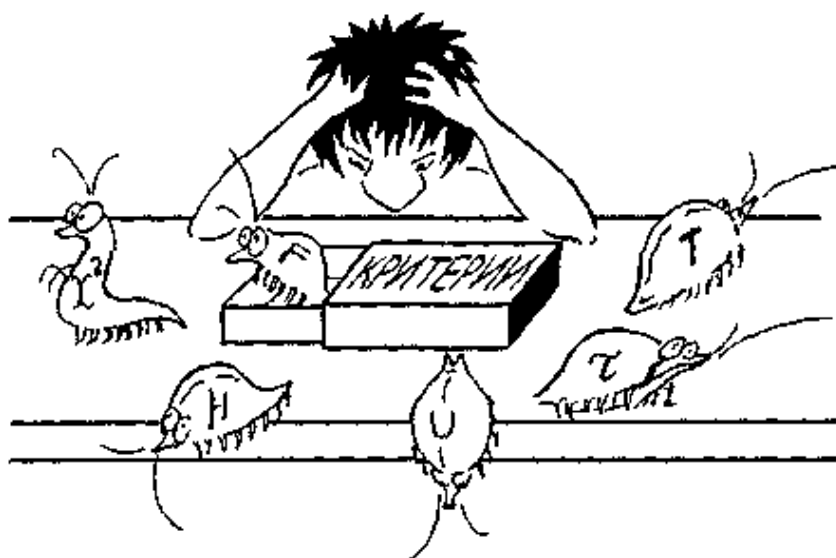


МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ, МОЛОДЕЖИ И СПОРТА
УКРАИНЫ
ОДЕССКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
имени И.И. МЕЧНИКОВА
ИНСТИТУТ МАТЕМАТИКИ, ЭКОНОМИКИ И МЕХАНИКИ

КАФЕДРА ОБЩЕЙ ПСИХОЛОГИИ И ПСИХОЛОГИИ РАЗВИТИЯ ЛИЧНОСТИ

Л.Н. АКИМОВА

ОСНОВЫ ПСИХОМЕТРИИ
МЕТОДИЧЕСКОЕ ПОСОБИЕ К КУРСУ



Одесса – 2012

Рекомендовано к печати Ученым советом ИМЭМ ОНУ имени И.И. Мечникова.
Протокол № 5 от 15.06.2012 г.

Рецензенты:

доктор психологических наук, профессор О.П. Санникова
кандидат физико-математических наук, доцент Г.М. Вартамян

**Акимова Л.Н. Основы психометрии. Методические указания к курсу / Л.Н. Акимова.
– Одесса, Студия «Негоциант», 2012. – 81 с.**

В методических указаниях к курсу «Основы психометрии» излагаются основные психометрические процедуры по разработке и адаптации психодиагностического инструментария. Описывается последовательность действий в статистических программах для решения основных психометрических задач.

Адресуется студентам старших курсов для выполнения лабораторных и зачетных работ по курсу «Основы психометрии», магистрам, аспирантам и психологам, по роду деятельности связанных с разработкой и адаптацией психодиагностических методик.

СОДЕРЖАНИЕ

Предисловие.....	4
Раздел 1. Основные психометрические процедуры.....	5
1.1. Последовательность стандартизации психологической методики.....	5
1.2. Методы оценки надежности.....	7
1.3. Методы оценки дискриминативности.....	10
1.4. Методы оценки валидности.....	12
1.4.1. Способы оценки валидности.....	19
Раздел 2. Документирование психодиагностических данных.....	24
2.1. Работа с программой Microsoft Excel.....	24
2.1.1. Общие положения.....	24
2.1.2. Виды курсора.....	25
2.1.3. Работа со списками, функции фильтра и сортировки.....	26
2.1.4. Создание и редактирование диаграммы в программе Microsoft Excel.....	27
Раздел 3. Статистические процедуры в психометрике.....	28
3.1. Ввод данных.....	28
3.1.2. Описательная статистика.....	29
3.1.3. Проверка нормальности распределения.....	32
3.1.4. Нормирование сырых баллов.....	34
3.1.5. Методы корреляционного анализа.....	38
3.1.6. Методы анализа надежности.....	40
3.1.7. Методы анализа выборок.....	41
3.1.8. Эксплораторный факторный анализ многих переменных в SPSS.....	44
Раздел 4. Содержание и структура лабораторных работ курса «Основы психометрии».....	46
Занятие 1.....	46
Занятие 2.....	46
Занятие 3.....	46
Занятие 4.....	46
Занятие 5.....	47
Занятие 6.....	47
Занятие 7.....	47
Занятие 8.....	48
Занятие 9.....	48
Занятие 10.....	48
Занятие 11.....	49
Литература.....	49
Основная.....	49
Электронные ресурсы.....	50
Дополнительная.....	50

Предисловие

Психометрика – это наука, лежащая в основе создания психодиагностических измерительных инструментов, а также в основе анализа качества и оптимизации существующих психологических тестов. За годы своего существования она выросла из занятия в основном академического в развитую индустрию, с обилием прикладных задач, измерительного инструментария (тестов) и хорошо обоснованными «правилами игры». Для эффективного развития практической психодиагностики сегодня требуется повышение психометрической культуры всех психологов, использующих измерительные психодиагностические методики. Методами рестандартизации теста должны владеть все психологи.

Целью курса «Основы психометрии» является ознакомление студентов с методами стандартизации и рестандартизации теста, простейшими приёмами проверки тестов на надежность и валидность, установления необъективных пунктов теста.

Задачами данного курса являются: ознакомление студентов с принципами организации, хранения и обработки психодиагностических данных, с основными методами статистического анализа, наиболее распространенными в прикладной и исследовательской работе психолога; формирование навыков работы с базами данных; планирование и осуществление этапов психометрической адаптации психодиагностических методик: проверки надежности и валидности отдельных пунктов теста, отсева ненадежных и невалидных пунктов теста, построение и анализ распределения тестовых баллов.

Данные методические указания предназначены для успешного усвоения дисциплины «Основы психометрии» студентами факультета психологии, состоят из четырех разделов и трех приложений. Первый раздел посвящен описанию содержания занятий, в остальных приведены методические рекомендации для ряда тем данного курса. В приложениях представлены материалы, необходимые для практической работы студентов.

Раздел 1. Основные психометрические процедуры

Основными психометрическими процедурами при разработке и адаптации психологических тестов являются:

1. Стандартизация и нормирование данных.
2. Оценка надежности шкалы теста.
3. Оценка дискриминативности шкалы теста.
4. Оценка валидности шкалы теста.

1.1. Последовательность стандартизации психологической методики

Последовательность стандартизации психологической методики выстраивается следующим образом:

1. Определяется генеральная совокупность, для которой предназначена методика.
2. Формируется репрезентативная выборка.
3. По результатам ответов испытуемых из выборки строится эмпирическое распределение.
4. Общий принцип построения стандартных шкал; разбиение выборки на группы, про которые известно, какой процент испытуемых из выборки они включают. Установление границ групп дает возможность отнести каждый индивидуальный результат в некоторую группу и точно установить, сколько процентов испытуемых имеют большую и меньшую степень выраженности психологического свойства по отношению к той группе, в которую попадает индивидуальный результат.
5. Пересчет «сырых» оценок в стандартные осуществляется различными способами в зависимости от того, является ли эмпирическое распределение нормальным.

Поэтому:

6. Определяют, является ли полученное распределение нормальным (проверяется статистическими критериями).
7. На нормальном распределении проводится стандартизация. Для этого в качестве точки отсчета выбирается нулевой результат, полученный после z-преобразования эмпирических значений, а границы групп стандартизации высчитываются на основе целочисленного ряда или ряда, кратного 2/3.

Или

8. Выполняется линейное преобразование z-оценок, при этом в качестве точки отсчета выбирается новое назначенное среднее значение, а границы групп стандартизации высчитываются на основе целочисленного ряда, кратного стандартному отклонению.
9. Если распределение отличается от нормального, то проводится нелинейная трансформация или процентильная стандартизация (Митина О.В., 2011).

Репрезентативность выборки – это представительность экспериментальной выборки, которая отражает основные свойства генеральной совокупности. Репрезентативность определяется путем сопоставления качественных и количественных характеристик экспериментальной выборки с характеристиками исследуемой популяции с целью определения их однородности.

К качественным характеристикам могут быть отнесены социально-демографические, видовые, интеллектуальные (образовательные), профессиональные и другие характеристики популяции (группы), которые являются предметом исследования или существенно влияют на него.

Формирование репрезентативной выборки по количественному признаку осуществляется при помощи теоремы П.А. Чебышева о вероятности ошибки репрезентативности, которая гласит: «С вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом числе независимых наблюдений выборочная средняя будет сколь угодно мало отличаться от генеральной средней», и далее (следствие 1 из теоремы П. А. Чебышева) «...что с вероятностью, как угодно близкой к единице («практически достоверно»), выборочная доля будет как угодно мало

отличаться от генеральной доли, если объем выборки достаточно велик» (Носс И.С., 2003, Карасев А.И., 1962).

Объем минимальной репрезентативной выборки определяется с помощью статистических таблиц больших чисел, по номограммам достаточно больших чисел, а также с помощью специальных расчетов.

Если средние показатели и дисперсия для генеральной и выборочной совокупностей известны, то для определения минимального объема выборки можно воспользоваться специальным расчетом. При этом вначале определяется желательный уровень точности измерения, выражаемый в долях измеряемой величины или процентах доверительной вероятности, например $\bar{x} \pm 0,01x$ или $P \pm 5\%$. Для непрерывно изменяющейся нормально распределенной величины объем выборки (n) будет определяться по формуле:

$$n = \frac{\left(\frac{t\sigma}{\Delta}\right)^2}{1 + \frac{1}{N}\left(\frac{t\sigma}{\Delta}\right)^2}, \quad \text{где} \quad (1.1)$$

n – объем минимальной репрезентативной выборки испытуемых;

N – объем генеральной совокупности;

t – значение абсциссы для кривой нормального распределения, определяемое желаемой точностью оценки или выбора (для $P=0,95$ $t=1,96$; $P=0,99$ $t=2,58$),

Δ – уровень точности в долях от среднего значения в выборке;

σ – величина стандартного отклонения.

В качестве примера определим величину выборки для генеральной совокупности 5000 человек: $\bar{x} = 10$, заданная ошибка не превышает 5% от среднего значения, вероятность попадания значений переменной за пределы доверительного интервала $\bar{x} \pm t\sigma$ не более 5% ($P=0,95$), стандартное отклонение $\sigma=2$.

$$n = \frac{\left(\frac{1,96 * 2}{0,05}\right)^2}{1 + \frac{1}{5000}\left(\frac{1,96 * 2}{0,05}\right)^2} \approx 2757$$

Формула расчета объема минимальной репрезентативной выборки при нахождении средней признака (интервальная измерительная шкала):

$$n = \frac{Nt^2\sigma^2}{N\sigma^2 + t^2\sigma^2} \quad (1.2)$$

Формула расчета объема минимальной репрезентативной выборки при нахождении доли элементов генеральной совокупности, обладающих некоторым признаком (номинальная измерительная шкала):

$$n = \frac{Nt^2pq}{N\Delta^2 + t^2pq}, \quad \text{где} \quad (1.3)$$

N – объем генеральной совокупности;

t – значение абсциссы для кривой нормального распределения, определяемое желаемой точностью оценки или выбора (для $P=0,95$ $t=1,96$; $P=0,99$ $t=2,58$),

Δ – уровень точности в долях от среднего значения в выборке, т.е. D – размер ошибки доверительного интервала измерения признаков;

σ – величина стандартного отклонения.

p – частота проявления данного признака; $q = 1 - p$.

Приведем пример расчета объема минимально-репрезентативной выборки испытуемых, если объем генеральной совокупности 5000 чел., а вероятность проявления исследуемого признака («угадывания» при помощи теста) – 0,8.

Дано: t – коэффициент кратности ошибки, связанный с вероятностью (P), с которой требуется гарантировать результаты выборочного наблюдения = 1,98;

Δ – размер ошибки доверительного интервала измерения признаков = 0,05;

N – объем генеральной совокупности = 5000;

p – частота проявления данного признака = 0,8; $q = 1 - p = 0,2$.

Определить: n – минимально необходимую численность экспериментальной выборки.

$$n = \frac{5000 * 1,98^2 * 0,8 * (1 - 0,8)}{5000 * 0,05^2 + 1,98^2 * 0,8 * (1 - 0,8)} \approx 239$$

Ответ: объем минимальной репрезентативной выборки составляет 239 человек.

Вывод: При повышении достоверности измерений происходит заметное уменьшение объема минимально репрезентативной выборки. Аналогичное влияние отказывает точность измерений.

1.2. Методы оценки надежности

Традиционно выделяют: надёжность по внутренней согласованности (к ней же относится надёжность эквивалентных половин теста), надёжность взаимозаменяемых форм, надёжность оценщика и, наконец, ретестовую надёжность (Анастаси А., 2001; Клайн П., 1994; Furr, R.M., 2008).

Распространено заблуждение, что при психометрической проверке теста достаточно вычислить какой-либо один показатель надёжности. Однако важно понимать, что перечисленные виды надёжности *не заменяют друг друга*. Они отличаются друг от друга различными источниками дисперсии ошибок. Поэтому, если того требуют особенности теста, он должен сопровождаться несколькими коэффициентами надёжности.

Надёжность как *внутренняя согласованность* (англ. – internal consistency) определяется связью каждого конкретного элемента теста с общим результатом, тем, насколько каждый элемент входит в противоречие с остальными, насколько каждый отдельный вопрос измеряет признак, на который направлен весь тест.

Для проверки внутренней согласованности применяются:

- метод расщепления или метод автономных частей
- метод эквивалентных бланков
- расщепление альфа Кронбаха, Кюдера – Ричардсона

Метод расщепления (Split-half reliability) заключается в разделении теста на две равные части (например, четные и нечетные вопросы, первая и вторая половина), а затем находится корреляция между ними. Для определения общей надёжности (гомогенности) теста полученные коэффициенты корреляции вводятся в формулу Спирмена – Брауна:

$$r = \frac{2r_{xy}}{1 + r_{xy}}, \text{ где} \tag{1.4}$$

r – корреляция «половин» теста.

Если корреляция высокая, тест можно считать надежным

Метод эквивалентных бланков (МЭБ) состоит в применении двух сопоставимых друг с другом форм теста для большой выборки (например, формы L и M для измерения в шкале интеллекта Стэнфорда – Бине). Результаты, полученные при выполнении двух форм, сравнивают и высчитывают корреляцию. Если коэффициент корреляции высокий, следовательно, тест надежен. Недостаток этого метода в том, что он подразумевает такой длительный и трудоемкий процесс, как создание двух эквивалентных форм.

Расщепление альфа Кронбаха (α Кронбаха) может быть рассмотрено как расширение расщепления Кьюдера – Ричардсона, применяемого в случае, когда пункты шкалы теста оцениваются дихотомически, т.е. принимают только два значения (например, ответы истинно/ложно). α Кронбаха теоретически связано с формулой прогнозирования Спирмана – Брауна. И обе эти формулы вытекают из классической теории тестов, заключающейся в том, что достоверность результатов тестирования может быть выражена как отношение дисперсий истинной и общей оценок (ошибки и истинной оценки).

Формула Кьюдера – Ричардсона (KR – 20)

$$KR = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k p_i q_i}{\sigma^2} \right), \quad \text{где} \quad (1.5)$$

p_i – доля лиц, ответивших на пункт i 1-ым вариантом ответа,

$q_i = 1 - p_i$,

k – количество пунктов теста,

σ^2 – дисперсия первичных баллов опросника.

В методе, предложенном Ли Кронбахом, сравнивается разброс каждого элемента с общим разбросом всей шкалы. Если разброс результатов теста меньше, чем разброс результатов для каждого отдельного вопроса, следовательно, каждый отдельный вопрос направлен на исследование одного и того же общего основания. Результат теста, полученный на основании таких вопросов можно считать истинным. Если такое истинное значение выработать нельзя, то есть получается случайный разброс при ответе на вопросы, тест ненадежен и коэффициент альфа Кронбаха будет равен 0. Если же все вопросы измеряют один и тот же признак, то тест надежен и коэффициент альфа Кронбаха в этом случае будет равен 1. Формула Кронбаха

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{S^2} \right), \quad \text{где} \quad (1.6)$$

k – количество пунктов теста;

s – дисперсия первичных баллов по пункту i ;

S – дисперсия итогового балла теста.

Коэффициент α Кронбаха лежит в пределах от 0 до 1.

$\alpha > 0,9$ – отличная

$\alpha > 0,8$ – хорошая

$\alpha > 0,7$ – приемлемая

$\alpha > 0,6$ – сомнительная

$\alpha > 0,5$ – малопригодная

$\alpha < 0,5$ – недопустимая

Коэффициент альфа может применяться и для решения другого типа задач. Так, с его помощью можно измерять степень согласованности экспертов, оценивающих тот или иной объект, стабильность данных при многократных измерениях и т.д.

Для проверки стабильности, устойчивости результатов диагностируемого свойства используется процедура «тест – ретест». Эта процедура заключается в проведении повторного опроса испытуемых по той же методике (ретеста). О стабильности признака судят по коэффициенту корреляции между результатами первого и второго обследования. Высокое значение свидетельствует о наличии синхронных изменений признака в выборке. На степень устойчивости, стабильности диагностируемого свойства могут влиять и многочисленные посторонние, ситуационные факторы. Поэтому важно соблюдать

требования единообразия процедуры проведения эксперимента в обоих случаях (Митина О.В., 2011).

Одной из причин невысокого показателя стабильности признака является то, что в период опроса этот признак находился в процессе интенсивного развития. В этом случае вычисленный низкий коэффициент корреляции может быть интерпретирован как свидетельство развития исследуемого свойства. Чтобы это установить, рекомендуется проверить выборку на однородность, и в случае отсутствия однородности попробовать выделить латентные группы. Например, одна часть испытуемых проходит путь развития в умеренном темпе, другая часть – развивается существенно быстрее, третья часть не развивается совсем. Каждую группу необходимо анализировать отдельно. Следовательно, получение низкого показателя стабильности может свидетельствовать не об отсутствии устойчивости свойства, а о наличии различных тенденций в его динамике.

Если исходя из теоретических соображений известно, что измеряемое свойство уже сформировано и должно быть достаточно устойчивым, то коэффициент корреляции между результатами теста и ретеста должен быть достаточно высоким (не ниже 0,80).

