УДК 303.092.52, 377.44

# СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНКИ ЭФФЕКТИВНОСТИ ОБУЧЕНИЯ

© 2017 Л.Н. Александровская[1], А.В. Кириллин[1], В.В. Медведев[2]

[1] Московский авиационный институт (национальный исследовательский университет)
[2] Московский государственный университет технологий и управления имени К.Г. Разумовского

Статья поступила в редакцию 08.01.2017

В статье на основе краткого обзора математического аппарата двухфакторного дисперсионного анализа в дискретной бинарной и непрерывной шкалах проведен сравнительный анализ ограничений области возможного применения каждого из этих двух вариантов. Для решения задач раздельной оценки степени подготовленности обучающихся и степени сложности заданий, используемых для их тестирования и контроля, предложена логистическая модель Г. Раша, имеющая универсальный характер и позволяющая как производить искомые оценки в обеих шкалах, так и экстраполировать полученные результаты для планирования необходимого и достаточного для достижения заданной эффективности циклов обучения.
*Ключевые слова:* эффективность обучения, степень подготовленности обучающихся, степень сложности контрольных заданий, двухфакторный дисперсионный анализ, вероятность успеха, логистическая модель Г. Раша.

## ВВЕДЕНИЕ

Процесс создания современной авиационной техники, а также ее дальнейшая безопасная эксплуатация требует повышенного внимания к компетенциям специалистов, задействованных работах на различных этапах жизненного цикла изделий. В ряде работ [1, 2], в которых освещены вопросы повышения уровня компетенций работников авиационной промышленности, наибольшее внимание уделено техническим и методическим аспектам обучения. С другой стороны вопросы связанные с контролем знаний не менее важны.

Традиционными методами контроля степени освоения обучающимися преподаваемого курса являются промежуточные зачеты с оценками «зачет»/«незачет» и экзамены с оценками чаще всего в пятибалльной системе: 5 – отлично; 4 – хорошо; 3 – удовлетворительной; 2,1 – плохо и очень плохо.

Однако более детальный анализ эффективности обучения показывает, что на результаты тестирования и контроля оказывает большое количество факторов [3,4]: здесь и степень освоения обучающимися предшествующих курсов, связанных с преподаваемой дисциплиной и их индивидуальные особенности (способность к аналитическому мышлению, внимательность и сосредоточенность, навыки работы с литературой и пр.). Большое значение имеет профессионализм преподавателя, его умение производить декомпозицию сложной проблемы на более простые составные части, контакт с аудиторией и пр. Поэтому простейшего анализа результатов проведения зачетов и экзаменов недостаточно – необходим более детальный анализ отдельно уровня сложности преподаваемой дисциплины и согласованным с ним уровнем подготовленности обучающихся. При этом наиболее часто используется математический аппарат дисперсионного анализа, как минимум – двухфакторного.

## 1. ОСНОВНЫЕ ПОЛОЖЕНИЯ ДВУХФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА

В табл. 1 приведено расположение данных в двухфакторном плане проведения эксперимента [5-7].

Здесь введены обозначения:

$$y_{i..}=\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}\ ,\ \bar{y}_{i..}=\frac{1}{bn}y_{i..}\ ,\ i=1,2,\ldots,a\ ;$$

$$y_{.j.}=\sum_{i=1}^{a}\sum_{k=1}^{n}y_{ijk}\ ,\ \bar{y}_{.j.}=\frac{1}{an}y_{.j.}\ ,\ j=1,2,\ldots,b\ ;$$

$$y_{ij.}=\sum_{k=1}^{n}y_{ijk}\ ,\ \bar{y}_{ij.}=\frac{1}{n}y_{ij.}\ ,\ i=1,2,\ldots,a\ ,$$

$$j=1,2,\ldots,b\ ;$$

$$y_{...}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}\ ,\ \bar{y}_{...}=\frac{1}{abn}y_{...}\ ,\ k=1,2,\ldots,n\,.$$

*Александровская Лидия Николаевна, доктор технических наук, профессор кафедры «Испытания летательных аппаратов». E-mail: aleksandrovskayaln@mati.ru*
*Кириллин Андрей Викторович, старший преподаватель кафедры «Испытания летательных аппаратов». E-mail: kirillinav@mati.ru*
*Медведев Виктор Васильевич, кандидат технических наук, доцент, доцент кафедры «Управление качеством инновационных наукоемких производств». E-mail: medvik@hotbox.ru*

