На правах рукописи

Колданов Петр Александрович

ПОСТРОЕНИЕ МОДЕЛИ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ ИТОГОВ ПРИЕМА В ФИЛИАЛЫ ВУЗА

Специальность 05.13.18 – Математическое моделирование, численные методы и комплексы программ

Автореферат

диссертации на соискание ученой степени кандидата технических наук

Петрозаводск 2009

Работа выполнена в Нижегородском государственном университете им. Н.И. Лобачевского

Научный руководитель:
доктор физико-математических наук, профессор
Тихов Михаил Семенович

Официальные оппоненты:
доктор технических наук, профессор
Питухин Евгений Александрович

доктор физико-математических наук, профессор
Хохлов Юрий Степанович

Ведущая организация –
Уфимский государственный авиационный технический университет.

Защита состоится "___"_________ 20___ г. в "_____" на заседании диссертационного совета Д 212.190.03 в Петрозаводском государственном университете (185910, Республика Карелия, Петрозаводск, пр. Ленина, 33).
С диссертацией можно ознакомиться в библиотеке Петрозаводского государственного университета.

Автореферат разослан "___"_________ 20___ г.

Ученый секретарь
диссертационного совета
д.т.н., профессор Рогов А.А.

Общая характеристика работы

Актуальность работы. Болонская "Декларация о Европейском пространстве для высшего образования", подписанная Россией в 2003 г., и "Концепция модернизации российского образования на период до 2010 г." основными целями определяют повышение качества и доступности образования. В связи с существенным увеличением количества ВУЗов и их филиалов, обеспечивающих доступность, в литературе появилось большое количество работ, посвященных анализу качества образования (Арефьев В.П., Михальчук А.А.), исследованию систем функционирования ВУЗов (Клещевский Ю.Н., Питухин Е.А.), оценке эффективности их деятельности (Абдулгалимов А.), механизмов формирования образовательных услуг (Бельчик Т.А.). Однако в указанных работах не затрагиваются вопросы сравнительного анализа деятельности подразделений ВУЗа и составления рейтинга филиалов.

При составлении рейтинга ВУЗов учитывается множество различных параметров, таких как количество преподавателей, имеющих ученую степень, библиотечный фонд и т.д. Однако в последнее время высказывается мнение (Филиппов В.М.) о важности внешней оценки ВУЗов, в частности, работодателями. Очевидно, к числу наиболее важных показателей престижности ВУЗа относится количество поступающих (абитуриентов).

В диссертации построена математическая модель приема в образовательные учреждения, где в качестве основного показателя выступает численность абитуриентов. На основе этой модели проводится многоальтернативный статистический анализ итогов приема в филиалы, что позволяет в конечном итоге составить их рейтинг. Вместе с тем, полученный результат может быть использован для составления рейтинга ВУЗов.

Заметим, что в литературе известны математические модели, описывающие деятельность различных организаций (Аркин В.И., Бияков О.А., Вальц О.В.), к исследованию которых применяются методы математической статистики (Бородин И.А., Виноградов О.П.). Однако указанные модели не учитывают специфику задачи анализа итогов приема в филиалы ВУЗа и поэтому не могут быть применены к нашей проблеме.

Рассмотренная в диссертации статистическая задача может быть отнесена к задаче межлабораторного сравнения результатов эксперимента (Biggerstaff B.J., Carlinet Y., Leigh S., Mills K.L.,Rukhin A.L., Vangel M.G.). Однако в работах этих авторов исследуется задача *оценки общего для всех значения параметра*. В настоящей диссертации решается *задача различения гипотез* сравнительного анализа для малых выборок. Такую задачу не удается решить как методами дисперсионного анализа, так как не выполняется условие однородности дисперсий, так и байесовскими методами, поскольку из-за небольшого периода наблюдений (8 лет) не представляется возможным указать априорное распределение различаемых гипотез.

