла вычисляется по формуле:

$$\tau = \frac{P - Q}{n \cdot (n - 1) / 2}$$

где P – число совпадений, Q – число инверсий, а $n \cdot (n - 1) / 2 = P + Q$.

При ручном расчете данные упорядочиваются по первому столбцу (шкале). Для каждого объекта подсчитывается, сколько раз его ранг по второй шкале оказывается меньше, чем ранг объектов, находящихся ниже. Это есть совпадений P . Затем для каждого объекта подсчитывается, сколько раз его ранг по второй шкале оказывается больше, чем ранг объектов, находящихся ниже. Это есть число инверсий Q .

Ранговый коэффициент корреляции τ -Кендалла на первый взгляд может показаться довольно схожим с коэффициентом r -Спирмана, однако они различаются способом вычисления и своей внутренней логикой. Коэффициент корреляции τ -Кендалла имеет вероятностную основу, то есть зная значение коэффициента мы можем определить вероятность совпадений и инверсий между сравниваемыми шкалами. Для этого существуют соотношения:

$$p(P) = \frac{1 + \tau}{2};$$
$$p(Q) = \frac{1 - \tau}{2}.$$

Например, если коэффициент корреляции $\tau = 0,50$, то значит, что вероятность совпадений равна 0,75 (или 75%), а вероятность инверсий – 0,25 (или 25%).

Коэффициент корреляции τ -Кендалла также вычисляется по формулам:

$$\tau = 1 - \frac{4 \cdot Q}{n \cdot (n-1)},$$

$$\tau = \frac{4 \cdot P}{n \cdot (n-1)} - 1,$$

где Q – сумма инверсий, P – сумма сочетаний, n – число пар объектов.

В данном пособии рассматривается пример вычисления коэффициента корреляции τ -Кендалла с помощью подсчета числа инверсий.

Пример. Исследователь протестировал по методике «Индекс отношения к здоровью» С. Дерябо, В. Ясвина 17 учащихся и их родителей. Цель: выяснить существует ли связь между показателями сформированности отношения к здоровью между старшеклассниками и их родителями?

Таблица 5

Показатели сформированности отношения к здоровью по методике «Индекс отношения к здоровью»

| № | Учащиеся | Родители учащихся |
|----|----------|-------------------|
| 1 | 14 | 23 |
| 2 | 11 | 27 |
| 3 | 18 | 23 |
| 4 | 17 | 21 |
| 5 | 15 | 27 |
| 6 | 24 | 19 |
| 7 | 18 | 28 |
| 8 | 18 | 19 |
| 9 | 18 | 21 |
| 10 | 17 | 19 |
| 11 | 13 | 19 |
| 12 | 22 | 19 |
| 13 | 15 | 28 |
| 14 | 17 | 14 |
| 15 | 14 | 19 |
| 16 | 27 | 21 |
| 17 | 19 | 23 |

Решение. Построим дополнительные столбцы необходимые для дальнейшей работы с методом (см. Таблицу 6).

Таблица 6

Показатели сформированности отношения к здоровью

| № | Учащиеся | Ранг | Родители учащихся | Ранг | Инверсии |
|----|----------|------|-------------------|------|----------|
| 1 | 11 | 1 | 27 | 14,5 | 13 |
| 2 | 13 | 2 | 19 | 4,5 | 1 |
| 3 | 14 | 3 | 23 | 12 | 9 |
| 4 | 14 | 4 | 19 | 4,5 | 1 |
| 5 | 15 | 5 | 27 | 14,5 | 10 |
| 6 | 15 | 6 | 28 | 16,5 | 10 |
| 7 | 17 | 7 | 21 | 9 | 5 |
| 8 | 17 | 8 | 19 | 4,5 | 1 |
| 9 | 17 | 9 | 14 | 1 | 0 |
| 10 | 18 | 10 | 23 | 12 | 5 |
| 11 | 18 | 11 | 28 | 16,5 | 6 |
| 12 | 18 | 12 | 19 | 4,5 | 0 |
| 13 | 18 | 13 | 21 | 9 | 2 |
| 14 | 19 | 14 | 23 | 12 | 3 |
| 15 | 22 | 15 | 19 | 4,5 | 0 |
| 16 | 24 | 16 | 19 | 4,5 | 0 |
| 17 | 27 | 17 | 21 | 9 | 0 |

Полученные показатели сформированности отношения к здоровью были проранжированы: меньшему значению по шкале присваивается ранг 1, следующему большему по величине значению присваивается ранг 2 и т.д.

В таблице 6 показатели учащихся были упорядочены по возрастанию. В соответствии с этим и поменялись местами показатели родителей учащихся. В дальнейшем работа будет идти только с рангами родителей учащихся.

Подсчет числа инверсий происходит следующим образом:

Самое верхнее число предпоследнего столбца «Ранг» – 14,5. Подсчитаем сколько всего чисел ниже по столбцу меньше чем 5. Таких чисел 13, поэтому ставим число 13 в последний столбец «инверсия» рядом с 14,5. Следующий ранг 4,5. Ниже него по столбцу только ранг 1, поэтому в столбце «инверсия» ставим число 1. Далее ниже ранга 12 находим 9 рангов, которые меньше 12. В столбце «инверсия» ставим число 9. И так далее.

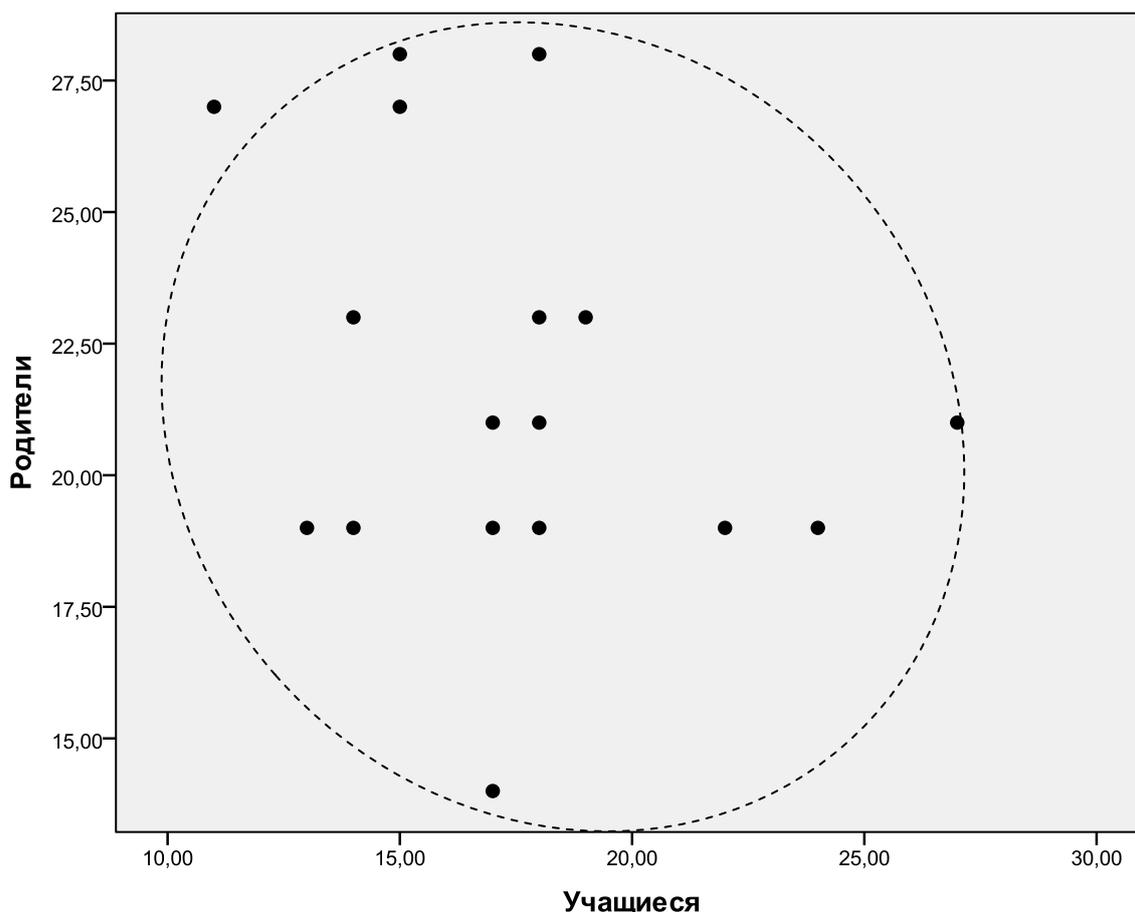


Рис.5. График связи показателей сформированности отношения к здоровью

Сумма всех инверсий равна $Q = 66$. Получаем:

$$\tau_{эмп} = 1 - \frac{4 \cdot Q}{n \cdot (n-1)} = 1 - \frac{4 \cdot 66}{17 \cdot (17-1)} = 1 - \frac{264}{272} \approx 0,03$$

Уровень значимости коэффициента корреляции проверяется по формуле:

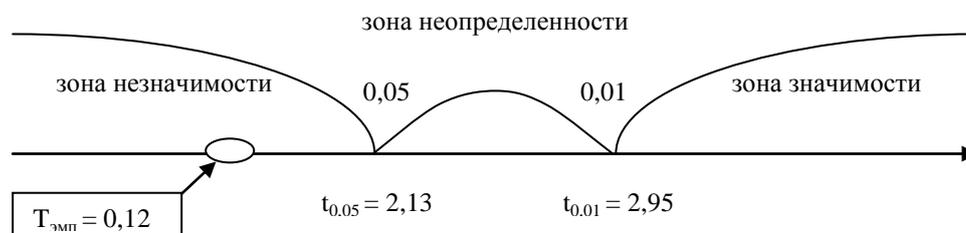
$$T_{эмп} = |\tau_{эмп}| \cdot \sqrt{\frac{n-2}{1-\tau_{эмп}^2}} \approx 0,12$$

где n – число коррелируемых признаков, а $\tau_{эмп}$ – коэффициент корреляции τ -Кендалла.

Критические значения для коэффициента корреляции находим по таблице критических значений критерия t-Стьюдента (см. Таблицу 3 Приложения). В нашем примере число степеней свободы будет $k = n-2 = 17-2 = 15$, поэтому часть статистической таблицы выглядит следующим образом:

| k | p | |
|----|-------|-------|
| | 0,05 | 0,01 |
| 15 | 2,131 | 2,947 |

На «оси значимости» расположим критические значения $t_{0,05} = 2,13$, $t_{0,01} = 2,95$ и эмпирическое значение $T_{эмп} = 0,12$.



Полученная величина $T_{эмп}$ попала в зону незначимости. Принимается гипотеза H_0 о том, что отсутствуют статистически достоверная корреляционная связь между показателями сформированности отношения к здоровью учащихся и их родителей.

Как указывалось выше, зная значение коэффициента корреляции τ -Кендалла мы можем определить вероятность совпадений и инверсий между сравниваемыми шкалами с помощью формул: $p(P) = (1 + \tau)/2$; $p(Q) = (1 - \tau)/2$.

В нашем случае коэффициент корреляции $\tau = 0,03$, следовательно, вероятность совпадений $p(P) = (1 + 0,03)/2 = 0,515$ (или 52%), а вероятность инверсий $p(Q) = (1 - 0,03)/2 = 0,485$ (или 48%).

Коэффициент корреляции τ -Кендалла: обработка в SPSS. Решим предыдущую задачу с помощью компьютерной программы SPSS.

1. Введем данные в таблицу в два столбца.
2. В верхнем меню выбираем *Анализ* → *Корреляции* → *Парные*.
3. В открывшемся окне переносим два столбца в правую часть для переменных. В этом же окне в разделе **Коэффициенты корреляции** снимаем флажок с коэффициента *Пирсона*, который стоит по умолчанию, и ставим флажок на коэффициенте *Кендалла* (см. Рис.7).

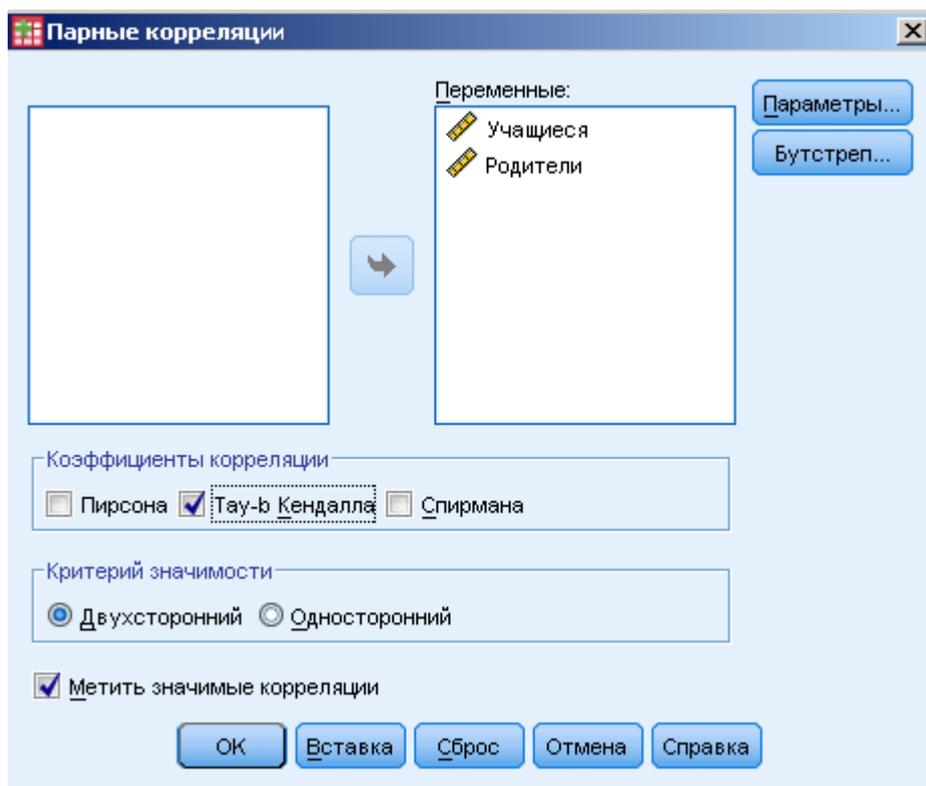


Рис.7. Окно программы SPSS

4. Нажимаем **OK** и получаем следующий результат:

Таблица 7

Корреляции

| | | | Учащиеся | Родители |
|----------------|----------|------------------------|----------|----------|
| тау-в Кендалла | Учащиеся | Кoeffициент корреляции | 1,000 | -,118 |
| | | Знч. (2-сторон) | | ,547 |
| | | N | 17 | 17 |
| | Родители | Кoeffициент корреляции | -,118 | 1,000 |
| | | Знч. (2-сторон) | ,547 | |
| | | N | 17 | 17 |

В результате получаем симметричную матрицу, в которой содержится коэффициент корреляции равный – 0,118 и р-уровень значимости (**Знч. (2-сторон)**). В данном случае он равен 0,547. Уровень $0,547 > 0,05$, поэтому принимается гипотеза H_0 об отсутствии статистически достоверной корреляционной связи.