**Таблица 1.** Расположение данных в двухфакторном плане

| | | *Фактор B* | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | $\cdots$ | $b$ |
| *Фактор A* | 1 | $y_{111}\cdots y_{11n}$ | $y_{121}\cdots y_{12n}$ | $\cdots$ | $y_{1b1}\cdots y_{1bn}$ |
| | 2 | $y_{211}\cdots y_{21n}$ | $y_{221}\cdots y_{22n}$ | $\cdots$ | $y_{2b1}\cdots y_{2bn}$ |
| | $\vdots$ | | | | |
| | $a$ | $y_{a11}\cdots y_{a1n}$ | $y_{a21}\cdots y_{a2n}$ | $\cdots$ | $y_{ab1}\cdots y_{abn}$ |

Общая сумма квадратов разлагается на следующие слагаемые:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\overline{y}_{...}\right)^2 = bn\sum_{i=1}^{a}\left(\overline{y}_{i..}-\overline{y}_{...}\right)^2 + an\sum_{j=1}^{b}\left(\overline{y}_{.j.}-\overline{y}_{...}\right)^2 + $$
$$+n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{y}_{ij.}-\overline{y}_{i..}-\overline{y}_{.j.}+\overline{y}_{...}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\overline{y}_{ij.}\right)^2. \qquad (1)$$

или в новых обозначениях:

$$S^2_{общ.} = S^2_A + S^2_B + S^2_{AB} + S^2_{ошибк.},$$

т.е. происходит разбиение на суммы квадратов, обусловленное «строками» или факторами А, «столбцами» или факторами *B*, взаимодействием между факторами А и В, ошибкой.

Число степеней свободы, соответствующих каждой сумме квадратов, составляет:

| Фактор | Число степеней свободы |
|---|---|
| А | $a-1$ |
| В | $b-1$ |
| АВ | $(a-1)(b-1)$ |
| Ошибка | $ab(n-1)$ |
| Сумма | $abn-1$ |

Из анализа числа степеней свободы видно, что для различения между эффектом взаимодействия и ошибкой $n \geq 2$.

Наблюдения в двухфакторном дисперсионном анализе описываются линейной статистической моделью

$$y_{ijk} = \mu + \theta_i + \beta_j + (\theta\beta)_{ij} + \zeta_{ijk};\ i = 1,2,\ldots,a;$$
$$j = 1,2,\ldots,b;\ k = 1,2,\ldots,n \qquad (2)$$

$\mu$ – математическое ожидание эффекта;
$\theta_i$ – эффект *i*-го уровня фактора А;
$\beta_j$ – эффект *j*-го уровня фактора В;
$\zeta_{ijk}$ – случайная ошибка.

Различают два вида модели (2): модель постоянных эффектов и модель случайных эффектов. При использовании модели постоянных эффектов результаты дисперсионного анализа применимы только к рассматриваемым уровням факторов; при использовании модели случайных эффектов результаты распространяются на всю область изменения факторов. Возможны также смешанные модели, когда уровни одного фактора выбираются случайным образом, а уровни другого – фиксированы. Различия в моделях имеет значение при расчете мощности критериев дисперсионного анализа, оценки компонент дисперсии; решающие правила о значимости или не значимости влияния каждого из факторов и их взаимодействий одинаковы для всех моделей. До сих пор не делалось никаких допущений о законах распределения вероятностей применимых в модели (2).

В классическом дисперсионном анализе применимы следующие допущения [6]:

- в модели постоянных эффектов $\sum_{i=1}^{a}\theta_i = 0$; $\sum_{j=1}^{b}\beta_i = 0$, т.е. для фиксированных факторов эффекты определяются как отклонения от математического ожидания $\mu$;
- в модели случайных эффектов принимаются нормальные законы распределения вероятностей $\theta_i \sim N\left(0;\sigma^2_\theta\right)$; $\beta_j \sim N\left(0;\sigma^2_\beta\right)$; $(\theta\beta)_{ij} \sim N\left(0;\sigma^2_{\theta\beta}\right)$;
- в обеих моделях ошибка считается нормально распределенной случайной величиной

$$\varepsilon_{ijn} \sim N\left(0;\sigma^2_0\right). \qquad (3)$$

Однако, общий принцип проведения дисперсионного анализа на основе разложения общей суммы квадратов на компоненты не требует указанных выше ограничений.

## 2. ДВУХФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ ОБУЧЕНИЯ В ДИСКРЕТНОЙ БИНАРНОЙ ШКАЛЕ

При оценивании результатов контроля обучающихся (тестировании в бинарной шкале) случайные величины $y_{ijk}$ могут принимать только два значения: 1 – при правильном ответе на контрольный вопрос и 0 – при неправильном.