Задача создания математического аппарата, позволяющего проводить анализ итогов приема, актуальна, по крайней мере, с двух точек зрения: 1) с прикладной – для получения объективной сравнительной оценки деятельности ВУЗов и их филиалов, как информация для принятия управленческих решений с целью повышения эффективности будущего приема; 2) с теоретической – построенная в диссертации модель приводит к задаче многоальтернативного различения гипотез (в математической статистике эта проблема изучена недостаточно полно). Кроме того, реальные данные представляют собой малую выборку (не более 8 наблюдений в каждом из филиалов), а проблема анализа малых выборок является актуальной и мало изученной.

Цель и задачи исследования. Целью диссертационной работы является разработка, а также исследование математической модели и основанного на ней многоальтернативного статистического анализа итогов приема абитуриентов на различные программы и формы обучения, и его применение для построения рейтинга филиалов конкретного ВУЗа по реальным данным.

Задачами исследования являются:

1. Построение математической модели результатов приемных кампаний в образовательные учреждения.

2. Проверка непротиворечивости реальных данных основным предположениям модели и выделение групп филиалов, работающих в одинаковых условиях.

3. Построение и анализ статистических тестов для проверки гипотезы однородности итогов приема в различные филиалы и их применение.

4. Построение и анализ оптимального статистического теста для сравнения двух совокупностей как основы многоальтернативного сравнительного анализа N совокупностей.

5. Построение и анализ теста комбинированной структуры для многоальтернативного различения гипотез.

6. Получение выводов о рейтинге филиалов на основе применения построенного теста по реальным данным приема в филиалы Нижегородского государственного университета (ННГУ) за 2000 − 2007 г.г.

Объект и инструмент исследования. Объектом исследования являются итоги приемных кампаний в филиалы ВУЗа. Для примера анализируются реальные данные приемных кампаний в филиалы ННГУ за 2000 − 2007 г.г. Инструментом исследования являются вероятностные методы построения математических моделей и статистические методы различения двух и более гипотез.

Методология и методы проведенного исследования. Для проверки непротиворечивости реальных данных основным предположениям модели и выделения групп филиалов, работающих в одинаковых условиях, использовались две методики. Первая основана на применении модификации теста Фишера к каждой паре городов. Во второй применяется нетривиальная модификация теста Саммиудина проверки равенства нескольких дисперсий. Модификация заключается в проверке гипотезы пропорциональности дисперсий с известным коэффициентом пропорциональности.

Для проверки гипотезы однородности нескольких совокупностей использовались тесты, основанные на пересечении доверительных интервалов для параметров, характеризующих математическое ожидание численности наборов на различные филиалы, программы, формы обучения.

Для многоальтернативного различения гипотез в качестве основы используется метод Лемана построения тестов комбинированной структуры.

Научная новизна и научная значимость полученных результатов. Все основные результаты диссертационной работы являются новыми и впервые опубликованы в работах диссертанта.

Предложена математическая модель, описывающая результаты приемных кампаний в образовательные учреждения, исследован метод построения тестов комбинированной структуры и специфика его применения для анализа предложенной модели. Впервые методы математической статистики применяются с целью построения статистических тестов для сравнительного анализа итогов приема в филиалы ВУЗа.

Основной теоретический результат заключается в соединении метода конструкции тестов комбинированной структуры с методом анализа совокупности малых выборок, применительно к разработанной математической модели.

Практическая значимость полученных результатов. Результаты, полученные в диссертационной работе, были применены для сравнительного анализа спроса на получение высшего образования в различных филиалах ННГУ, по различным программам, формам обучения и специальностям и были использованы для разработки стратегии управления филиалами.

Методика анализа, описанная в диссертационной работе, может быть использована как для анализа итогов приема в конкретный ВУЗ, так и для сравнительного анализа итогов приема в различные образовательные учреждения.

Основные положения, выносимые на защиту:

1. Математическая модель, описывающая результаты приема в образовательные учреждения.

2. Анализ адекватности построенной модели реальным данным итогов приема в филиалы ННГУ на различные программы и формы обучения.