Коэффициент корреляции r-Пирсона

Назначение. Коэффициент корреляции r-Пирсона характеризует наличие линейной связи между признаками. Его применение требует предварительной проверки данных на соответствие нормальному распределению. Формула для подсчета коэффициента корреляции r-Пирсона выглядит следующим образом:

$$r_{xy} = \frac{\sum(x_i - \bar{X}) \cdot (y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{X})^2 \cdot \sum(y_i - \bar{Y})^2}}$$

где \bar{X} , \bar{Y} – средние арифметические элементов ряда x и y соответственно.

Вычисление коэффициента корреляции r-Пирсона – процесс достаточно емкий, поэтому рассмотрим алгоритм его расчета в программах MS Excel и SPSS.

Коэффициент корреляции r-Пирсона: обработка в Excel. Решим предыдущую задачу с помощью компьютерной программы Microsoft Excel (см. Рис.8).

| Учащиеся | Родители | | |
|----------|----------|------------------------|--|
| 11 | 27 | =ПИРСОН(B2:B18;C2:C18) | |
| 13 | 19 | | |
| 14 | 23 | | |
| 14 | 19 | | |
| 15 | 27 | | |
| 15 | 28 | | |
| 17 | 21 | | |
| 17 | 19 | | |
| 17 | 14 | | |
| 18 | 23 | | |
| 18 | 28 | | |
| 18 | 19 | | |
| 18 | 21 | | |
| 19 | 23 | | |
| 22 | 19 | | |
| 24 | 19 | | |
| 27 | 21 | | |

Рис.8. Рабочее окно MS Excel

Произвольно выбираем свободную ячейку в таблице. В ячейке прописываем « = ПИРСОН (...) ». В окне выделяем первый столбец «учащиеся» и через точку с запятой выделяем второй столбец «родители». Закрываем скобку и нажимаем **Enter**. В выбранной ячейке получим значение коэффициента корреляции r-Пирсона: – 0,29. Далее полученное значение сравнивается с критическими по таблице 1 (см. Приложение) и делается соответствующий вывод.

Коэффициент корреляции r-Пирсона: обработка в SPSS.
Решим предыдущую задачу с помощью компьютерной программы SPSS.

1. Введем данные в таблицу в два столбца.
2. В верхнем меню выбираем **Анализ** → **Корреляции** → **Парные**.
3. В открывшемся окне переносим два столбца в правую часть для переменных. В этом же окне в разделе **Коэффициенты корреляции** проверяем, что по умолчанию стоит флажок на **Пирсона** (см. Рис.9).

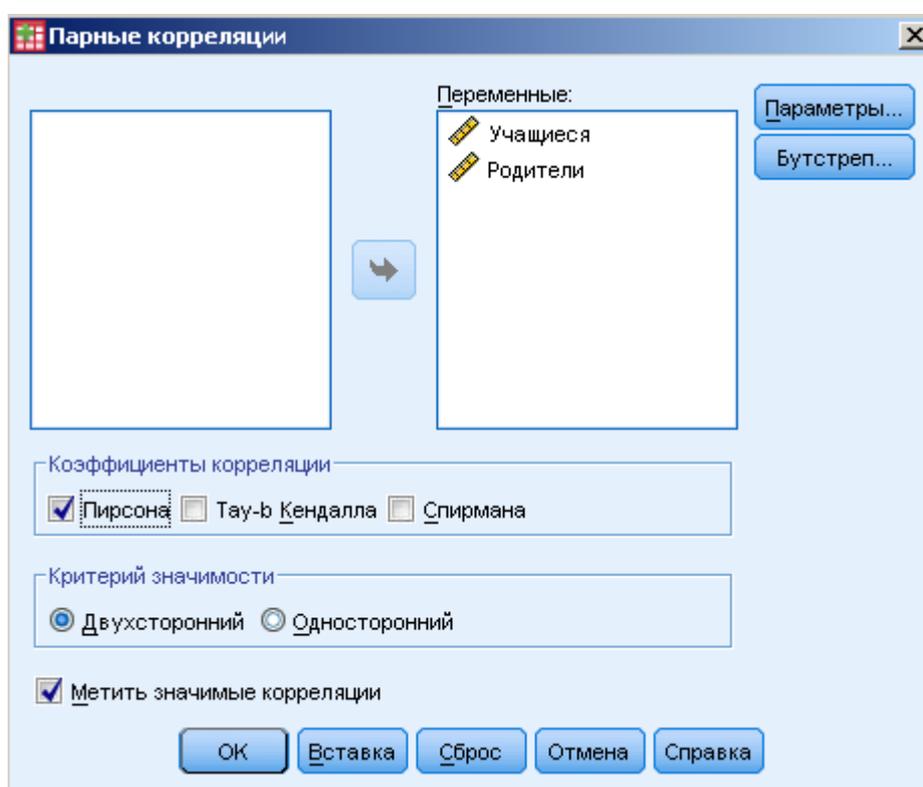


Рис.9. Окно программы SPSS

4. Нажимаем **OK** и получаем следующий результат:

Таблица 8

Корреляции

| | | Учащиеся | Родители |
|----------|--------------------|----------|----------|
| Учащиеся | Корреляция Пирсона | 1 | -,292 |
| | Знч.(2-сторон) | | ,256 |
| | N | 17 | 17 |
| Родители | Корреляция Пирсона | -,292 | 1 |
| | Знч.(2-сторон) | ,256 | |
| | N | 17 | 17 |

В результате получаем симметричную матрицу, в которой содержится коэффициент корреляции равный – 0,292 и р-уровень значимости (**Знч.(2-сторон)**). В данном случае он равен 0,256. Значение $0,256 > 0,05$, поэтому гипотеза H_0 принимается.

Анализ корреляционных матриц

Если корреляционный анализ включает в себя изучение связей не двух, а нескольких переменных, то корреляции вычисляются попарно между этими переменными, а результаты заносятся в *корреляционную матрицу*.

Корреляционная матрица – это результат вычисления корреляций для каждой пары из множества переменных. Корреляционная матрица является квадратной: число строк равно числу столбцов. Она симметрична относительно главной диагонали, на которой располагаются единицы. Главная задача анализа корреляционной матрицы – выявление структуры взаимосвязей множества признаков.

Статистически значимые элементы корреляционной матрицы можно представить графически в виде *корреляционного графа*. Корреляционный граф – это фигура, состоящая из вершин и соединяющих их линий. Вершины соответствуют признакам, а линии соответствуют статистически достоверным связям. Построение корреляционного графа обычно начинают с переменной имеющей наибольшее число значимых связей.

Задача. Выяснить, как связаны между собой факторы личностного дифференциала и показатели социальной фрустрированности юношей и девушек.

Таблица 9

| Показатели | | | | |
|------------|------------------------------|----------|--------|--------------|
| № | Социальная фрустрированность | «Оценка» | «Сила» | «Активность» |
| 1 | 2 | -1 | 6 | -3 |
| 2 | 2,6 | 3 | 5 | 3 |
| 3 | 1,9 | -16 | -8 | -6 |
| 4 | 1,2 | -4 | 12 | 2 |
| 5 | 1,5 | 5 | 13 | 10 |
| 6 | 2 | -4 | -1 | 9 |
| 7 | 0,2 | 18 | 18 | 9 |
| 8 | 0,8 | 21 | 17 | 17 |
| 9 | 2 | 2 | 1 | 2 |
| 10 | 0,9 | 10 | 12 | 3 |
| 11 | 0,6 | 21 | 6 | 11 |
| 12 | 0,5 | 21 | 13 | 9 |
| 13 | 2,2 | -10 | 2 | 5 |
| 14 | 2 | -19 | -5 | 4 |
| 15 | 1,8 | -18 | -15 | 1 |
| 16 | 1,2 | 4 | 5 | 7 |
| 17 | 1,9 | 1 | 9 | 9 |
| 18 | 2,2 | -3 | -1 | 2 |
| 19 | 1,9 | 4 | 12 | 6 |
| 20 | 1,9 | 4 | 4 | 5 |

Решение. Для решения задачи воспользуемся коэффициентом корреляции r -Пирсона. Парно вычислим корреляции между столбцами с помощью программы MS Excel (см. алгоритм на стр. 37) и занесем результаты в корреляционную матрицу (см. Таблицу 10).

Таблица 10

| Корреляционная матрица | | | | |
|------------------------|-------|--------|------|------------|
| | УСФ | Оценка | Сила | Активность |
| УСФ | 1,00 | | | |
| Оценка | -0,72 | 1,00 | | |
| Сила | -0,60 | 0,81 | 1,00 | |
| Активность | -0,52 | 0,68 | 0,59 | 1,00 |

Обозначения:

УСФ – уровень социальной фрустрированности; Оценка – фактор оценки; Сила – фактор силы; Активность – фактор активности.

Критические значения для коэффициента корреляции r -Пирсона находим по таблице 1 (см. Приложение). Поиск критических величин ведется по числу $k = n - 2$, где n – число пар испытуемых, т.е. $k = 18$. Наша часть таблицы выглядит следующим образом:

| k | p | |
|-----|------|------|
| | 0,05 | 0,01 |
| 18 | 0,44 | 0,56 |

Выделим статистически значимые элементы корреляционной матрицы (см. Таблицу 11):

Таблица 11

| | УСФ | Оценка | Сила | Активность |
|--------------|------------|--------|------|------------|
| УСФ | 1,00 | | | |
| Оценка | -0,72 | 1,00 | | |
| Сила | -0,60 | 0,81 | 1,00 | |
| Активность | -0,52 | 0,68 | 0,59 | 1,00 |
| Обозначение: | | | | |
| | $P < 0,05$ | | | |
| | $P < 0,01$ | | | |

Наиболее сильные положительные связи выявлены между факторами «Оценки» и «Силы» ($p < 0,01$), «Оценки» и «Активности» ($p < 0,01$), «Активности» и «Силы» ($p < 0,01$).

Наиболее сильные отрицательные связи выявлены между фактором «Оценки» и социальной фрустрированностью ($p < 0,01$), фактором «Силы» и социальной фрустрированностью. Менее сильная – между фактором «Активности» и социальной фрустрированностью ($p < 0,05$).

Статистически значимые элементы корреляционной матрицы можно представить графически в виде корреляционного графа: (см. Рис. 10):

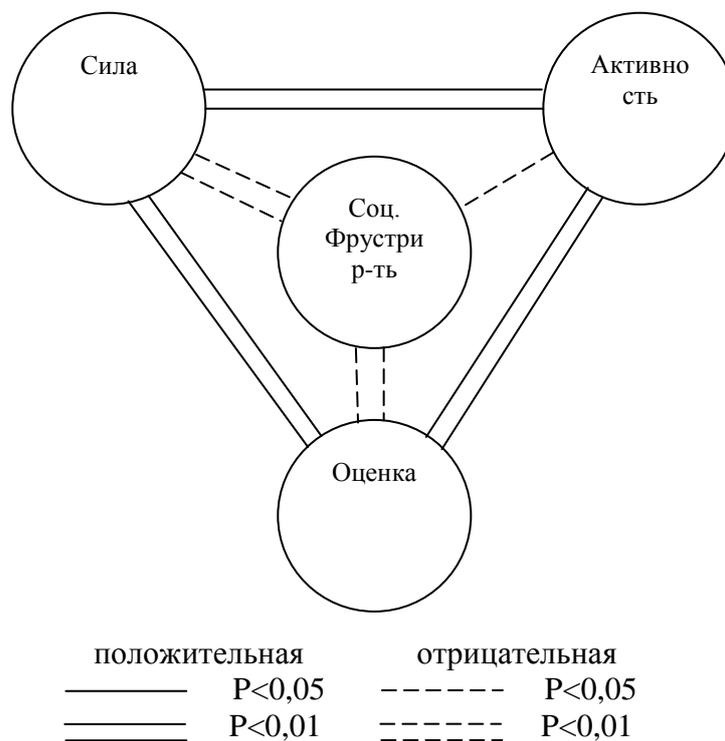


Рис 10. Корреляционный граф

Корреляционный анализ: обработка в SPSS. Решим предыдущую задачу с помощью компьютерной программы SPSS.

1. Введем данные в таблицу в четыре столбца.
2. В верхнем меню выбираем *Анализ* → *Корреляции* → *Парные*.
3. В открывшемся окне переносим два столбца в правую часть для переменных. В этом же окне в разделе *Коэффициенты корреляции* выбираем необходимый коэффициент корреляции (см. Рис.11).

В результате получаем симметричную матрицу, в которой содержатся коэффициенты корреляции, а ниже их р-уровни значимости (**Знч.(2-сторон)**).

Коэффициенты, справа от которых расположены скобки (**) являются статистически значимыми на уровне $p < 0,01$. Коэффициенты, справа от которых расположены скобки (*) являются статистически значимыми на уровне $p < 0,05$.

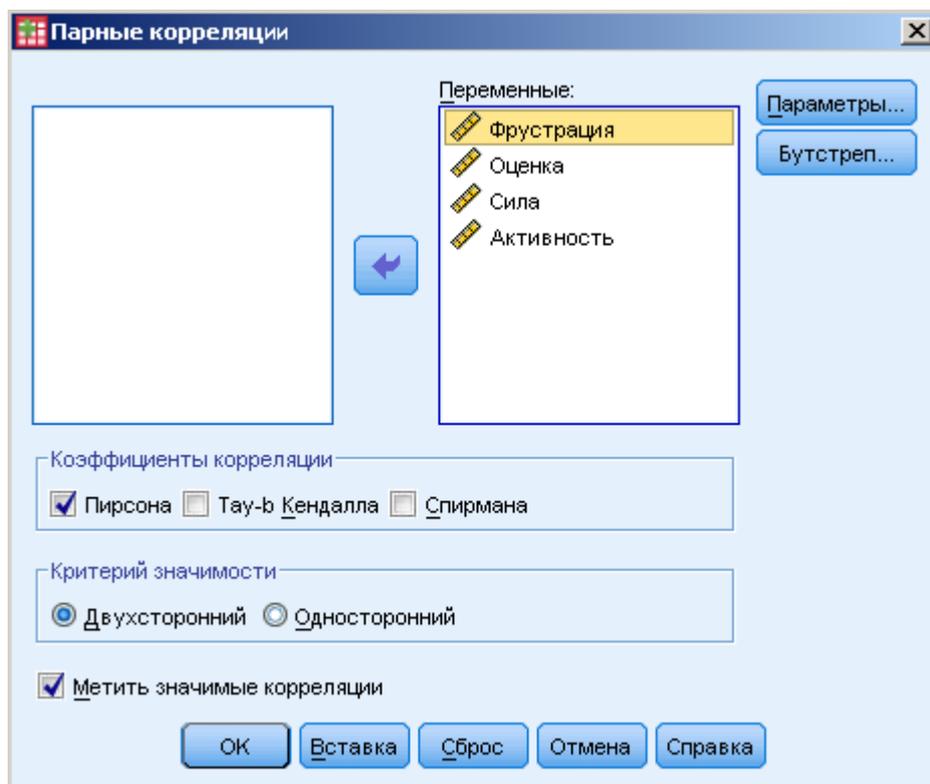


Рис.11. Окно программы SPSS

4. Нажимаем **OK** и получаем следующий результат:

Таблица 12

Корреляции

| | | Фрустрация | Оценка | Сила | Активность |
|------------|--------------------|------------|---------|---------|------------|
| Фрустрация | Корреляция Пирсона | 1 | -,723** | -,596** | -,516* |
| | Знч.(2-сторон) | | ,000 | ,006 | ,020 |
| | N | 20 | 20 | 20 | 20 |
| Оценка | Корреляция Пирсона | -,723** | 1 | ,809** | ,676** |
| | Знч.(2-сторон) | ,000 | | ,000 | ,001 |
| | N | 20 | 20 | 20 | 20 |
| Сила | Корреляция Пирсона | -,596** | ,809** | 1 | ,588** |
| | Знч.(2-сторон) | ,006 | ,000 | | ,006 |
| | N | 20 | 20 | 20 | 20 |
| Активность | Корреляция Пирсона | -,516* | ,676** | ,588** | 1 |
| | Знч.(2-сторон) | ,020 | ,001 | ,006 | |
| | N | 20 | 20 | 20 | 20 |

** . Корреляция значима на уровне 0.01 (2-сторон.).

* . Корреляция значима на уровне 0.05 (2-сторон.).

