РЕФЕРАТ

**«Основные принципы тестирования»**

**Содержание**

1. Математическая статистика
2. Понятие «валидность», «надежность»
3. Основные статистические понятия (меры центральной тенденции)
4. Понятие математической статистики

Математическая статистика занимается математическим описанием случайных явлений, т. е. построением вероятностных моделей, а также проверкой их пригодности. Математическая статистика – прикладная отрасль математики, основанная на теории вероятности и предназначенная для систематизации и анализа эмпирических данных, полученных при изучении массовых явлений.

Существует 3 раздела статистики:

* описательная
* индуктивная
* корреляционный анализ.

Описательная статистика позволяет описывать, подытоживать и воспроизводить в виде таблиц или графиков данные того или иного распределения, вычислять среднее, его размах и дисперсию.

Индуктивная статистика (проверяющая). Ее задача – проверка того, можно ли рассматривать результаты, полученные на данной выборке на всю популяцию, из которой взята эта выборка; позволяет выяснить до какой степени путем индукции на большое количество объектов обобщить закономерности, полученные при изучении их ограниченной группы в ходе конкретного эксперимента.

Корреляционный анализ позволяет узнать, на сколько связаны две переменные с целью предсказания возможных значений одной из них, если известны значения другой.

Статистические методы, которые позволяют делать обобщение и измерять корреляцию.

* параметрические, которые служат для проверки гипотез о параметрах некоторой функции или для их оценивания, и в которых используются такие параметры среднего значения или дисперсия данных.
* непараметрические используются, когда исследователь имеет дело с очень малыми выборками, некоторые функции от функции распределения изученных случайных явлений. Непараметрические критерии служат для проверки гипотез о функции распределения наблюдаемых значений.

Выделяют 3 типа данных:

* количественные, полученные при измерениях;
* качественные, свойства элементов в выборке или популяции, их нельзя измерить и единственной их количественной оценкой служит частота встречаемости
* порядковые – данные соответствуют местам этих методов, полученных при их расположении в возрастающем порядке.

1. Понятие «валидность», «надежность»

Важнейшими показателями качества психодиагностических методик являются надежность и валидность. Надежность психодиагностических методик свидетельствует о повторяемости, стабильности результатов, об их постоянстве и устойчивости. Оно показывает, на сколько точно производятся психологические измерения, насколько можно доверять получаемым результатам.

Известный специалист в области психодиагностики К.М. Гуревич предлагает выделить три типа надежности: надежность самого измерительного инструмента, стабильность изучаемого признака и константность, т.е. независимость результатов от личности экспериментатора.

Надежность теста показывает независимость его результатов от действия разных случайных факторов. Разнообразные внешние и внутренние факторы могут вызывать отклонение результатов теста.

Для проверки стабильности изучаемого признака используется метод, получивший название «mecm – pemecm», который заключается в проведении повторного психодиагностического испытания той же выборки испытуемых (не менее 30 человек) через определенный промежуток времени, в вычислении коэффициента корреляции между результатами первого (X) и второго (Y) тестирования. Этот коэффициент и представляет собой показатель стабильности исследуемого признака, т.е. надежность теста проявляется в том, что одни и те же люди при повторном обследовании дают те же результаты.

Как правило, повторное обследование проводится через несколько месяцев (но не более чем через полугода). Нельзя проводить повторное испытание слишком быстро после первого, поскольку есть опасность, что испытуемые будут воспроизводить свои ответы по памяти. Однако этот срок не может быть слишком большим, по сколько в этом случае возможно изменение, развитие самой исследуемой функции. Коэффициент стабильности считается приемлемым в том случае, когда его величина не ниже 0,8.

Коэффициент константности определяется путем корреляции результатов двух психодиагностических испытаний, проведенных на одной и той же выборке испытуемых с соблюдением идентичности условий, но разными экспериментаторами. Он должен быть не ниже 0,8.

