

С.В. Клименко, В.В. Огоренко

ОЦЕНКА ОБЪЕКТИВНОСТИ РЕЗУЛЬТАТОВ КОМПЬЮТЕРНОГО ПСИХОМЕТРИЧЕСКОГО ТЕСТИРОВАНИЯ

Аннотация. Предложены информационные технологии обработки измерений психометрического тестирования личности в задаче оценки объективности результатов и определения их диагностической ценности.

Ключевые слова: психометрическое тестирование, оценка нейротизма, оценка экстраверсии-интроверсии, ранговый метод, выборка измерений, корреляция, статистические связи, вероятность ошибок.

Постановка задачи. Человек в любых управленческих структурах и человеко-машинных автоматизированных системах представляет собой самое ненадежное звено. Именно от него чаще всего зависит состояние объектов повышенной опасности. Безопасность их эксплуатации определяется не только надежностью механизмов и машин, но и надежностью человека, как звена управления, принимающего решения. Его надежность зависит от психического состояния. Таким образом, при эксплуатации автоматизированных систем кроме технической безопасности должна решаться задача поддержки на высоком уровне психологической безопасности путем контроля психического состояния персонала, участвующего в управлении и особенно тех, кто принимает решения [1]. Задача психоконтроля решается путем компьютерного психометрического тестирования [2]. Компьютерное психометрическое тестирование – это один из многих способов получения информации о психике и личностях людей, которая не может быть получена путем прямого опроса человека о его психическом состоянии, о мотивах поведения, что он думает и знает о своей психике. Большинство людей не склонно задумываться о том, какое у них психическое состояние, какие особенности их личности. Они не способны оценивать у себя степень выраженности тех или иных факторов, характеризующих психическое состояние. Люди не всегда хотят

давать о себе правдивую информацию в соответствии с тем, что они о себе знают или думают, что знают. У человека могут быть особые причины давать о себе искаженную информацию. Есть большая категория людей, которые в принципе недоверчиво относятся к психологам и их вопросам. Если они даже соглашаются на собеседование или тестирование, то нет уверенности в том, что они сообщают о себе правдивые сведения или правдивые ответы на вопросы.

Поэтому результаты компьютерного психометрического тестирования, по которым принимается то или иное решение о состоянии психики, должны контролироваться на объективность. В классическом тесте Айзенка при определении темперамента это задача решается следующим образом. Из 57 вопросов теста 48 предназначены для оценки нейротизма и экстраверсии-интроверсии, а 9 вопросов для определения искренности (правдивости) тестируемых, которые оцениваются числом ответов, совпадающих с ключом теста: если число совпадающих ответов меньше 5, то результаты тестирования считаются ненадежными. Субъективность такого подхода к оценке объективности тестирования очевидна. Представляется, что информация об искренности тестируемых содержится в результатах тестирования, в функциях теста $S(k)$, решений $R(k)$ и времени $T(k)$ [2].

Статистическая обработка измерений компьютерного тестирования. Оценка корреляционной связи между значениями информативности стимулов и временными показателями их обработки пациентом может служить их косвенным подтверждением его правдивости: более длинным стимулам должны соответствовать более длительные промежутки времени на их восприятие и принятие решений. В медицинской практике чаще всего для этих целей используется ранговый коэффициент корреляции Спирмена. Это связано с тем, что, как правило, статистические закономерности исследуемых последовательностей неизвестны. Под рангом выборочного значения измерения случайной величины понимается его номер в упорядоченной по возрастанию выборке этих измерений. Для оценки степени связи случайных величин используются не их числа, а соответствующие им ранги. Корреляционная связь означает, что изменения $S(k)$ в ту или в иную сторону будет вызывать подобные изменения в измерениях $T(k)$. Эта

же закономерность имеет место и для рангов двух выборок случайных величин.