Статистики двухфакторного дисперсионного анализа могут быть построены на основе статистики $\chi^2$ Пирсона, применяемой в основном как критерий согласия между эмпирическими и теоретическими распределениями [5].

Пусть $m_C$ – число наблюдаемых событий, а $p_C$ – гипотетические вероятности мультиномиального распределения

$$P\left(m_1,\ldots,m_C / p_1,\ldots,p_C,n\right)\sim$$
$$\sim P_1^{m_1} P_2^{m_2}\ldots P_C^{m_C}\left(1-P_1-P_2-\ldots-P_C\right)^{n-m_1-\ldots-m_C}.$$

Тогда $\chi^2=\sum_{k=1}^{C}\frac{\left(m_k-np_k\right)^2}{np_k}$ – $\chi^2$-распределение с ($C$–1) числом степеней свободы.

Оценками вероятностей $p_k$ являются оценки вида $\overline{p}_k=m_k/n$.

Нулевая гипотеза принимается при $\chi^2>\chi^2_{1-\alpha}\left(C-1\right)$, где $\chi^2_{1-\alpha}\left(C-1\right)$ – квантиль $\chi^2$-распределения уровня значимости $\alpha$ и ($C$–1) числом степеней свободы.

При измерениях результатов ответов обучающихся в номинальной шкале табл. 1 примет вид, представленный в табл. 2.

Здесь в каждой ячейке содержится одно измерение 0 или 1 $\left(n=1\right)$; $m_{i\cdot}$ – число правильных ответов $i$-ого обучающегося на все вопросы; $m_{\cdot j}$ – число правильных ответов всех обучающихся на $j$ вопрос.

Как было отмечено выше при $n=1$ выделить эффект взаимодействия не удается, и двухфакторный эксперимент сводится к двум однофакторным, т.е. вычисляются суммы

$$S_a^2=b\sum_{i=1}^{a}\left(\overline{y}_{i\cdot}-\overline{y}_n\right)^2,\; S_b^2=a\sum_{j=1}^{b}\left(\overline{y}_{\cdot j}-\overline{y}_n\right)^2.$$

В рассматриваемом случае

$$S_a^2=b\sum_{i=1}^{a}\left(\overline{m}_{i\cdot}-\frac{1}{a}\sum_{i=1}^{a}\overline{m}_{i\cdot}\right)^2=\frac{1}{b}\sum_{i=1}^{a}\left(m_{i\cdot}-\frac{1}{a}\sum_{i=1}^{a}m_{i\cdot}\right)^2,$$

$$S_b^2=a\sum_{j=1}^{b}\left(\overline{m}_{\cdot j}-\frac{1}{b}\sum_{j=1}^{b}\overline{m}_{\cdot j}\right)^2=\frac{1}{a}\sum_{j=1}^{b}\left(m_{\cdot j}-\frac{1}{b}\sum_{j=1}^{b}m_{\cdot j}\right)^2,$$

где $\sum_{a}m_{i\cdot}=\sum_{b}m_{\cdot j}=m$.

Таким образом $m_{i\cdot}$, $m_{\cdot j}$ представляют собой экспериментальные числа событий соответственно в $i$-ой строке (для $i$-ого обучающегося) и в $j$-ом столбце (для $j$-го вопроса), а числа $m/a$ и $m/b$ ожидаемые числа событий при условии равномерного их распределения между строками и столбцами.

Для построения статистики дисперсионного анализа при измерениях в номинальной шкале необходимо разделить суммы $S_a^2$ и $S_b^2$ на ожидаемое число событий

$$\chi^2_{1-\alpha}\left(C-1\right)\;\chi^2_A=\frac{S_a^2}{m/a};\;\chi^2_B=\frac{S_b^2}{m/b}.$$

Числа $m/a$ и $m/b$ могут получаться дробными, поэтому более удобная для практического использования вероятностная интерпретация:

$$p_{i\cdot}=\frac{m_{i\cdot}}{b}=\overline{m}_{i\cdot};\; p_{\cdot j}=\frac{m_{\cdot j}}{a}=\overline{m}_{\cdot j};\; p_{\cdot\cdot}=\frac{m}{ab}.$$

Тогда $\chi^2_A=\frac{b\left(\sum_{i=1}^{a}p_{i\cdot}-p_{\cdot\cdot}\right)^2}{p_{\cdot\cdot}}$ – с числом степеней свободы $\left(a-1\right)$, и $\chi^2_B=\frac{a\left(\sum_{j=1}^{b}p_{\cdot j}-p_{\cdot\cdot}\right)^2}{p_{\cdot\cdot}}$ – с числом степеней свободы $\left(b-1\right)$.