Независимость результатов от личности диагноста называется константностью. Константность как показатель надежности необходим для оценки проективных методик. Например, если в присутствии нового экспериментатора сдвиг произошел у всех испытуемых в одинаковой степени в одном направлении, то сам по себе этот факт на надежность методики не окажет влияния. Надежность изменится в том случае, когда воздействие экспериментатора на испытуемых различно: у одних результаты уменьшились, у других увеличились, у третьих остались примерно такими же, вследствие чего могут измениться порядковые места результатов испытуемых в выборке.

Показатель константности вычисляется как коэффициент корреляции результатов двух опытов, проведенных в относительно одинаковых условиях на одной и той же выборке испытуемых, но разными экспериментаторами: Коэффициент корреляции не должен быть ниже 0,80.

Таблица 1.1

Методы измерения надежности в зависимости от количества сеансов тестирования и форм теста (по О.В. Митиной)

Число сеансов тестирования	Число форм теста	
	одна	две
Один	Расщепление альфа Кронбаха, Кьюдера – Ричардсона	Метод взаимозаменяемых форм (непосредственный)
Два	Тест – ретест	Метод взаимозаменяемых форм (отсроченный)

Таблица 1.2

Методы измерения надежности и связываемые с ними факторы ее нарушения

Способы проверки надежности	Устанавливаемый фактор нарушения
Тест – ретест	Время
Метод взаимозаменяемых форм (непосредственный)	Содержание
Метод взаимозаменяемых форм (отсроченный)	Время и содержание
Расщепление альфа Кронбаха, Кьюдера – Ричардсона	Содержание
Смена диагностов	Диагносты

1.3. Методы оценки дискриминативности

Трудность задания в классической теории тестов определяется через соотношение количества испытуемых, справившихся с данным заданием, и общего количества испытуемых, т.е. трудность задания – это доля учащихся, которые справились с заданием.

Оценкой трудности задачи служит измеряемый в процентах индекс трудности.

$$I = \frac{T}{N} \cdot 100, \quad \text{где} \quad (1.7)$$

I – индекс трудности задания;

T – количество испытуемых, правильно решивших задачу;

N – общее количество испытуемых.

Из формулы видно, что чем выше показатель трудности, тем задание легче, и соответственно, чем меньше показатель трудности задания, тем задание сложнее. Например, если $p = 30\%$, то это значит, что только 30% испытуемых справились с этим заданием, а если $p = 70\%$, то 70% справилось с заданием, и получается, что первое задание сложнее, чем второе.

Показатель трудности очень важен для определения характеристики тестового задания и помогает проранжировать задания, входящие в тест по степени сложности. Благодаря этому можно определить место задания в тесте.

В правильно сконструированном тесте задания должны располагаться по нарастанию сложности, т.е. сначала даются самые легкие, далее все сложнее и сложнее. В хорошо сбалансированном по трудности тесте есть несколько самых трудных заданий со значением $p=20\%$ и несколько самых легких со значением $p=80\%$. Остальные задания по значениям p занимают промежуточное положение между крайними ситуациями и имеют в основном трудность 60-70% в критериально-ориентированном тесте и 40-60% в нормативно-ориентированном.

В рамках нормативно-ориентированного подхода наиболее удачными считаются задания средней трудности $p=q=0,5$, которые обеспечивают максимальную дисперсию теста.

Индекс трудности может измеряться и не в процентном отношении. В этом случае показателем трудности является величина

$$U_T = 1 - \frac{N_n}{N}, \quad \text{где} \quad (1.8)$$

N_n – число испытуемых, правильно решивших задачу; для личностных опросников – число испытуемых, ответивших на пункт в соответствии с ключом;

N – общее число испытуемых.

Чтобы учесть вероятность случайного успеха, достигнутого в результате угадывания ответа задачи с предписанным ответом, используют скорректированную формулу

$$U_T = 1 - \frac{N_n - \frac{N}{m-1}}{N}, \quad \text{где} \quad (1.9)$$

N_n – число испытуемых, неправильно решивших задачу;

m – число вариантов ответов.

Показатель U_T изменяется от нуля (простейший – все ответили правильно) до 1 (максимальная трудность – ни один испытуемый не дал правильного ответа).

При политомичных ответах формула вычисления трудности пункта приобретает следующий вид:

$$U_T = 1 - \frac{\sum_{i=1}^N [R_i(w_{\max} + w_{\min}) + (-1)^{R_i} w_i] - Nw_{\min}}{N(w_{\max} - w_{\min})}, \quad \text{где} \quad (1.10)$$

w_{\max} и w_{\min} – соответственно максимальный и минимальный балл выраженности согласия или числовой эквивалент выраженности;

w_i – балл по данному пункту, полученный i -м испытуемым;

$R_i = 0$, если i -й пункт прямой, $R_i = 1$, если i -й пункт обратный.

Дискриминативность – дифференцирующая, различающая способность теста в целом или отдельного тестового задания, указывающая на их способность разделять отдельных испытуемых по уровню выполнения. Таким образом, это способность отделять испытуемых с высоким общим баллом по тесту от тех, кто получил низкий балл, т.е. это способность отдельных пунктов (заданий) теста дифференцировать обследуемых относительно «максимального» или «минимального» результата теста.

Дискриминативность задания определяется обычно как разность между относительной численностью испытуемых, справившихся с заданием, из высокопродуктивной и низкопродуктивной группы, т.е. разностью индексов трудности заданий в высокопродуктивной и низкопродуктивной группах. Определенным функциональным синонимом дискриминативности является информативность.

Дискриминативность определяется подсчетом точечно-бисериального коэффициента корреляции (для дихотомических ответов) или подсчетом коэффициента корреляции ответов на пункт с общим показателем по шкале в целом. Коэффициент дискриминации может принимать значения от -1 до +1. Чем выше значение того или другого показателя, тем более дискриминативен пункт.

$$r_p b = \frac{\bar{x}_n - \bar{x}}{\sigma_x} \sqrt{\frac{N_n}{N - N_n}}, \quad \text{где} \quad (1.11)$$

\bar{x} – среднее арифметическое всех индивидуальных оценок по тесту;

\bar{x}_n – среднее арифметическое оценок по тесту у испытуемых, правильно выполнивших задание (в случае опросника личностного – соответствие с «ключом»);

σ_x – среднее квадратическое отклонение индивидуальных оценок по тесту для выборки;

N_n – число испытуемых, правильно решивших задачу (или тех, чей ответ на данный пункт опросника соответствует «ключу»);

N – общее число испытуемых.

Коэффициент дискриминативности задания также является, по сути, показателем критериальной валидности отдельного пункта, поскольку определяется по отношению к внешнему критерию – суммарному результату или оценкам продуктивности реальной деятельности испытуемых.

Индекс дискриминативности вычисляется с применением метода контрастных групп. Необходимым условием применения метода в этом случае является наличие близкого к нормальному распределению оценок по валидации критерия.

Доля членов контрастных групп может изменяться в широких пределах в зависимости от величины выборки. Чем больше выборка, тем меньшей долей испытуемых можно ограничиться при выделении групп с высоким и низким результатами. Нижняя граница «отсечения групп» составляет 10% от общего числа испытуемых в выборке, верхняя – 33%. Десятипроцентные группы берут редко, поскольку их малочисленность

снижает статистическую надежность индексов дискриминации. Чаще из выборки «извлекают» по 27 или 33% испытуемых.

Индекс дискриминации вычисляется как разность между долей лиц, правильно решивших задачу, из «высокопродуктивной» и «низкопродуктивной» групп и обозначается D :

$$D = \frac{N_{n_{\max}}}{N_{\max}} - \frac{N_{n_{\min}}}{N_{\min}}, \text{ где} \quad (1.12)$$

$N_{n_{\max}}$ – число испытуемых, правильно решивших задачу в высокопродуктивной группе; для личностных опросников – число испытуемых, ответивших на пункт в соответствии с ключом в высокопродуктивной группе;

N_{\max} – общее число испытуемых в высокопродуктивной группе;

$N_{n_{\min}}$ – число испытуемых, правильно решивших задачу в низкопродуктивной группе; для личностных опросников – число испытуемых ответивших на пункт в соответствии с ключом в низкопродуктивной группе;

N_{\min} – общее число испытуемых в низкопродуктивной группе.

Дискриминативность может определяться и методом Л. Фергюсона. Коэффициент Фергюсона (в формуле читается как дельта) – это отношение между показателем дискриминативности, полученным для некоторого теста и максимальным значением дискриминативности, которое может обеспечить такой тест.

$$\delta = \frac{(n+1) \left(N^2 - \sum_{i=1}^N F_i^2 \right)}{nN^2} \quad \text{где} \quad (1.13)$$

N – количество испытуемых,

n – количество заданий теста,

F_i – частота встречаемости показателя по пункту теста (показатель $F_i = T$ по формуле трудности заданий теста).

$\delta = 0$, когда все испытуемые получили одинаковые показатели, то есть когда нет дискриминативности. $\delta = 1$ при равномерном распределении заданий, в которых наиболее полно были реализованы все возможные проявления измеряемого свойства.

1.4. Методы оценки валидности

Надежность характеризует устойчивость результатов диагностической процедуры относительно объектов исследования (испытуемых) и является необходимым, но недостаточным условием того, что методика является хорошим измерительным инструментом. Устойчивость результатов относительно измеряемых свойств объектов характеризуется **валидностью** – другим ключевым критерием оценки качества методики. Понятие валидности связано с однозначностью, устойчивостью результатов относительно предмета измерения (конструкта). Валидность теста не может качественно и количественно превышать надежность (валидная, но ненадежная методика является практически бесполезной). В то же время вовсе не обязательно, что повышение надежности должно сопровождаться повышением валидности.

Количественно валидность теста может выражаться через корреляции результатов, полученных с его помощью, с другими показателями, например с успешностью выполнения соответствующей деятельности.

Валидность теста – понятие, указывающее, что тест измеряет и насколько хорошо он это делает (Анастаси, Урбина, 2001). Валидность включает сведения о том, пригодна ли

методика для измерения того показателя, ради которого она создавалась, а также меру ее эффективности. В стандартных требованиях к педагогическим и психологическим тестам (Стандарты тестирования, 1985) валидность определяется как обязательная и наиболее важная часть информации о методике, включающая описание того, относительно каких групп психологических свойств личности могут быть сделаны выводы (и их обоснования), опирающиеся на конкретные тестовые оценки или другие формы оценивания. Доказательство валидности психодиагностической методики также предполагает указание данных о степени согласованности результатов опроса с другими сведениями об исследуемом человеке, полученными из различных источников (теоретические ожидания, наблюдения, экспертные оценки, результаты других методик, достоверность которых установлена, и т.д.), суждение об обоснованности прогноза развития исследуемого качества, связь изучаемой области поведения и/или личностных особенностей с определенными психологическими конструктами. Валидность описывает также конкретную направленность методики (контингент испытуемых по возрасту, уровню образования, социально-культурной принадлежности и т.д.) и степень обоснованности выводов в контекстных условиях использования теста. В совокупности сведений, характеризующих валидность теста, должна содержаться информация об адекватности применяемой модели деятельности с точки зрения отражения в ней изучаемой психологической особенности, степени однородности заданий (субтестов), включенных в тест, их сопоставимости при количественной оценке результатов теста в целом. Определение области изучаемых свойств играет главную роль при выборе методики исследования и интерпретации ее данных. Содержащаяся в названии информация, как правило, недостаточна для суждения о сфере применения и является лишь составляющей валидности. Очерчивая сферу применения методики, валидность отражает и уровень обоснованности результатов измерения. Чем больше количество сопутствующих неучтенных факторов, которые могут влиять на результат исследования, чем более значительно их воздействие на результат, тем ниже достоверность показателей. В еще большей степени достоверность данных теста определяется набором измеряемых свойств, их значимостью для осуществления диагностируемой сложной деятельности, полнотой и существенностью отражения в материале методики предмета измерения. Так, для удовлетворения требованиям валидности диагностическая методика, предназначенная для практического применения, должна включать анализ широкого круга различных по своей природе показателей, наиболее важных для достижения успеха в данной деятельности.

Так как понятие валидности достаточно широкое и охватывает целый комплекс свойств, характеризующих ту или иную методику, не существует единого подхода к определению валидности. В зависимости от того, какую именно сторону валидности необходимо обосновать, выбираются различные способы этого обоснования, т.е. разные виды валидности. В руководствах и справочниках по психодиагностике и психометрике упоминаются многочисленные виды валидности. Рассмотрим некоторые из них.

Очевидная (внешняя, доверительная) валидность – другой вид представлений о тесте, сфере его применения, результативности и прогностической ценности, которые возникают у испытуемого или другого лица, не располагающего специальными сведениями о характере использования и целях методики. Эти представления испытуемых и пользователей психодиагностической информации, не имеющих психологического образования, формируются на основании названия методики и вводной информации на опросных бланках, выдаваемых испытуемым и наиболее доступных неспециалистам.

Хотя данный вид валидности и не является обязательным компонентом объективно устанавливаемой валидности, ее наличие достаточно часто является весьма желательным. Она выступает в качестве фактора, побуждающего испытуемых к обследованию, способствует более серьезному и ответственному отношению к работе по выполнению заданий теста и к заключениям, формулируемым психологом. Оценка очевидной валидности существенно улучшает применение понятных формулировок и терминов, а

также заданий, по содержанию являющихся наиболее естественными с учетом возрастной, половой, профессиональной специфики испытуемых.

Если содержание тестовых заданий представляется несерьезным, чрезмерно легким, не соответствующим сути изучаемой деятельности, это может привести к ироническому, негативному отношению испытуемых к ситуации обследования. С другой стороны, слишком высокая очевидная валидность возникает в результате переоценки возможностей методики, ее направленности и прогностической значимости и может вызвать избыточную мотивацию, нежелательную эмоциональную напряженность при обследовании, выраженное установочное поведение, что приводит к снижению других видов валидности. Для достижения оптимального уровня очевидной валидности рекомендуется вводить в инструкцию краткую характеристику целей исследования, избегать общих названий, которые можно неправильно истолковать (например: «Тест личности», «Тест умственных способностей», «Тест оценки достижений» и т.д.).

В современной психометрии выделяются три основных вида валидности: 1) содержательная (логическая); 2) эмпирическая и 3) концептуальная (В.М. Мельников, Л.Т. Ямпольский, 1985).

Содержательная (логическая) валидность (content validation). Один из основных типов валидности методики, характеризующий степень репрезентативности содержания пунктов измеряемой области психических свойств. Валидность по содержанию закладывается уже при подборе пунктов будущей методики. Разработке конкретного содержания пунктов методики предшествуют максимально полная систематическая проверка соответствующей литературы, а также консультации со специалистами по данному предмету. На основе собранной таким путем информации составляется спецификация методики, где указываются тестируемые области, стоящие за ними психологические свойства и характеристики, а также их относительное значение для описания конструкта на данном этапе. Конкретные пункты оцениваются по принципу их близости к реальным требованиям, методика в целом проверяется на репрезентативность исследуемой области.

Очевидная внешняя валидность может находиться в разных отношениях с содержательной валидностью. В одних случаях содержательная и внешняя валидность совпадают, в других внешняя валидность используется для маскировки содержательной валидности. С этой целью содержанию заданий и инструкции придается специальная форма, маскирующая истинную цель исследования. Тем самым предупреждается возникновение установочного поведения при тестировании.

В отличие от очевидной валидности содержательная валидность всегда связана с оценкой отношения содержания тестового материала к истинной, а не мнимой цели измерения. Для определения содержательной валидности используются экспертные методы. Как и во всякой экспертизе, необходимо решить следующие вопросы: отобрать компетентных экспертов, сформировать экспертную группу, организовать процедуру экспертизы и т. п. Специфичным является объект экспертизы – содержание теста. Эксперты должны оценить содержание заданий теста по их соответствию психическому свойству, объявляемому в качестве содержания валидизируемого теста. С этой целью экспертам предъявляются спецификация к тесту и список заданий. Если конкретное задание полностью соответствует спецификации, то эксперт обозначает его как соответствующее содержанию теста. В противном случае он бракует задание и письменно обосновывает свою оценку. Если несколько экспертов бракуют одни и те же задания, они признаются несоответствующими спецификации к тесту. Если между экспертами существуют расхождения в оценке заданий, их анонимно знакомят с письменными объяснениями экспертов, забраковавших задания, после чего экспертиза повторяется.

Содержательная валидность является одной из важных форм валидации критериально-ориентированных тестов и тестов достижений, а также методик,

предназначенных для профотбора, анализа успешности овладения профессией. Личностные опросники, а также тесты интеллекта связаны с проявлением более индивидуализированных форм деятельности, не обладают тем сходством с исследуемой областью поведения, как это имеет место в случае тестов достижений, и позволяют лишь косвенно судить о реальной деятельности испытуемого. Поэтому данный вид валидности имеет ограниченное применение и используется лишь на начальных стадиях составления методики.

Эмпирическая валидность (empirical validity). Идея эмпирической валидности заключается в определении способности теста служить индикатором или предсказателем строго определенной психической особенности или формы поведения человека. Для измерения этого свойства теста вычисляется коэффициент корреляции тестового результата с внешним критерием – $R(X, K)$. Величина корреляции между показателем теста и мерой критерия отражает коэффициент валидности. В качестве коэффициента валидации могут использоваться: линейный коэффициент корреляции (Пирсона), ранговые коэффициенты корреляции Спирмена, Кендала, фи-коэффициент (ϕ) сопряженности (результат применения корреляции Пирсона к двум бинарным переменным).

Коэффициент валидности должен быть статистически значим. Низким признается коэффициент валидности порядка 0,20–0,30, средним – 0,30–0,50 и высоким – свыше 0,60. Значимость коэффициента валидности зависит от объема выборки или контрольных групп (при методе контрастных групп) или от количества степеней свободы (df).

В качестве критерия может выступать любой показатель, независимо и бесспорно измеряющий ту же психологическую характеристику, что и валидируемый тест. Вопрос о выборе критерия является в этом виде валидности основным. От выбора критерия зависят качественная и количественная оценки валидности. Условно можно выделить три группы критериев: экспертные, экспериментальные и «жизненные».

Экспертный критерий предполагает использование экспертных оценок для независимого измерения свойств личности. Экспериментальный – использование результатов одновременного и независимого тестирования испытуемых другим тестом, предположительно измеряющим то же свойство личности, что и валидируемый тест. Коэффициент корреляции между результатами двух измерений называется эмпирической взаимной валидностью. Его величина зависит как от степени совпадения содержания измерения, так и от надежности тестов. Поэтому максимальные коэффициенты эмпирической взаимной валидности имеют параллельные тесты.