3. Статистические тесты для многоальтернативного принятия решений, и их исследование численными методами.

4. Методика использования статистического анализа совокупности малых выборок.

5. Методика получения выводов по результатам применения тестов комбинированной структуры по совокупности малых выборок и сравнительный анализ итогов приема в филиалы ННГУ.

Опубликованность результатов. Результаты диссертации опубликованы в девяти работах, из них две статьи – в журнале "Нелинейный мир", рекомендованном ВАК; три работы опубликованы в сборниках статей по материалам международных конференций (две – в трудах конференции "Прикладная статистика в социально-экономических проблемах" (Н.Новгород), одна – в трудах конференции "Математические методы и информационные технологии в экономике, социологии, образовании" (Пенза)); четыре публикации – в журнале "Обозрение прикладной и промышленной математики". В совместной работе автору принадлежит использование методов анализа совокупности малых выборок, исследование тестов, и выполнение конкретного анализа, а соавтору – идея работы.

Апробация работы. Результаты диссертации докладывались на Международной конференции "Прикладная статистика в социально-экономических проблемах" (Н.Новгород, 2003 г.), Восьмом Всероссийском симпозиуме по прикладной и промышленной математики (Кисловодск, 2008 г.), Девятом Всероссийском симпозиуме по прикладной и промышленной математике (Волгоград, 2008 г.), на семинарах кафедры теории статистических решений и кафедры прикладной теории вероятностей ННГУ.

Структура диссертации. Диссертация состоит из введения, четырех глав, заключения и приложения. Объем диссертации – 140 страниц текста. Объем приложений – 75 страниц. Список литературы – 114 наименований на 15 страницах. Общий объем – 230 страниц.

Краткое содержание работы

Во введении обоснована актуальность темы, описаны основные задачи диссертации и методы их решения. В частности, отмечено, что для решения поставленной задачи используется метод, теоретические основы которого изложены в работах Лемана. Однако для его применения к анализу конкретной ситуации потребовались дополнительные математические исследования.

Вопросам построения тестов однородности нескольких совокупностей и использования информации в малых выборках уделялось достаточно много внимания (Samiudin M., Atiqullah M.A., Hanif M., Asad P.), однако основным предположением в этих работах являлось предположение равенства дисперсий. В нашем исследовании мы строим тесты для случая, когда дисперсии связаны соотношением $\left(\frac{(\sigma^1)^2}{n^1} = \frac{(\sigma^2)^2}{n^2} = \ldots = \frac{(\sigma^N)^2}{n^N}\right)$, где n^i – известные коэффициенты ($i = \overline{1, N}$). Кроме того, тест Фишера, используемый в указанных работах, недостаточно устойчив при отклонении от нормальности, что отмечалось Г. Шеффе и другими авторами. В данной работе мы строим тесты устойчивые при отклонении от нормальности.

В работах (Rukhin A.L., Vangel M.G.) для модели дисперсионного анализа (так называемая задача межлабораторного сравнения) рассматривались задачи *рекуррентного оценивания* параметра в случае, когда дисперсия известна с точностью до коэффициента $\sigma^i = \frac{\sigma^0}{k^i}$. Аналитическое решение таких задач представляет большие трудности. Мы рассматриваем задачи *проверки гипотез* о параметрах.

В работе строится тест различения *многих* ($L \geq 3$) гипотез. Исследуются его характеристики численными методами. Задача многоальтернативного сравнения нескольких совокупностей, рассматриваемая здесь, является новой.

В главе 1 построена математическая модель итогов приема в филиалы ВУЗа и сформулированы основные задачи, решаемые в диссертации.

Будем считать, что численность итогов приема в филиал города i – случайная величина $X^i = \sum\limits_{k=1}^{n^i} \xi_k^i$ (ξ_k^i – индикатор того, что k-й житель из i-го города будет поступать в филиал), имеющая нормальное распределение $X^i \in N\left(a^i, (\sigma^i)^2\right)$, $a^i = \sum\limits_{k=1}^{n^i} p_k^i$, $(\sigma^i)^2 = \sum\limits_{i=1}^{n^i} p_k^i(1-p_k^i)$, $p_k^i = P(\xi_k^i = 1)$, (если $p_k^i = p^i$ для всех k, то $(\sigma^i)^2 = n^i p^i(1-p^i) = n^i(\sigma_0^i)^2$).