Качество методики определяется тем, на сколько хорошо она составлена, насколько однородна, что свидетельствует о ее направленности на диагностику одного и того же свойства, признака.

Для проверки надежности инструмента по показателю однородности используют метод расщепления – для этого все задания психодиагностического инструмента делятся на четные и нечетные (по нумерации), отдельно обрабатываются, а затем подсчитываются коэффициенты корреляции между этими рядами. Чем выше величина коэффициента корреляции, тем однороднее методика, тем выше ее надежность.

Другим показателем качества методики является ее валидность. По определению видного американского тестолога А. Анастази, «валидность теста – понятие, указывающее нам, что тест измеряет и насколько хорошо он это делает». Валидность свидетельствует о том, пригодна ли методика для измерения определенных качеств, особенностей и на сколько эффективно она это делает.

Наиболее распространенным способом нахождения теоретической валидности методики является конвергентная валидность, т.е. сопоставление данной методики с авторитетными родственными методиками т доказательства значимых связей с ними. Сопоставление с методиками, имеющими другое теоретическое основание и констатация отсутствия значимых связей с ними, называется дискриминантной валидностью.

Другой вид валидности – прагматическая валидность – проверка методики с точки зрения ее практической значимости, эффективности, полезности. Для проведения такой проверки, как правило, используются так называемые независимые внешние критерии, т.е. используется независимый от теста, внешний источник информации о проявлении в реальной жизни и деятельности людей измеряемого психического свойства. Среди таких внешних критериев могут быть успеваемость, профессиональные достижения, успехи в разных видах деятельности, субъективные оценки (или самооценки). Если, например, методика измеряет особенности развития профессионально-важных качеств, то для критерия необходимо, найти такую деятельность или отдельные операции, где именно эти качества реализуются.

Для проверки валидности теста можно использовать метод «известных групп», когда приглашаются люди, про которых известно, к какой группе по критерию они относятся (например, группа «высокоуспешных, дисциплинированных студентов» – высокий критерий и группа «неуспевающих, недисциплинированных студентов» – низкий критерий, а студенты со средними значениями не участвуют в тестировании), проводят тестирование и находят корреляцию между результатами теста и критерием.

|  |  |  |
| --- | --- | --- |
|  | Высокий критерий | Низкий критерий |
| Высокий результат теста | а | в |
| Низкий результат теста | с | d |

Здесь а – число испытуемых, попавших в высокую группу по тесту и по критерию, с – число испытуемых, попавших в высокую группу по критерию и имеющих низкие результаты теста. При полной валидности теста элементы в и с должны быть равны нулю. Меру совпадения, корреляции между крайними группами по тесту и критерию оценивают с помощью фи-коэффициента Гилфорда.

Если было 30 человек испытуемых, то статистически значимая связь теста с критерием можно констатировать если Р больше 0,36.

Что касается величины коэффициентов валидности, то она по разным причинам всегда ниже, чем коэффициент надежности. По мнению ведущих психодианостов, низким признается коэффициент валидности порядка 0,20 – 0,30, средним – 0,30 – 0,50, высоким – выше 0,60.

В зависимости от специфики самой диагностической методики и внешнего критерия различают несколько ее видов.

1. Валидность «по одновременности», или текущая валидность. Определяется с помощью внешнего критерия, информация по которому собирается в период проведения испытаний (например, успеваемость, производительность труда и др.).

2. Прогностическая, или «предсказывающая», валидность отличается тем, что информация по внешнему критерию собирается некоторое время спустя после проведения обследования. Такая валидность свидетельствует о вожможностях методики предсказывать успешность испытуемого в каких-либо видах деятельностию

Для проверки прогностической валидности теста нужно обследовать более 300 человек, например, абитуриентов, а через 2-3 года выделить среди этих людей критериальные группы «отличников» и «двоечников» и посчитать корреляцию с прежними показателями теста. Если корреляция выше 0,4-0,6, то тест можно использовать для профотбора абитуриентов и прогноза их учебной успешности.