Пример 1

В табл. 3 приведены результаты статистического моделирования при $a=10$; $b=10$; вероятность успешного ответа всех обучающихся на все вопросы 0,5. Общее число положительных ответов меньше 50, что определяется статистической погрешность. Общее число положительных ответов распределяется между вопросами и обучающимися в соответствии с равномерно распределенными случайными числами $x_{ij}<0,5$ принимается $y_{ij}=0$, при $x_{ij}>0,5$ принимается $y_{ij}=1$. Ожидаемый ответ после проведения дисперсионного анализа: сложность вопросов и подготовленность обучающихся не влияет на результаты контроля.

$$\chi^2_A=\frac{10\sum_{i=1}^{a}\left(m_{i\cdot}-3,9\right)^2}{3,9}\approx 7,923<\chi^2_{0,9}\left(9\right)=14,7,$$

**Таблица 2.** Результаты тестирования обучающихся в двухфакторном плане в номинальной шкале

| | | *Вопросы* | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | $\cdots j$ | $b$ | $m_{i\cdot}$ |
| *Обучающиеся* | 1 | | | | | $m_{1\cdot}$ |
| | 2 | | | | | $m_{2\cdot}$ |
| | $\vdots i$ | | | | | $\vdots$ |
| | $a$ | | | | | $m_{a\cdot}$ |
| | $m_{\cdot j}$ | $m_{\cdot 1}$ | $m_{\cdot 2}$ | $\cdots$ | $m_{\cdot b}$ | $\sum_{a}m_{i\cdot}=\sum_{b}m_{\cdot j}=m$ |

**Таблица 3.** Результаты статистического моделирования задачи двухфакторного эксперимента

| | | *Вопросы* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $m_{i\cdot}$ |
| *Обучающиеся* | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 4 |
| | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 3 |
| | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 4 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 4 |
| | 5 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 5 |
| | 6 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| | 7 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| | 8 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 8 |
| | 9 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 4 |
| | 10 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 4 |
| | $m_{\cdot j}$ | 5 | 4 | 3 | 6 | 3 | 3 | 7 | 2 | 2 | 4 | 39 |

$$\chi_B^2 = \frac{10\sum_{j=1}^{b}\left(m_{\cdot j} - 3,9\right)^2}{3,9} \approx 6,385 < \chi_{0,9}^2\left(9\right) = 14,7 .$$

Таким образом, полученный результата анализа совпадает с ожидаемым, т.е. экспериментальные данные согласуются с их ожидаемыми значениями и ни один из факторов не оказывает влияния на результаты контроля обучающихся.

Однако даже в данном примере видно, что значение $\chi_B^2$ дальше отстоит от процентной точки $\chi_{1-\alpha}^2$, чем значение $\chi_A^2$, т.е. данные столбцов (фактор сложности вопроса) более согласован с ожидаемыми значениями, чем данные строк (фактор подготовленности обучающихся).

В общем случае обработки экспериментальных данных возможны решения о значительности влияния как сложности вопросов, так и подготовленности обучающихся. При этом традиционно возможны лишь два решения: наличие или отсутствие такого взаимодействия.

При оценке эффективности обучения нас интересует не только наличие влияния исследуемого фактора, но и степень этого влияния.

Для оценки степени влияния может быть предложен подход, описанный в [8].

$\chi^2$ – критерий согласия относится к группе критериев значимости, построенных на следующих общих принципах:

- выбирается некоторая статистика (функция от экспериментальных данных);

- на основе теоретических исследований принимается нулевая гипотеза о теоретическом законе распределения вероятностей данной статистики;

- определяются области малых вероятностей (критические области), в которых согласно принципу практической невозможности осуществления маловероятного события не должно попасть экспериментальное значение статистики;

- если значение статистике не попадается в область малых вероятностей, определяемой выбранным уровнем значимости, то такое событие не противоречит принятой нулевой гипотезе и последняя может быть принята;

- выбор условия значимости в практических приложениях достаточно произволен и составляет $\alpha = 0,05 \div 0,1$.

Предлагаемый подход состоит в следующем. Используемые статистики (в т.ч. и $\chi^2$ -критерий) достаточно часто имеют бесконечные шлейфы ($\chi^2$ имеет бесконечный правый шлейф).

В результате всегда можно подобрать такое значение $\alpha$, назовем его критическим $\alpha_{кр}$, для которого нулевая гипотеза принимается. При этом малым значениям $\alpha_{кр}$ соответствует большая область принятия нулевой гипотезы и наоборот. Широкая область принятия нулевой гипотезы соответствует большому допустимому рассогласованию между экспериментальными и теоретическими данными, что сказывается на достоверности принимаемого решения. Таким образом, величина $\alpha_{кр}$ может быть принята как оценка адекватности теоретической модели экспериментальным данным.