2.2. Регрессионный анализ

Парный регрессионный анализ

В предыдущей главе, мы рассмотрели коэффициент корреляции, как показатель оценки статистической связи. Одной из задач является поиск функциональной зависимости между двумя переменными, т.е. аргументом X и функцией Y . Переменная X – независимая переменная (предиктор), переменная Y – зависимая переменная (отклик).

Регрессионный анализ – количественное представление связи или зависимости между X и Y .

Прямолинейную регрессию можно отразить уравнением прямой линии:

$$Y = k \cdot X + b,$$

где: Y – значения признака по линии регрессии, т. е. теоретические значения,

k – угловой коэффициент регрессии,

X – значения признака-фактора.

b – свободный член.

Если независимая переменная одна, то регрессия называется парной. Рассмотрим простейшую парную регрессионную модель – линейную.

Пример. Исследователь предположил, что спортсмены, находящиеся в состоянии фрустрации (негативном психическом состоянии, обусловленном невозможностью удовлетворения тех или иных потребностей) чаще проявляют агрессивность в поведении. В исследовании принимали участие спортсмены в количестве 96 человек. Задача сводится к нахождению зависимости между фрустрацией (X) и агрессивностью (Y) среди испытуемых.

В этом случае линейное уравнение регрессии будет иметь общий вид:

$$Y = k \cdot X + b \text{ или } \langle \text{Агрессивность} \rangle = k \cdot \langle \text{Фрустрация} \rangle + b,$$

где:

k – угловой коэффициент регрессии,

X – значения признака-фактора.

b – свободный член.

Зависимость между фрустрацией и агрессивностью спортсменов показана на рисунке 1. Коэффициент корреляции г-Пирсона, вычисленный с помощью SPSS, указывает на наличие положительной статистически достоверной связи между показателями фрустрации и агрессивности на уровне значимости $p < 0,0001$ (см. Табл. 1).

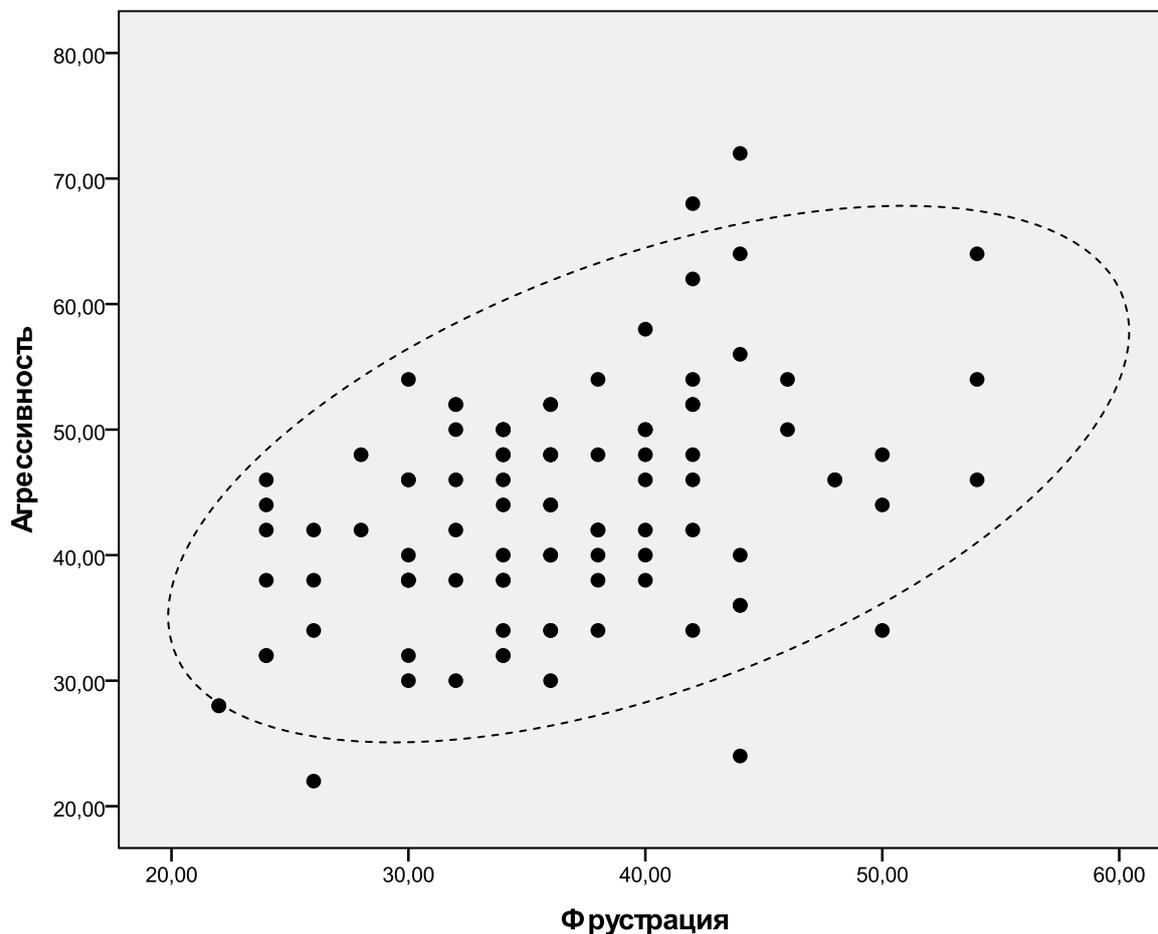


Рис.1.Связь фрустрации и агрессивности

Таблица 1

Корреляция

| | | Фрустрация | Агрессивность |
|---------------|--------------------|------------|---------------|
| Фрустрация | Корреляция Пирсона | 1 | ,418** |
| | Знч.(2-сторон) | | ,000 |
| | N | 96 | 96 |
| Агрессивность | Корреляция Пирсона | ,418** | 1 |
| | Знч.(2-сторон) | ,000 | |
| | N | 96 | 96 |

** . Корреляция значима на уровне 0.01 (2-сторон.).

Для обработки данных методом регрессионного анализа с помощью SPSS в верхнем меню выбираем *Анализ* → *Регрессия* → *Линейная...* В появившемся окне (см. Рис.2) переносим из левой части окна зависимую переменную (агрессивность) в верхнюю правую часть, независимую переменную (фрустрация) в нижнюю правую часть окна.

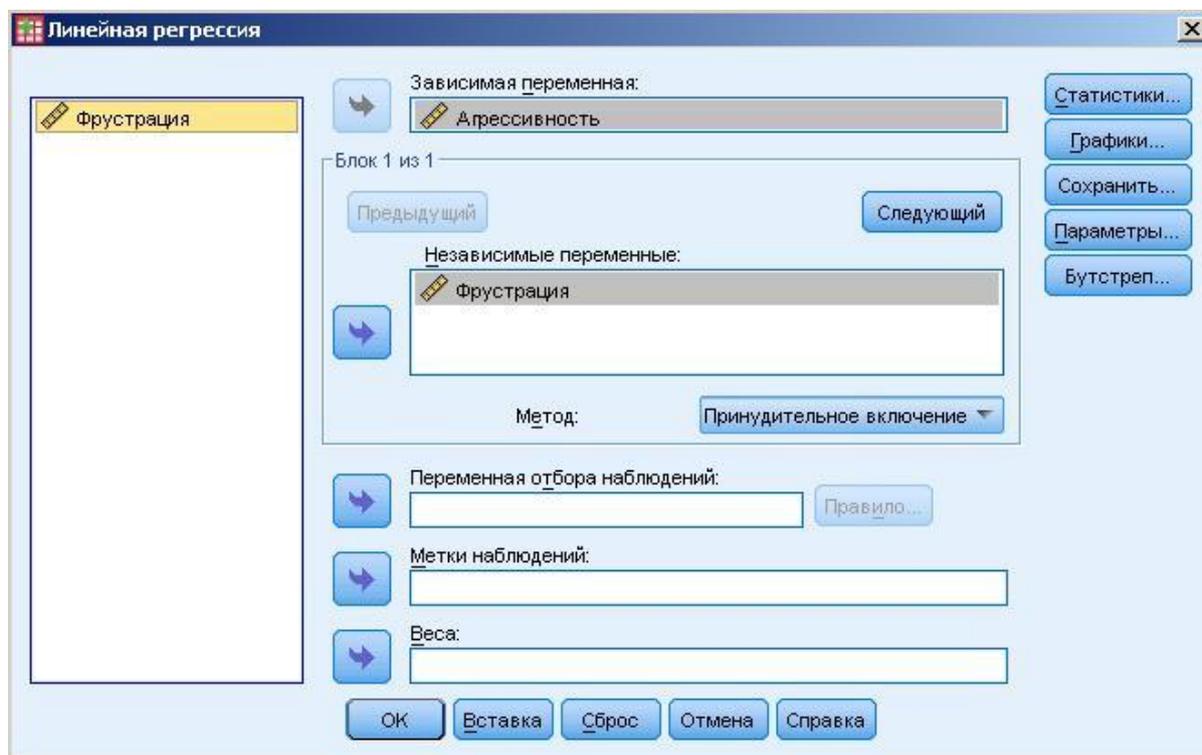


Рис.2. Регрессия

После нажатия на кнопку **OK**, на выходе получаем четыре таблицы (см. Табл.2-5).

Таблица 2
Введенные или удаленные переменные^b

| Модель | Включенные переменные | Исключенные переменные | Метод |
|--------|-------------------------|------------------------|--------------------------|
| 1 | Фрустрация ^a | | Принудительное включение |

a. Включены все запрошенные переменные

b. Зависимая переменная: Агрессивность

Таблица 3

Сводка для модели

| Модель | Н | R-квадрат | Скорректированный R-квадрат | Стд. ошибка оценки |
|--------|-------------------|-----------|-----------------------------|--------------------|
| 1 | ,418 ^а | ,174 | ,166 | 8,47990 |

а. Предикторы: (конст) Фрустрация

В таблице 3 величина Н – это коэффициент корреляции между зависимой и независимой переменными ($r = 0,418$), а R-квадрат называется коэффициентом детерминации ($R^2 = 0,174$). Он определяет долю вариации одной из переменных, которая объясняется вариацией другой переменной. В нашем случае $R^2 = 0,174$, т.е. доля вариации агрессивности объясняется вариацией фрустрации на 17%, или 17% изменчивости в агрессивности могут быть объяснены различиями во фрустрации среди спортсменов. Остальные 83% объясняются воздействиями других факторов.

Таблица 4

Дисперсионный анализ^б

| Модель | Сумма квадратов | ст.св. | Средний квадрат | F | Знч. |
|-------------|-----------------|--------|-----------------|--------|-------------------|
| 1 Регрессия | 1427,206 | 1 | 1427,206 | 19,847 | ,000 ^а |
| Остаток | 6759,419 | 94 | 71,909 | | |
| Всего | 8186,625 | 95 | | | |

а. Предикторы: (конст) Фрустрация

б. Зависимая переменная: Агрессивность

В таблице 4 указано значение критерия F-Фишера равно 19,847 и его р-уровень значимости – 0,000. Это означает, что коэффициент корреляции между зависимой и независимой переменными статистически значим и модель регрессии может быть содержательно интерпретирована.

Коэффициент регрессии k и свободный член b находятся в последней таблице 5 в столбце «нестандартизированные коэффициенты».

Коэффициенты^а

| Модель | Нестандартизованные коэффициенты | | Стандартизованные коэффициенты | t | Знч. |
|---------------|----------------------------------|-------------|--------------------------------|-------|------|
| | B | Стд. Ошибка | Бета | | |
| 1 (Константа) | 24,721 | 4,344 | | 5,690 | ,000 |
| Фрустрация | ,522 | ,117 | ,418 | 4,455 | ,000 |

а. Зависимая переменная: Агрессивность

В данном случае $k = 0,522$, а $b = 24,721$. Линейное уравнение регрессии будет иметь вид:

$$\text{«Агрессивность»} = 0,522 \cdot \text{«Фрустрация»} + 24,721.$$

В таблице 5 содержатся стандартизованные коэффициенты регрессии (Бета), а также значение критерия t-Стьюдента и уровень значимости, позволяющие определить их статистическую значимость. В уравнение могут быть приняты только те регрессионные коэффициенты, которые являются статистически значимыми. Стандартизованные коэффициенты служат показателями вклада каждой переменной в регрессионную модель (в нашем случае такая переменная одна). В парной регрессии стандартизованный коэффициент есть коэффициент корреляции между зависимой и независимой переменными.

Множественная регрессия

Множественная регрессия – это метод анализа связи между несколькими независимыми переменными (предикторами) и зависимой переменной (откликом).

В предыдущем примере мы рассмотрели влияние одной независимой переменной (фрустрация) на зависимую (агрессивность). Предположим, что на проявление агрессивности спортсменов также влияет и тревожность.

В этом случае линейное уравнение множественной регрессии будет общий иметь вид:

$$Y = k \cdot X_1 + b \cdot X_2 + c \text{ или } \langle \text{Агрессивность} \rangle = k \cdot \langle \text{Фрустрация} \rangle + b \cdot \langle \text{Тревожность} \rangle + c,$$

где:

k – угловой коэффициент регрессии,

b – угловой коэффициент регрессии,

c – свободный член.