3. Ретроспективная валидность определяется на основе критерия, отражающего события или состояния качества в прошлом. Она также может свидетельствовать о предсказательных возможностях методики.

Стандартизированность психодиагонсотической методики предполагает, с одной стороны, единую стандартную процедуру ее проведени и обработки результтатов, а с другой стороны, включает перевод «сырых первоначальных результатов теста» в стандартную шкалу тестовых баллов, в результате чего можно сопоставлять результаты у разных испытуемых в разных тестах. Выборка, на которой определяются статистические тестовые нормы, называется выборкой стандартизации» (она должна быть не менее 200 человек).

Обобщенная формула перевода сырого балла в заданную стандартную шкалу имеет вид:

Y = ДZ + M,

где Y – стандартный балл по произвольной стандартной шкале с центром М и отклонением Д.

Для многих тестов используют более сложную процедуру нелинейной нормализации (переход к нормальному распределению), в результате чего создают «конверсионную таблицу» для перевода сырых баллов в стандартные баллы по заданной шкале, в которой каждому «сырому баллу» указан соответствующий стандартный балл по определенной шкале, например, шкале «стенов», причем диагносту уже не дано делать самостоятельных вычислений по нормированию баллов.

Следует учитывать репрезентативность тестовых норм – правомерность применения тестовых норм в большой группе людей. Если тестовые нормы были вычислены на выборке школьников, то эти нормы нельзя автоматически переносить на студентов – надо провести рестандартизацию теста, проверить его на студентах и вычислить тестовые нормы для студентов.

Кроме статистических тестовых норм часто используют критериальные нормы, т.е. достиг ли человек критического требуемого уровня развития профессионально важного психологического качества. В критериальных тестах учитывают не степень отклонения баллов от центра шкалы, а достижение или недостижение какого-то критического уровня на шкале.

Следует учитывать и достоверность теста – способность теста защищать информацию от мотивационных (сознательных и бессознательных) искажений и социальной желательности ответов (это достигается через введение в тест проверочной шкалы лжи).

В случае, если тестирование проводится в интересах и по просьбе самого человека (ситуация консультирования), то ответы человека более правдивы. Если тестирование проводят по инициативе администрации, например в целях профотбора (ситуация экспертизы), то испытуемый склонен давать «социально-желательные», нужные, выгодные «ответы», т.е. эти ответы являются сознательно или бессознательно искаженными, и в этих случаях нельзя использовать тесты без шкалы достоверности.

Измерение психометрических характеристик теста осуществляют разработчики тест, но пользователи должны знать и понимать, на сколько тот или иной тест надежен, валиден, стандартизирован, достоверен и репрезентативен.

Таким образом, качество любой психодиагностической методики зависит от степени ее стандартизованности, надежности и валидности. При разработке любой диагностической методики ее авторы должны проводить соответствующую проверку и сообщать в руководстве к ее применению полученные результаты.

Главная цель статистического метода – представить количественные данные в систематизированной и сжатой форме с тем, чтобы облегчить их понимание. Колонка из 1000 тестовых показателей может производить внушительное, даже ошеломляющее впечатление. Но в таком виде она мало что говорит. Чтобы навести порядок в этом хаосе цифр, нужно прежде всего составить таблицу частного распределения.

|  |  |
| --- | --- |
| Классы (интервалы) | Частота |
| 52 - 55  48 - 51  44 - 47  40 - 43  36 - 39  32 - 35  28 - 31  24 - 37  20 - 23  16 - 19  12 - 15  8 - 11  Всего | 1  1  20  73  156  328  244  136  28  8  3  2  1000 |