В рассматриваемом выше примере с точностью до табличных значений процентных точек $\chi^2$-распределения для фактора А $\alpha_{кр}$= 0,5; для фактора В $\alpha_{кр}$= 0,7. Таким образом фактор А более сильно влияет на результат тестирования обучающихся, чем фактор В.

## 3. ДВУХФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ ОБУЧЕНИЯ В НЕПРЕРЫВНОЙ ШКАЛЕ

В случае, если ответы обучающихся на вопросы оцениваются в баллах т.е. имеется более полная информация, чем в бинарном случае, для оценки эффективности обучаемых используется традиционный двухфакторный дисперсионный анализ, методика проведения которого широко известна и опубликована в ряде фундаментальных работ [5-7].

Поэтому в данном разделе остановимся лишь на нарушении исходных предпосылок при проведении этого анализа.

В начале обучения степень подготовленности обучающихся принимается обычно нормально-распределенной, при небольшом количестве как слабо подготовленных, так и сильно подготовленных слушателей. При этом программы обучения подготовленная для среднего обучающегося будет слишком сложной для первых и слишком простой для вторых. По мере обучения число хорошо подготовленных слушателей смещается вправо, т.е. происходит переход от нормального распределения к асимметричному распределению с положительным показателем асимметрии.

Далее при контроле стандартного курса обучения вопросы обычно подбираются одинаковой сложности. Для оценки сложности, при этому отсутствуют какие либо рекомендации по составлению числа вопросов уменьшенной, средней и повышенной сложности. В перечисленных условиях ограничения (3) в виде нормальности распределений явно нарушается.

Ниже прилагается универсальный подход двухфакторного эксперимента, не связанный с какими-либо предположениями о законах распределения исследуемых факторов и погрешности измерений, основанный на использовании эмпирических вероятностях результатов контроля обучающихся.

Для большей наглядности изложим предлагаемый подход на данных примера 1.

Пример 2.

Пусть оценивание знаний обучающихся производится в 10-ти балльной шкале и гипотетические результаты (в баллах) будут:

- по столбцу: $m_{i.} = \{4,3,1,4,5,4,2,8,4,4\}$;

- по строке: $m_{.j} = \{5,4,36,3,3,7,2,2,4\}$.

Приравняем значения баллов приближенные эмпирические вероятности, полученные из условия равномерного распределения:

- по столбцу: 4/10; 3/10; 1/10; 4/10; 5/10; 4/10; 2/10; 8/10; 4/10; 4/10;

- по строке: 5/10; 4/10; 3/10; 6/10; 3/10; 3/10; 7/10; 2/10; 2/10; 4/10.

В результате приходим к предыдущему плану двухфакторного дисперсионного анализа в бинарной шкале.

Более строго по данным столбца и строки можно построить эмпирические функции распределения.

Для этого располагаем данные в вариационный ряд:

- по столбцу: 1; 2; 3; 4; 4; 4; 4; 4; 5; 8;

- по строке: 2; 2; 3; 3; 3; 4; 4; 5; 6; 7.

Каждому $i$-ому значению столбца или $j$-ому значению строки будет соответствовать несмещенная оценка вероятности $p_i = \dfrac{i}{n+1}$ и $p_j = \dfrac{j}{n+1}$, где $n$ – объем выборки, в данном примере $n = 10$.

Одинаковым значениям приписывается средняя вероятность. Возвращаясь к первоначальному расположению данных, получим:

- по столбцу: 6/11; 3/11; 1/11; 6/11; 9/11; 6/11; 2/11; 10/11; 6/11; 6/11;

- по строке: 8/11; 6,5/11; 4/11; 9/11; 4/11; 4/11; 10/11; 1,5/11; 1,5/11; 6,5/11;

Проводя далее расчеты по методике двухфакторного плана в бинарной шкале, получим:

$$\chi^2_A \approx 11{,}98 < 14{,}7,$$

$$\chi^2_B \approx 13{,}36 < 14{,}7.$$

Таким образом, гипотеза по незначимости влияния уровня сложности и уровня подготовленности обучающихся принимается, однако с меньшим критическим уровнем значимости, чем в предыдущем примере, что свидетельствует о большой чувствительности данного подхода к возможным погрешностям.

Однако более эффективной является логистическая модель Г. Раша.