Коэффициент корреляции r -Пирсона, вычисленный с помощью SPSS, указывает на наличие положительных статистически достоверных связей между показателями фрустрации и агрессивности, фрустрации и тревожности, агрессивности и тревожности на уровне значимости $p < 0,0001$ (см. Таблицу 6).

Таблица 6

Корреляция

| | | Фрустрация | Агрессивность | Тревожность |
|---------------|--------------------|------------|---------------|-------------|
| Фрустрация | Корреляция Пирсона | 1 | ,418** | ,683** |
| | Знч.(2-сторон) | | ,000 | ,000 |
| | N | 96 | 96 | 96 |
| Агрессивность | Корреляция Пирсона | ,418** | 1 | ,432** |
| | Знч.(2-сторон) | ,000 | | ,000 |
| | N | 96 | 96 | 96 |
| Тревожность | Корреляция Пирсона | ,683** | ,432** | 1 |
| | Знч.(2-сторон) | ,000 | ,000 | |
| | N | 96 | 96 | 96 |

** . Корреляция значима на уровне 0.01 (2-сторон.).

Для обработки данных методом множественного регрессионного анализа с помощью SPSS в верхнем меню выбираем *Анализ* → *Регрессия* → *Линейная...* В появившемся окне (см. Рис.3) переносим из левой части окна зависимую переменную (агрессивность) в верхнюю правую часть, независимые переменные (фрустрация и тревожность) в нижнюю правую часть окна.

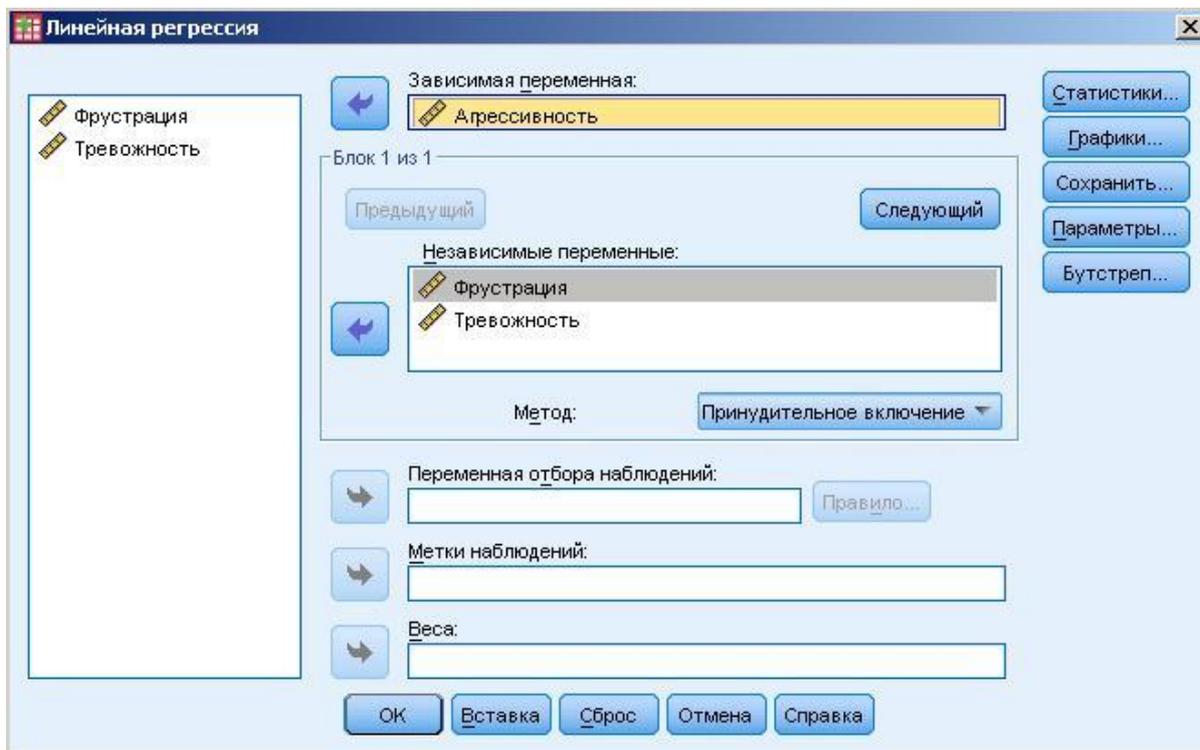


Рис.3. Окно программы SPSS

После нажатия на кнопку **OK**, на выходе получаем четыре таблицы (см.Табл.7-10). Угловые коэффициенты k , b и свободный член c находятся в последней таблице 10 в столбце «нестандартизированные коэффициенты».

В данном случае $k = 0,287$, $b = 0,320$, а $c = 19,432$. Линейное уравнение регрессии будет иметь вид:

$$\text{«Агрессивность»} = 0,287 \cdot \text{«Фрустрация»} + 0,320 \cdot \text{«Тревожность»} + 19,432.$$

Таблица 7
Введенные или удаленные переменные^b

| Модель | Включенные переменные | Исключенные переменные | Метод |
|--------|---|------------------------|--------------------------|
| 1 | Фрустрация, Тревожность ^a | | Принудительное включение |

a. Включены все запрошенные переменные

b. Зависимая переменная: Агрессивность

Таблица 8

Сводка для модели

| Модель | N | R-квадрат | Скорректированный R-квадрат | Стд. ошибка оценки |
|--------|-------------------|-----------|-----------------------------|--------------------|
| 1 | ,464 ^а | ,215 | ,198 | 8,31284 |

а. Предикторы: (конст) Фрустрация, Тревожность

В таблице 8 величина N – это коэффициент множественной корреляции между зависимой и набором независимых переменных ($r = 0,464$), а R-квадрат называется коэффициентом множественной детерминации ($R^2 = 0,215$). Он определяет долю вариации одной из переменных, которая объясняется вариацией других переменных, т.е. доля вариации агрессивности объясняется вариацией тревожности и фрустрации на 22%. Остальные 78% объясняются воздействиями других факторов.

Таблица 9

Дисперсионный анализ^b

| Модель | Сумма квадратов | ст.св. | Средний квадрат | F | Знч. |
|-------------|-----------------|--------|-----------------|--------|-------------------|
| 1 Регрессия | 1760,024 | 2 | 880,012 | 12,735 | ,000 ^а |
| Остаток | 6426,601 | 93 | 69,103 | | |
| Всего | 8186,625 | 95 | | | |

а. Предикторы: (конст) Фрустрация, Тревожность

б. Зависимая переменная: Агрессивность

В таблице 9 указано значение критерия F-Фишера равно 12,735 и его р-уровень значимости – 0,000. Это означает, что коэффициент множественной корреляции между зависимой и двумя независимыми переменными статистически значим и модель регрессии может быть содержательно интерпретирована.

Угловой коэффициент регрессии k и свободный член b находятся в последней таблице 10 в столбце «нестандартизированные коэффициенты».

Коэффициенты^а

| Модель | Нестандартизованные коэффициенты | | Стандартизованные коэффициенты | t | Знч. |
|---------------|----------------------------------|-------------|--------------------------------|-------|------|
| | B | Стд. Ошибка | Бета | | |
| 1 (Константа) | 19,432 | 4,893 | | 3,971 | ,000 |
| Тревожность | ,320 | ,146 | ,276 | 2,195 | ,031 |
| Фрустрация | ,287 | ,157 | ,229 | 1,824 | ,071 |

а. Зависимая переменная: Агрессивность

В таблице 10 содержатся стандартизированные коэффициенты регрессии (Бета) – 0,276 и 0,229, значения критерия t-Стьюдента (2,195 и 1,824) и уровни значимости (0,031 и 0,071). Регрессионный коэффициент, показывающий вклад фрустрации в изменчивость агрессивности статистически не значим – $p = 0,071 > 0,05$, следовательно, он может быть исключен из модели. Тогда линейное уравнение примет вид парной регрессии:

$$\langle \text{Агрессивность} \rangle = 0,320 \cdot \langle \text{Тревожность} \rangle + 19,432.$$

Стандартизированные коэффициенты служат показателями вклада двух переменных в регрессионную модель. Используя стандартизированные коэффициенты регрессии можно сопоставить влияние двух факторов «тревожность» и «фрустрация» на агрессивность спортсменов путем сравнения этих коэффициентов. $0,276 / 0,229 = 1,2$, следовательно, вес фактора тревожности выше в 1,2 раза фактора фрустрации в изменчивость показателей агрессивности спортсменов.

При регрессионном анализе следует помнить, что регрессионное уравнение, как и все показатели тесноты корреляционной связи, свидетельствует лишь о связи между изучаемыми признаками, проявляющейся в их совместном варьировании. Уравнение связи не может рассматриваться с точки зрения причинно-следственных отношений в отрыве от теоретических предположений, гипотез.

2.3. Эксплораторный факторный анализ

Факторный анализ – это группа методов, направленных на выявление, математическое выражение и определенную причинную интерпретацию латентных структур.

Факторный анализ возник и первоначально разрабатывался в задачах психологии. На основе факторно-аналитической концепции, заложенной выдающимся английским психологом Чарльзом Спирменом возник целый ряд психологических теорий: модели Г. Айзенка, Дж. Гилфорда, С. Барта, Р. Б. Кеттела, Л. Терстоуна и других исследователей. Позднее факторный анализ нашел применение при решении различных практических задач в социологии, медицине, экономике и т.д.

Считается, что причиной совместной изменчивости какого-либо множества переменных является некий скрытый фактор или латентная переменная. Например, общительный, отзывчивый и оптимистичный человек, который имеет широкий круг знакомых и друзей, и часто бывает импульсивным, психологи называют *экстравертом*. Специалиста, умеющего устанавливать контакт с собеседником, поддерживать разговор, умеющего слушать и выходить из контакта, можно охарактеризовать как обладающего *компетентностью* в общении.

При наличии статистически значимых линейных корреляционных связей между наблюдаемыми переменными предполагается, что они обусловлены влиянием какого-то одного латентного фактора или группы факторов. Другими словами, сильно коррелирующие переменные объединяются в факторы.

Факторный анализ делится на эксплораторный и конфирматорный. Эксплораторный или разведочный факторный анализ используется при поиске факторов, позволяющих описать исходную корреляционную матрицу. Конфирматорный или подтверждающий факторный анализ предназначен для проверки гипотезы о числе факторов и их нагрузках.

Несмотря на некоторую туманность самих постановок задач факторного анализа в более сложных случаях и критику данного метода со стороны математиков, факторный анализ в настоящее время развивается и популярен среди психологов, педагогов и других специалистов.

Основные идеи факторного анализа наиболее ясно можно представить при помощи геометрической интерпретации. Рассмотрим самый простой пример однофакторного анализа двух шкал тревожности по методике Спилберга (см. Табл.1 и Рис.1).

Таблица 1

Результаты по методике Спилберга

| № | Испытуемый | СТ | ЛТ |
|----|---------------|----|----|
| 1 | Александра М. | 35 | 36 |
| 2 | Анастасия В. | 40 | 39 |
| 3 | Анжелика К. | 39 | 47 |
| 4 | Виктория Ц. | 37 | 54 |
| 5 | Диана Ю. | 43 | 52 |
| 6 | Екатерина Б. | 43 | 54 |
| 7 | Елена В. | 38 | 34 |
| 8 | Елизавета К. | 37 | 42 |
| 9 | Кристина Б. | 44 | 51 |
| 10 | Марина В. | 42 | 44 |
| 11 | Марина С. | 45 | 47 |
| 12 | Марина Т. | 56 | 67 |
| 13 | Мария Ал. | 60 | 57 |
| 14 | Мария Ш. | 42 | 41 |

Коэффициент корреляции r -Пирсона, полученный между показателями шкал «Личностная тревожность» и «Ситуативная тревожность» равен 0,733 (см. Табл.2). Коэффициент статистически значим на уровне $p < 0,01$.

Таблица 2

Корреляционная матрица

| | | Ситуативная_тревожность | Личностная_тревожность |
|-------------------------|--------------------|-------------------------|------------------------|
| Ситуативная_тревожность | Корреляция Пирсона | 1 | ,733** |
| | Знч.(2-сторон) | | ,003 |
| | N | 14 | 14 |
| Личностная_тревожность | Корреляция Пирсона | ,733** | 1 |
| | Знч.(2-сторон) | ,003 | |
| | N | 14 | 14 |

** . Корреляция значима на уровне 0.01 (2-сторон.).

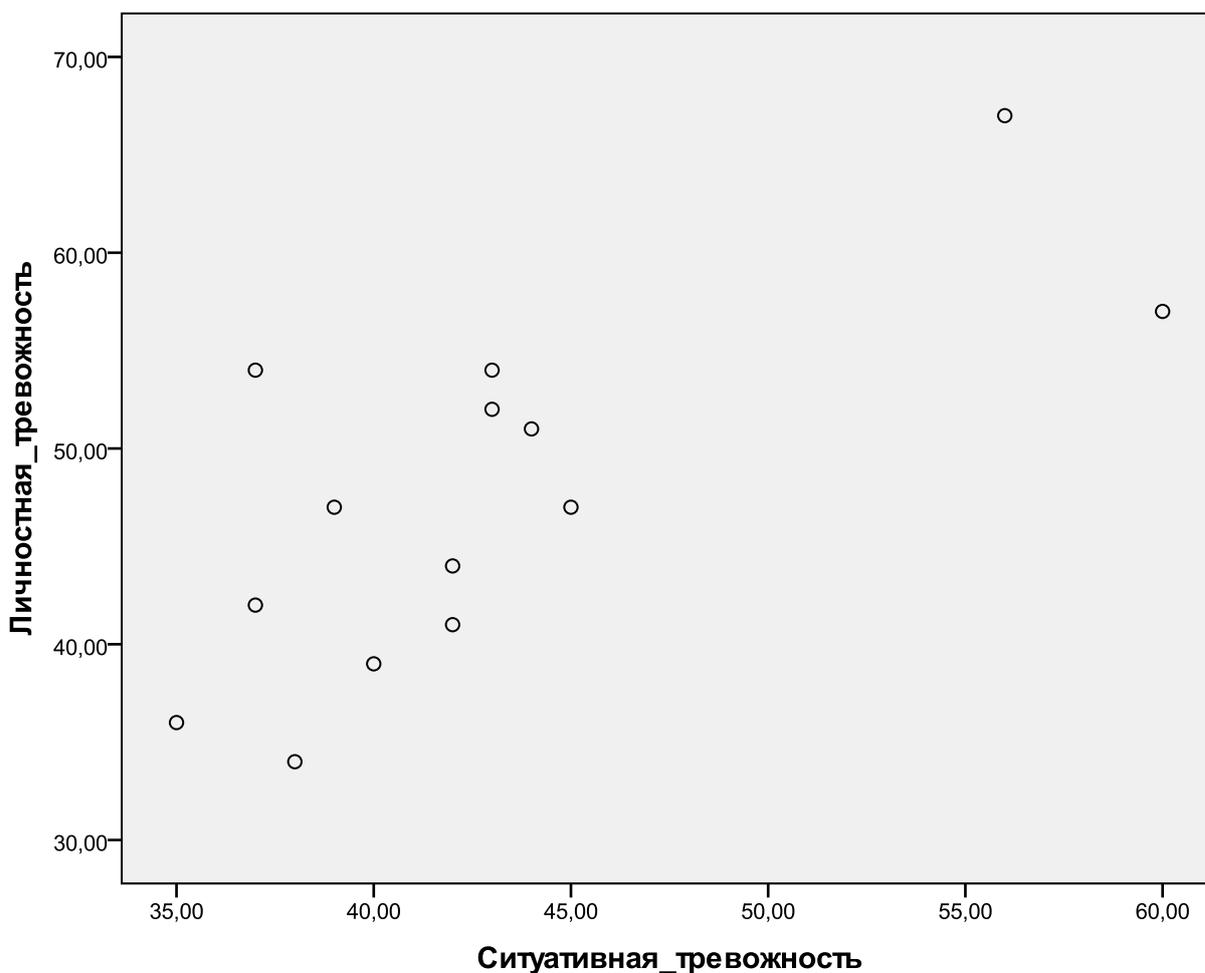


Рис.1. Связь личностной и ситуативной тревожности

В данном случае мы полагаем, что причиной совместной изменчивости личностной и ситуативной тревожности является один фактор, который можно просто определить как *тревожность*.