В качестве основного параметра, характеризующего эффективность приемных кампаний (или спрос на получение высшего образования), использован параметр $p^i = \dfrac{a^i}{n^i}$, где n^i – число жителей города i, которые потенциально ($p_k^i > 0$) могут обучаться в филиале.

Существенным предположением для построения и исследования модели является: ${\sigma^j}^2 = n^j \sigma_0^2$, $\forall j = \overline{1, N}$ или $\dfrac{{\sigma^i}^2}{{\sigma^j}^2} = \dfrac{n^i}{n^j}$, где $i, j \in \{1, \ldots, N\}$. Такое предположение означает, что разброс численности наборов зависит лишь от численности городов.

Итоги приема удобно представить в виде матрицы $\|x_i^j\|$, где x_i^j – численность набора в j-ый филиал в год i ($j = \overline{1, N}$, $i = \overline{1, m_j}$), N – количество филиалов, m_j – количество наборов.

В качестве примера в диссертации рассматриваются реальные данные об итогах приема в 8 филиалов ННГУ в течение $2000 - 2007$ гг. В частности, итоги приемных кампаний на все программы получения образования имеют вид:

год\ филиал	1Ф	2Ф	3Ф	4Ф	5Ф	6Ф	7Ф	8Ф
2000	254,00	199,00	187,00	182,00	-	-	-	-
2001	238,00	381,00	305,00	148,00	193,00	99,00	235,00	-
2002	386,00	447,00	612,00	205,00	233,00	396,00	444,00	101,00
2003	328,00	445,00	700,00	247,00	223,00	563,00	400,00	129,00
2004	393,00	388,00	706,00	261,00	223,00	570,00	465,00	119,00
2005	447,00	376,00	685,00	284,00	235,00	479,00	364,00	102,00
2006	568,00	390,00	726,00	374,00	205,00	404,00	348,00	111,00
2007	432,00	302,00	651,00	391,00	260,00	532,00	438,00	54,00

где 1Ф – филиал №1, 2Ф – филиал №2,..., 8Ф – филиал №8.

Классическая задача однофакторного анализа состоит в проверке гипотезы однородности H_1: $p^1 = p^2 = \ldots = p^N$ в предположении, что дисперсии наблюдений *одинаковы*. Это соответствует тому, что спрос на высшее образование в различных филиалах одинаков.

Основной задачей диссертации является задача сравнительного анализа эффективностей итогов приема в филиалы, которая формулируется следующим образом: требуется построить правило (тест со многими решениями) для различения гипотез:

$$\begin{gathered} H_1 : p^1 = p^2 = \ldots = p^N; \ldots H_{i_1} : p^1 > p^2 = \ldots = p^N; \ldots \\ H_{i_2} : p^1 > p^2 > p^3 = p^4 = \ldots = p^N; \ldots H_L : p^1 < p^2 < \ldots < p^N. \end{gathered} \tag{1}$$

Гипотеза H_1 означает, что спрос на высшее образование в городах, где работают филиалы, одинаков; гипотеза H_{i_1} означает, что спрос на высшее образование в филиалах $2, 3, \ldots, N$ одинаков и ниже, чем спрос на высшее образование в филиале 1; и, наконец, гипотеза H_L означает, что спрос в филиале 1 ниже, чем в филиале 2, в филиале 2 ниже, чем в филиале 3,..., спрос в филиале $N-1$ ниже, чем в филиале N.