Если известны выборки измерений $S(k)$ и $T(j)$, то их ранги вычисляются следующим способом. По выборкам $S(k)$ и $T(k)$ сначала формируют упорядоченные выборки $Y_S(j)$ и $Y_T(k)$. Упорядоченные выборки обладают следующим свойством $Y(1) < Y(2) < \dots < Y(j-1) < \dots < Y(n-1) < Y(n)$, где $Y(1) = S_{\min}$ (или T_{\min}) и $Y(n) = S_{\max}$ (или T_{\max}). Ранги $R(S_k)$ и $R(T_k)$ вычисляются по формулам

$$R(S_k) = \sum_{j=1}^n \operatorname{sgn}(S(k) - Y_S(j)), \quad R(T_k) = \sum_{j=1}^n \operatorname{sgn}(T(k) - Y_T(j)),$$

где $\operatorname{sgn}(x)$ – функция единичного скачка.

Коэффициент корреляции Спирмена определяется как нормированная сумма рангов двух выборок измерений

$$r^* = 1 - \frac{6 \sum_{k=1}^n (R(S(k)) - R(T(k)))^2}{n(n^2 - 1)}.$$

Показатель Спирмена является случайной величиной, так как $S(k)$ и $T(k)$ изменяются случайным образом при каждом тестировании. Однако, если $S(k)$ и $T(k)$ не коррелированы, то статистические закономерности коэффициента Спирмена известны: при $n > 10$, т.е. если число измерений в выборках больше 10, его распределение почти нормальное с нулевым математическим ожиданием и разбросом около нуля $\pm 3\sqrt{(n-1)^{-1}}$ (с вероятностью 0,997) или $\pm 2\sqrt{(n-1)^{-1}}$ (с вероятностью 0,956). Таким образом, корреляция между $S(k)$ и $T(k)$ признается не значимой, близкой к нулю, если рассчитанное значение показателя Спирмена r^* удовлетворяют неравенству $(-\frac{2 \div 3}{\sqrt{n-1}} \leq r^* \leq \frac{2 \div 3}{\sqrt{n-1}})$. Например, если сумма квадратов разности 120 и $n = 10$, то эмпирический коэффициент корреляции Спирмена равен

$$\sum_{k=1}^{10} \Delta R^2(k) = 120; \quad r^* = 1 - \frac{6 \cdot 120}{10 \cdot 99} = 0,273.$$

Так как $r^* \leq \frac{2}{\sqrt{n-1}} = 0,66$, то корреляция между $S(k)$ и $T(k)$

близка к нулю, т.е. они не коррелированы и следовательно, результаты тестирования сомнительны.

При малых размерах выборок измерений ($10 \leq n \leq 25$) более эффективным показателем корреляции является ранговый критерий Ширахатэ. Это аналог критерия Спирмена, асимптотически ему эквивалентен при больших выборках, но при малых у него меньше вероятности принятия ошибочных решений.

Проведем сначала нормировку измерений $S(k)$ и $T(k)$, разделив их на средние значения \bar{S} и \bar{T} .

$$X(k) = \frac{S(k)}{\bar{S}}, \quad \bar{S} = \frac{1}{n} \sum_{k=1}^n S(k); \quad Y(k) = \frac{T(k)}{\bar{T}}, \quad \bar{T} = \frac{1}{n} \sum_{k=1}^n T(k).$$

Их ранги определяются как сумма собственных и взаимных рангов

$$R(X(k)) = \sum_{i=1}^n \operatorname{sgn}(X(k) - X(i)) + \sum_{i=1}^n \operatorname{sgn}(X(k) - Y(i)),$$

$$R(Y(k)) = \sum_{i=1}^n \operatorname{sgn}(Y(k) - Y(i)) + \sum_{i=1}^n \operatorname{sgn}(Y(k) - X(i)).$$