Исходным объектом для факторного анализа служит корреляционная матрица (см. Табл.5), но интерпретируется как конфигурация векторов, число которых равно числу скоррелированных переменных. Взаимное расположение векторов определяется значениями коэффициентами корреляции, пропорциональными косинусам углов между векторами:

$$r_{ij} = \cos \alpha_{ij}$$

Следовательно, зная коэффициент корреляции между двумя переменными, мы можем узнать угол между парой векторов максимально объясняющих расположение координат на плоскости, и наоборот (см. Рис.2).

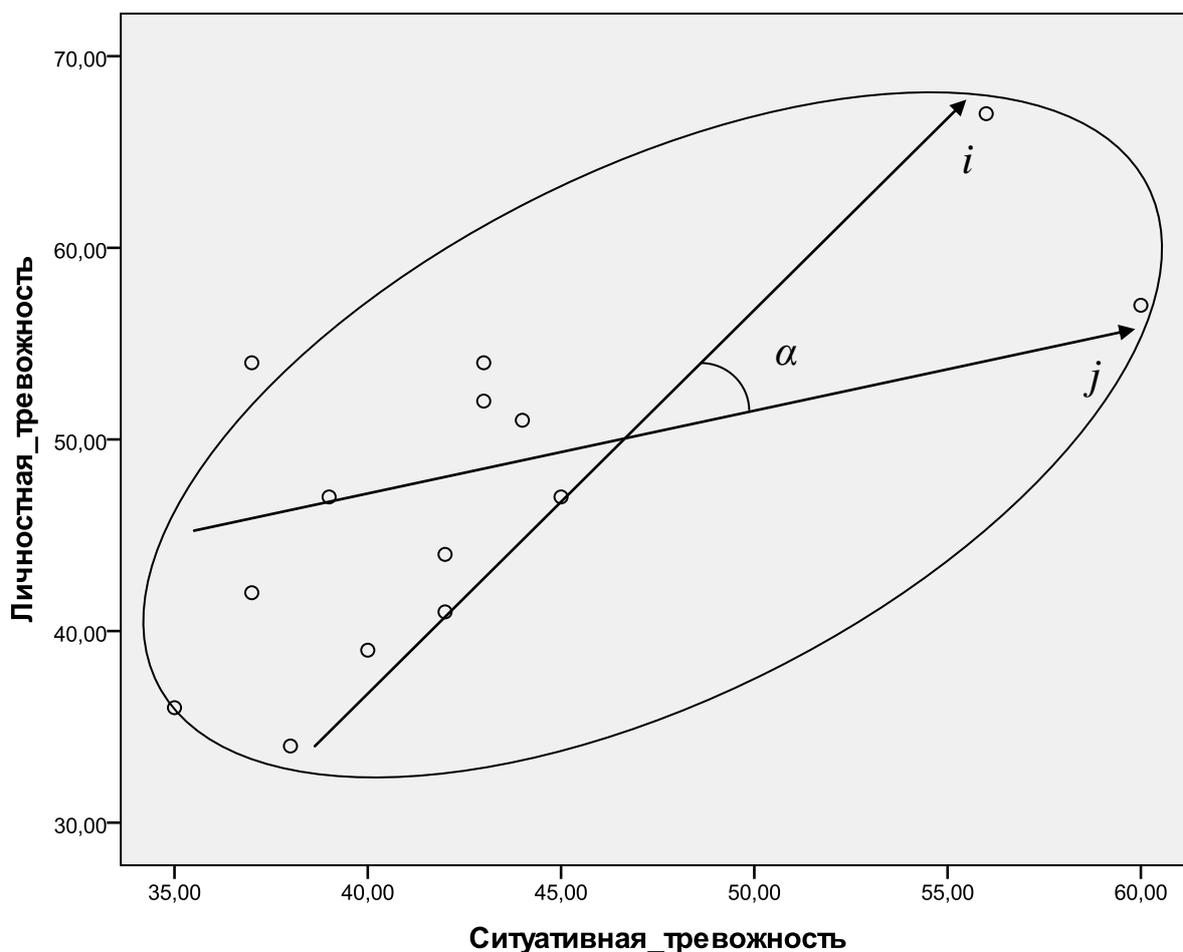


Рис.2. Конфигурация векторов

В нашем случае коэффициент корреляции $r = 0,733$, т.е. $0,733 = \cos \alpha_{ij}$, или $\alpha_{ij} = \arccos(0,733) = 0,748$. Величина $0,748$ представлена в радианах, поэтому чтобы получить угол в градусах: $\alpha_{ij} = 0,748 \cdot 57,3^\circ = 42,88^\circ$. Угол между векторами i и j равен приблизительно 43° .

Далее необходимо построить один вектор, хорошо объясняющий вышеуказанную корреляцию. Для этого создаются два вектора F_1 и F_2 , которые являются независимыми друг от друга, корреляция между которыми равна нулю (косинус угла в 90° равен нулю), и которые проходят через точку пересечения векторов i и j . В данном случае вектор F_1 принимает наиболее объяснительную и информативную форму, когда векторная ось представляет собой биссектрису угла между векторами i и j (см. Рис.3).

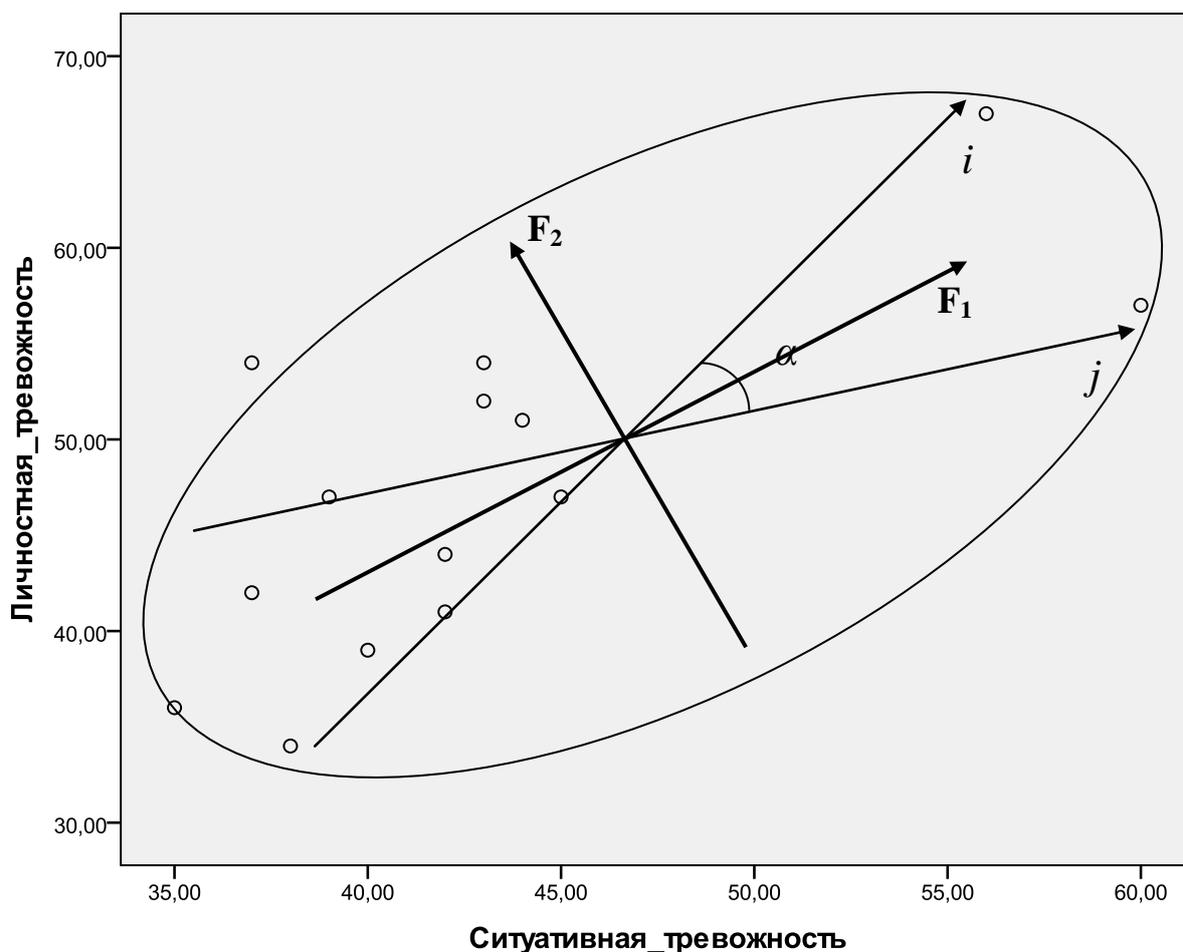


Рис.3. Конфигурация векторов

Так как векторная ось F_1 – биссектриса, то угол между вектором i и вектором F_1 $\alpha_1 = 42,88^\circ / 2 = 21,44^\circ$ и угол между вектором j и вектором F_1 $\alpha_2 = 42,88^\circ / 2 = 21,44^\circ$. Используя формулу $r_{ij} = \cos \alpha_{ij}$, найдем коэффициенты корреляции между вектором F_1 и векторами i и j : $\cos (21,44^\circ) = 0,931$.

Запишем полученный результат в виде факторной матрицы:

| Переменные | F1 |
|-------------------------|-------|
| Ситуативная тревожность | 0,931 |
| Личностная тревожность | 0,931 |

Коэффициенты корреляции между факторами и переменными в литературе называют *нагрузками*. Извлеченный фактор F_1 интерпретируется как тревожность. При наличии трех и более коррелирующих переменных для объяснения исходной матрицы одного фактора F_1 возможно было бы недостаточно.

Например, для трех переменных необходимо построение трехмерной факторной оси, для четырех – четырехмерной и т.д. Геометрическая интерпретация факторного анализа имеет ограничения, поэтому в реальных исследованиях поиск факторов осуществляется аналитически.

Факторный анализ имеет множество проблем, главной из которых является выделение и интерпретация главных факторов. Не существует однозначного критерия выделения факторов, и потому здесь неизбежен субъективизм исследователей в интерпретации результатов.

Покажем алгоритм извлечения факторов с помощью программы SPSS на более сложном примере.

Задача. Исследователя интересуют факторы, лежащие в основе структуры полученных им данных. В эксперименте принимало участие 96 спортсменов-легкоатлетов различной категории. Он получил об испытуемых следующие данные: возраст, стаж, разряд, показатели соперничества, агрессивности и конфликтности.

Полученные данные представлены на рис.4, где под цифрами 1,2,3... закодированы испытуемые. Так, испытуемый под номером 1 в возрасте 19 лет, имеет: спортивный стаж 10 лет, 3-й разряд, показатель соперничества – 10 баллов, агрессивности – 48 баллов, конфликтности – 56 баллов и т.д.



| | Возраст | Стаж | Разряд | Соперничество | Агрессивность | Конфликтность |
|---|---------|-------|--------|---------------|---------------|---------------|
| 1 | 19,00 | 10,00 | 3,00 | 10,00 | 48,00 | 56,00 |
| 2 | 24,00 | 13,00 | 1,00 | ,0 | 22,00 | 40,00 |
| 3 | 19,00 | 10,00 | 3,00 | 10,00 | 38,00 | 39,00 |
| 4 | 21,00 | 9,00 | 1,00 | 8,00 | 52,00 | 44,00 |
| 5 | 17,00 | 8,00 | 2,00 | 10,00 | 52,00 | 52,00 |
| 6 | 21,00 | 6,00 | ,0 | 8,00 | 46,00 | 44,00 |
| 7 | 17,00 | 6,00 | 3,00 | 12,00 | 48,00 | 60,00 |
| 8 | 20,00 | 4,00 | 2,00 | 3,00 | 46,00 | 50,00 |

Рис.4. Исходные данные

На начальном этапе проведем корреляционный анализ исходных данных. Отметим, что в факторном анализе и других методах многомерного анализа при выборе коэффициента корреляции, как правило, предпочтение отдается параметрической формуле r-Пирсона. Результаты корреляционного анализа отражены в таблице 3. Из таблицы видно, что соперничество, агрессивность и конфликтность попарно статистически достоверно связаны между собой положительными коэффициентами корреляции. Стаж спортсменов положительно коррелирует с их возрастом ($p < 0,01$) и разрядом ($p < 0,05$). Других статистически достоверных связей не выявлено.

Таблица 3

Корреляционная матрица

| | | Соперн-во | Агрес-ть | Конфл-ть | Возраст | Стаж | Разряд |
|---------------|-------------|-----------|----------|----------|---------|--------|--------|
| Соперничество | К.Пирсона | 1 | ,349** | ,358** | -,032 | -,032 | ,129 |
| | Знч.(2-ст.) | | ,000 | ,000 | ,754 | ,758 | ,209 |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |
| Агрессивность | К.Пирсона | ,349** | 1 | ,211* | -,050 | -,071 | -,001 |
| | Знч.(2-ст.) | ,000 | | ,039 | ,627 | ,494 | ,990 |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |
| Конфликтность | К.Пирсона | ,358** | ,211* | 1 | ,012 | -,009 | ,007 |
| | Знч.(2-ст.) | ,000 | ,039 | | ,908 | ,932 | ,948 |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |
| Возраст | К.Пирсона | -,032 | -,050 | ,012 | 1 | ,618** | ,014 |
| | Знч.(2-ст.) | ,754 | ,627 | ,908 | | ,000 | ,889 |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |
| Стаж | К.Пирсона | -,032 | -,071 | -,009 | ,618** | 1 | ,218* |
| | Знч.(2-ст.) | ,758 | ,494 | ,932 | ,000 | | ,033 |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |
| Разряд | К.Пирсона | ,129 | -,001 | ,007 | ,014 | ,218* | 1 |
| | Знч.(2-ст.) | ,209 | ,990 | ,948 | ,889 | ,033 | |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |

** . Корреляция значима на уровне 0.01 (2-сторон.).

* . Корреляция значима на уровне 0.05 (2-сторон.).