Эмпирическая валидность теста – это всегда валидность по отношению к конкретному критерию. Поэтому процесс конструирования теста часто сопровождается поиском подходящего критерия. Если к моменту проверки валидности такого экспериментального критерия не оказывается, в качестве него используются характеристики реального поведения, которые на основе существующих психологических теорий предполагаются связанными с измеряемым свойством. Например, в качестве *жизненных критериев* для тестов интеллекта используются показатели успешности обучения, экстравертированности – успешность административной деятельности, тревожности – частота заболеваний неврозом и т. п.

Приведем пример критериев:

- успеваемость (для тестов способностей к обучению, тестов достижений, тестов интеллекта);
- производственные достижения (для методик профессиональной направленности);
- эффективность реальной деятельности – рисования, моделирования и т.д. (для тестов специальных способностей);
- субъективные оценки (для тестов личности);

- объективные социально-демографические и биографические данные (стаж, образование, профессия, успешность реализации в профессиональной и иной деятельности);
- заключения специалистов о результатах обследований и экспертиз (например, врачебный диагноз и т.п.);
- результаты экзаменов;
- данные других методик и тестов, валидность которых считается установленной.

Внешний критерий должен отвечать трем основным требованиям – он должен быть **релевантным, свободным от помех (контаминации) и надежным.**

Под **релевантностью** имеется в виду, что в критерии задействованы именно те особенности индивидуальной психики, которые измеряются данной диагностической методикой. Внешний критерий и диагностическая методика должны находиться между собой во внутреннем смысловом соответствии, быть качественно однородными по психологической сущности (Гуревич, 1985). Например, для теста, измеряющего индивидуальные особенности мышления, умение выполнять логические действия с определенными объектами, понятиями, нужно искать критерии проявления именно этих умений. Профессиональная деятельность, как правило, имеет различные цели и задачи со своей спецификой и условиями выполнения. Поэтому не следует проводить сопоставление успешности по диагностическим методикам с производственной эффективностью в целом. Необходимо найти именно соотносимый с методикой критерий. Не имеет смысла пользоваться критерием, про релевантность которого ничего не известно.

Требования **свободы от контаминации** обусловлены тем, что успешность в реальной деятельности зависит не только от индивидуальных особенностей самого человека, но и от факторов, обусловленных ситуацией и внешним влиянием среды. Чтобы в какой-то мере избежать этого, следует отбирать для исследования такие группы людей, которые находятся в более или менее одинаковых условиях. Можно также корректировать помехи влияния среды с помощью статистических методов. Например, производительность определять не по абсолютным значениям, а в отношении к средней производительности рабочих, работающих в аналогичных условиях.

Когда говорят, что критерий должен иметь статистически достоверную надежность, это означает, что он должен отражать постоянство и устойчивость исследуемой функции.

Поиски адекватного и легко выявляемого критерия относятся к очень важным и сложным задачам валидации. Выбор должен иметь психологическое обоснование. Широта и степень общности критерия прямо зависят от сложности и комплексности конструкта, измеряемого валидируемой методикой. Более гомогенные и узкие конструкты требуют более простых и частных критериев.

Валидация по какому-либо критерию наиболее эффективно осуществляется методом контрастных групп.

Эмпирическая валидность включает как комплекс связей с текущим состоянием изучаемого свойства, так и вероятность и обоснованность прогноза на будущее. Таким образом, эмпирическая валидность охватывает и текущую, и прогностическую валидности.

Следует помнить, что если при сопоставлении с критерием получен значимый показатель валидности, то можно считать, что с той или иной степенью достоверности задача решена. Отсутствие корреляции может быть следствием как невалидности самой методики, так и ошибочности гипотезы о наличии причинно-следственной связи между диагностируемым конструктом и критериальным показателем.

Концептуальная валидность (construct validity). Она устанавливается путем доказательства правильности теоретических концепций, положенных в основу теста. Обращение к концептуальной валидности необходимо в тех случаях, когда результаты тестовых измерений используются не просто для предсказания поведения, а как основа

для выводов о том, в какой степени испытуемые обладают некоторой определенной психологической характеристикой. Причем измеряемая психологическая характеристика не может быть отождествлена с какой-либо наблюдаемой особенностью поведения, а представляет собой теоретическую концепцию. Примерами таких концепций являются интеллект, черты личности, мотивы, установки и т. д. Концептуальная валидность имеет значение также при разработке принципиально новых тестов, для которых не определены внешние критерии валидности. По существу, проверка концептуальной валидности теста проходит через три основных этапа:

1. Определяется некоторая теоретическая концепция, которая предположительно объясняет выполнение валидизируемого теста.

2. Из теоретической концепции выводятся одна или несколько гипотез, связанных с валидизируемым тестом.

3. Выдвинутые гипотезы подвергаются эмпирической проверке.

Если эмпирические данные подтверждают гипотезу, то тем самым подтверждается психологическая концепция, положенная в основу теста, и способность теста служить инструментом измерения этой концепции. Чем убедительнее эмпирическое подтверждение, тем определеннее можно говорить о валидности теста по отношению к психологической концепции, положенной в его основу. Если эмпирические данные не подтверждают выдвинутых гипотез, то ошибка могла возникнуть из-за:

- а) неправильности теоретической концепции, положенной в основу теста;
- б) отсутствия соответствия между тестом и теоретической концепцией;
- в) ошибок при выдвижении гипотез, относящихся к квалифицируемому тесту.

В отличие от эмпирической валидности, для которой обычно имеется единственный внешний критерий, при оценке концептуальной валидности проверка подвергаются по возможности все гипотезы, вытекающие из теоретической концепции, лежащей в основе валидизируемого теста.

На практике для определения концептуальной валидности широко используется предложенная Д. Т. Кэмпбеллом и Д. В. Фиске процедура оценки «конвергентной и дискриминантной валидности» [3, 69]. Эта процедура предполагает использование наряду с валидизируемым тестом специальной многомерной батареи контрольных тестов, подобранной таким образом, чтобы в нее входили как тесты, предположительно связанные с валидизируемым тестом, так и не связанные с ним. Экспериментатор должен заранее предсказать, какие тесты будут высоко коррелировать с валидизируемым тестом, а с какими тестами корреляция будет низкой. Тесты, которые, по предположению, высоко коррелируют с валидизируемым тестом, называются *конвергирующими*, а не коррелирующие – *дискриминантными*. Концептуальная валидность может считаться удовлетворительной, если коэффициенты корреляций валидизируемого теста с группой конвергирующих тестов статистически значимо выше коэффициентов корреляций с группой дискриминантных тестов.

Когда батарея контрольных тестов относительно малочисленна, для сравнительного сопоставления коэффициентов корреляций в группах конвергирующих и дискриминантных тестов достаточно визуального анализа матрицы корреляций. При анализе связей большого числа тестов, а именно десять и более, для оценки концептуальной валидности следует использовать процедуры группировки признаков: метод корреляционных плеяд, факторный анализ, компонентный анализ и т.д.

Так как понятие валидности достаточно широкое и охватывает целый комплекс свойств, характеризующих психологический инструментарий, не существует единого подхода к определению валидности. В теории психометрии определяют два основных типа и по крайней мере одиннадцать видов валидности методик (по И.С. Носс, 2003)



Рис. 1.1. Структура валидности (по И.С. Носс)

Внешняя валидность есть мера возможности распространения результатов тестирования на генеральную совокупность. Внешняя валидизация осуществляется путем расчета репрезентативности экспериментальной выборки при отработке теста и перекрестной валидизации, то есть определения валидности теста на выборке, отличающейся от той, на которой отработывался стимульный материал и задания к тесту.

Внутренняя валидность – это мера соответствия тестовых оценок уровню развития измеряемого свойства. Она состоит по крайней мере из пяти элементов:

а) концептуальная (теоретическая) валидность – теоретическое обоснование возможности измерения исследуемого свойства данным психодиагностическим средством;

б) содержательная валидность – степень репрезентативности содержания заданий теста измеряемой области психических свойств;

в) конструктная валидность – степень репрезентативности исследуемого психологического конструкта в результатах теста («насколько результаты теста рассматриваются в качестве меры исследуемого психологического конструкта – фактора, свойства»);

г) операционная валидность – степень репрезентативности в конкретных экспериментальных операциях реальных операциональных условий психической деятельности;

д) эмпирическая валидность – совокупность характеристик валидности теста, полученная сравнительным статистическим способом оценивания. К этим характеристикам относятся так называемые очевидная и критериальная виды валидности.

Очевидная валидность – это представление о тесте, сфере его применения, результативности и прогностической ценности, которое возникает у испытуемого и другого лица, не располагающего специальными сведениями о характере использования и целях методики.

Критериальная валидность – комплекс характеристик, включающий текущую и прогностическую валидности методики и отражающий соответствие диагноза и прогноза определенному кругу критериев развития измеряемого свойства. Критериальная валидность оценивается величиной связи, корреляции показателя теста с «внешним критерием».

Текущая [диагностическая (конкурентная)] *валидность* – мера способности теста дифференцировать испытуемых по изучаемому признаку. Текущая валидность теста определяется исходя из его «трудности» и мер изменчивости измеряемого признака.

Прогностическая валидность является элементом содержания достоверности тестов и определяется как степень точности и обоснованности суждения о диагностируемом психическом свойстве по его результату спустя определенное время после измерения.

Показателем прогностичности теста является степень регрессии тестовых данных к объективным критериям. Регрессия представляет собой функцию $f(x_1, x_2, x_3, \dots, x_n)$, описывающую зависимость среднего показателя теста, измеряющего данное свойство от заданных фиксированных значений реального проявления этого свойства (внешнего критерия – y). Эта функция может носить линейный и нелинейный характер (линейная регрессия: $y = b + ax$; параболическая зависимость: $y = b + ax + cx^2$; гиперболическая зависимость: $y = b + a/x$; показательная функция: $y = b \cdot a^x$).

В практической психодиагностике для прогноза развития или проявления измеренного свойства чаще применяют линейную или множественную линейную регрессию ($y = b + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$). При использовании множественной линейной функции при прогнозе развития измеряемого свойства повышается уровень прогностичности тестирования и его достоверность за счет перекрытия разными методиками различных сторон измеряемого свойства.

1.4.1. Способы оценки валидности

Экспертные оценки. Метод экспертных оценок заключается в том, что эксперты дают заключение на основе оценочной шкалы, предложенной им разработчиком теста. Надежность такого критерия достигается увеличением числа экспертов.

Можно указать два различных способа учета мнений экспертов для выработки критерия.

1. *Коллективная оценка.* Эксперты совместно вырабатывают компромиссное суждение о каждом испытуемом. В этом случае результат весьма зависим от личностных особенностей и характера группового взаимодействия экспертов.

2. *Независимая оценка.* Оценивание испытуемых каждым экспертом происходит независимо, в закрытом режиме (то есть не оглашается). Затем полученные оценки усредняются с учетом весов, приписываемых каждому эксперту в соответствии с его экспертным статусом.

В обоих случаях оценивание может выполняться с использованием следующих методов:

- Балльной шкалы (испытуемым независимо друг от друга приписываются баллы, отражающие степень выраженности критерия).
- Ранжирования (все испытуемые упорядочиваются экспертами по степени выраженности критерия).
- Парных сравнений (для каждой пары испытуемых определяется обладатель более сильной выраженности критерия). Этот способ имеет преимущества при оценивании плохо дифференцированных признаков или при низкой квалификации экспертов. Задача экспертов облегчается благодаря отсутствию количественной оценки испытуемого. Показателем места, занимаемого объектом в ряду других, служит общее число предпочтений этого объекта по сравнению с остальными по данным всех экспертов.

Примером использования балльного оценивания может служить экспертная оценка пунктов опросника, где каждый вопрос оценивается экспертами (тремя в приведенной схеме) по 3-балльной системе: 1 балл – соответствует, 2 балла – частично соответствует, 3 балла – не соответствует.

Таблица 1.3

Схема применения экспертной оценки с использованием балльной шкалы

Вопрос	Эксперты									Σ	M^1
	1			2			3				
	а	б	в	а	б	в	а	б	в		
1											
...											

Каждый вопрос оценивается по трем критериям:

- критерий «а» – соответствие концепту (конструкту – психологическое свойство, определение и описание которого разрабатывается авторами шкалы);
- критерий «б» – конкретность, однозначность в формулировании вопросов, адекватная лингвистическая форма;
- критерий «в» – пригодность вариантов ответов.

Сделать вывод о валидности пунктов и шкалы на основе экспертных оценок можно следующим образом.

M – от 1 – 1,5 – достаточная валидность пункта (шкалы).

M – свыше 1,5 – 2 недостаточная валидность пункта (шкалы).

M – свыше 2 – пункт (шкала) не валиден.

Корреляционный анализ. В качестве меры валидности наиболее часто на практике применяются разные виды корреляционного анализа связи между индивидуальными оценками по методике и оценками по критерию валидизации. Однако необходимо помнить, что использовать коэффициент корреляции корректно только при наличии теоретического обоснования линейной связи между сопоставляемыми показателями. Если критериальным показателем являются результаты выполнения другой (эталонной) методики и оба набора данных нормально распределены и могут быть признаны интервальными, то в качестве показателя валидности используют коэффициент корреляции Пирсона. Однако чаще всего оценки по критерию бывают дихотомическими или ранговыми. В этих случаях вычисляют коэффициент корреляции точечно-бисериальной или ранговой соответственно.

Еще один способ установить значимость линейной связи может применяться при любом типе данных, получаемых как в ходе выполнения валидируемой методики, так и внешнего критерия – подсчет ϕ -коэффициента (хи-квадрат коэффициента сопряженности в кросс-табуляционных таблицах). Для этого оба сопоставляемых набора данных перекодируются в категориальные (дихотомические или полихотомические). Данный показатель может служить дополнительным подтверждением значимой корреляционной связи (в случае совпадения выводов) или, наоборот, предостеречь исследователя от поспешных заключений (в случае если статистические заключения будут противоречить друг другу).

Для расчета ϕ -коэффициента строится таблица сопряженности.

Таблица 1.4

Примеры таблиц сопряженности 2×2

		Признак X		Итог
		0	1	
Признак Y	0	a	b	a+b
	1	c	d	c+d
Итог		a+c	b+d	N

	Высокий итоговый балл по шкале	Низкий итоговый балл по шкале
Высокие показатели по	a	b

¹ M – математическое среднее

критерию		
Низкие показатели по критерию	c	d

Если данные представлены в четырехклеточной таблице сопряженности, то применяется формула

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \quad \text{где} \quad (1.14)$$

a, b, c, d – соответствующие данные в таблице сопряженности.

Анализ выборок. Как уже указывалось, корреляционный анализ уместен только в ситуации, когда существует обоснование именно линейной связи между сопоставляемыми переменными. Однако достаточно часто успех в какой-то деятельности, используемый в качестве внешнего критерия, не связан линейно с успехом в выполнении валидизируемой методики. Может существовать оптимальный уровень выраженности свойства. Более высокие результаты, полученные при выполнении методики, либо никак не сказываются на выраженности критерия, либо ведут к ее ухудшению. Например, считает ли кассир в магазине с обычной скоростью или очень быстро, не имеет значения с точки зрения его профессиональной успешности. Слишком высокая мотивация к достижению результата также может явиться препятствием на пути к этому результату (закон Йеркса – Дотсона).

В ситуации отсутствия обоснования линейной связи между показателями рекомендуется использовать метод контрастных групп. В зависимости от численности общей выборки в контрастную группу может входить от 10 до 33% общей выборки, но не менее 30 человек. По критерию валидации выделяются две контрастные группы: испытуемые, обладающие наиболее высокими значениями критерия валидации, и обладающие наиболее низкими. В случае нормальной распределенности первичных баллов по методике сопоставление контрастных групп наиболее адекватно проверять по t-критерию значимых различий в средних. Если нормальность отсутствует или данные по методике нельзя считать интервальными, можно использовать непараметрический критерий сравнения двух выборок Манна – Уитни.

Если внешний критерий имеет более двух уровней реализации, следует использовать однофакторный дисперсионный анализ (параметрический или непараметрический – в зависимости от типа данных по валидизируемой методике). Установленные и содержательно интерпретированные межгрупповые различия взаимосвязи между валидизируемой методикой и внешним критерием позволяют более аргументированно обосновать валидность. С этой целью может быть использован двухфакторный дисперсионный анализ (параметрический или непараметрический в зависимости от типа данных методики), где вторым независимым фактором вслед за внешним критерием могут выступать пол, возраст, профессия, страна проживания и т.д.

Множественный регрессионный анализ. В случае, когда можно не просто выделить интегральный внешний критерий, а составить перечень входящих в него элементов и оценить их относительную значимость для осуществления критериальной деятельности, анализ проводится на базе множественного регрессионного анализа оценок методики на совокупность элементов внешнего критерия с учетом их удельных весов. Для каждого элемента его корреляция с критерием в целом умножается на вес в тесте, и полученные произведения суммируются по всем элементам критерия.

Факторный анализ позволяет операционализировать трактовку методики как совокупности латентных факторов (шкал) и наблюдаемых переменных (пунктов), выявить и проанализировать структуру связей латентных переменных исследуемой методики с критериальными латентными и измеряемыми переменными, выявить общие и специфические для группы сопоставляемых методик факторы, степень их представленности в результатах, то есть определить факторный состав и факторные нагрузки результата методики.

Данный вид анализа проводится в два этапа. Выборка расщепляется случайным образом на две статистически равные части, и вначале структура латентных переменных (состав детерминируемых ими измеряемых переменных и корреляции друг с другом) определяется эмпирическим путем с помощью **эксплораторного факторного анализа** на первой подвыборке. На следующем этапе, используя вторую подвыборку, необходимо уточнить структурную и измеряемую модели и оценить степень ее пригодности статистически с использованием **конфирматорного факторного анализа**.

Если при этом методика находится в стадии адаптации и прошла психометрический анализ на других языках, то конфирматорный факторный анализ позволит выявить статистически значимые различия в измеряемой и структурной моделях адаптируемого перевода и оригинала, а также уже адаптированных переводов на другие языки.