Показатели Ширахатэ – это сумма произведений рангов $R(X(k))$ и $R(Y(k))$

$$S = \sum_{k=1}^n R(X(k))R(Y(k)).$$

Если сравнивать некоррелированные выборки, то S , как случайная величина, может изменяться в интервале от $S_1 = \frac{n(n+1)(2n+1)}{3}$ до $S_2 = \frac{n(n+1)(4n-1)}{3}$ и пороги сравнения для принятия решения о некоррелированности исследуемых измерений $S(k)$ и $T(k)$ при выполнении неравенства $S_1 \leq S \leq S_2$ находятся по табличным значениям. При $n > 20$ можно использовать для оценки коррелированности как критерий Спирмена, так и критерий Ширахатэ [3].

Некоторую информацию о состоянии психики пациента во время тестирования можно получить, если проанализировать флуктуации измерений времени и скорости обработки им стимулов. Эта

информация содержится в выборочных дисперсиях $D^*(T)$ и $D^*(V)$. Коэффициенты разброса равны отношению корня квадратного из выборочных дисперсий и средних значений \bar{T} и \bar{V}

$$K_T = \frac{\sqrt{D_T^*}}{\bar{T}}, \quad K_V = \frac{\sqrt{D_V^*}}{\bar{V}}, \quad \text{где} \quad D_T^* = \frac{1}{n-1} \sum_{k=1}^n (T_k - \bar{T})^2,$$

$$D_V^* = \frac{1}{n-1} \sum_{k=1}^n (V_k - \bar{V})^2.$$

Можно предположить, что эти коэффициенты характеризуют волнение пациента в процессе тестирования, его нерешительность при принятии решений. Очень малые значения коэффициента K_V при больших средних значениях \bar{V} и малых дисперсиях D_V^* ставят под сомнение результаты тестирования как и малые, близкие к нулю значения коэффициента корреляции.

Представляет интерес еще один показатель текущего тестирования – это выбросы измерений $T(k)$. В статистике под выбросами понимают измерения, значения которых явно отличаются от остальных в данной последовательности в большую или меньшую сторону от среднего значения. Они должны быть обнаружены с указанием к каким стимулам они относятся. Трудности решения этой задачи в том, что статистические закономерности измерений $T(k)$ неизвестны. Первый шаг ее решения – визуальный просмотр выборки измерений $T(k)$. Выборка ранжируется от минимального до максимального значения

$$\tau(1) < \tau(2) < \dots < \tau(i-1) < \tau(i) < \dots < \tau(n-1) < \tau(n),$$

где $\tau(1) = T_{\min}$ и $\tau(n) = T_{\max}$.

Из визуального анализа ранжированных данных можно получить информацию об отношении пациента к стимулам, на которые затрачено максимальное и минимальное время или которые обрабатывались с минимальной или максимальной скоростью.

Для исследования, могут ли они быть выбросами, воспользуемся критерием Диксона для ассиметричных случайных величин [3]. Для проверки является $\tau_1 = \tau(1)$ и $\tau_n = \tau(n)$ выбросом, вычисляется показатель Диксона

$$\delta_1 = \frac{\tau_2 - \tau_1}{\tau_n - \tau_1} \text{ или } \delta_n = \frac{\tau_n - \tau_{n-1}}{\tau_n - \tau_1}.$$

Для проверки, что оба они выбросы, вычисляются два показателя

$$\delta_{11} = \frac{\tau_2 - \tau_1}{\tau_{n-1} - \tau_1} \text{ и } \delta_{1n} = \frac{\tau_n - \tau_{n-1}}{\tau_n - \tau_2}.$$

Измерения признаются выбросами, если их значения больше пороговых (или критических) $\delta_1(0)$, $\delta_n(0)$, $\delta_{11}(0)$ и $\delta_{1n}(0)$. Значения порогов при надежности $P = 0,95$ даны в таблице 1.