## 4. ЛОГИСТИЧЕСКАЯ МОДЕЛЬ ГЕОРГА РАША

Бельгийский математик Г. Раш предложил в качестве модели оценки эффективности и сбалансированности обучения логистическую модель [9]

$$P(\theta - \beta) = \frac{\exp\{(\theta - \beta)\}}{1 + \exp\{(\theta - \beta)\}}, \qquad (4)$$

где β – уровень сложности контрольного задания, θ – уровень подготовленности слушателей.

При организации активного эксперимента возможна оценка степени подготовленности группы обучающихся при фиксированной степени сложности и наоборот при фиксированном составе группы. При этом компоненты модели Раша могут быть заменены эмпирической оценкой вероятности по частоте $\overline{\beta} = m/a$, $\overline{\theta} = m/b$. Формирование этих оценок было ранее описано в докладе в разделе 2.

Покажем, что эта модель является универсальной и может быть использована для анализа результатов тестирования как в бинарном случае, так и при оценивании в баллах.

В [8] для эмпирической функции распределения предложено следующее нормализующее преобразование:

$$u = \ln \frac{p}{1-p}. \qquad (5)$$

Обозначив случайную величину u = θ−β из (5) получим:

$$\frac{p}{1-p} = \exp\{(\theta-\beta)\};$$

$$p = (1-p)\exp\{(\theta-\beta)\};$$

$$p = \frac{\exp\{(\theta-\beta)\}}{1+\exp\{(\theta-\beta)\}},$$

т.е. модель Раша.

В процессе подготовки операторов к работе со сложной техникой к обучающимся предъявляются чрезвычайно высокие требования к качеству освоения преподаваемого материала. При этом одноразового обучения зачастую недостаточно и возникает вопрос о планировании всего процесса подготовки, а в частности в определении оптимального объема циклов обучения.

Данная задача может быть решена с использованием хорошо известных моделей роста надежности. Задача оценки динамики эффективности обучения по постановке близка к задаче оценки динамики надежности изделия в зависимости от числа циклов испытаний, после каждого из которых производится доработка изделия с целью устранения выявленных в процессе испытаний дефектов, повышая его надежность. Действительно, контроль знаний обучающихся может рассматриваться как их испытание, а выявленные неосвоенные фрагменты курса как дефекты обучения, требующего их повторного освоения и доработки.

При использовании моделей роста надежности остановим свой было на логистической модели роста, что является логичным и естестныввенным с учетом логистической природы модели Раша. Логистическая модель роста надежности имеет вид:

$$R(n) = \frac{R_0}{R_0 + (1-R_0)\exp(-\Omega n)}. \qquad (6)$$

При числе циклов $n=0$, $R=R_0$; при $n \to \infty$, $R \to 1$.

Нетрудно видеть, что модель (6) является частным случаем модели (4), когда $R_0 = 0,5$ , $n=1$, $\Omega = \theta - \beta$ и рассматриваются только положительные значения Ω, приводящие к росту надежности *R*. Кроме этого в модели (6), параметр Ω = *const*, а в модели (4) $\Omega = \theta - \beta$ – случайная величина, косвенное измерение которой осуществляется на основе оценки вероятности *P* по частоте.

Модель (6) может быть записана в эквивалентном дискретном рекуррентном виде

$$R_i = \frac{R_{i-1}\exp\{\Omega\}}{R_{i-1}\exp\{\Omega\} + (1-R_{i-1})},$$

откуда после несложных преобразований получим

$$\ln \frac{R_i}{q_i} = \ln \frac{R_{i-1}}{q_{i-1}} + \Omega,$$

и окончательно переходя к обозначениям модели (4), получим уравнение прогнозирования на 1 шаг:

$$\ln \frac{p_i}{q_i} = \ln \frac{p_{i-1}}{q_{i-1}} + (\theta_{i-1} - \beta_{i-1}), \qquad (7)$$

где $(\theta_{i-1} - \beta_{i-1})$ – оценки, полученные путем обработки информации на предыдущих циклах вплоть до $(i-1)$ включительно.

Заметим, что здесь используется простейшая линейная модель зависимости показателя роста $\Omega_i$ от числа циклов:

$$\Omega_i = (\theta - \beta) n_i.$$

В качестве оценок случайной величины $(\theta_{i-1} - \beta_{i-1})$ рационально использовать среднеарифметическую оценку

$$\Omega_{i-1} = \frac{1}{n-1}\sum_{j=1}^{n-1} \ln \frac{p_j}{q_j}.$$

Уравнение (7) может быть расширено для прогнозирования на более, чем один шаг вперед, например:

$$\ln \frac{p_{i+1}}{q_{i+1}} = \ln \frac{p_{i-1}}{q_{i-1}} + (\theta_{i-1} - \beta_{i-1}) \cdot 2$$

и т.д.