Структурное моделирование. Конфирматорный анализ является одним из видов более общего подхода, позволяющего выявлять и анализировать более сложные и обобщенные взаимосвязи между переменными (латентными и измеряемыми, зависимыми и независимыми). Его использование помогает определить состав шкал, статистически оценить, насколько задаваемая ключом структура опросника согласуется с экспериментальными данными. В то время как проверка на одномоментную надежность с помощью показателя альфа Кронбаха, во-первых, не дает статистических критериев, а во-вторых, анализирует каждую шкалу вместе с составляющими ее индикаторами изолированно, конфирматорный факторный анализ преодолевает оба этих существенных недостатка. Помимо показателей, позволяющих провести сравнительное сопоставление, насколько тот или иной пункт методики действительно согласуется с измеряемым шкалой конструктом, КФА выдает статистические критерии согласованности модели с экспериментальными данными и одновременно определяет структуру взаимосвязей всех латентных характеристик, определяемых данной методикой.

В частности, с помощью структурного моделирования можно выявить характеристики выборки испытуемых, значимо повышающие силу взаимосвязи между результатами выполнения методики и внешним критерием, называемые модераторами. Это могут быть социально-демографические, профессиональные характеристики, а также переменные, имеющие психологическую природу (интересы, мотивы, темпераментные характеристики).

Методика может иметь недостаточную валидность для всей выборки испытуемых, что не позволяет использовать ее в прогностических целях. Однако возможно, что, выявив модератор взаимосвязи между показателями методики и внешним критерием, можно установить выборку, характеризующую соответствующим значением модератора, для которого методика будет признана прогностически эффективной.

Таблица 1.5

Классификация статистических задач и методов их решения

Задачи	Условия	Методы
1. Оценка нормальности распределения или выявление различий в распределении признака 2. Нормализация тестовых баллов с выделением классификационных (уровневых) границ	а) при сопоставлении эмпирического распределения с теоретическим	χ^2 – критерий Пирсона; λ – критерий Колмогорова – Смирнова; W – критерий Шапиро – Уилки; m – биномиальный критерий. Кластерный анализ
	б) при сопоставлении двух эмпирических распределений	χ^2 – критерий Пирсона; λ – критерий Колмогорова – Смирнова; ϕ (phi) коэффициент сопряженности Пирсона. Кластерный анализ
3. Оценка эмпирической (критериальной) валидности или выявление степени согласованности изменений	а) сопоставление двух метрических шкал	r – коэффициент корреляции Пирсона R² – коэффициент регрессии (детерминации). Регрессионный анализ
	б) сопоставление двух шкал в различных единицах измерения	r – коэффициент ранговой корреляции Спирмена. τ – коэффициент ранговой корреляции Кендала

4. Оценка критериальной валидности или внутренней согласованности теста	а) сопоставление множества пунктов теста с их итоговым баллом	Множественный регрессионный анализ Факторный анализ
	б) выделение существенных пунктов	Факторный анализ
5. Оценка надежности как внутренней согласованности теста	а) выделение пунктов теста, имеющих одинаковое значение	Альфа Кронбаха Множественный регрессионный анализ

Раздел 2. Документирование психодиагностических данных

Документ – специально созданный человеком предмет, предназначенный для передачи и хранения информации, по форме фиксации могущий быть письменным (тексты, рисунки), иконографическим (кино- и видеоматериалы), фонетическим (аудиозаписи) и пр. «Документирование» – широкое понятие, которое охватывает создание и первичных списков, и сводных таблиц «сырых» психодиагностических данных (в ходе либо по окончании этапа сбора эмпирических данных), и подготовку данных к следующему этапу исследования – обработке, и создание архива, базы психодиагностических данных, и т.д.

В данном разделе представлены основные операции в трех программах Microsoft Excel, с помощью которых производится документирование и презентация сведений, полученных в ходе психодиагностического исследования.

2.1. Работа с программой Microsoft Excel

Программа Microsoft Excel обладает большими возможностями с точки зрения оперирования полученными в ходе психодиагностики сведениями, позволяет осуществлять множество операций по обработке данных. Во многих статистических программах существует возможность экспорта данных в Excel; программа удобна еще и тем, что в нее, в свою очередь, легко импортируются текстовые файлы.

2.1.1. Общие положения

Файлы программы Microsoft Excel являются электронными таблицами, соответственно, построены как столбцы и строки (рис. 1).

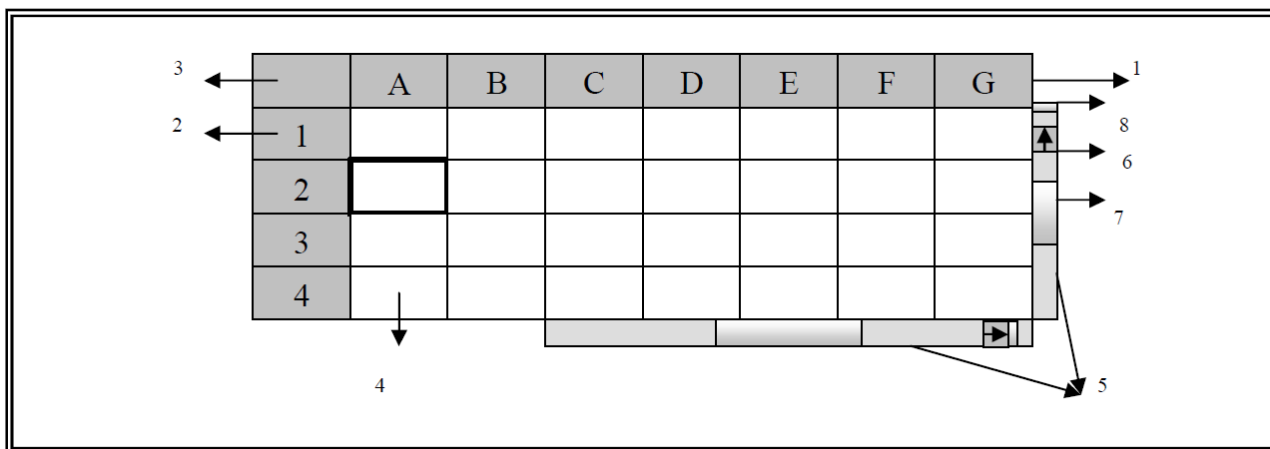


Рис. 2.1. Окно программы Microsoft Excel

На рисунке цифрами обозначены:

- 1 – названия столбцов (буквы английского алфавита, всего допускается 256 столбцов),
- 2 – названия строк (нумерация, количество не ограничено),
- 3 – область, при нажатии на которую выделяются все столбцы и строки одновременно,
- 4 – ячейки таблицы (кодируются как в шахматах, например, выделенная ячейка называется A2),
- 5 – вертикальная и горизонтальная «полосы прокрутки», позволяющие передвигаться по документу,
- 6 – указатель со стрелкой (таже функция),
- 7 – «бегунок», потянув мышкой за который, можно быстро переместиться по документу,
- 8 – горизонтальный (внизу справа – соответствующий вертикальный) разделитель текста.

При этом сами файлы программы Microsoft Excel называются «книгами», поскольку содержат «листы» (показаны в нижнем левом углу окна программы, стандартный набор –


3 листа, возможно добавление (Вставка → Лист)² и удаление (Правка → Удалить лист) из главного меню), каждый из которых является самостоятельным документом. Запись данных осуществляется в ячейки, после ввода данных нажимают Enter.

2.1.2. Виды курсора

В программе Microsoft Excel при определенном положении мышки курсор может приобретать 5 различных видов, каждый из которых имеет определенное функциональное назначение (табл. 1).

Таблица 2.1



Виды и функции курсора в программе Microsoft Excel.

Вид курсора: 	
<i>Функция</i>	<i>Задание</i>
1. Активирование ячейки, делающее возможным запись в ней	1. Активировать ячейку A1, записать фамилию первого респондента, перейти к A2, записать следующую фамилию
2. При одинарном нажатии на ячейку и последующем вводе данных предыдущее содержание ячейки стирается	2. Вернуться к первой фамилии, щелкнув один раз, попробовать исправить надпись
3. При двойном щелчке курсор приобретает другой вид (I), что делает возможным внесение в ячейку исправлений	3. Произвести подобную операцию – исправить букву в фамилии первого респондента ³
4. Служит для выделения «области» (совокупности ячеек)	4. Выделить область A1:C3 (подвести курсор к A1 и не отпуская клавишу, довести до C3)
5. Служит для выделения столбцов и строк (для этого курсор должен находиться на названии столбца или строки).	5. Выделить столбец D (навести на букву D, нажать один раз). Теперь возможно редактирование столбца (команды контекстного меню – правая клавиша).
Вид курсора: I	
1. Служит для исправления текста в ячейках (это же осуществляется в строке формул).	1. Активировать ячейку C1 предыдущим типом курсора, затем через строку формул и тип курсора (I) внести в нее изменения.
Вид курсора: + появляется при наведении мышки на нижний правый угол ячейки либо выделенной области	
1. Копирование	1. Активировать ячейку A1 первым типом курсора, повести в нижний правый бок до приобретения курсором вида +, потянуть вниз и вбо;
2. Автоматическая нумерация	2. Добавить столбец перед списком фамилий (выделить столбец A, в контекстном меню правой клавиши выбрать «добавить ячейки»); поставить в ячейку A1

² Здесь и далее символ → обозначает последовательность операций.

³ Исправление возможно в так называемой «строке формул» – находится над таблицей и под панелью инструментов.

⁴ Таким образом задается «шаг» – разница между соседними ячейками, таким образом, можно автоматически формировать интервальную шкалу (например, A1-«3», A2-«6», протяжка → в

	цифру «1», в ячейку A2 цифру «2» ⁴ ; выделить область A1:A2 первым типом курсора
3. «Протягивание» формул	Подвести к нижнему правому краю, потянуть вниз до конца списка фамилий (автоматически проставятся номера респондентов)
4. Вид курсора:  появляется при наведении мышки на промежуток между строками либо столбцами (область названий)	
1. Изменение высоты строки либо ширины столбца	1. Подвести мышку к промежутку между столбцами B и C так, чтобы курсор принял этот вид, нажать и потянуть вправо (столбец B расширится); то же – со строками
2. Придание совокупности столбцов или строк одинаковой высоты / ширины	2. Выделить столбцы C, D и E (первым типом курсора – нажать на C и, не отпуская, потянуть вправо до E), затем подвести мышку между столбцами D и E (до принятия 4-го вида), сузить столбец D (остальные примут такую же ширину); аналогично – со строками
5. Вид курсора: 	
1. Перемещение выделенной области	1. Выделить область A1: C3, подвести курсор к границе выделения так, чтобы он принял соответствующий вид и стрелка при этом касалась границы области выделения, перетащить в правую часть файла.

2.1.3. Работа со списками, функции фильтра и сортировки

При большом объеме сведений, полученных в ходе психодиагностического исследования, работа со списками затрудняется из-за небольшого размера окна файла (так, переходя к 100-му респонденту в списке, исследователь «теряет из вида» первую строку файла с заголовками-обозначениями шкал методик). Для этого служат «разделители текста» (см. рис. 1): их можно «взять» мышкой и перетянуть (например, горизонтальный разделитель – к нижней границе строки заголовков) в нужное место, после чего документ разбивается на две части («верх» можно оставить фиксированным, а в нижней части передвигаться по списку). Возможно использование горизонтального и вертикального разделителей по отдельности – тогда вернуть их на прежнее место можно аналогичным путем (обратные действия), возможно совместное – тогда убрать их можно, щелкнув дважды мышкой на их пересечении.

Еще одна функция, часто востребованная при работе со списками, – сортировка данных (например, расположение фамилий в алфавитном порядке). Немаловажным является выделение соответствующей области перед сортировкой (чтобы фамилии «поменялись местами» вместе со всей остальной информацией на респондентов). В главном меню (Данные → Сортировка) возможно выбрать параметры сортировки (столбец, по которому производится сортировка, условия – по возрастанию либо по убыванию).

следующих ячейках появится 9, 12, 15 и т. д.), порядок может начинаться с любой цифры, последовательность может быть прямой или обратной.

Одной из наиболее удобных возможностей программы Microsoft Excel является возможность выбирать из общего списка людей, соответствующих одному (например, высокая тревожность) или нескольким (например, высокий нейротизм и высокая экстраверсия) условиям с помощью функции фильтра (табл. 2).

Таблица 2.2

Работа с функцией фильтра в программе Microsoft Excel

<i>Ф у н к ц и я</i>	<i>З а д а н и е</i>
1. Фильтр всегда ставится на столбец, который, соответственно, необходимо выделить	1. Выделить столбцы со шкалами «интернальность» (ИО), «макиавеллизм» (М) в файле «ЛР»
2. Применение фильтра (отмена – аналогично): Данные → Фильтр → Автофильтр	2. Применить фильтры (под названиями столбцов появятся кнопки раскрывающихся списков)
3. Выбор условий фильтра («вход» в фильтр): нажатие стрелки, далее выбор нужного параметра	3. В фильтре столбца со шкалой «интернальность» выбрать слово «условия», задать условия – «больше или равно 18»; ок; Аналогичную процедуру проделать в столбце со шкалой «макиавеллизм»
4. Копирование «отфильтрованных» данных производится только при выделении области (не столбцов, не строк и не всего файла (см. рис. 1).	4. Выделить область с данными с высоким уровнем интернальности, копировать, перейти на Лист 2, вставить.

2.1.4. Создание и редактирование диаграммы в программе Microsoft Excel

Таблица 2.3

Создание диаграммы в программе Microsoft Excel

<i>Ф у н к ц и я</i>	<i>З а д а н и е</i>
1. Вставка диаграммы осуществляется одинарным нажатием на соответствующий значок панели инструментов	Вставить диаграмму
2. Первый шаг: выбор типа диаграммы, «далее»	Выбрать тип «гистограмма», «далее»
3. Второй шаг: а) выбор диапазона данных, по которым должна строиться диаграмма, производится путем выделения мышкой; б) выбор параметра (строки или столбцы), по которому должно вестись сравнение; в) в закладке «ряд» устанавливаются: 1) подписи легенды, 2) подписи по оси x; «далее»	а) поставить курсор в графу «диапазон», мышкой выделить область файла (шкала «циклотимность» и шкала «экзальтированность» – от 1-й до 10-й цифры); б) поставить «флажок» мышкой на слове «столбцы»; в): 1) установить подписи легенды: в окне «ряд» выделить «ряд1», поставить курсор в графу «имя», мышкой выделить название 1-й шкалы – «ЦТ», аналогично «ряд2» – «ЭК», 2) установить подписи по оси x – поставить курсор в графу, мышкой выделить с 1-го по 10-й фамилии респондентов; «далее»
4. Третий шаг: ввод параметров диаграммы	Ввести название диаграммы, подписи осей, «далее», «готово»

Раздел 3. Статистические процедуры в психометрике

3.1. Ввод данных

В подавляющем большинстве случаев ввод данных осуществляется в программе Microsoft **Excel**, установленной практически на всех компьютерах. Перед тем как начать ввод данных, необходимо определить структуру будущего файла. Структура файла должна соответствовать плану исследования, который должен быть составлен так, чтобы его исходные данные можно было бы обработать в соответствии с задачами исследования.

Структура данных может быть задана в виде предварительного списка переменных с указанием их типов и диапазонов возможных значений.

Таблица 3.1

№	Название	Тип	Диапазон возможных значений
1	Пол	Номинативная	1 – мужской, 2 – женский
2	Возраст	Ранговая	19–55
3	Пункт 1 (опросника)	Ранговая	1–7
...	(Далее все пункты опросника)		
35	Показатель теста СЖО	Количественная	50–190

После того как вы опросили испытуемых по опроснику, необходимо информацию об участниках исследования и все ответы по пунктам обследованных ввести в таблицу Microsoft **Excel**. Если ответы давались испытуемыми с использованием числовой шкалы, то именно эти баллы по каждому пункту необходимо ввести в компьютер. Если для фиксации ответов на вопросы использовалась вербальная шкала, то кодировка ответов числами происходит согласно таблице 8.

Таблица 3.2

Альтернативы ответов и их формулировки

Количество альтернатив	Возможные формулировки ответов
2	(2) <i>Да</i> , (1) <i>Нет</i>
3	(3) <i>Да</i> , (2) <i>Затрудняюсь ответить</i> (или <i>Не знаю</i> , <i>Не уверен</i>), (1) <i>Нет</i>
4	(4) <i>Согласен</i> , (3) <i>Скорее согласен</i> , (2) <i>Скорее не согласен</i> , (1) <i>Не согласен</i>
5	(5) <i>Согласен</i> , (4) <i>Скорее согласен</i> , (3) <i>Затрудняюсь ответить</i> (или <i>Не знаю</i> , <i>Не уверен</i>), (2) <i>Скорее не согласен</i> , (1) <i>Не согласен</i>
7	(7) <i>Абсолютно согласен</i> , (6) <i>Согласен</i> , (5) <i>Скорее согласен</i> , (4) <i>Затрудняюсь ответить</i> , (3) <i>Скорее не согласен</i> , (2) <i>Не согласен</i> , (1) <i>Абсолютно не согласен</i>

Категорически не рекомендуется суммировать предварительно с помощью калькулятора ответы по ключам и вводить в компьютер шкальные баллы. Это касается не только разрабатываемой методики, но и используемых для валидации уже известных шкал. Во-первых, ошибки при устном подсчете шкальных значений неизбежны, а во-вторых, теряется много информации, которая впоследствии может оказаться полезной. Например, может представлять интерес анализ отдельных пунктов различных методик.

Очень часто для подсчета значения по шкале необходимо инвертировать фиксированные значения. Для обращения пунктов шкалы в Microsoft **Excel** после каждого столбца, содержащего ответы респондентов по обратному пункту (в соответствии с ключом), вставить пустой столбец. В первой строке вставленных столбцов поставить номер и букву R рядом. Например, наименование столбца, в котором предполагается

разместить результаты по пункту, обратному номеру 3 из опросной методики, может быть «Пункт_3R». Далее в каждой строке, начиная со второй, вставленного пустого столбца ввести значение, равное:

$$\text{максимальный балл ответа} + \text{минимальный балл ответа} - \text{текущий балл ответа}.$$

Для бинарной шкалы, состоящей из двух возможных значений 1 и 2, это преобразование выглядит как $3 - x_i$, где x_i – ответ i -го испытуемого на вопрос, для 5-балльной шкалы, использующей оценки от 1 до 5, используется формула $6 - x_i$, для 7-балльной – формула $8 - x_i$ и т.д.