Исходная корреляционная матрица может быть представлена в виде графа (см. Рис.5):

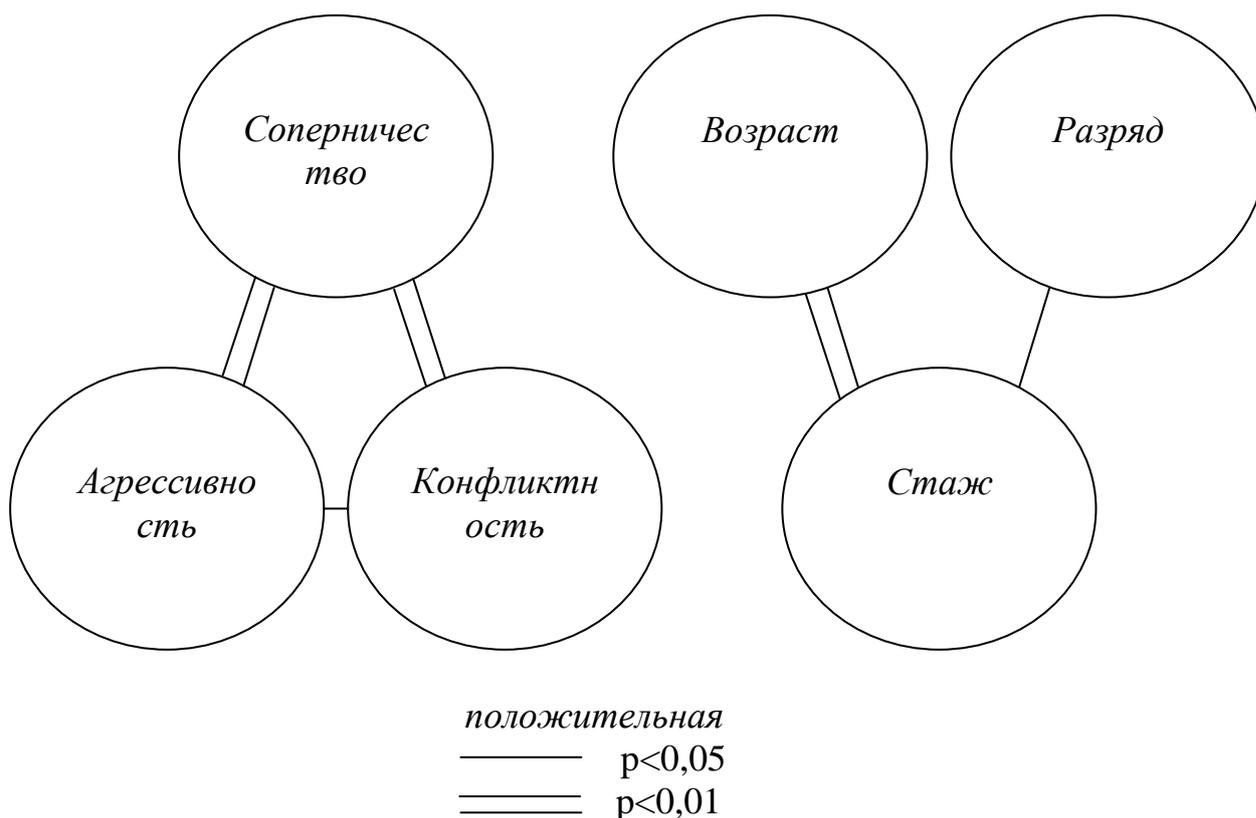


Рис.5. Корреляционный граф

Для запуска процедуры факторного анализа в SPSS выбираем в верхнем меню **Анализ** → **Снижение размерности** → **Факторный анализ...**

В появившемся (см. Рис.6) окне переносим из левой части окна в правую часть необходимые переменные.

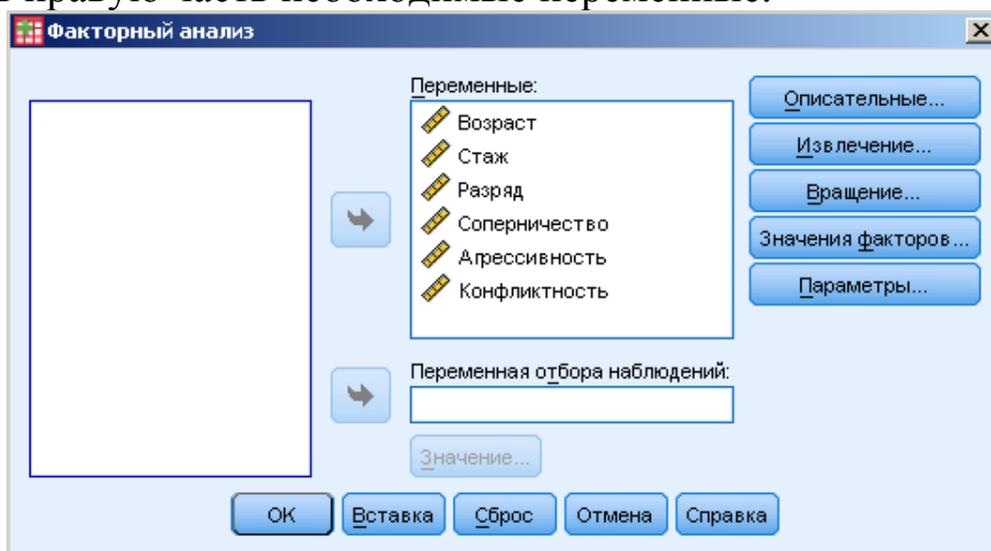


Рис.6. Ввод данных

Нажав на кнопку **Описательные...** (см. Рис.7) можно вывести в результаты дополнительные характеристики (среднее, стандартное отклонение, корреляционную матрицу и т.д.).

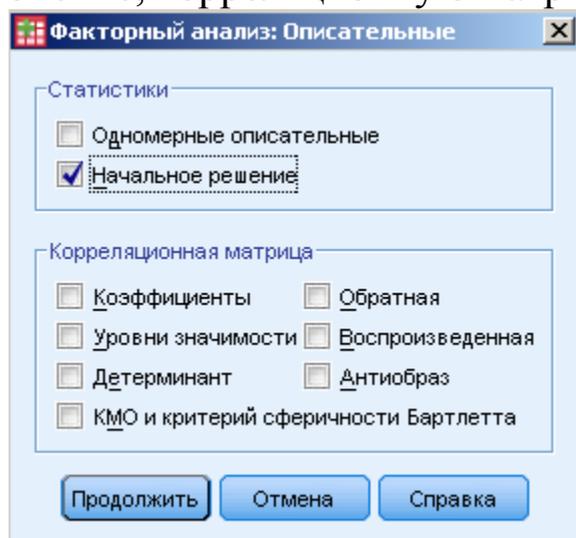


Рис.7. Описательные характеристики

Нажав на кнопку **Извлечение...** (см. Рис.8) можно выбрать метод получения факторов (по умолчанию указан метод главных компонент). Также можно самостоятельно указать необходимое число извлекаемых факторов (по умолчанию программа сама определит число факторов с помощью специальных критериев).

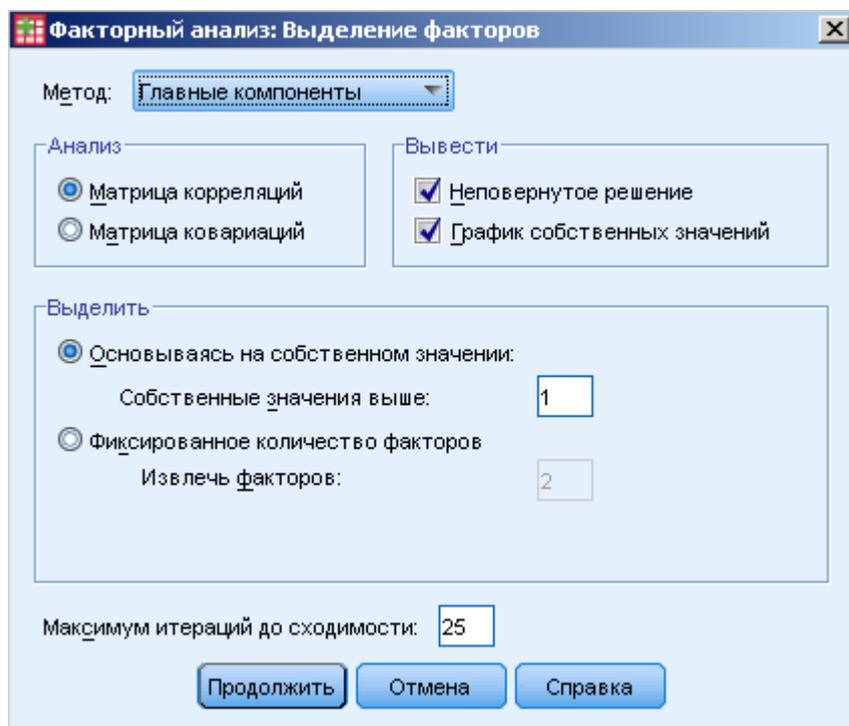


Рис.8. Извлечение факторов

Нажав на кнопку **Вращение...** (см. Рис.9) можно выбрать процедуру вращения факторов, т.е. способов расположения факторов максимально объясняющих исходные корреляции (по умолчанию вращение не указано). Также можно вывести график нагрузок (график нагрузок в SPSS наиболее ясно показывает двухфакторные решения).

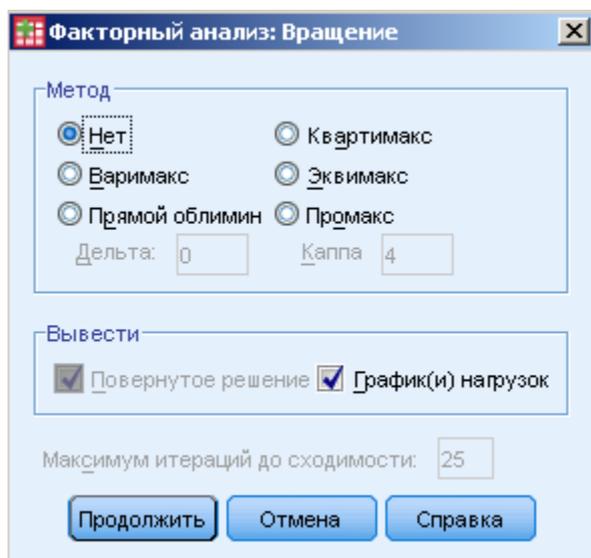


Рис.9. Вращение факторов

После нажатия кнопки **ОК** на выходе получим следующие результаты (см. Табл. 4-6 и Рис.10):

Таблица 4

| Общности | | |
|---------------|-----------|-------------|
| | Начальные | Извлеченные |
| Возраст | 1,000 | ,728 |
| Стаж | 1,000 | ,822 |
| Разряд | 1,000 | ,144 |
| Соперничество | 1,000 | ,656 |
| Агрессивность | 1,000 | ,470 |
| Конфликтность | 1,000 | ,478 |

Метод выделения: Анализ главных компонент.

В таблице 4 под общностью понимается часть вариации (дисперсии) переменной, объясняемая факторами (в нашем случае двумя).

Наибольшие общности имеют переменные «стаж», «возраст» и «соперничество». Минимальную общность с извлеченными

факторами имеет переменная «разряд», которая возможно связана с другими, не учтенными факторами.

Таблица 5

Полная объясненная дисперсия

| Компонента | Начальные собственные значения | | | Суммы квадратов нагрузок извлечения | | |
|------------|--------------------------------|-------------|----------------|-------------------------------------|-------------|----------------|
| | Итого | % Дисперсии | Кумулятивный % | Итого | % Дисперсии | Кумулятивный % |
| 1 | 1,701 | 28,355 | 28,355 | 1,701 | 28,355 | 28,355 |
| 2 | 1,595 | 26,590 | 54,944 | 1,595 | 26,590 | 54,944 |
| 3 | ,999 | 16,651 | 71,596 | | | |
| 4 | ,784 | 13,071 | 84,667 | | | |
| 5 | ,572 | 9,536 | 94,202 | | | |
| 6 | ,348 | 5,798 | 100,000 | | | |

Метод выделения: Анализ главных компонент.

Собственные значения это показатели, характеризующие объяснительную мощность факторов, причем собственные значения выше единицы означают, что существуют тесные корреляция между исходными данными. Из таблицы 5 видно, что первый фактор объясняет 28% вариации (дисперсии) всех переменных, второй фактор – 26% и т.д.

Таблица 6

Матрица компонент^а

| | Компонента | |
|---------------|------------|------|
| | 1 | 2 |
| Возраст | ,721 | ,456 |
| Стаж | ,759 | ,496 |
| Разряд | ,162 | ,342 |
| Соперничество | -,465 | ,663 |
| Агрессивность | -,473 | ,497 |
| Конфликтность | -,375 | ,581 |

Метод выделения: Анализ методом главных компонент.

а. Извлеченных компонент: 2

В таблице 6 показаны нагрузки по двум извлеченным компонентам или факторам. Первый фактор биполярный, т.к. имеет как положительные, так и отрицательные коэффициенты корреляции с исходными переменными. Второй фактор –

униполярный, т.е. имеет коэффициенты одного знака. Из таблицы 8 можно заключить, что извлеченные факторы плохо поддаются интерпретации, так как большинство переменных имеет значимые нагрузки (выше 0,4) по обоим компонентам.

На рисунке 10 график компонент представляет собой прямоугольную систему координат, где каждая переменная есть точка, а величины, определяющие ее положение – нагрузки по факторам.