Показано, что количество возникающих гипотез при анализе итогов приема в N филиалов $L = \sum\limits_{r=1}^{N} \sum\limits_{k_1+k_2+\ldots+k_r=N} \dfrac{N!}{k_1!k_2!\ldots k_r!}$ (или асимптотически $L \sim N!\left(2(\ln 2)^{N+1}\right)^{-1}$), которое быстро растет с ростом N (при $N = 3$,

$L = 13$; при $N = 5$, $L = 541$). Поэтому целесообразно использовать метод, позволяющий редуцировать задачу к возможно меньшему числу выполняемых сравнений.

В главе 2 проводится проверка непротиворечивости реальных данных основному предположению модели о пропорциональности разброса численностей наборов в различных городах ($\sigma^{j^2} = n^j \sigma_0^2$, $\forall j = \overline{1, N}$). Проверка основана на применении модифицированных тестов Фишера и Саммиудина. Показано, что можно выделить однородные группы филиалов, для которых это предположение не отвергается.

Для проверки гипотезы однородности предложен статистический тест, основанный на пересечении доверительных интервалов (l_1^i, l_2^i) для параметров p^i $(i = \overline{1, N})$, который имеет вид:

$$\phi(x) = \begin{cases} 1, & \text{если} \quad \bigcap\limits_{i=1}^{N} (l_1^i, l_2^i) = \varnothing, \\ 0, & \text{если} \quad \bigcap\limits_{i=1}^{N} (l_1^i, l_2^i) \neq \varnothing. \end{cases} \tag{2}$$

Методом статистического моделирования показано, что тест (2) имеет удовлетворительную мощность.

При анализе итогов приема в филиалы из однородных групп для построения доверительных интервалов использованы методы анализа совокупности малых выборок. Численными методами показано, что при использовании таких доверительных интервалов мощность теста (2) увеличивается.

Основным результатом главы 2 является построение теста проверки гипотезы H: $p^i \geq p^j$ против альтернативы K: $p^i < p^j$, который имеет вид:

$$\varphi(X^i, X^j) = \begin{cases} 1, & T_{ij}(X^i, X^j) \leq c, \\ 0, & T_{ij}(X^i, X^j) > c, \end{cases} \tag{3}$$

где

$$T_{ij}(X^i, X^j) = \frac{\left(\dfrac{\overline{X^i}}{n^i} - \dfrac{\overline{X^j}}{n^j}\right) \Big/ \sqrt{\dfrac{1}{m_i} + \dfrac{1}{m_j}}}{\sqrt{\left(\dfrac{1}{n^{i2}} \sum\limits_{k=1}^{m_i} \left(X_k^i - \overline{X^i}\right)^2 + \dfrac{1}{n^i n^j} \sum\limits_{k=1}^{m_j} \left(X_k^j - \overline{X^j}\right)^2\right) / (m_i + m_j - 2)}}, \tag{4}$$

константа $c = t_{\alpha, m_1+m_2-2}$ – квантиль распределения Стьюдента порядка $1-\alpha$. Доказано, что этот тест (обобщающий известный (Э.Леман)) является равномерно наиболее мощным в классе несмещенных.

Предложена модификация теста (3) за счет использования методов анализа совокупности малых выборок. При этом для получения выводов об эффективности приемных кампаний в два филиала используются результаты приемных кампаний во все филиалы из выделенной однородной группы. Модифицированный тест имеет вид:

$$\varphi'(x) = \begin{cases} 1, & T'_{ij}(x) \le c, \\ 0, & T'_{ij}(x) > c, \end{cases} \tag{5}$$

$$T'_{ij}(x) = \frac{\left(\dfrac{\overline{X^i}}{n^i} - \dfrac{\overline{X^j}}{n^j}\right) \Big/ \sqrt{\dfrac{1}{m_i} + \dfrac{1}{m_j}}}{\sqrt{\left(\dfrac{1}{n^{i2}} \sum\limits_{l=1}^{m_i} \left(X_l^i - \overline{X^i}\right)^2 + \sum\limits_{\substack{k=1 \\ k \ne i}}^{N} \dfrac{1}{n^i n^k} \sum\limits_{l=1}^{m_k} \left(X_l^k - \overline{X^k}\right)^2\right) \Big/ (n-r)}}, \tag{6}$$

$c = t_{\alpha, n-N}$, $n = \sum\limits_{i=1}^{N} m_i$. Численный анализ показал, что мощность теста (5) больше, чем мощность теста (3).