Чтобы перенести полученные значения переменных столбцов с буквой R в новую матрицу данных, лучше всего пользоваться окном меню *Специальная вставка*, которое можно вызвать *Правка* → *Специальная вставка* или щелчком правой кнопки мыши, предварительно скопировав расчетные значения трансформированной переменной. Затем вставить скопированные значения в новую матрицу данных, используя меню *Специальная вставка*, установив метку в строке значения.

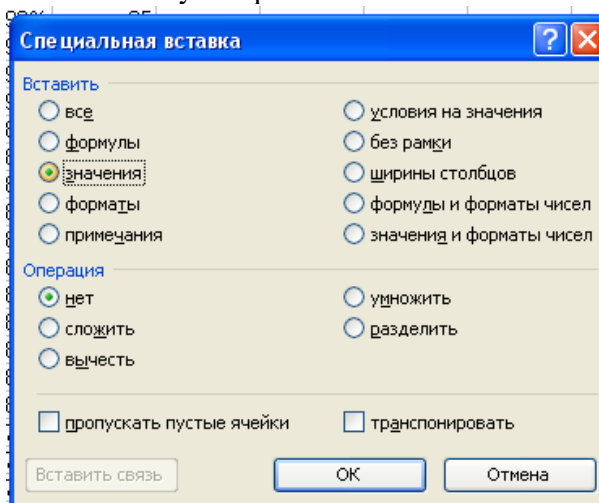


Рис. 3.1. Окно *Специальная вставка*

Для ввода данных в программе **SPSS** необходимо:

- открыть программу **SPSS**. Программа открывается окном с двумя вкладками внизу: *Данные (Data View)* и *Переменные (Variable View)*. Щелкаем лист *Переменные (Variable View)* и заполняем лист, вписывая наименования переменных, указывая их тип и другую необходимую информацию;
- скопировать файл данных, введенный в **Microsoft Excel**, и вставить его в лист *Данные (Data View)*, используя меню *Правка* → *Вставка*.

3.1.2. Описательная статистика

Показатели описательной статистики в программе **Microsoft Excel** можно рассчитать с помощью *Мастера функций* и пакета *Анализа данных*. Мастер функций является встроенным пакетом обработки данных, а *Анализ данных* – специальной надстройкой.

Для открытия окна *Мастера функций* необходимо щелкнуть мышью на значок f_x (*вставка функции*). В открывшемся окне в строке *Категория* выбрать *Статистические* и в строке *Выберите функцию* выбрать необходимую функцию.

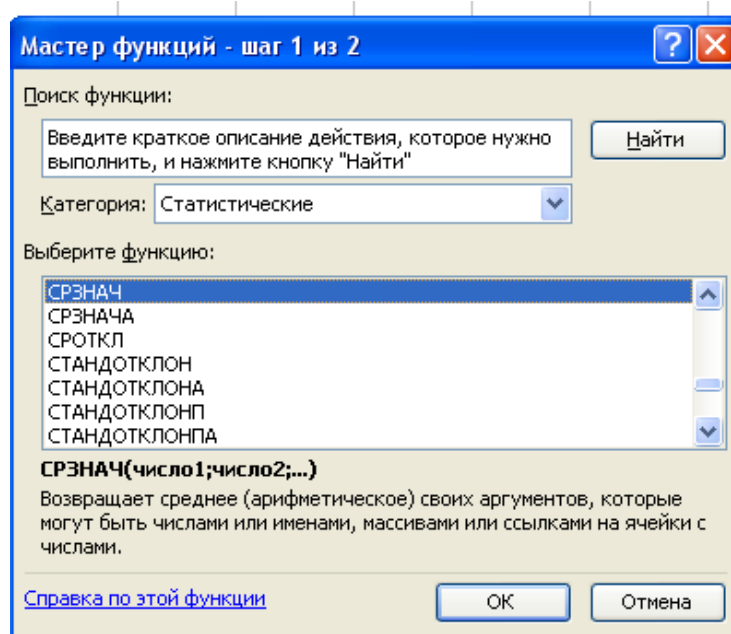


Рис. 3.2. Окно *Мастера функций*

Чтобы установить специальную настройку *Анализ данных*, которая существенно облегчает и ускоряет работу с данными, необходимо в строке меню в верхней части окна программы выбрать *Сервис*, в открывающемся меню пункт *Настройки*, щелкнув по нему мышью. Для установки *Пакета анализа* в открывшемся диалоговом окне установите курсор рядом с этим пунктом и нажмите *ОК*. При повторном обращении к меню *Сервис* появится пункт *Анализ данных*.

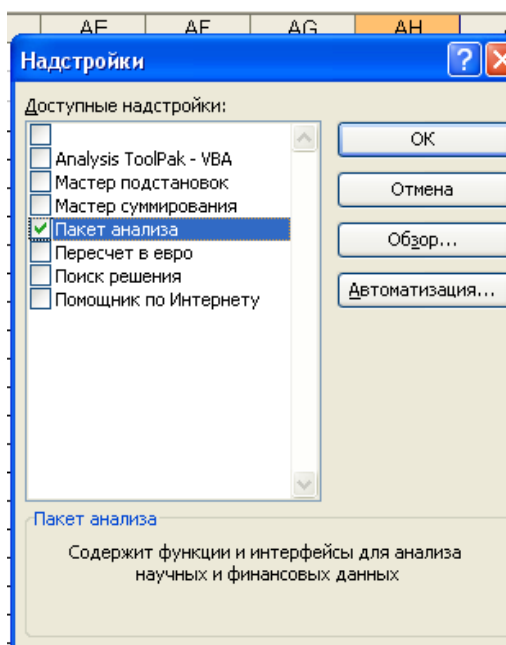


Рис. 3.3. Окно *настроек*

Для расчета показателей описательной статистики необходимо последовательно осуществить следующие шаги:

- войти в меню *Сервис* → *Анализ данных*;
- в открывшемся окне *Анализ данных* выбрать *Описательная статистика* и нажать *ОК*;

- в открывшемся окне *Описательная статистика* в строке *Входной интервал* ввести входной интервал, выделив вариационный ряд числовых значений;
- установить курсор в ячейке *Итоговая статистика*;
- нажать ОК.

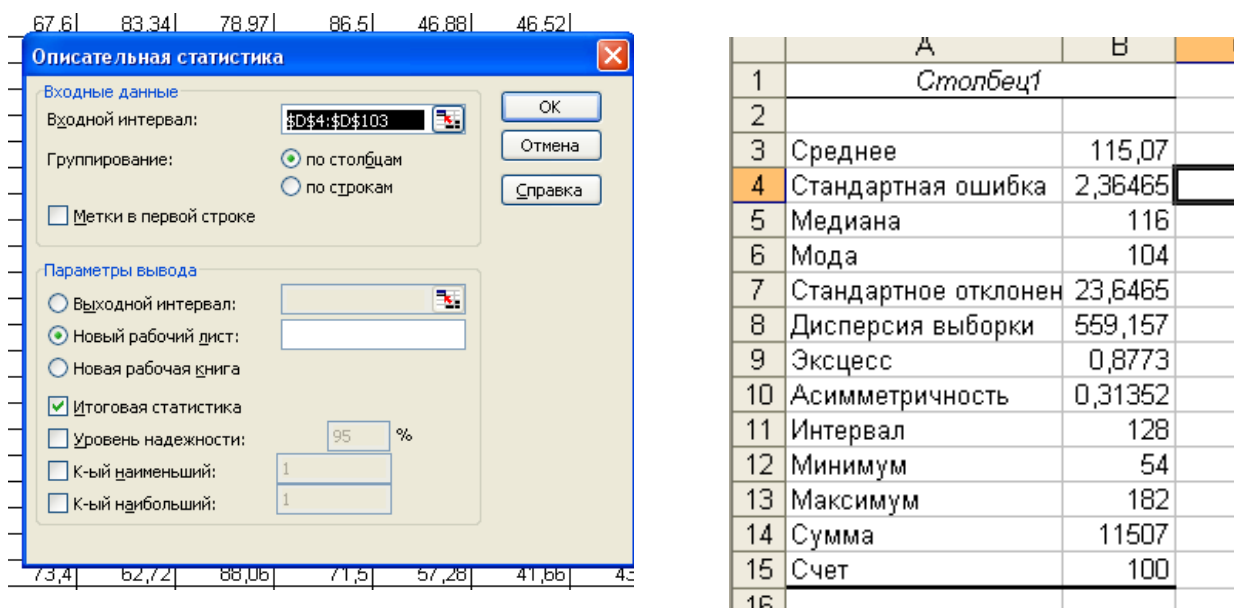


Рис. 3.4. Окно *Описательной статистики* и вывод данных

Итоговые показатели описательной статистики включают меры центральной тенденции и меры изменчивости (разброса), эксцесса и асимметрии. Они дают возможность создать представление о параметрах нормальности распределения косвенным способом.

Коэффициент асимметрии – это мера отклонения распределения частоты от симметричного распределения, то есть такого, у которого на одинаковом удалении от среднего значения по обе стороны выборки данных располагается одинаковое количество значений. Если наблюдения подчиняются нормальному распределению, то асимметрия равна нулю. Для проверки на нормальное распределение можно применять следующее правило. Если асимметрия значительно отличается от нуля (больше доверительного интервала), то гипотезу о том, что данные взяты из нормально распределенной генеральной совокупности, следует отвергнуть. Если вершина асимметричного распределения сдвинута к меньшим значениям, то говорят о положительной асимметрии, в противном случае – об отрицательной.

Коэффициент вариации или эксцесс указывает, является ли распределение пологим (при большом значении коэффициента) или крутым. Коэффициент вариации равен нулю, если наблюдения подчиняются нормальному распределению. Поэтому для проверки на нормальное распределение можно применять еще одно правило. Если коэффициент вариации значительно отличается от нуля, то гипотезу о том, что данные взяты из нормально распределенной генеральной совокупности, следует отвергнуть.

Как правило, для переменных, относящихся к интервальной шкале и подчиняющихся нормальному распределению, в качестве основной характеристики используют среднее значение, а в качестве меры разброса – стандартное отклонение или стандартную ошибку. Для порядковых или интервальных переменных, не подчиняющихся нормальному распределению, – соответственно медиану или первый и третий квартили. Для переменных, относящихся к номинальной шкале, нельзя дать других значимых характеристик, кроме моды.

Распределение, статистически значимо не отличающееся от нормального, характеризуется тем, что его асимметрия и эксцесс статистически значимо не отличаются

от нуля. При этом именно симметрия (то есть близость к нулю асимметрии) является особенно важным моментом. Эффект отрицательного эксцесса пропадает на выборке более 100, а при положительном эксцессе – на выборке более 200.

Самым простым способом проверки симметричности является проверка статистического совпадения медианы и среднего. Это условие можно считать выполненным с уровнем доверия альфа, если медиана попадает в доверительный интервал с уровнем доверия альфа для вычисленного среднего. Этот интервал нетрудно подсчитать в **Microsoft Excel** с помощью *Мастера функций*, категории *Статистические*, функции **ДОВЕРИТ**.

Поскольку в описательной статистике основные показатели – среднее, медиана, стандартное отклонение – уже были определены, в открытом окне *Аргументы функции* Мастера функций в поле *Альфа* вводится выбранный уровень значимости, в поле *Станд_откл* – величина стандартного отклонения, в поле *Размер* – объем выборки. Если разность между средним и медианой по абсолютной величине не превосходит доверительного интервала, распределение можно считать симметричным.

Для вычисления показателей центральной тенденции и изменчивости в программе **SPSS** необходимо:

- выбрать вкладку *Анализ* → *Описательная статистика* → *Частоты*;
- в открывшемся окне *Частоты* нажатием стрелки перевести переменные для статистического анализа из левого поля в правое; щелкнуть поле *Статистики*;
- в открывшемся окне *Частоты: Статистики* устанавливаются метки в полях *Квартили*, *Стандартное отклонение*, *Дисперсия*, *Максимум*, *Минимум*, *Стандартная ошибка среднего*, *Среднее*, *Медиана*, *Мода*, *Асимметрия*, *Эксцесс*, и нажимаем *Продолжить*;
- в окне *Частоты* щелкаем в поле *Диаграммы*;
- в открывшемся окне *Частоты: Диаграммы* выбираем тип диаграммы *Гистограммы* и устанавливаем метку *С нормальной кривой*; щелкаем *Продолжить*;
- в окне *Частоты* щелкаем *ОК*.

Для получения доверительного интервала для среднего в программе **SPSS** необходимо:

- выбрать вкладку *Анализ* → *Описательная статистика* → *Разведочный анализ*;
- в открывшемся окне *Исследовать* нажатием стрелки перевести переменные для статистического анализа в поле *Список зависимых переменных*; щелкнуть поле *Статистики*;
- в открывшемся окне *Исследовать: Статистики* установить метку в поле *Описательные статистики*, доверительный интервал для среднего 95%, нажать *Продолжить*;
- в окне *Исследовать* нажимаем *ОК*.

3.1.3. Проверка нормальности распределения

Как правило, анализ данных начинается с изучения частоты встречаемости значений исследуемого признака в имеющемся множестве наблюдений. Не все значения встречаются с одинаковой частотой, некоторые встречаются реже, некоторые чаще. Для анализа распределения частот составляются таблицы и графики, которые дают информацию о форме распределения признака. Обычно выделяют следующие типичные формы распределения множества значений.

Равномерное распределение – когда все значения встречаются одинаково или почти одинаково.

Симметричное распределение – когда одинаково часто встречаются крайние значения.

Нормальное распределение – симметричное распределение, у которого крайние значения встречаются редко и частота постепенно повышается от крайних к серединным значениям признака.

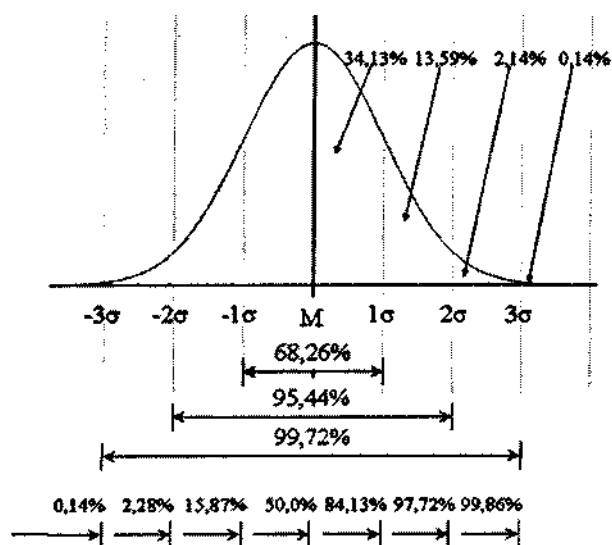


Рис. 3.5. Кривая нормального распределения и доля выборки, попадающая в каждый интервал

Асимметричные распределения: левосторонние – с преобладанием частот малых значений, правосторонние – с преобладанием частот больших значений.

Оценка нормальности распределения проводится на основе критериев Колмогорова – Смирнова (λ), Шапиро – Уилки (W), хи-квадрат (χ^2).

В программе **Microsoft Excel** нет встроенных функций для расчета критериев нормальности распределения.

Для проверки нормальности распределения в программе **SPSS** необходимо:

- выбрать вкладку *Анализ* → *Описательная статистика* → *Разведочный анализ*;
- в открывшемся окне *Исследовать* нажатием стрелки перевести переменные для статистического анализа в поле *Список зависимых переменных*; щелкнуть поле *Статистики*;
- в открывшемся окне *Исследовать: Статистики* установить метку в поле *Описательные статистики*, доверительный интервал для среднего 95%, нажать *Продолжить*;
- в окне *Исследовать* щелкнуть в поле *Графики*;
- в открывшемся окне *Исследовать: Графики* установить метку в поле *Графики и критерии для проверки нормальности* и нажать *Продолжить*;
- в окне *Исследовать* нажимаем *ОК*.

После обработки данных программа выводит результаты в таблице.

Таблица 3.3

Критерий нормальности

	Колмогоров – Смирнов ^а			Шапиро – Уилки		
	Статистика	ст. св.	Значимость	Статистика	ст. св.	Значимость
СЖО	0,081	94	0,160	0,976	94	0,088

а. Поправка значимости Лильефорса

В таблице, в колонках «Статистика» представлены полученные значения критериев, в колонках «ст.св.» – степени свободы (объем выборки), в колонках «Значимость» – уровень значимости. Полученный уровень значимости критерия К-С приводится с поправкой Лильефорса. Если уровень значимости $p > 0,05$, то распределение нормально.

Значение $p < 0,05$ критерия указывает на значимое отклонение распределения от нормального и предпочтение непараметрических методов статистического анализа.

3.1.4. Нормирование сырых баллов

Большинство методов стандартизации основано на нормальном распределении. Поэтому для выделения норм тестовых шкал необходимо проверить полученные распределения на нормальность. Если эмпирическое распределение статистически значимо не отличается от нормального, тогда для нормирования такого распределения достаточно знать два показателя – среднее арифметическое (M) и выборочное стандартное отклонение (σ).

Наиболее простейший способ нормирования данных – разделение диапазона значений на три части или определение низкого, среднего и высокого уровней.

При нормальном распределении статистической нормой считается диапазон значений в границах $[M-\sigma; M+\sigma]$. Доля значений в указанном диапазоне составляет 68,26%.

Эмпирически границы уровней определяются следующим образом:

граница низких баллов – $M - \sigma$,

граница высоких баллов – $M + \sigma$.

Для проверки достоверности диапазонов выделенных границ и нормальности эмпирического распределения может быть использован критерий хи-квадрат (χ^2). Чтобы рассчитать критерий χ^2 , необходимо:

- подсчитать процентную долю количества испытуемых в эмпирическом распределении по каждому уровню и внести их в соответствующую расчетную таблицу;
- внести данные по процентной доле испытуемых теоретического распределения в расчетную таблицу;
- рассчитать критерий χ^2 по формуле

$$\chi^2 = \sum_{i=1}^k \frac{(E_i - T_i)^2}{T_i} \quad (3.1)$$

Таблица 3.6

Расчетная таблица для критерия χ^2

Уровни	Эмпирическое	Теоретическое	$(E_i - T_i)^2 / T_i$
Значительно ниже нормы		2,28	
Несколько ниже нормы		13,59	
Средний		68,26	
Несколько выше нормы		13,59	
Значительно выше нормы		2,28	
		Σ	

Полученное значение является эмпирическим значением критерия. Его необходимо сравнить с критическим, которое находят в таблице критических значений. Для поиска нам понадобится значение df (количество степеней свободы).