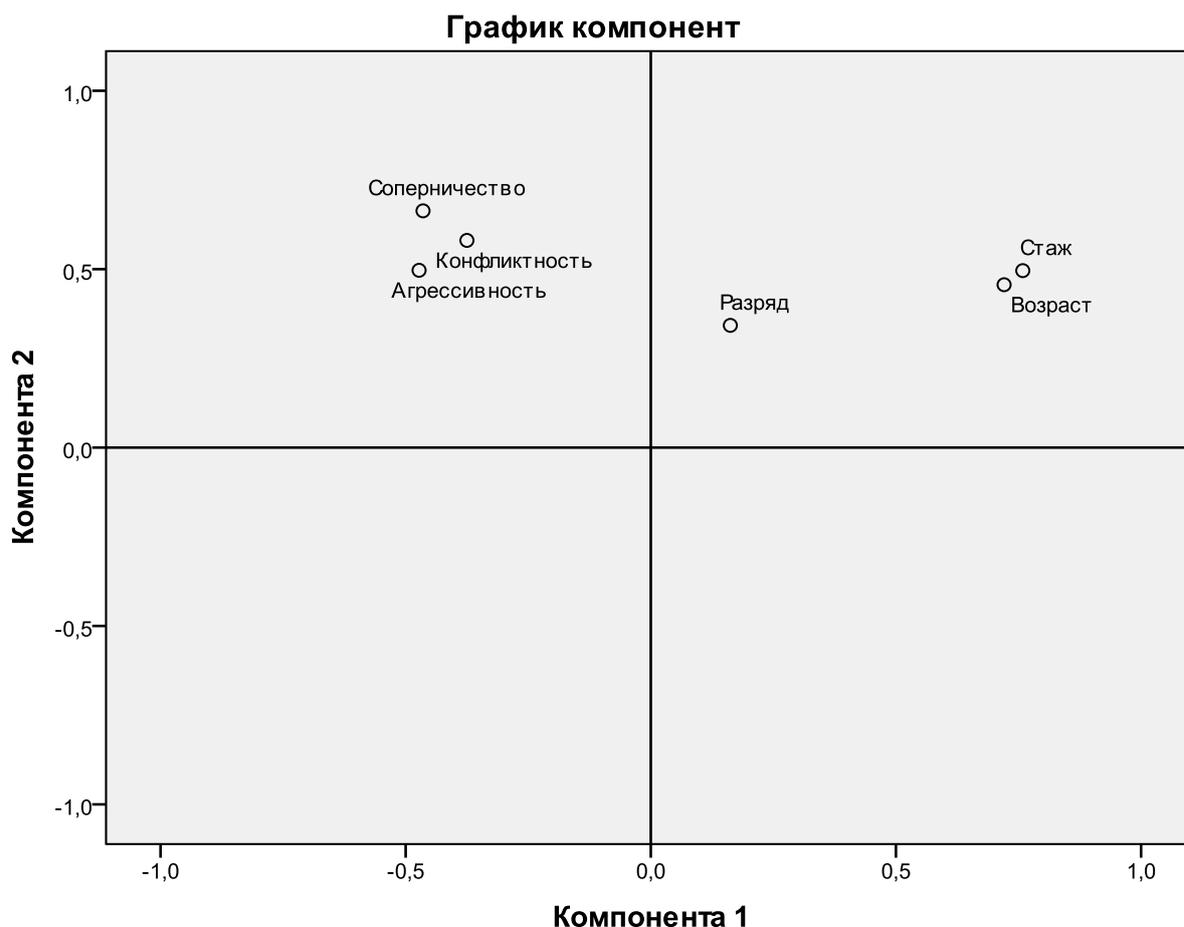


Рис.10. График компонент

На полученном выше рисунке видно, что два извлеченных фактора (горизонтальный и вертикальный) дает не очень хорошее объяснение корреляционной матрицы.

Для получения более ясного решения два фактора необходимо расположить, как показано на рисунке 11. Для этого необходимо осуществить процедуру вращения. В программе SPSS 18.0 доступно пять видов вращения: три ортогональных (варимакс, квартимакс, эквимакс), одно косоугольное (прямой

облимин) и одно комбинированное (промакс).

Ортогональный вращение методом варимакс – наиболее распространенный способ максимизации разброса квадратов нагрузок для каждого фактора, что приводит к увеличению больших и уменьшению малых значений факторных нагрузок. В результате простая структура получается для каждого фактора в отдельности.

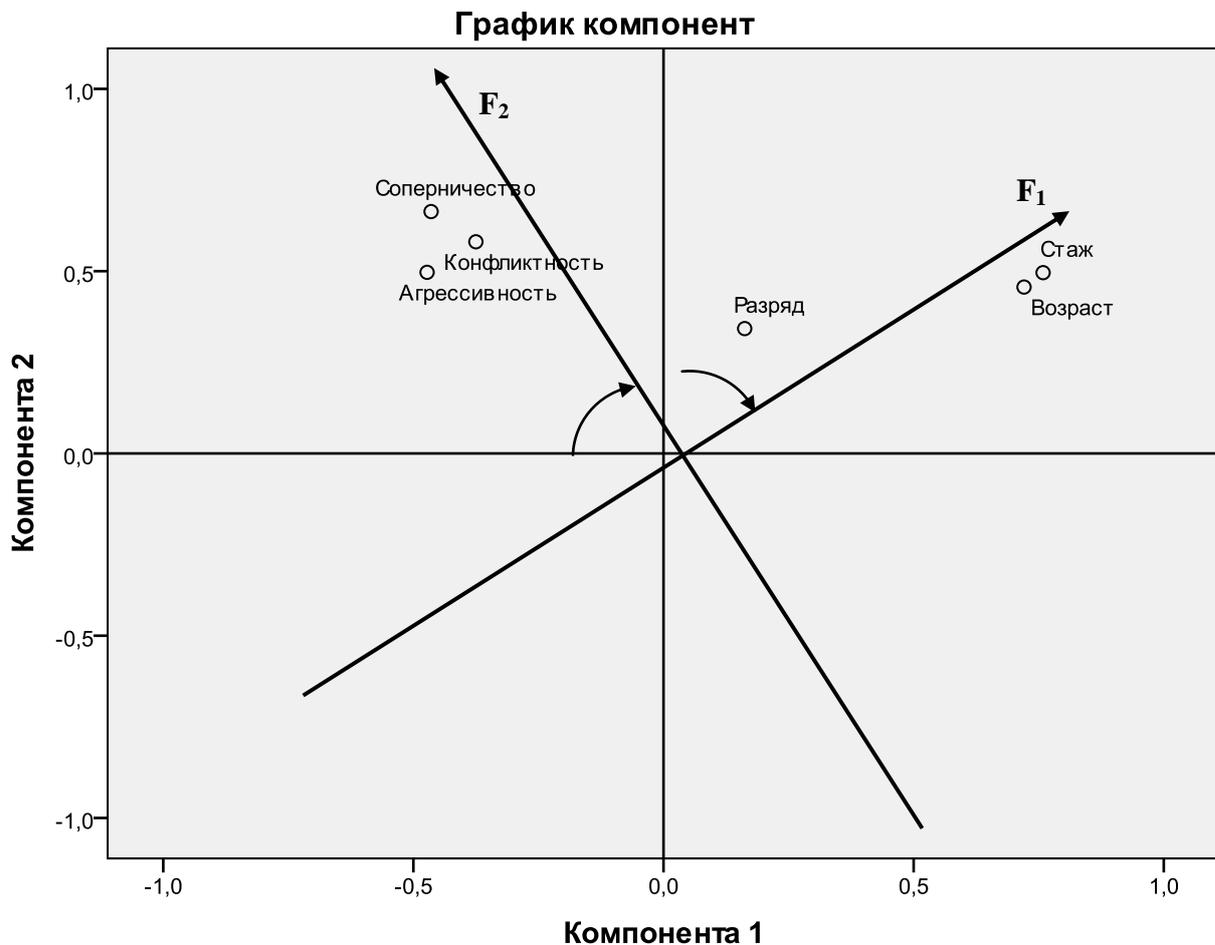


Рис.11. График вращения компонент

Повторим процедуру факторного анализа, как указывалось выше, и получим следующие результаты в таблицах 7,8 и на рисунке 12.

Таблица 7

Матрица повернутых компонент^a

| | Компонента | |
|---------------|-------------|-------------|
| | 1 | 2 |
| Возраст | ,848 | -,090 |
| Стаж | ,903 | -,083 |
| Разряд | ,340 | ,167 |
| Соперничество | ,048 | ,808 |
| Агрессивность | -,061 | ,683 |
| Конфликтность | ,067 | ,688 |

Метод выделения: Анализ методом главных компонент.

Метод вращения: Варимакс с нормализацией Кайзера.

а. Вращение сошлось за 3 итераций.

Процесс вращения компонент осуществлялся не произвольно как показано на рисунке 8, а путем умножения исходной матрицы компонент (см. Табл.6) на матрицу преобразования (см. Табл.8).

Таблица 8

Матрица преобразования компонент

| Компонента | 1 | 2 |
|------------|------|-------|
| 1 | ,783 | -,622 |
| 2 | ,622 | ,783 |

Метод выделения: Анализ методом главных компонент.

Метод вращения: Варимакс с нормализацией Кайзера.

В таблице 8 нагрузки по двум извлеченным компонентам после вращения изменились. Оба фактора униполярные (коэффициентами ниже 0,3 можно пренебречь). Первый фактор имеет максимальные нагрузки по переменным «возраст», «стаж» и «разряд», а второй фактор по переменным «соперничество», «агрессивность» и «конфликтность».

Первый фактор, имеющий высокие нагрузки по переменным «возраст», «стаж» и «разряд» можно интерпретировать как *спортивный опыт*, а второй фактор, имеющий высокие нагрузки по переменным «соперничество», «агрессивность» и «конфликтность» можно интерпретировать как *спортивное соперничество*.

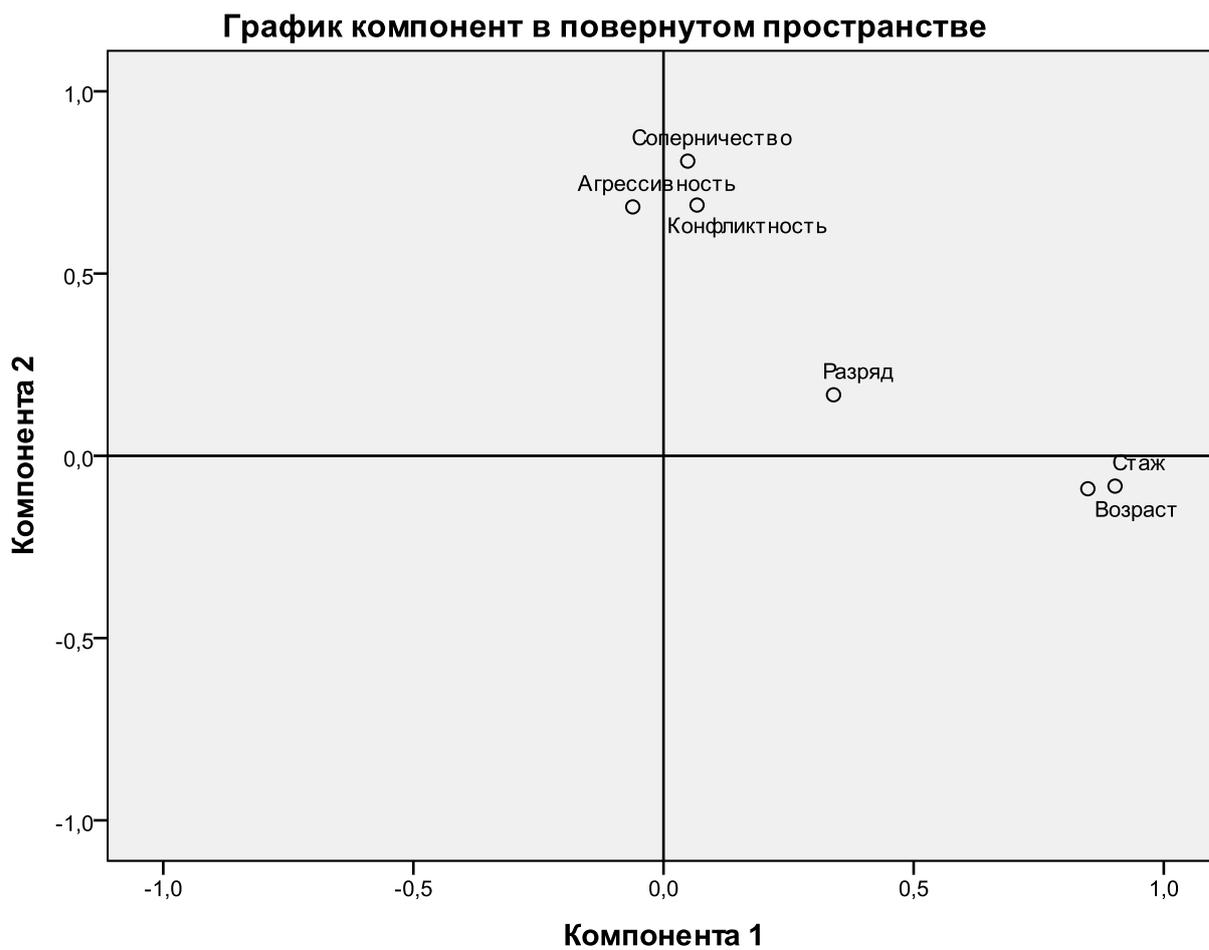


Рис.12. График компонент после вращения

Факторы F_1 и F_2 можно представить в виде формул:

«Спортивный опыт» = $0,903 \cdot$ «Стаж» + $0,848 \cdot$ «Возраст»;

«Спортивное соперничество» = $0,808 \cdot$ «Соперничество» + $0,688 \cdot$ «Конфликтность» + $0,683 \cdot$ «Агрессивность».

Таким образом, мы видим, что эксплораторный факторный анализ является эффективным инструментом классификации данных. Однако данный метод не стоит абсолютизировать и рекомендуется его проводить совместно с кластерным анализом или методом многомерного шкалирования (см. учебник А. Д. Наследова [7]) для получения наиболее обоснованной картины.

ЗАКЛЮЧЕНИЕ

Общий технический прогресс и бурная информатизация учебных и научно-исследовательских институтов существенно повлияли на развитие количественной методологии: разрабатываются и совершенствуются программы обработки данных, в том числе он-лайн в сети Интернет. В профессиональной деятельности психологов, педагогов и других специалистов, занимающихся анализом экспериментальных данных, процедура использования математических методов является важной составной частью психолого-педагогических исследований. Здесь особое внимание уделяется организации и процедуре сбора данных, а также совершенствованию диагностического инструментария, в том числе с помощью методов статистики.

Анализ данных, проводимый с помощью компьютера, освобождает от утомительных расчетов и доступен сейчас практически каждому. Исходя из этого, возрастает необходимость более осознанного подхода к обработке данных различными вычислительными методами и необходимость хотя бы на минимальном уровне знать, какие операции происходят с данными при их преобразовании. Другими словами, особую важность приобретает адекватность метода исходным данным.

Использование многомерных методов в сочетании с высококачественным программным обеспечением, имеющим интуитивно понятный интерфейс для статистического анализа данных, позволяет специалисту сконцентрироваться на наиболее важной стороне своей исследовательской работы – качественной, содержательной интерпретации полученных результатов.