В главе 3 исследуется методика Лемана различения $L \ge 3$ гипотез, на основе которой строится тест для различения (1). Ее конкретное применение потребовало нетривиальных исследований.

Именно, пусть задано семейство распределений $\Im = \{f(x, \theta) : \theta \in \Omega\}$ и требуется построить статистическую процедуру для различения L основных гипотез:

$$H_i : \theta \in \Omega_i, \quad i = 1, \dots, L; \quad \Omega_i \bigcap \Omega_j = 0, \quad \bigcup_{i=1}^{L} \Omega_i = \Omega. \tag{7}$$

Метод построения тестов комбинированной структуры для различения (7) основан на трех основных положениях:

1. Многоальтернативная задача выбора одной из L гипотез (7) эквивалентна совокупности M задач проверки гипотез $H'_j : \theta \in \omega_j$ против альтернатив $K'_j : \theta \in \omega_j^{-1}$, $\left(\bigcup\limits_{i=1}^{L} \Omega_i = \omega_j \bigcup \omega_j^{-1} = \Omega, j = 1, \dots, M, \omega_j \bigcap \omega_j^{-1} = \emptyset\right)$.

Гипотезы H'_j называются порождающими задачу различения основных гипотез H_i. Число M и области ω_i могут быть выбраны неоднозначно.

2. При выполнении определенного условия существует взаимно-однозначное соответствие между решающей процедурой $\delta(x)$ различения L гипотез и семейством $\{\delta'_\gamma : \gamma = 1 \dots M\}$ тестов порождающих гипотез. Такое условие называется условием совместимости и имеет вид:

$$P_\theta\left(x \in \bigcup_{i=1}^{L} A_i\right) = 1 \quad \text{или} \quad (\text{при } M = 2) \quad p_\theta\{(\delta'_1(x), \delta'_2(x)) \in D\} = 1 \\ \text{при} \quad \forall \theta \in \Omega, \tag{8}$$

где A_i – область принятия H_i, $\delta, \delta'_1, \delta'_2$ – решающие функции с пространствами решений $D = \{d_i : i = 1 \dots L\}$, $D'_1 = \{d'_{1i} : i = 1 \dots K\}$, $D'_2 = \{d'_{2j} : j = 1 \dots S\}$, связанные соотношением $\delta = (\delta'_1, \delta'_2)$, которое означает, что если $\delta'_1 = d'_{1i}$, $\delta'_2 = d'_{2j}$, то $\delta = d_{ij}$, причем d_{ij} равносильно одновременному принятию решений d'_{1i} и d'_{2j}, т.е. $d_{ij} = (d'_{1i}, d'_{2j})$, $d_{ij} \in D$.

При выполнении (8) невозможно такое сочетание d'_γ и d'_j (решений об истинности H'_γ, H'_j), которое приводит к не существующему в D решению.

Так, для трехальтернативной задачи:

$$H_1 : \theta < \theta_0, \quad H_2 : \theta = \theta_0, \quad H_1 : \theta > \theta_0, \tag{9}$$

используем две порождающие гипотезы: $H'_1 : \theta \geq \theta_0$ против $K'_1 : \theta < \theta_0$ и $H'_2 : \theta \leq \theta_0$ против $K'_2 : \theta > \theta_0$.

Условие (8) будет выполняться, если вероятность одновременного отвержения $H'_1 : \theta \geq \theta_0$ и $H'_2 : \theta \leq \theta_0$ будет равна 0 при любом θ.