$df = (k-1) \cdot (l - 1)$ где,

k – число уровней в сопоставляемых распределениях (число строк в таблице);

l – количество сопоставляемых распределений (число столбцов в таблице).

$df = (5-1) \cdot (2 - 1) = 4$

По таблице находим для уровня значимости $p=0,05$ и $df=4$ критическое значение 9,488. Если эмпирическое значение меньше критического, то делается вывод о достоверности границ и незначительном отклонении эмпирического распределения от теоретического.

Таблица 3.7

Схема разбиения диапазона на три, пять и семь интервалов (по О.В. Митиной)

а) Схема разбиения на три интервала

Диапазон значений	Диапазон значений после z-преобразования	Интерпретация	Процент испытуемых
Более $M+\sigma$	Более 1	Выше нормы	~16%
$[M-\sigma; M+\sigma]$	$[-1; 1]$	Норма	~68%
Менее $M-\sigma$	Менее -1	Ниже нормы	~16%

б) Схема разбиения на пять интервалов

Диапазон значений	Диапазон значений после z-преобразования	Интерпретация	Процент испытуемых
Более $M+2\sigma$	Более 2	Значительно выше нормы	~2,3%
$[M+\sigma; M+2\sigma]$	$[1; 2]$	Несколько выше нормы	~13,7%
$[M-\sigma; M+\sigma]$	$[-1; +1]$	Норма	~68%
$[M-2\sigma; M-\sigma]$	$[-2; -1]$	Несколько ниже нормы	~13,7%
Менее $M-2\sigma$	Менее -2	Значительно ниже нормы	~2,3%

в) Схема разбиения на семь интервалов

Диапазон значений	Диапазон значений после z-преобразования	Интерпретация	Процент испытуемых
Более $+2\sigma$	Более 2	Очень сильно выше нормы	~2,3%
$[M+4/3\sigma; M+2\sigma]$	$[4/3; 2]$	Сильно выше нормы	~6,8%
$[M+2/3\sigma; M+4/3\sigma]$	$[2/3; 4/3]$	Несколько выше нормы	~16%
$[M-2/3\sigma; M+2/3\sigma]$	$[-2/3; 2/3]$	Норма	~49,8%
$[M-4/3\sigma; M-2/3\sigma]$	$[-4/3; -2/3]$	Несколько ниже нормы	~16%
$[M-2\sigma; M-4/3\sigma]$	$[-2; -4/3]$	Сильно ниже нормы	~6,8%
Менее $M-2\sigma$	Менее -2	Очень сильно ниже нормы	~2,3%

Еще одним способом нормирования данных является перевод сырых баллов в z-взвешенные результаты и последующий перевод на их основе в другие стандартные единицы: стены, T-баллы, IQ-баллы.

Стандартные взвешенные результаты выражают отдаленность отдельных индивидов от среднего нормализационной совокупности при помощи стандартной (нормированной случайной) величины z. Взвешенные результаты (z) определяют так, что вычисляют различия отдельных общих взвешенных результатов от средней нормализационной совокупности, а эти различия делят на стандартное отклонение совокупности. Средним общим взвешенным результатам отвечает $z = 0$, общий взвешенный результат, который на 1 стандартное отклонение превышает среднее, равен 1, а если он ниже него, то -1 и т. д.

$$z_i = \frac{x_i - M}{\sigma}, \quad \text{где} \quad (3.2)$$

x_i – сырой балл i-го испытуемого,

M – среднее значение,

σ – стандартное отклонение.

Расчет в Excel. Для расчета z-взвешенного результата в Microsoft Excel необходимо:
– вызвать функцию НОРМАЛИЗАЦИЯ (Мастер функций (f_x) → Статистические → Нормализация);

- в открывшемся окне «Аргументы» функции ввести x – нормализуемое значение, Среднее – среднее арифметическое распределения, Стандартное_откл – стандартное отклонение распределения;
- нажать ОК;
- подвести курсор в правый нижний угол и растянуть на все значения в выборке.

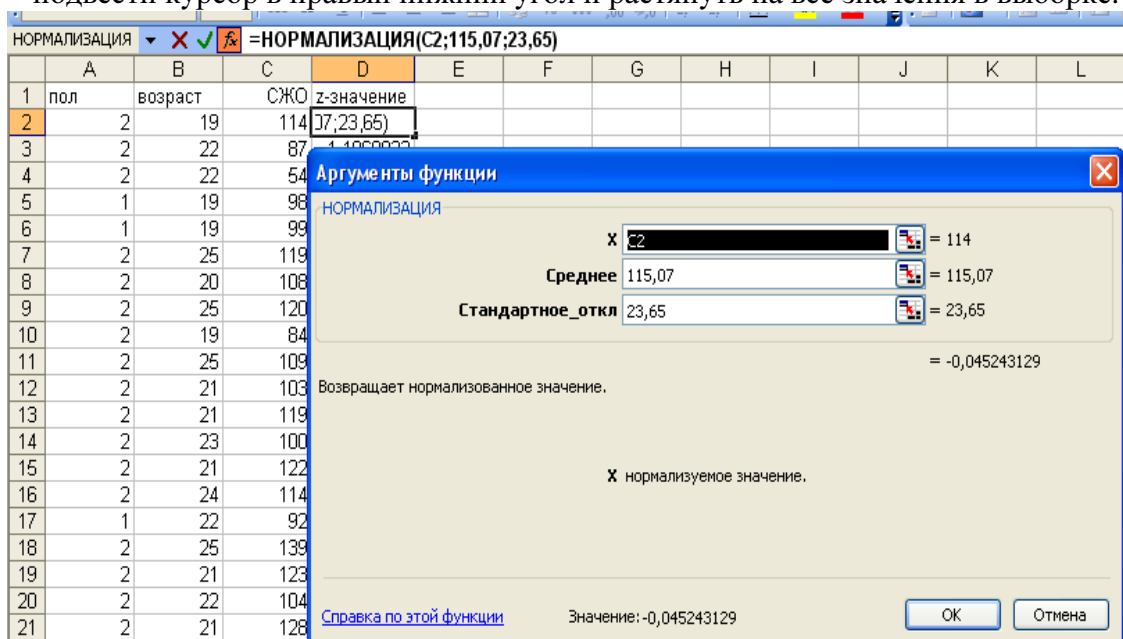


Рис. 3.7. Функция в Microsoft Excel, позволяющая вычислять z-взвешенный результат

Перевод сырых баллов в *стэны* можно производить двумя путями.

Известно, что *стэн* – это стандартизированное значение со средним 5,5 и стандартным отклонением 2.

Исходя из этого, для перевода в *стэны* можно использовать формулу:

$Стэн = 2z + 5,5$, где z – z-взвешенный результат.

Второй способ перевода баллов в *стэны*:

- найти среднее арифметическое и стандартное отклонение;
- от среднего отсчитать по пять интервалов в $1/2\sigma$ в обе стороны;
- пронумеровать 10 полученных интервалов слева направо. Номер интервала и будет *стэном*.

T-баллы имеют тот же принцип перевода, что и *стэны*. Чаще всего *T-баллы* используются для стандартизации сырых баллов, имеющих большой размах (разницу между мин. и макс. баллами). К примеру, *T-баллы* используются в тесте ММРІ и его отечественной адаптации СМІЛ.

В *T-балах* среднее = 50, стандартное отклонение = 10.

Формула для перевода в *T-баллы*.

$T-балл = 10z + 50$, где z – z-взвешенный результат.

Для *IQ-баллов* среднее арифметическое = 100, стандартное отклонение 15.

Формула для перевода в *IQ баллы*:

$IQ = 15z + 100$, где z – z-взвешенный результат.

Процентили (перцентили) и квантили. Процентиль может использоваться для стандартизации как нормально распределенных баллов, так и данных с ненормальным распределением. В шкале процентилей нет понятия среднего арифметического и стандартного отклонения. Процентиль показывает ранг испытуемого в выборке, который выражен в процентной доле. Частным случаем процентильной меры являются квантили

распределения – 25%, 50%, 75% процентиля. Таким образом, процентиль показывает процент испытуемых, получивших балл меньше либо равный данному. Он является достаточно распространенной шкалой стандартизации, т.к. очень удобен и понятен. Его диапазон от 1 до 100.

Расчет в **Excel**. Для расчета процентилей понадобится функции *Сортировка* и *СЧЕТЕСЛИ*. Функция *Сортировка* позволяет ранжировать ряд значений от минимального к максимальному или наоборот. Для выполнения сортировки нужно войти в меню *Данные* → *Сортировка*. После того, как значения проранжированы, их вводят в формулу: $=(СЧЕТЕСЛИ(диапазон;условие)*100)/N$, где N – количество человек.

	A	B	C	D	E	F	G	H	I	J	K	L
1	пол	возраст	СЖО	Процентиль	формула				Точка	Столбец1	Ранг	Процент
2	2	19	54	1	=СЧЕТЕСЛИ(С\$2:С\$101;"<="&C2)*100/100				79	182	1	100,00%
3	2	22	60	2					81	181	2	98,90%
4	2	22	73	3					66	176	3	97,90%
5	1	19	74	4					95	164	4	96,90%
6	1	19	76	5					78	162	5	95,90%
7	2	25	77	6					82	156	6	94,90%
8	2	21	79	7					98	154	7	93,90%

Рис. 3.8. Окно расчета процентиля⁵

Для расчета процентиля используется также пакет *Анализ данных*:

- выбираем *Сервис* → *Анализ данных*;
- в открывшемся окне *Анализ данных* выбираем *Ранг и перцентиль*;
- в открывшемся окне *Ранг и перцентиль* в поле *Входной интервал* вводится вариационный ряд значений выделением столбца (строки); если необходимо, вводится *Выходной интервал* щелчком мыши в нужную ячейку вывода;
- нажимаем ОК.

Результаты вывода представлены 4-мя столбцами.

- Столбец *Точка* показывает порядковый номер значения.
- Столбец *Столбец1* – значения.
- Столбец *Ранг* – ранговое значение сырого бала.
- Столбец *Процент* – значение процентиля.

Расчет в **SPSS**. В программе **SPSS** используется функции:

- Входим в меню *Преобразовать* → *Ранжировать наблюдения*;
- в открывшемся окне *Ранжировать наблюдения* переводим тестируемую переменную в поле *Переменные*; щелкаем в закладку *Типы рангов*;
- в открывшемся окне *Ранжировать наблюдения: Типы рангов* устанавливаем метку в строке *Дробный ранг в %*⁶; нажимаем *Продолжить* и далее нажимаем ОК.

⁵ Символ \$ в адресе ячейки обозначает, что он не будет меняться. Дело в том, что Excel, при копировании формулы в следующую (в нашем примере нижнюю) ячейку автоматически изменяет диапазоны, заданные в этой формуле. Если у адреса D2 нет знака \$, то при копировании функции ниже, D2 превратится в D3 и т.д. – будут изменяться номера строк. При подсчете функции СЧЕТЕСЛИ, нам необходимо без всяких сдвигов считать каждый раз по всему диапазону, поэтому мы ставим знак \$ перед номером строки. К примеру, адрес ячейки D1, где D – обозначение столбца, 1 – обозначение строки. \$D1 – фиксируется столбец, D\$1 фиксируется строка, \$D\$1 фиксируется вся ячейка.

⁶Если установить метку в строке *нормальные значения*, то можно выбрать формулы моделирования для нормального распределения.

3.1.5. Методы корреляционного анализа

Корреляционный анализ используется для проверки надежности, а также валидности в тех случаях, когда критерий валидизации является результатом измерения какого-либо показателя с использованием другой, уже апробированной методики.

Расчет в **Excel**.

- Щелкнув пустую ячейку, в которую необходимо разместить значение коэффициента корреляции, вызываем *Мастер функций* щелчком в значок f_x ;
- в открывшемся окне *Аргументы функции* в строке *Массив1* вводим параметры первого вариационного ряда выделением столбца, в котором расположены значения исследуемой переменной; в строке *Массив2* то же для второй переменной;
- нажимаем ОК.

Способ с использованием пакета *Анализа данных*.

- Входим в меню *Сервис* → *Анализ данных*;
- в открывшемся окне *Анализ данных* в поле *Инструменты анализа* выбираем *Корреляция* и нажимаем ОК;
- в открывшемся окне *Корреляция* в строке *Входной интервал* выделяем расположенные рядом переменные, предназначенные для статистической обработки;
- устанавливаем метку в строке *Выходной интервал* и щелчком в поле, а затем в ячейку вывода вводим место вывода значения(ий);
- нажимаем ОК.

The screenshot shows an Excel spreadsheet with a data table and a 'Function Arguments' dialog box for the CORREL function. The data table has columns V, W, X, and Y. The dialog box shows the following information:

Массив1	Массив2	Результат
W3:W115	X3:X115	= 0,304557905

Возвращает коэффициент корреляции между двумя множествами данных.

Массив2 второй диапазон значений. Значениями могут быть числа, имена, массивы или ссылки с именами.

Значение: 0,304557905

Buttons: [Справка по этой функции](#), ,

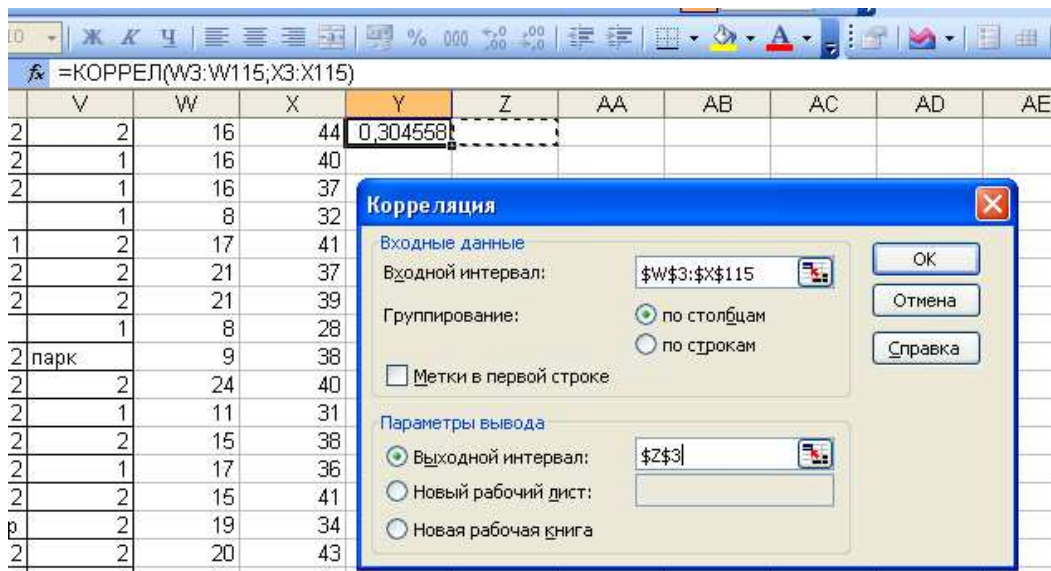


Рис. 3.9. Окна расчета коэффициента корреляции в программе Microsoft Excel

В программе **Microsoft Excel** предусмотрен расчет коэффициента корреляции Пирсона, но предусмотрен и расчет ранговых коэффициентов корреляции Спирмена и Кэндала.

Рассмотрим расчет коэффициентов корреляции в программе **SPSS**.

- Входим в меню *Анализ* → *Корреляции* → *Парные*;
- в открывшемся окне *Парные корреляции* переводим тестируемые переменные в поле *Переменные* щелчком по фигурной стрелке;
- в поле *Коэффициенты корреляции* устанавливаем метку нужного коэффициента;
- в поле *Критерии значимости* устанавливаем метку *Двухсторонний*;
- устанавливаем метку в строке *Метить значимые корреляции*;
- нажимаем ОК.

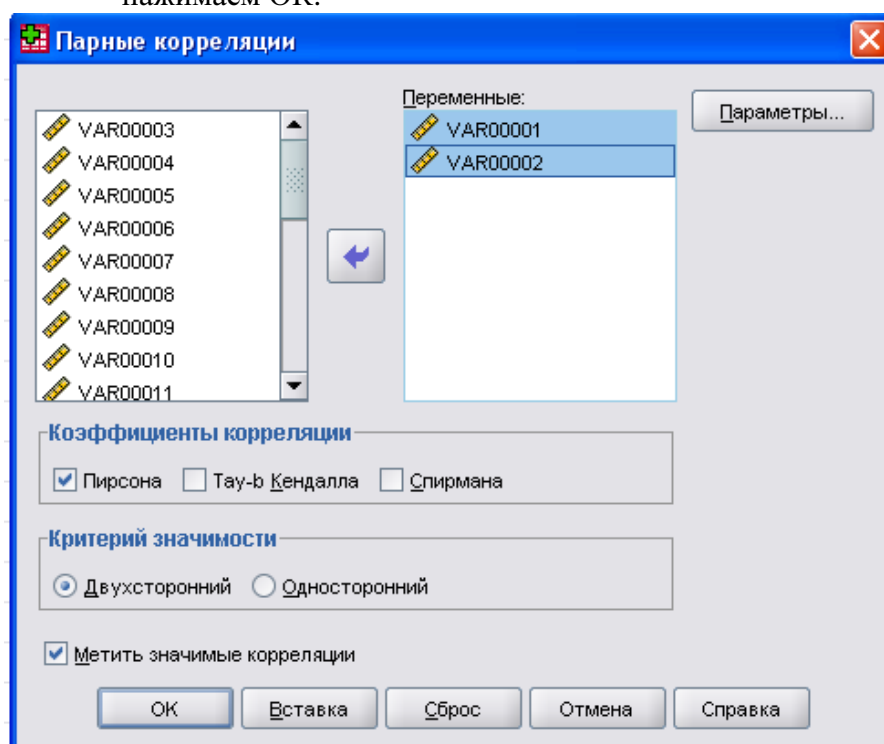


Рис. 3.10. Окно расчета корреляции в программе SPSS

Ассоциативный тест хи-квадрат. Еще один способ установления взаимосвязи между переменными – ассоциативный тест хи-квадрат. Он может быть использован для проверки валидности в ситуации, когда критерий валидации представляет собой категориальную переменную (чаще всего дихотомическую) и распределение баллов по валидируемой шкале не является нормальным, а также для проверки надежности в аналогичных случаях. Существенным преимуществом этого способа определения взаимосвязей по сравнению с подсчетом коэффициентов корреляции является возможность устанавливать наличие не только линейной, но и нелинейной связи. Чтобы выполнить тест хи-квадрат, необходимо создать дополнительную переменную, являющуюся категориальным аналогом переменной, содержащей ответы по шкале. Можно разделить всю выборку на две равные группы в зависимости от того, показатель по шкале меньше или больше выборочной медианы. В новой переменной наблюдения, соответствующие 1-ой группе, имеют значение 1, а второй – 2. Аналогично выборку можно разбить на три-четыре подгруппы.