Далее модель Раша может быть использована для анализа влияния факторов сложности контрольных заданий и подготовленности обучающихся как в дискретной, так и в непрерывной шкалах, в зависимости от вида получаемой информации.

Пример 3.

Пусть заданное значение $P_{заданное}$= *0,99*, т.е. $\ln \frac{0,99}{0,01} = 4,595$. Далее, пусть, значение $\ln \frac{p_1}{q_1}$, полученное после проведения первого ознакомительного контроля знаний составляет:

$$\ln \frac{p_1}{q_1} = \ln \frac{0,6}{0,4} = 0,405 = \theta_1 - \beta_1, \text{ т.к. в (7)}$$

$$\ln \frac{p_0}{q_0} = \ln \frac{0,5}{0,5} = 0.$$

Тогда после второго цикла обучения ожидается достижение значения $\ln \frac{p_2}{q_2} = 0,405 \cdot 2$,

$\frac{p_2}{q_2} = 2,25$, откуда $P_2 = 0,692$.

При таком темпе роста надежности для достижения $P_{заданное}= 0,99$ потребуется $4,595 - 0,405 = 0,405 \cdot n$ циклов обучения, откуда $n \approx 10$.

Пример 4. Пусть подготовленность группы обучающихся выше и составляет $p_1 = 0,8$. Тогда $\ln\frac{0,8}{0,2} = 1,386$ и для достижения $P_{заданное}= 0,99$ потребуется уже $n \approx 3$ циклов обучения.

При практическом использовании предложенного подхода возникает вопрос о достоверности принимаемых решений. Вероятность в модели (4) является монотонно возрастающей функцией величины θ–β. Таким образом, нижняя доверительная граница оценки этой вероятности по частоте соответствует нижней доверительной границе косвенного измерения θ–β. В результате приходим к простой практической методике учета погрешностей при замене истинного значения вероятности *P* его оценкой $\hat{P} = \frac{M}{N}$, показанной ниже на конкретном примере.

Пример 5. Пусть в результате ознакомительного тестирования 10 обучающихся по 10 контрольным вопросам ($N = n \cdot k = 10 \cdot 10 = 100$) получено 80 правильных ответов, т.е. $\hat{P} = \frac{80}{100} = 0,8$.

Используя нормальную аппроксимацию биномиального распределения, получим нижнюю 90% доверительную границу при γ = 0,9:

$$P_H = \hat{P} - u_{1-\gamma}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}} =$$

$$= 0,8 - 1,28\sqrt{\frac{0,8 \cdot 0,2}{100}} \approx 0,75,$$

где $u_{1-\gamma}$ – квантиль стандартного нормального распределения.

Далее расчет проводится аналогично примеру 3, т.е. $\ln\frac{0,75}{0,25} = \theta_1 - \beta_1 = 1,099$, и верхнее значение числа циклов $n_B = \frac{13,8 - 1,832}{0,916} \approx 13$.

При этом прогнозируемое значение после 2-ого цикла обучения составит $\ln\frac{p_2}{q_2} = 1,099 \cdot 2$, откуда $p_2 = 0,9$.

Пусть далее, при проведении тестирования на 2-ом цикле получено 90 правильных ответов, т.е. $\hat{P} = \frac{90}{100} = 0,9$.

Нижняя доверительная граница при γ = 0,9 составит:

$$P_H = 0,9 - 1,28\sqrt{\frac{0,9 \cdot 0,1}{100}} \approx 0,862,$$

откуда

$$\theta_2 - \beta_2 = \ln\frac{0,862}{0,138} - \ln\frac{0,75}{0,25} = 1,832 - 1,099 = 0,733.$$

Средний темп роста составит при этом $\frac{1,099 - 0,733}{2} = 0,916$ и прогнозируемых число циклов для достижения $P_H = 0,999999$ будет $n_B = \frac{13,8 - 1,832}{0,916} \approx 13$, т.е. уже погрешности оценок $\hat{P}$ уже существенно влияют на количество циклов обучения.

Для сокращения необходимого числа циклов обучения необходимо повысить его эффективность.