Для семейства с монотонным отношением правдоподобия и достаточной статистикой T гипотеза $H'_1(H'_2)$ отвергается, если $T < c_1$ $(T > c_2)$, и константы определяются из уравнений: $P_{\theta_0}(T < c_1) = \alpha_1$ $(P_{\theta_0}(T > c_2) = \alpha_2)$. Гипотезы H'_1 и H'_2 одновременно отвергаются, если $c_2 < T < c_1$. Так как функция распределения – неубывающая функция, то из $\alpha_1 + \alpha_2 < 1$ следует $c_1 < c_2$, и $P_\theta(c_2 < T < c_1) = 0$ при любом θ. Таким образом, условие совместимости (8) в этом случае эквивалентно условию $\alpha_1 + \alpha_2 < 1$.

3. Если ввести функции потерь $w(\theta, \delta)$, то при определенном условии комбинация оптимальных решений двухальтернативных задач приводит к оптимальному решению многоальтернативной задачи. Такое условие называется условием аддитивности функции потерь и формулируется следующим образом:

$$w(\theta, \delta) = w_1(\theta, \delta_1') + w_2(\theta, \delta_2'), \quad \delta = (\delta_1', \delta_2'). \tag{10}$$

Известно (Леман), что при выполнении условий (8), (10) справедливо:

1) комбинация байесовских функций $\delta_1'(x)$ и $\delta_2'(x)$ приводит к байесовской функции $\delta(x)$, если они построены при одном априорном распределении.

2) если $\delta_1'(x)$ и $\delta_2'(x)$ – минимаксные, то $\delta(x)$ – минимаксная.

3) если $\delta_1'(x)$ и $\delta_2'(x)$ – несмещенные, то $\delta(x)$ – несмещенная.

4) если функции мощности оптимальных в классе несмещенных тестов $\delta_\gamma'(x)$ – непрерывны по θ для всех $\gamma = 1, \ldots, M$, то $\delta(x)$ оптимальна в том же классе.

Это позволяет конструировать процедуры с L решениями, используя соответствующим образом подобранные процедуры проверки порождающих гипотез, причем при комбинации оптимальных решений двухальтернативных задач полученное решение также будет оптимальным, если выполняются условия совместимости и аддитивности функции потерь.

Исследованы особенности, возникающие при построении теста с помощью метода комбинированной структуры. Показано, что условие аддитивности (10) приводит к ограничениям на компоненты функции потерь. Найдены ограничения на вероятности ошибочных решений и их линейные комбинации (при различении $L \geq 3$ гипотез), вытекающие из условий несмещенности и аддитивности функции потерь. Исследована редукция числа проверок, необходимых для различения L гипотез при конструировании решающего правила методом комбинированной структуры. Для различения L гипотез максимальная редукция заключается в применении

$$M = \begin{cases} \log_2 L, & L = 2^M \\ [\log_2 L] + 1, & 2^{M-1} < L < 2^M \end{cases}$$

двухальтернативных правил, и

$$M = \begin{cases} \log_3 L, & L = 3^M \\ [\log_3 L] + 1, & 3^{M-1} < L < 3^M \end{cases}$$

трехальтернативных решающих правил.

В разделе 3.5 построен статистический тест различения 3-х гипотез (9)

$$H_1 : p^i > p^j, \quad H_2 : p^i = p^j, \quad H_3 : p^i < p^j,$$

который имеет вид:

$$\delta_1(x) = \begin{cases} d_1, & T_{ji}(x) > c'_{ji} \\ d_2, & c_{ij} < T_{ji}(x) < c'_{ji} \\ d_3, & T_{ji}(x) < c_{ji} \end{cases} \tag{11}$$

где d_s $(s = 1, 2, 3)$ – решение принять гипотезу H_s, $T_{ji}(x)$ находится из (4).

$$P\left(T_{ji} < c_{ji} | p^i = p^j\right) = \alpha_{ij}, \quad P\left(T_{ji} > c'_{ji} | p^i = p^j\right) = \alpha_{ji}. \tag{12}$$

Заметим, что T_{ji} можно заменить на T'_{ji} из (6). Тест (11) основан на комбинации двух равномерно наиболее мощных в классе несмещенных тестов (3) (или 5). Он является оптимальным в классе несмещенных при выполнении условий (8), (10), для чего достаточно выполнения $\alpha_{ij} + \alpha_{ji} < 1$ при всех $i, j = \overline{1, N}$. Тест (11) является центральным в конструкции решения основной задачи диссертации.