Расчет в **SPSS**.

- Входим в меню *Анализ* → *Описательные статистики* → *Таблицы сопряженности*;
- в открывшемся окне *Таблицы сопряженности* щелчком по фигурной стрелке переносим в поле *Строки* переменную, содержащую значение по категориальному аналогу валидируемой переменной, и в поле *Столбцы* значения переменной по критерию валидации;
- нажимаем кнопку *Статистика*; в открывшемся окне устанавливаем метку *Хи-квадрат*;
- нажимаем *Продолжить*, затем *ОК*

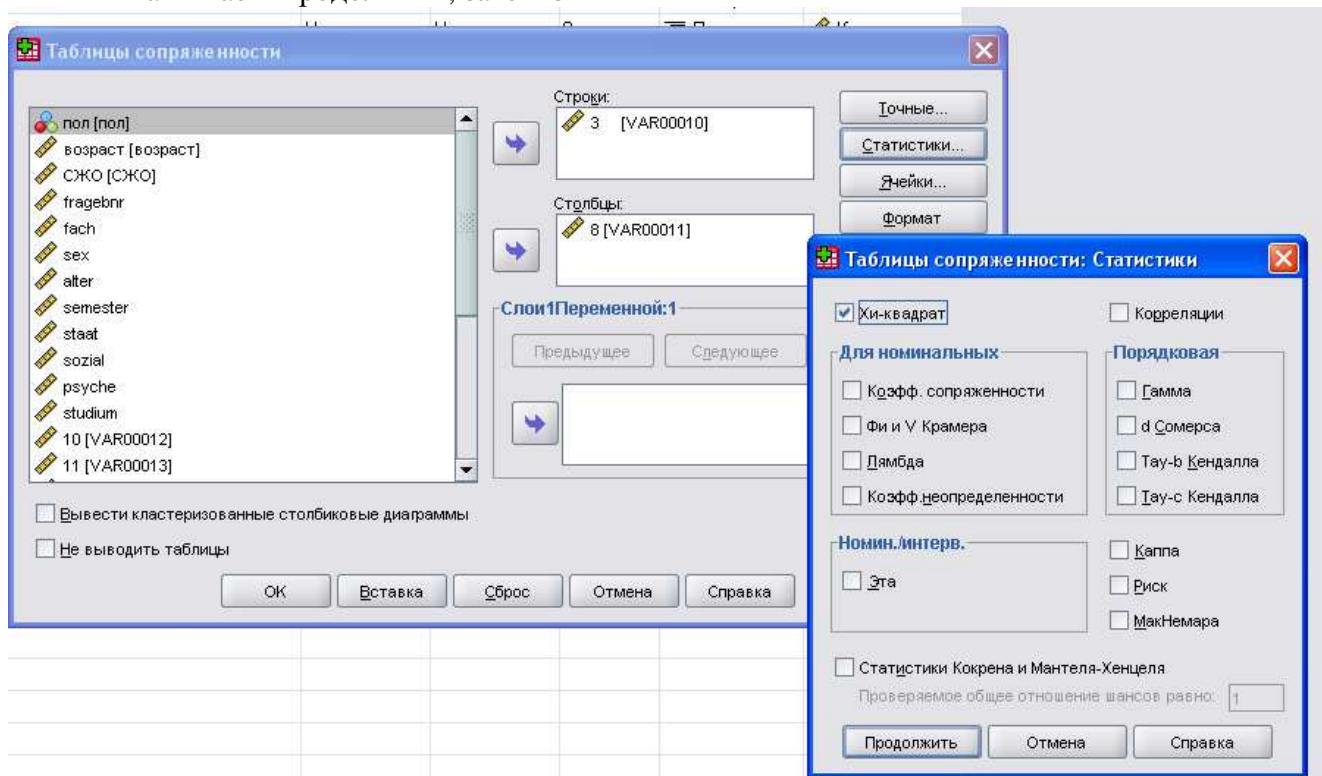


Рис. 3.11. Окна расчета ассоциативного теста хи-квадрат в SPSS

3.1.6. Методы анализа надежности

Анализ согласованности пунктов (вопросов) шкалы теста является одним из методов оценки надежности. Определяется с помощью критерия α -Кронбаха.

Расчет в программе **Microsoft Excel** осуществляется по формулам Кьюдера – Ричардсона или Кронбаха.

Расчет в **SPSS**.

- Входим в меню *Анализ* → *Шкалирование* → *Анализ надежности*;
- в открывшемся окне *Анализ пригодности* щелчком мыши по фигурной стрелке переводим тестируемую переменную в поле *Пункты*;
- в строке *Модель*⁷ выбираем *Альфа* и щелкаем в поле *Статистики*;
- в поле *Описательные для* устанавливаем метки в строке *Шкалы* и *Шкалы, если пункт удален*;
- нажимаем *Продолжить*, затем *ОК*.

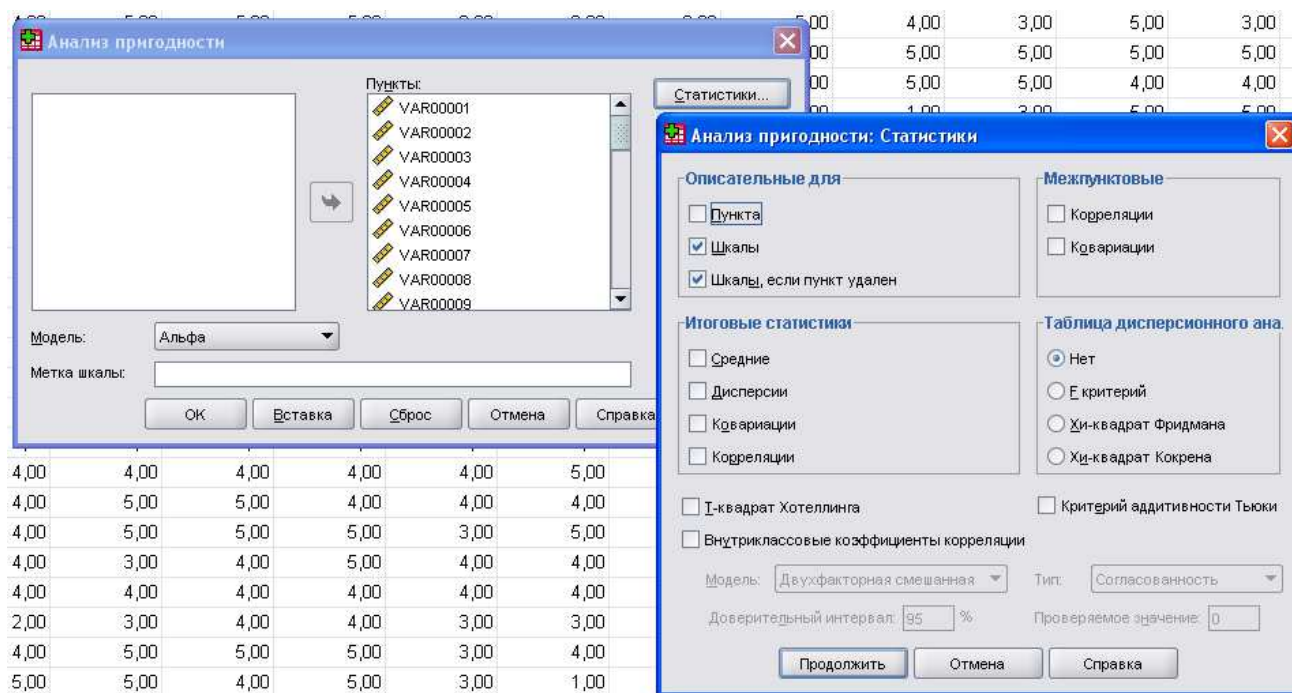


Рис. 3.12. Окна расчета надежности методом Кронбаха в SPSS

3.1.7. Методы анализа выборок

Сравнение выборок на однородность – статистический прием в психометрике, используемый при валидации по методу контрастных групп, а также для установления эмпирического различия между какими-то группами. В первом случае речь идет о подтверждении априорного различия, предполагаемого исходя из теоретических соображений, во втором – апостериорного, полученного в результате анализа конкретных выборочных данных. Апостериорные различия нуждаются в дальнейшем теоретическом осмыслении и обосновании.

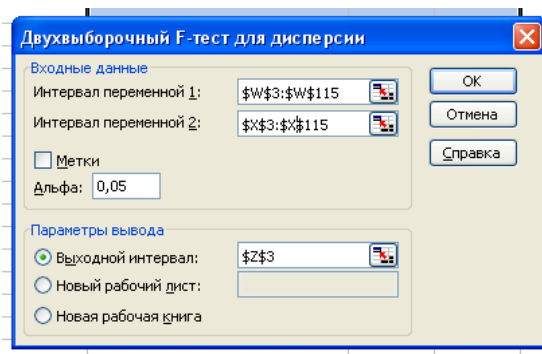
В программе **Microsoft Excel** сравнение выборок на однородность проводится в два этапа и только параметрическими методами сравнения средних (t-критерием Стьюдента). На первом этапе анализируются дисперсии с использованием F-теста (критерия Фишера для сравнения дисперсий) и на его основе выбирается метод сравнения средних с помощью t-теста (критерия Стьюдента).

В программе **SPSS** статистический анализ однородности дисперсий (F – критерий Ливина) и статистический анализ сравнения средних проводятся одновременно. Программа имеет возможности расчетов непараметрических критериев сравнения выборок.

⁷ В строке *Модель* можно выбрать и другой метод для оценки надежности, например

Расчет в Excel.

- Входим в меню *Сервис* → *Анализ данных*;
- в открывшемся окне *Анализ данных* выбираем *Двухвыборочный F-тест для дисперсии*;
- в открывшемся окне *Двухвыборочный F-тест для дисперсии* в строку *Интервал переменной 1* вводим ряд значений первой подвыборки выделением его в соответствующем столбце (строке); аналогично в поле *Интервал переменной 2* вводим ссылку на данные второй подвыборки;
- в строке *Альфа* вводим уровень значимости теста (по умолчанию он равен 0,05);
- устанавливаем параметры *Выходного интервала*;
- нажимаем ОК.



	Переменная 1	Переменная 2
Среднее	0,458566372	0,468929204
Дисперсия	0,078822034	0,084393781
Наблюдения	113	113
df	112	112
F	0,933979173	
P(F<=f) одностороннее	0,359208631	
F критическое одностороннее	0,731855888	

Рис. 3.13. Окно *Двухвыборочный F-тест* и вывод данных

Если уровень значимости критерия ниже установленного значения *Альфа*, то нулевая гипотеза о равенстве дисперсий отклоняется и при дальнейшем статистическом анализе используется *Двухвыборочный t-тест* с различными дисперсиями.

В случае если дисперсии равны, т.е. эмпирическое значение превышает критическое, необходимо проводить *Двухвыборочный t-тест с одинаковыми дисперсиями*.

Для первого случая в строке меню в верхней части программы выбираем *Сервис* → *Анализ данных*, в открывшемся окне *Анализ данных* – *Двухвыборочный t-тест с различными дисперсиями*. Для второго – *Двухвыборочный t-тест с одинаковыми дисперсиями*.

В открывшемся диалоговом окне *Двухвыборочный t-тест с различными (одинаковыми) дисперсиями* в строке *Интервал переменной 1* вводим ссылку на массив данных, содержащий первую подвыборку. Аналогично в строке *Интервал переменной 2* вводим ссылку на массив данных, содержащий вторую подвыборку. В строке *Гипотетическая средняя разность* нужно указать предполагаемую нулевой гипотезой разность средних. В случае если нулевая гипотеза предполагает их равенство, эту ячейку можно оставить пустой. Уровень значимости теста указывается в строке *Альфа*. По умолчанию он равен 0,05. Вводим выходной интервал.

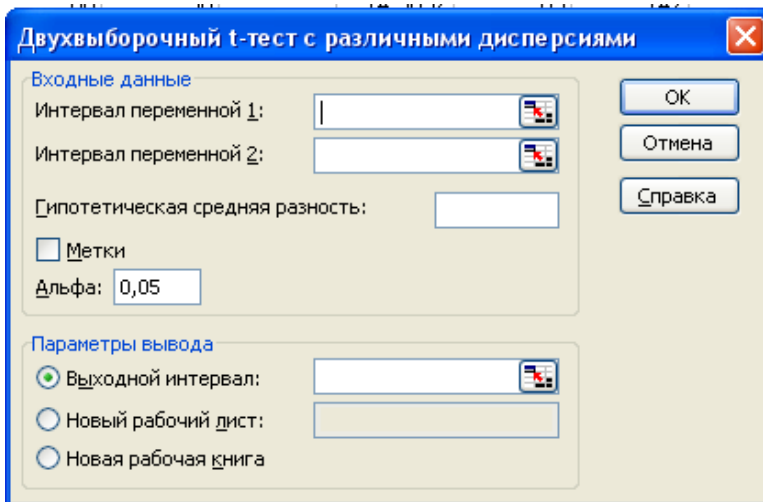


Рис. 3.14. Окно *Двухвыборочный t-тест*

Расчет в **SPSS** по параметрическому критерию.

- Входим в меню *Анализ* → *Сравнение средних* → *T-критерий для независимых выборок*;
- в открывшемся окне *T-критерий для независимых выборок* переводим щелчком по фигурной стрелке тестируемую переменную из левого поля в правое; в поле *Группировать по* щелчком по фигурной стрелке переводим группирующую переменную, которая характеризует подвыборки и щелкаем в клавишу *Задать группы*;
- в открывшемся окне *Задать группы* в строки *Группа 1* и *Группа 2* вводим числовые значения группирующей переменной (контрастных групп);
- нажимаем *Продолжить* и затем *ОК*.

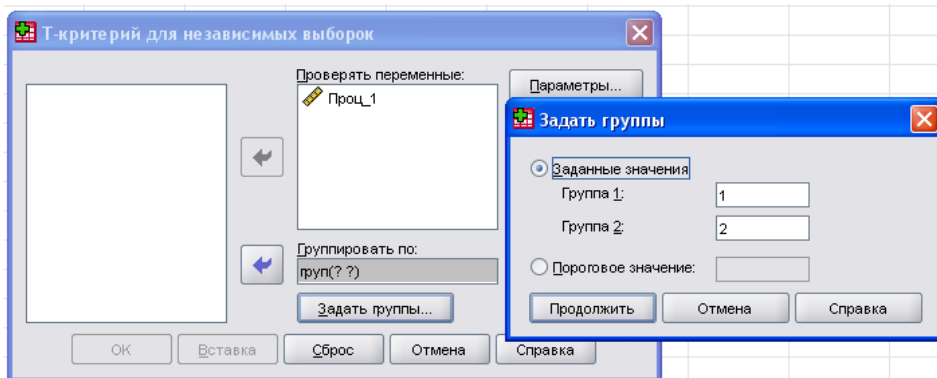


Рис. 3.15. Окно для расчета критерия *Стьюдента* для независимых выборок в программе SPSS

Расчет в **SPSS** по непараметрическому критерию.

- Входим в меню *Анализ* → *Непараметрические критерии* → *Для двух независимых выборок*;
- в открывшемся окне *Критерии для двух независимых выборок* переводим щелчком по фигурной стрелке тестируемую переменную из левого поля в правое; в поле *Группирующая переменная* щелчком по фигурной стрелке переводим группирующую переменную, которая характеризует подвыборки, и щелкаем клавишу *Задать группы*;

- в открывшемся окне *Задать группы* в строки *Группа 1* и *Группа 2* вводим числовые значения группирующей переменной (контрастных групп); нажимаем *Продолжить*;
- в поле *Критерии* устанавливаем метку в строке *U Манна – Уитни* и щелкаем *ОК*.

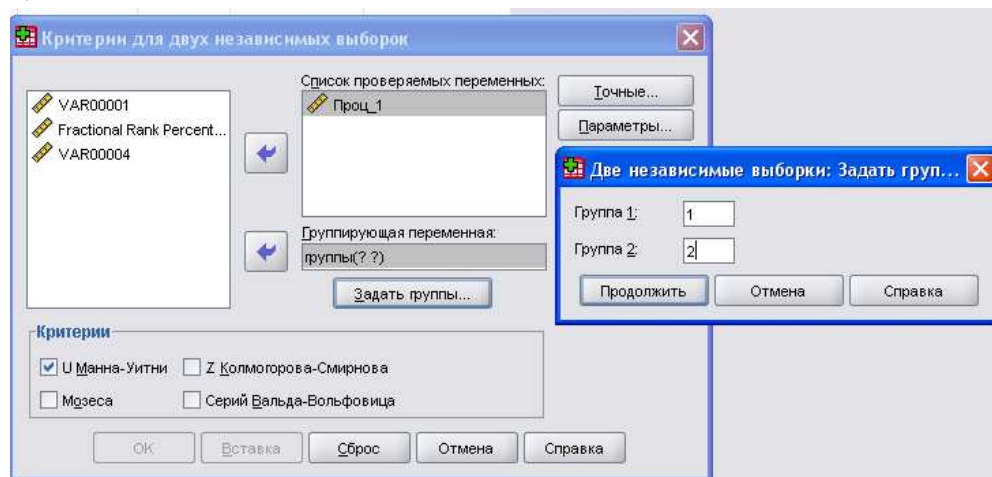


Рис. 3.16. Окно для расчета непараметрических критериев (критерия *U Манна – Уитни*)

3.1.8. Эксплораторный факторный анализ многих переменных в SPSS.

Эксплораторный или исследовательский, разведочный факторный анализ – это такая схема факторного анализа, при которой исследователь изначально не знает, какая система факторов позволяет описать матрицу корреляционных связей. Для проведения эксплораторного факторного анализа с помощью SPSS вставьте переменные в окно *Данные*. В строке меню в верхней части окна программы выберите пункт *Анализ*, в открывающемся меню пункт *Снижение размерности* → *Факторный анализ*, что приведет к открытию диалогового окна *Факторный анализ*. Перенесите щелчком по фигурной стрелке из общего окна с переменными (слева) в верхнее правое окно *Переменные* переменные, содержащие значения баллов по всем пунктам.