Пример 6. Пусть при остальных условиях примера 5 число правильных ответов после 2-ого цикла обучения повысилось до 95. Тогда $\hat{P} = 0,95$, нижняя доверительная граница составит

$$P_H = 0,95 - 1,28\sqrt{\frac{0,95 \cdot 0,05}{100}} \approx 0,922,$$

откуда

$$\theta_2 - \beta_2 = \ln\frac{0,922}{0,078} - \ln\frac{0,75}{0,25} = 2,47 - 1,099 = 1,371.$$

Средний темп роста будет $\frac{1,099 - 1,371}{2} = 1,235$ и потребуется $n_B = \frac{13,8 - 2,47}{1,235} \approx 9$, т.е. повышение эффективности обучение на 5% приводит к снижению объема обучения более чем на 30%.

Таким образом, разработанная методика позволяет планировать и управлять многоэтапным процессом обучения, что обеспечивает его эффективность.

## ЗАКЛЮЧЕНИЕ

В отличие от традиционных моделей двухфакторного дисперсионного анализа модель Г. Раша позволяет не только оценить влияние факторов сложности, но и рассчитать вероятность успешного обучения как в дискретной, так и в непрерывной шкалах, провести прогноз и планировать число циклов обучения, что невозможно при использовании традиционных моделей дисперсионного анализа. Эта особенность предложенного подхода в значительной степени помогает оценить и проанализировать процесс подготовки специалистов в области эксплуатации авиационных и космических систем в динамике, предоставляя информацию для своевременной коллекции процесса обучения и выработки наиболее оптимальных решений.

## СПИСОК ЛИТЕРАТУРЫ


1. *Алымов В.Н., Теренин С.С., Щербак В.В.* Технические средства обучения инженерно-технического персонала воздушных судов гражданской авиации // Труды МИЭА. Навигация и управление летательными аппаратами. 2015. № 11. С. 29-33.
2. *Федоренко В.С., Галушка С.А., Семоненко Ю.Ф.* К вопросу об оценке уровня профессиональной подготовки авиационного персонала с применением технических средств обучения // Фундаментальные исследования. 2015. № 7-2. С. 348-353.
3. *Александровская Л.Н., Кириллин А.В., Шумская Л.П.* Математические модели в оценивании эффективности повышения квалификации // Качество. Инновации. Образование. 2014. № 8 (111). С. 30-34.
4. Модель оценки эффективности переподготовки специалистов при многоуровневом обучении / *Л.Н. Александровская, А.В. Кириллин, П.А. Иосифов, И.П. Митрофанова* // Качество. Инновации. Образование. 2015. №1 (116). С. 3-10.
5. *Рао С.Р.* Линейные статистические методы и их применение. М.: Наука, 1968. 548 с.
6. *Монтгомери Д.К.* Планирование эксперимента и анализ данных. Л.: Судостроение, 1980. 384 с.
7. *Джонсон Н., Лион Ф.* Статистика и планирование эксперимента в технике и науке. Методы планирования эксперимента. М.: Мир, 1981. 520 с.
8. *Крюков С.П., Бодрунов С.Д., Александровская Л.Н. и др.* Методы анализа и оценивания рисков в задачах менеджмента безопасности сложных технических систем. СПб.: Корпорация «Аэрокосмическое оборудование», 2007. 460 с.
9. *Ким В.С.* Тестирование учебных достижений. Монография. Уссурийск: Издательство УГПИ. 2007. 214 с.


## COMPARATIVE ANALYSIS OF THE METHODS OF EVALUATING THE EFFICIENCY OF LEARNING


© 2017 L.N. Aleksandrovskaya[1], A.V. Kirillin[1], V.V. Medvedev[2]

[1] Moscow Aviation Institute (National Research University)
[2] K.G. Razumovsky Moscow State University of Technologies and Management



In the article, based on a brief review of the mathematical apparatus of two-way ANOVA in a discrete binary and continuous scales comparative analysis is made in the areas of possible limitations of each of these two options. To solve the problems of split students preparedness assess and the complexity of tasks used for their test and control, a logistic model of G. Rasch, that has a universal character and allows both to produce the desired assessment in both scales, and to extrapolate results planning for the necessary and sufficient to achieve the desired effectiveness of the training cycle is proposed.


*Keywords:* learning efficiency, the degree of preparedness of students, the degree of complexity of control tasks, two-way ANOVA, probability of success, logistic model G. Rasch.



*Lidiya Aleksandrovskaya, Doctor of Technics, Professor at the Flight Vehicles Tests Department.*
*E-mail: aleksandrovskayaln@mati.ru*
*Andrey Kirillin, Senior Lecturer at the Flight Vehicles Tests Department. E-mail: kirillinav@mati.ru*
*Victor Medvedev, Candidate of Technics, Associate Professor at the Quality Management of Innovative High-Tech Industries Department. E-mail: medvik@hotbox.ru*