Таблица 1

n	5	6	7	8	9	10	11	12	пороги
$\delta_1(0)$	0,74	0,68	0,64	0,6	0,58	0,56	0,55	0,53	Н
$\delta_n(0)$	0,55	0,46	0,38	0,34	0,32	0,29	0,27	0,25	В
$\delta_{11}(0)$	0,86	0,77	0,71	0,67	0,64	0,61	0,59	0,56	Н
$\delta_{1n}(0)$	0,76	0,62	0,52	0,45	0,40	0,37	0,34	0,32	В
n	13	14	15	16	17	18	19	20	пороги
$\delta_1(0)$	0,52	0,51	0,50	0,50	0,49	0,48	0,47	0,46	Н
$\delta_n(0)$	0,24	0,22	0,21	0,21	0,20	0,19	0,18	0,18	В
$\delta_{11}(0)$	0,56	0,54	0,53	0,52	0,51	0,50	0,50	0,49	Н
$\delta_{1n}(0)$	0,30	0,28	0,26	0,25	0,24	0,23	0,22	0,21	В

Здесь Н – пороги для оценки нижних (минимальных) выбросов, В – пороги для оценки верхних (максимальных) выбросов.

При числе измерений больше двадцати $n > 25$ следует использовать используется критерий Груббса [3]

$$t_n = \frac{\tau_n - \bar{T}}{\sqrt{D_T^*}} \text{ и } t_1 = \frac{\bar{T} - \tau_1}{\sqrt{D_T^*}},$$

где D_T^* – выборочная дисперсия, вычисляемая по формуле

$$D_T^* = \frac{1}{n-1} \sum_{k=1}^n (T(k) - \bar{T})^2.$$

Если $t_1 \geq t_{01}$ или $t_n \geq t_{0n}$, то $\tau(1)$ или $\tau(n)$ признаются выбросами. Для оценки порогов используются формулы

$$t_{0n} = \Psi\left(1 - \frac{1-P}{n}\right), \quad t_{10} = \begin{cases} 1,31 + 0,435 \ln(n - 2,7), & 20 \leq n \leq 35, \\ 1,962 + 0,281 \ln(n - 15), & 35 < n \leq 500, \end{cases}$$

где $\Psi(z)$ – функция, обратное интегралу вероятности Гаусса [3], вычисляемая по формуле

$$\Psi(z) = 4,91 \left[z^{0,14} - (1-z)^{0,14} \right].$$

Если $15 \leq n \leq 25$, то пороговые значения t_{01} и t_{0n} при надежности $P = 0,95$ следует выбирать из таблицы 2.

Таблица 2

n	15	16	17	18	19	20	21	22	23	24	25
t_{0n}	2,617	2,644	2,668	2,691	2,712	2,732	2,750	2,768	2,784	2,800	2,815
t_{01}	2,493	2,523	2,551	2,577	2,600	2,623	2,644	2,664	2,683	2,701	2,717

Выводы. На результаты психометрического тестирования оказывает влияние субъективный фактор: отношение личности к такому способу получения информации о тех или иных характеристиках и особенностях его психики, и показателях, зависящих от его психического состояния. Это отношение характеризуется правдивостью ответов испытуемых на стимулы-вопросы. Предложены два метода проверки объективности результатов тестирования: 1) путем оценки статистической связи между числом знаков в вопросах и измерениями времени на их чтение, осмысление и подготовку решений; 2) анализа разбросов измерений времени и скорости обработки стимулов в процессе тестирования. Если оценки не удовлетворяют статистическим критериям, то результаты тестирования считаются сомнительными и не могут использоваться при подготовке диагностических решений.

ЛИТЕРАТУРА

1. Малайчук В.П., Огоренко В.В. Введение в психологию принятия решений. – Системные технологии. – Днепропетровск. – 2009. – 130с.
2. Клименко С. В., Огоренко В. В. Информационные технологии индивидуального компьютерного психометрического тестирования. Системные технологии 6'(53).–Днепропетровск.–2007. – С.116-124.
3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816с.