ЛИТЕРАТУРА

1. Гусев, А. Н., Измайлов, Ч. А., Михалевская, М. Б. Измерение в психологии: общий психологический практикум. / 2-е изд. – М.: Смысл, 1998. – 286 с.
2. Ермолаев О.Ю. Математическая статистика для психологов. Учебник / О. Ю. Ермолаев – 2-е изд. испр. / О. Ю. Ермолаев. – М.: Московский психолого-социальный институт. – Флинта, 2003. – 336 с.
3. Митина О. В., Михайловская И. Б. Факторный анализ для психологов / О.В. Митина, И.Б. Михайловская. – М.: Учебно-методический коллектор «Психология», 2001. – 169 с.
4. Наследов А. Д. Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие / А. Д. Наследов. – СПб.: Речь, 2004. – 392 с.
5. Наследов А. Д. SPSS: Компьютерный анализ данных в психологии и социальных науках. 2-е изд. – СПб.: Питер, 2007. – 416 с.
6. Крамер Д. Математическая обработка данных в социальных науках: современные методы: учеб. для студ. высших учеб. заведений / Дункан Крамер; пер. с англ. И. В. Тимофеева, Я. И. Киселевой; науч. – М.: Издательский центр «Академия», 2007. – 288 с.
7. Остапенко Р. И. Математические основы психологии: учебно-методическое пособие для студентов и аспирантов психологических и педагогических специальностей вузов / Р. И. Остапенко. – Воронеж.: ВГПУ, 2010. – 76 с.
8. Суходольский Г. В. Математическая психология / Г. В. Суходольский. – Харьков.: Изд-во Гуманитарный центр, 2006. – 360 с.

ПРИЛОЖЕНИЯ

Таблица 1

Критические значения коэффициента корреляции r-Пирсона

| к = n-2 | р | | к = n-2 | р | |
|---------|------|------|---------|------|------|
| | 0,05 | 0,01 | | 0,05 | 0,01 |
| 5 | 0,75 | 0,87 | 23 | 0,40 | 0,51 |
| 6 | 0,71 | 0,83 | 24 | 0,39 | 0,50 |
| 7 | 0,67 | 0,8 | 25 | 0,38 | 0,49 |
| 8 | 0,63 | 0,77 | 26 | 0,37 | 0,48 |
| 9 | 0,60 | 0,74 | 27 | 0,37 | 0,47 |
| 10 | 0,58 | 0,71 | 28 | 0,36 | 0,46 |
| 11 | 0,55 | 0,68 | 29 | 0,36 | 0,46 |
| 12 | 0,53 | 0,66 | 30 | 0,35 | 0,45 |
| 13 | 0,51 | 0,64 | 38 | 0,33 | 0,42 |
| 14 | 0,50 | 0,62 | 40 | 0,30 | 0,39 |
| 15 | 0,48 | 0,61 | 45 | 0,29 | 0,37 |
| 16 | 0,47 | 0,59 | 50 | 0,27 | 0,35 |
| 17 | 0,46 | 0,58 | 60 | 0,25 | 0,33 |
| 18 | 0,44 | 0,56 | 70 | 0,23 | 0,3 |
| 19 | 0,43 | 0,55 | 80 | 0,22 | 0,28 |
| 20 | 0,42 | 0,54 | 90 | 0,21 | 0,27 |
| 21 | 0,41 | 0,53 | 100 | 0,20 | 0,25 |
| 22 | 0,40 | 0,52 | 125 | 0,17 | 0,23 |

Критические значения коэффициента корреляции r-Спирмена

| n | p | | n | p | | n | p | |
|----|------|------|----|------|------|----|------|------|
| | 0,05 | 0,01 | | 0,05 | 0,01 | | 0,05 | 0,01 |
| 5 | 0,94 | - | 17 | 0,48 | 0,62 | 29 | 0,37 | 0,48 |
| 6 | 0,85 | - | 18 | 0,47 | 0,60 | 30 | 0,36 | 0,47 |
| 7 | 0,78 | 0,94 | 19 | 0,46 | 0,58 | 31 | 0,36 | 0,46 |
| 8 | 0,72 | 0,88 | 20 | 0,45 | 0,57 | 32 | 0,36 | 0,45 |
| 9 | 0,68 | 0,83 | 21 | 0,44 | 0,56 | 33 | 0,34 | 0,45 |
| 10 | 0,64 | 0,79 | 22 | 0,43 | 0,54 | 34 | 0,34 | 0,44 |
| 11 | 0,61 | 0,76 | 23 | 0,42 | 0,53 | 35 | 0,33 | 0,43 |
| 12 | 0,58 | 0,73 | 24 | 0,41 | 0,52 | 36 | 0,33 | 0,43 |
| 13 | 0,56 | 0,70 | 25 | 0,49 | 0,51 | 37 | 0,33 | 0,43 |
| 14 | 0,54 | 0,68 | 26 | 0,39 | 0,50 | 38 | 0,32 | 0,41 |
| 15 | 0,52 | 0,66 | 27 | 0,38 | 0,49 | 39 | 0,32 | 0,41 |
| 16 | 0,50 | 0,64 | 28 | 0,38 | 0,48 | 40 | 0,31 | 0,40 |

Критические значения критерия t-Стьюдента

| k | p | | k | p | | k | p | |
|----|--------|--------|----|-------|-------|----|-------|-------|
| | 0,05 | 0,01 | | 0,05 | 0,01 | | 0,05 | 0,01 |
| 1 | 12,706 | 63,657 | 23 | 2,069 | 2,807 | 45 | 2,014 | 2,690 |
| 2 | 4,303 | 9,925 | 24 | 2,064 | 2,797 | 46 | 2,013 | 2,687 |
| 3 | 3,182 | 5,841 | 25 | 2,060 | 2,787 | 47 | 2,012 | 2,685 |
| 4 | 2,776 | 4,604 | 26 | 2,056 | 2,779 | 48 | 2,011 | 2,682 |
| 5 | 2,571 | 4,032 | 27 | 2,052 | 2,771 | 49 | 2,010 | 2,680 |
| 6 | 2,447 | 3,707 | 28 | 2,048 | 2,763 | 50 | 2,009 | 2,678 |
| 7 | 2,365 | 3,499 | 29 | 2,045 | 2,756 | 51 | 2,008 | 2,676 |
| 8 | 2,306 | 3,355 | 30 | 2,042 | 2,750 | 52 | 2,007 | 2,674 |
| 9 | 2,262 | 3,250 | 31 | 2,040 | 2,744 | 53 | 2,006 | 2,672 |
| 10 | 2,228 | 3,169 | 32 | 2,037 | 2,738 | 54 | 2,005 | 2,670 |
| 11 | 2,201 | 3,106 | 33 | 2,035 | 2,733 | 55 | 2,004 | 2,668 |
| 12 | 2,179 | 3,055 | 34 | 2,032 | 2,728 | 56 | 2,003 | 2,667 |
| 13 | 2,160 | 3,012 | 35 | 2,030 | 2,724 | 57 | 2,002 | 2,665 |
| 14 | 2,145 | 2,977 | 36 | 2,028 | 2,719 | 58 | 2,002 | 2,663 |
| 15 | 2,131 | 2,947 | 37 | 2,026 | 2,715 | 59 | 2,001 | 2,662 |
| 16 | 2,120 | 2,921 | 38 | 2,024 | 2,712 | 60 | 2,000 | 2,660 |
| 17 | 2,110 | 2,898 | 39 | 2,023 | 2,708 | 61 | 2,000 | 2,659 |
| 18 | 2,101 | 2,878 | 40 | 2,021 | 2,704 | 62 | 1,999 | 2,657 |
| 19 | 2,093 | 2,861 | 41 | 2,020 | 2,701 | 63 | 1,998 | 2,656 |
| 20 | 2,086 | 2,845 | 42 | 2,018 | 2,698 | 64 | 1,998 | 2,655 |
| 21 | 2,080 | 2,831 | 43 | 2,017 | 2,695 | 65 | 1,997 | 2,654 |
| 22 | 2,074 | 2,819 | 44 | 2,015 | 2,692 | 66 | 1,997 | 2,652 |

Таблица 4

«Сырые» данные для задачи на стр. 56

| № | Возраст | Стаж | Разряд | Соперничество | Агрессивность | Конфликтность |
|----|---------|------|--------|---------------|---------------|---------------|
| 1 | 19 | 10 | 5 | 11 | 48 | 56 |
| 2 | 24 | 13 | 2 | 1 | 22 | 40 |
| 3 | 19 | 10 | 5 | 11 | 38 | 39 |
| 4 | 21 | 9 | 2 | 9 | 52 | 44 |
| 5 | 17 | 8 | 3 | 11 | 52 | 52 |
| 6 | 21 | 6 | 1 | 9 | 46 | 44 |
| 7 | 17 | 6 | 5 | 13 | 48 | 60 |
| 8 | 20 | 4 | 3 | 4 | 46 | 50 |
| 9 | 18 | 5 | 1 | 7 | 44 | 35 |
| 10 | 20 | 7 | 5 | 8 | 50 | 44 |
| 11 | 17 | 9 | 2 | 9 | 48 | 50 |
| 12 | 23 | 16 | 2 | 6 | 46 | 44 |
| 13 | 22 | 5 | 3 | 11 | 64 | 43 |
| 14 | 21 | 5 | 5 | 5 | 30 | 31 |
| 15 | 20 | 11 | 5 | 5 | 44 | 50 |
| 16 | 19 | 9 | 1 | 9 | 72 | 46 |
| 17 | 19 | 6 | 2 | 5 | 42 | 54 |
| 18 | 17 | 5 | 1 | 5 | 30 | 52 |
| 19 | 17 | 7 | 3 | 5 | 34 | 36 |
| 20 | 19 | 5 | 1 | 2 | 32 | 43 |
| 21 | 19 | 2 | 1 | 10 | 42 | 44 |
| 22 | 20 | 10 | 1 | 12 | 30 | 45 |
| 23 | 18 | 3 | 3 | 8 | 54 | 53 |
| 24 | 21 | 9 | 1 | 1 | 36 | 33 |
| 25 | 20 | 5 | 1 | 6 | 52 | 54 |
| 26 | 17 | 2 | 1 | 10 | 50 | 42 |
| 27 | 21 | 7 | 3 | 6 | 46 | 47 |
| 28 | 21 | 12 | 1 | 5 | 44 | 59 |
| 29 | 17 | 6 | 2 | 1 | 44 | 44 |
| 30 | 18 | 8 | 1 | 3 | 46 | 40 |
| 31 | 17 | 5 | 1 | 3 | 34 | 56 |
| 32 | 19 | 8 | 2 | 5 | 58 | 51 |
| 33 | 23 | 10 | 5 | 10 | 48 | 49 |
| 34 | 19 | 8 | 5 | 7 | 46 | 39 |
| 35 | 17 | 2,5 | 3 | 3 | 34 | 47 |
| 36 | 18 | 10 | 5 | 6 | 38 | 45 |
| 37 | 20 | 5 | 2 | 3 | 46 | 47 |
| 38 | 14 | 5 | 2 | 10 | 50 | 40 |
| 39 | 17 | 5 | 3 | 6 | 40 | 47 |
| 40 | 19 | 5 | 2 | 4 | 34 | 37 |
| 41 | 22 | 9 | 3 | 7 | 46 | 43 |
| 42 | 21 | 11 | 5 | 3 | 40 | 41 |
| 43 | 17 | 3 | 1 | 2 | 48 | 40 |
| 44 | 21 | 7 | 4 | 2 | 48 | 50 |
| 45 | 24 | 5 | 1 | 6 | 40 | 45 |
| 46 | 22 | 10 | 3 | 5 | 48 | 42 |
| 47 | 19 | 5 | 1 | 1 | 34 | 44 |

| | | | | | | |
|----|----|-----|---|----|----|----|
| 48 | 25 | 16 | 2 | 2 | 32 | 32 |
| 49 | 19 | 10 | 2 | 2 | 38 | 43 |
| 50 | 17 | 7 | 2 | 10 | 54 | 50 |
| 51 | 20 | 7 | 2 | 8 | 52 | 52 |
| 52 | 21 | 8 | 2 | 8 | 52 | 54 |
| 53 | 22 | 7 | 3 | 7 | 54 | 49 |
| 54 | 21 | 12 | 1 | 9 | 68 | 51 |
| 55 | 21 | 8,5 | 2 | 4 | 40 | 38 |
| 56 | 21 | 11 | 2 | 10 | 32 | 62 |
| 57 | 18 | 5 | 1 | 9 | 38 | 48 |
| 58 | 20 | 7 | 3 | 9 | 28 | 46 |
| 59 | 19 | 7 | 3 | 11 | 32 | 63 |
| 60 | 19 | 3 | 1 | 3 | 54 | 52 |
| 61 | 20 | 5 | 1 | 3 | 50 | 40 |
| 62 | 18 | 4 | 1 | 2 | 42 | 39 |
| 63 | 20 | 10 | 1 | 7 | 38 | 47 |
| 64 | 21 | 4 | 1 | 5 | 46 | 47 |
| 65 | 17 | 3 | 1 | 7 | 48 | 38 |
| 66 | 19 | 3 | 1 | 5 | 62 | 46 |
| 67 | 20 | 5 | 1 | 9 | 38 | 41 |
| 68 | 19 | 8 | 3 | 2 | 40 | 36 |
| 69 | 18 | 7 | 3 | 3 | 38 | 39 |
| 70 | 18 | 6 | 2 | 4 | 34 | 40 |
| 71 | 20 | 7 | 1 | 10 | 46 | 53 |
| 72 | 19 | 10 | 2 | 4 | 38 | 51 |
| 73 | 21 | 10 | 2 | 7 | 42 | 47 |
| 74 | 23 | 1 | 1 | 9 | 40 | 50 |
| 75 | 21 | 13 | 3 | 3 | 38 | 44 |
| 76 | 18 | 10 | 2 | 2 | 42 | 53 |
| 77 | 17 | 5 | 4 | 4 | 32 | 30 |
| 78 | 19 | 9 | 3 | 6 | 56 | 44 |
| 79 | 22 | 12 | 2 | 5 | 54 | 42 |
| 80 | 22 | 12 | 2 | 9 | 48 | 32 |
| 81 | 18 | 9 | 3 | 5 | 36 | 26 |
| 82 | 17 | 2 | 1 | 7 | 48 | 52 |
| 83 | 21 | 9 | 1 | 2 | 38 | 40 |
| 84 | 19 | 9 | 3 | 4 | 42 | 45 |
| 85 | 17 | 4 | 1 | 3 | 40 | 32 |
| 86 | 18 | 2 | 1 | 2 | 28 | 43 |
| 87 | 22 | 12 | 1 | 4 | 34 | 48 |
| 88 | 19 | 7 | 1 | 4 | 24 | 39 |
| 89 | 30 | 22 | 2 | 6 | 44 | 47 |
| 90 | 14 | 3 | 2 | 7 | 64 | 34 |
| 91 | 20 | 6 | 2 | 6 | 50 | 48 |
| 92 | 18 | 4 | 1 | 10 | 42 | 50 |
| 93 | 17 | 4 | 3 | 8 | 50 | 57 |
| 94 | 17 | 8 | 2 | 4 | 42 | 47 |
| 95 | 21 | 9 | 1 | 8 | 52 | 43 |
| 96 | 20 | 8 | 3 | 8 | 50 | 56 |