Для этого показатели группируются по заранее выработанным интервалам значений. Когда же показатели распределены по группам, подсчитываются число групп и число показателей в каждой из них. Полученное таким способом число и есть частота (количество случаев) для соответствующего интервала. Сумма всех частот равняется N –общему числу случаев. В таблице даны результаты 1000 студентов по тесту на усвоение кода, в котором производилась замена искусственных слов или бессмысленных слогов из одного набора аналогичными элементами из другого набора. Значения первичного показателя (число правильных ответов, данных испытуемым за 2 минуты) уложились в пределы от 8 до 55. Этот диапазон был разбит на интервалы по 4 очка в каждом: от 8-11 до 52-55. Из колонки частот видно, что результаты двух испытуемых находятся в интервале между 8 и 11, трех – между 12 и 15 и т.д.

Информация, содержащаяся в частотном распределении, может быть так же представлена графически в виде кривой распределения.

Этот тип кривой обладает важными математическими свойствами, и на ней основаны многие виды статистического анализа. Кривая может принимать и другие формы по которым можно будет судить о сложности и легкости теста.

По существу наша кривая означает, что число случаев максимально в середине распределения и постепенно спадает к ее краям. Кривая симметрична и имеет единственный пик в центре. Большинство распределений численных показателей – от роста и веса до способностей и параметров личности – приближаются к нормальной кривой. Вообще говоря, чем больше группа, тем ближе распределение к теоретической нормальной кривой.

1. Основные статистические понятия (меры центральной тенденции)

Группа тестовых показателей может быть описана в терминах той или иной меры центральной тенденции. Такая мера указывает единственный, наиболее типичный или репрезентативный результат, характеризующий выполнение теста всей группой. Самой известной из таких мер является среднее (точнее среднеарифметическое) значение (М). Оно находится сложением всех результатов и делением получившейся суммы на число случаев (N). Другой мерой центральной тенденции является мода, или наиболее часто встречающийся результат. В частном распределении мода определяется как середина интервала, для которого частота максимальна. Например в нашей таблице мода находится посередине между 32 и 35, т.е. равна 33,5. Отметим, что этот результат соответствует самой высокой точке кривой распределения на рисунке. Третья мера центральной тенденции – это медиана, т.е. результат, находящийся в середине последовательности показателей, если их расположить в порядке возрастания или убывания. Медиана есть точка, делящая распределение ровно пополам, причем одна половина результатов лежит справа от нее, а другая слева.

Мода – Мо – соответствует либо наиболее частому, либо среднему значению класса с наибольшей частотой. Мо используют редко и в тех случаях, где нужно общее представление.

Правила вычисления Мо:

1. Когда все значения в группе встречаются одинаково, то считают, что группа значения Мо не имеет 3;3; 6;6; 7;7;

2. Когда два соседних значения имеют одинаковую частоту и эта частота больше частоты любого другого значения, то Мо есть среднее этих 2 значений. 0;1; 1; 2;2;2; 3; 3; 3; 4; Мо = 2,5

3. Когда два несмежных значения имеют равные частоты и эти частоты больше частот любого значения, то существуют 2 Мо, и группу называют бимодальной. 10; 11; 11; 11; 12; 13; 14; 14; 14; 17 Мо1 = 11; Мо2 = 14

Медиана (Ме) соответствует центральному значению в последовательном ряду полученных значений. Это значение, которое делит упорядоченное множество данных пополам.

Правила вычисления Ме:

1. Если данные содержат нечетное число знаков, то Ме есть среднее значение для случая, когда они упорядочены.

11 13 18 19 20

## Ме = 18

2. Если четное число, то Ме складывает 2 средних значения и они делятся пополам.

**Список литературы**

1. Алексеев А.С. Основы психологии. М.: 2008 г.

2. Бахматов И.К. Психология личностных отношений. М.: 2007 г.

3. Гуров Е.Г. Психодиагностика. Ростов-на-Дону: Феникс, 2007 г.

4. Грошев Е.В. Психология и психодиагностика. М.: 2006.