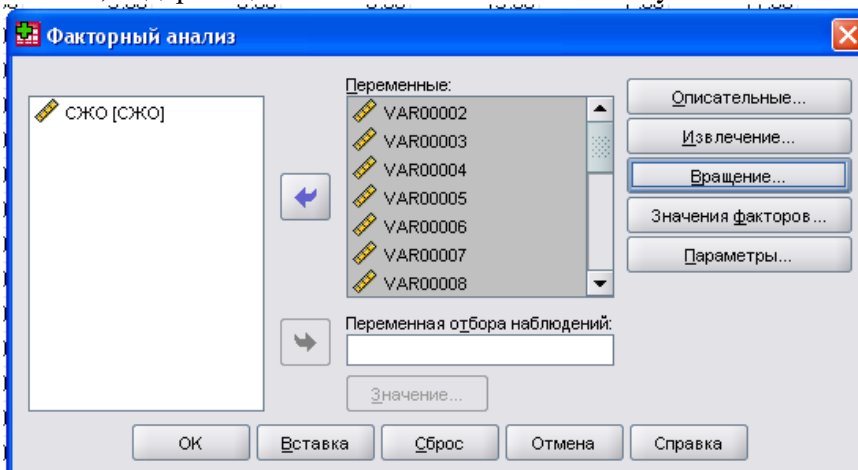


Рис. 3.17. Окно *Факторный анализ* в программе SPSS

Не закрывая диалоговое окно, нажмите кнопку *Описательные*, что приведет к открытию следующего диалогового окна.

Собственно опции этого окна не содержат команд, касающихся факторного анализа, однако некоторые из окошек полезно отметить, так как выполнение соответствующих команд является своеобразной формой контроля: все ли данные правильно прочитаны, нет

ли где ошибочных пропусков и т.д.⁸ Отметьте флажками в группе *Статистики – Одномерные описательные*, а в группе *Корреляционная матрица – Коэффициенты и Уровни значимости, Детерминант, КМО и критерий сферичности Бартлетта (Тест Кайзера–Мейера–Олкина и коэффициент сферичности Бартлетта)*. Нажмите кнопку *Продолжить* и вернитесь к предыдущему окну.

Нажмите кнопку *Извлечение*, чтобы открыть диалоговое окно *Факторный анализ: Извлечение факторов*. В строке *Метод* выбираем метод выполнения факторизации *Главные компоненты*. В группе *Вывести на дисплей* установите флажок *График собственных значений*, чтобы получить график собственных значений («каменистую осыпь»). По умолчанию в SPSS предполагается выполнять 25 итераций для выполнения процедуры извлечения факторов. В случае небольшого числа переменных этого вполне достаточно, однако при анализе наборов данных, содержащих большое число переменных, для сходимости алгоритма может потребоваться более 25 итераций. В этом случае следует увеличить значение в поле *Максимальное число итераций для сходимости* до 50 или 100 вначале, и более – в случае необходимости. В качестве критерия для выделения значимого количества факторов в SPSS по умолчанию используется критерий Кайзера, оставляющий факторы с собственными значениями, превосходящими единицу. Однако можно оставить меньше или больше факторов. Для этого нужно в группе *Выделить* указать число факторов в поле рядом с переключателем *Фиксированное количество факторов*. Нажмите кнопку *Продолжить*, чтобы вернуться в диалоговое окно Факторного анализа.

Нажатие кнопки *Вращение* приводит к открытию диалогового окна *Факторный анализ: Вращение*.

Как правило, косоугольное вращение является более адекватным при анализе данных психометрических исследований. Чтобы его выполнить, установите переключатель *Прямой облимин*. Чтобы вывести график размещения первичных переменных в факторном пространстве, в группе *Вывести* отметьте флажком *Графики нагрузок*. Нажмите кнопку *Продолжить*, чтобы вернуться в диалоговое окно *Факторного анализа*.

Чтобы вывести факторные значения (баллы) (factor scores), необходимо вызвать диалоговое окно *Факторный анализ: Значения факторов*. Для этого в диалоговом окне *Факторный анализ* нужно щелкнуть кнопку *Значения факторов*. В открывшемся окне установить метку *Сохранить как переменные*. Метод по умолчанию оставьте *Регрессия*, нажмите кнопку *Продолжить*.

Если вы хотите вывести факторные нагрузки в порядке убывания их величин⁹, нажмите кнопку *Параметры*, чтобы открыть диалоговое окно *Факторный анализ: Параметры*.

В группе *Формат вывода коэффициентов* установите флажок *Отсортировать по величине*. Нажмите кнопку *Продолжить*, чтобы вернуться в диалоговое окно *Факторного анализа*.

Нажмите на кнопку ОК, чтобы провести анализ¹⁰.

⁸ Однако предпочтительнее проводить специальную подготовку данных («чистку»), как описывается ниже.

⁹ Эта возможность бывает полезной, когда переменных много и выбирать имеющие максимальные факторные нагрузки по тому или иному фактору из большого массива данных оказывается затруднительно.

¹⁰ Как правило, эту процедуру приходится выполнять как минимум два раза. Вначале вы получаете график следа, на его оснований принимаете решение о числе значимых факторов и устанавливаете это число для вращения. Кроме того, если после косоугольного вращения коэффициенты интеркорреляций факторов оказались очень маленькими, целесообразно выполнить также ортогональное вращение varimax и сравнить получаемые в обоих случаях факторные нагрузки.

Раздел 4. Содержание и структура лабораторных работ курса «Основы психометрии»

Занятие 1.

1. Создание сводной матрицы данных и базы данных в программе Microsoft Excel.
2. Расчет описательной статистики в программе Microsoft Excel.
3. Создание таблиц в программе Microsoft Excel.

Тема: Ввод данных.

Цель: Получить навык в формировании файлов данных и матриц данных.

Порядок работы:

1. Дать характеристику программ статистической обработки данных: Excel, SPSS.
2. Создать файл данных, матрицу данных.
3. Рассчитать основные статистические показатели мер центральной тенденции, мер изменчивости, мер, характеризующих распределение.
4. Построить сводные таблицы статистической обработки данных. Описать и дать интерпретацию полученным показателям.

Занятие 2.

1. Создание таблиц распределения частот.
2. Построение графиков распределения частот.
3. Построение таблиц сопряженности.

Тема: Процедуры оценки нормальности распределения.

Цель: Получить навык в оценке нормальности распределения тестовых баллов и способах их интерпретации.

Порядок работы:

1. Дать характеристику параметрам нормального распределения.
2. Построить интервальную таблицу, отражающую кумуляты накопленных частот.
3. Оценить нормальность распределения с помощью статистических критериев Колмогорова – Смирнова, Шапиро – Уилки, хи-квадрат.
4. Сделать вывод о нормальности распределения на основе показателей статистических критериев.

Занятие 3.

1. Нормирование данных психологической диагностики в программе Microsoft Excel.

Тема: Процедуры создания норм теста

Цель: Получить навык нормирования тестовых баллов.

Порядок работы:

1. Охарактеризовать метрические шкалы и метрические единицы измерения психологических свойств.
2. Перевод сырых баллов в z-взвешенные результаты.
3. Перевод сырых баллов в процентиля (перцентили). Определение квартилей и квантилей распределения.
4. Перевод сырых баллов в стеноны.
5. Перевод сырых баллов в T-баллы и Q-баллы.
6. Составить таблицы перевода сырых баллов в метрические.

Занятие 4.

1. Концептуальная валидность одномерной шкалы теста.

Тема: Построение первичной формы теста

Цель: Получить навык разработки концептуальной валидности одномерной шкалы теста.

Порядок работы:

1. Дать определение понятиям «содержательная и концептуальная валидность».
2. Дать психологическое описание (феноменологию) измеряемой психологической характеристике согласно теоретической концепции, определить ее как понятие, выделить ее аспекты.
3. Составить тестовую шкалу из 12 вопросов, которые могут измерить заданную психологическую характеристику (свойство) и раскрывают выделенные аспекты (от 3 до 6 по каждому аспекту).
4. Составить психологический портрет сильновыраженного и слабовыраженного проявления измеряемой характеристики.

Занятие 5.

1. Метод экспертной оценки концептуальной валидности.

Тема: Построение первичной формы теста

Цель: Получить навык использования метода экспертной оценки в изучении валидности тестовой шкалы.

Порядок работы:

1. Охарактеризовать методы оценки валидности теста.
2. Разработать шкалу теста, включающую 12 вопросов, обозначив способ формулирования вопросов и систему оценки ответов.
3. Провести экспертную оценку сформулированных вопросов по предложенной схеме. Результаты оформить в виде таблицы (см. Таблица 6. Схема применения экспертной оценки с использованием балльной шкалы).
4. Сделать вывод о валидности на основе экспертных оценок.

Занятие 6.

1. Эмпирическая валидность одномерной шкалы теста.
2. Статистические процедуры оценки эмпирической валидности.

Тема: Эмпирическая валидность одномерной шкалы теста.

Цель: Получить навык в оценке эмпирической валидности одномерной тестовой шкалы.

Порядок работы:

1. Дать определение понятию «эмпирическая валидность».
2. Составить матрицу эмпирических данных по результатам тестирования в сырых баллах.
3. Составить таблицу сопряженности, используя результаты критериальной и валидируемой переменных.
4. Провести расчет коэффициента валидности.
5. Сделать вывод о степени валидности шкалы теста.

Занятие 7.

1. Использование факторного анализа для оценки валидности.
2. Процедура факторного анализа в программе SPSS for Windows 17 Russian version.
3. Интерпретация результатов факторизации.

Тема: Внутренняя валидность шкалы теста.

Цель: Получить навык в использовании факторного анализа для установления внутренней валидности теста.

Порядок работы:

1. Дать определение понятиям «внутренняя и внешняя валидность».

2. Экспортировать данные из программы Microsoft Excel в программу SPSS.
3. Выполнить процедуру эксплораторного факторного анализа.
4. Описать факторную структуру теста.
5. Сделать вывод о валидности факторной структуры.

Занятие 8.

1. Надежность как внутренняя согласованность шкалы теста.
2. Анализ надежности методом расщепления.
3. Анализ надежности методом Кронбаха.
4. Процедура анализа надежности в программе SPSS for Windows 17 Russian version.

Тема: Надежность тестовых шкал.

Цель: Получить навык в оценке надежности тестовых шкал.

Порядок работы:

1. Дать определение понятию «надежность как устойчивость и внутренняя согласованность теста». Описать методы определения надежности.
2. Провести расчет надежности методом расщепления.
3. Провести расчет методом Кронбаха.
4. Сделать вывод о надежности тестовой шкалы.

Занятие 9.

1. Анализ трудности заданий теста. Расчет индекса трудности в программе Microsoft Excel.
2. Порядок заданий (пунктов) теста в зависимости от индекса трудности заданий.

Тема: Первичный анализ тестовых заданий.

Цель: Получить навык в расчете индекса трудности задания и описания контрастных групп.

Порядок работы:

1. Дать определение понятию «статистическая трудность задания теста» и описать, для каких задач она используется.
2. Провести расчет индекса трудности заданий теста.
3. Построив ранжированный ряд индексов трудности, провести анализ пунктов теста, обосновав психологическое содержание пунктов, вызывающих трудности.
4. Выделить контрастные группы: высокопродуктивную и низкопродуктивную. Провести интерпретацию пунктов теста, имеющих наибольшую дифференцирующую способность.

Занятие 10.

1. Дискриминативность пунктов теста.
2. Расчет дискриминативности в программе Microsoft Excel.

Тема: Анализ тестовых заданий.

Цель: Получить навык в оценке дискриминативности пунктов теста.

Порядок работы:

1. Дать определение понятию «дискриминативность тестовых заданий».
2. Провести расчет ИД (индекса дискриминативности) как разности долей правильных ответов высокопродуктивной и низкопродуктивной групп.

$$r_{\text{дис}} = ((p_1)_j - (p_0)_j) / 100,$$

где p_{1j} – доля правильных ответов высокопродуктивной группы,

p_{0j} – доля правильных ответов низкопродуктивной группы.

3. Провести расчет коэффициента дискриминативности Фергюсона.

Полученные данные занести в таблицу:

№ задания	p_j (для всех испытуемых)	Группа		$\Gamma_{\text{дис}}$	δ
		p_{1j}	p_{0j}		

4. Дать интерпретацию полученных расчетов.

Занятие 11.

1. Виды тестов достижений.
2. Тесты достижений в практике образования.

Тема: Разработка тестовой шкалы в тестах достижений.

Цель: Получить навык в формулировании тестовых заданий в тестах достижений.

Порядок работы:

1. Требования к тестовым заданиям в тестах достижений.
2. Виды тестовых заданий в тестах достижений.
3. Разработать следующие виды тестовых заданий по «Общей психологии»:
 - 4 тестовых задания с выбором одного правильного ответа;
 - 4 тестовых задания с выбором нескольких правильных ответов;
 - 2 задания на установления соответствия;
 - 2 задания на установления правильной последовательности;
 - 2 задания на дополнение;
 - 1 задание с градуированным ответом.
4. Оформить разработанные задания в тестовую шкалу.
5. Описать систему оценки тестовых заданий и тестовой шкалы.

Литература

Основная

1. Анастаси А. Дифференциальная психология. Индивидуальные и групповые различия в поведении / Пер. с англ. – М.: Апрель Пресс, Изд-во ЭКСМО-Пресс, 2001. – 752 с.
2. Анастаси А., Урбина С. Психологическое тестирование. – СПб.: Питер, 2009. – 688 с.
3. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. – СПб.: Питер, 2001. – 528 с.
4. Клайн П. Справочное руководство по конструированию теста. – Киев: ПАН Лтд, 1994. – 288 с.
5. Кондаков И.М. Создание психологических опросников с помощью статистического пакета SPSS for Windows 11.5.0. Учебно-методическое пособие / Электронный ресурс [<http://www.matlab.mgppu.ru/work/0028.htm>]
6. Майкл Фер Р. Психометрика: Введение / Р. Майкл Фер, Верн Р. Бакарарк; пер. с англ. А.С. Науменко, А.Ю. Попова; под ред. Н.А. Батурина, Е.В. Эйдмана. – Челябинск: Издательский центр ЮУрГУ, 2010. – 445 с.
7. Мельников В. М., Ямпольский Л. Т. Введение в экспериментальную психологию личности: Учеб. пособие. – М.: Просвещение, 1985. – 319 с.
8. Митина О.В. Математические методы в психологии: Практикум / О.В. Митина. – М.: Аспект Пресс, 2008. – 238 с.
9. Митина О.В. Разработка и адаптация психологических опросников. – М.: Смысл, 2011. – 235 с.

10. Митина О.В., Михайловская И.Б. Факторный анализ для психологов. – М.: Учебно-методический коллектор «Психология», – 2001. – 169 с.
11. Наследов А.Д. SPSS: Компьютерный анализ данных в психологии и социальных науках, 2-е изд. – СПб.: Питер, 2007. – 416 с.
12. Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие. – СПб.: Речь, 2004. – 329 с.
13. Носс И.Н. Введение в технологию психодиагностики. – М.: Изд-во Института Психотерапии, 2003. – 251 с.
14. Общая психодиагностика. – СПб.: Изд-во «Речь», 2000. – 440 с.
15. Практикум по общей, экспериментальной и прикладной психологии: Учеб. пособие / В. Д. Балин, В. К. Гайда, В. К. Гербачевский и др. // Под общей ред. А.А. Крылова, С.А. Маничева. – СПб.: Питер, 2000. – 560 с.

Электронные ресурсы

- <http://psyfactor.org/lybr10.htm>
<http://www.ht.ru/cms/development>
http://lab.psytest.ru/ru/Poleznaja_informacija/Izbrannye_stati
<http://www.jullyaa.narod.ru/>
http://tukachev.flogiston.ru/blog/?page_id=2
<http://www.психометрика.рф/>

Дополнительная

1. Бешелев С.Д., Гуреич Ф.Г. Математико-статистические методы экспертных оценок. – М.: Статистика, 1980. – 263 с.
2. Бурлачук Л.Ф. Психодиагностика: Учебник для вузов / Л. Ф. Бурлачук. – СПб.; Питер, 2003. – 351 с.
3. Гайдышев И. Анализ и обработка данных: Специальный справочник. – СПб.; Питер, 2001. – 752 с.
4. Гильбух Ю.З. Актуальные проблемы валидизации психологических тестов // Вопросы психологии. – 1978. – № 5. – С. 108-117.
5. Гильбух Ю.З. Надежность психологических тестов и пути ее повышения // Вопросы психологии. – 1979. – № 3. – С. 96-105.
6. Гильбух Ю.З. Проблема валидности эксперимента в психологии // Психологический журнал. – 1987. – № 4. – Т. 8. – С. 117-125.
7. Гильбух Ю.З. Проблема теоретического обоснования предмета испытаний при разработке психологических тестов // Вопросы психологии. – 1982. – № 1. – С. 29-39.
8. Горбачева Е.М. Опыт конструирования диагностического критериально-ориентированного теста // Вопросы психологии. – 1985. – № 5. – С. 133-139.
9. Гринберг Я.З. Надежность психологического тестирования: индивидуальные и групповые данные // Психологический журнал. – 1990. – № 5. – Т. 11. – С. 50-55.
10. Дюк В.А. Компьютерная диагностика. – СПб.: «Братство», 1994. – 364 с.
11. Ермолаев О.Ю. Математическая статистика для психологов: Учебник /О.Ю. Ермолаев. – 2-е изд., испр. – М.: Московский психолого-социальный институт Флинта, 2003. – 336 с.
12. Забродин Ю.М., Похилько В.И. Статистические и семантические проблемы конструирования и адаптации многофакторных межличностных тест-опросников // Психологический журнал, 1987, № 6. – С. 79-89.
13. Калинин С.И. Компьютерная обработка данных для психологов / Под науч. ред. А.Л. Тулупьева. – СПб.: «Речь», 2002. – 134 с.

14. Нормативные предписания к разработчикам и пользователям психодиагностических методик // ВП/5, 1987. – С. 176-181.
15. Факторный, дискриминантный и кластерный анализ: Пер с англ./Дж. – О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; Под ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
16. Фишер Р.А., Статистические методы для исследователей, пер. с англ., – М., Госстатиздат, 1958. – 268 с.
17. Шмелев А.Г. Введение в экспериментальную психосемантику. – М.: Изд-во Моск. ун-та, 1983. – 157 с.
18. Шмелев А.Г. Психодиагностика личностных черт. – СПб.: Речь, 2002. – 480 с.