Рассмотрим статистический тест различения основных гипотез (1):

$$H_1 : p^1 = p^2 = \ldots = p^N; \ldots H_{i_1} : p^1 > p^2 = \ldots = p^N; \ldots$$
$$H_L : p^1 < p^2 < \ldots < p^N.$$

Показано, что для выполнения условия (8) необходимо, чтобы:

$$P(T_{kj} < c_{kj}, T_{ki} < c_{ki}, T_{ij} < c_{ij}) = 1.$$

Пусть $n^i = n^j = n^k$, $m_i = m_j = m_k$. Тогда, $T_{kj} = T_{ki} + T_{ij}$. Константы c_{ij}, c_{ki}, c_{kj} выбираются из уравнений $P\left(T_{ij} < c_{ij} | p^i = p^j\right) = \alpha_{ij}$, $P\left(T_{kj} < c_{kj} | p^k = p^j\right) = \alpha_{kj}$, $P\left(T_{ki} < c_{ki} | p^k = p^i\right) = \alpha_{ki}$, и, следовательно, при $\alpha_{ij} = \alpha_{kj} = \alpha_{ki}$ выполняется $c_{ij} = c_{ki} = c_{kj} = c$. В этом случае $P(T_{kj} < 2c, T_{ki} < c, T_{ij} < c) = 1$, а $P(T_{kj} < c, T_{ki} < c, T_{ij} < c) \leq 1$, т.е. при принятии гипотез $h_{ij} : p^i \leq p^j$ и $h_{ki} : p^k \leq p^i$ мы не можем гарантировать принятие гипотезы $h_{kj} : p^k \leq p^j$. Таким образом, при таком способе выбора системы порождающих гипотез условие совместимости (8) не выполняется.

Чтобы устранить это противоречие, будем различать гипотезы с точностью Δ, $(\Delta > 0)$, т.е. в качестве порождающих использовать гипотезы $h'_{ij} : p^i \leq p^j + \Delta$ против альтернатив $k'_{ij} : p^i > p^j + \Delta$ $(i, j = 1, \ldots, N)$. Показано, что в этом случае проблема совместимости может быть успешно разрешена.

В разделе 3.6 построен статистический тест различения 3-х гипотез $H'_1 : p^i + \Delta < p^j$; $H'_2 : |p^i - p^j| < \Delta$; $H'_3 : p^i > p^j + \Delta$ (или $H'_1 : p^i \overset{\Delta}{<} p^j$; $H'_2 : p^i \overset{\Delta}{=} p^j$; $H'_3 : p^i \overset{\Delta}{>} p^j$), который имеет вид (11), но константы c''_{ji}, c'''_{ji} определяются из уравнений:

$$\begin{aligned} &P\left(T_{ji}(x) > c''_{ji} | p^i + \Delta = p^j\right) = \alpha_{ij}, \\ &P\left(T_{ji}(x) < c'''_{ji} | p^i = p^j + \Delta\right) = \alpha_{ji}. \end{aligned} \tag{13}$$

Тест (11) с условием (13) отличается от теста (11) с условием (12) константами c''_{ji}, c'''_{ji} и является оптимальным в классе несмещенных при $\alpha_{ij} + \alpha_{ji} < 1$ $(i, j = \overline{1, N})$.

Построен статистический тест сравнения итогов приема в N филиалов, заключающийся в комбинировании правил (11) с условием (13). Показано, что решение основной задачи диссертации является оптимальным в классе несмещенных.

Проведен анализ эффективности построенных тестов различения двух и более гипотез. Для тестов различения двух гипотез в качестве показателя эффективности используется функция мощности. Для тестов различения трех гипотез в качестве показателя эффективности используется вектор

$\langle a_1, a_2, a_3 \rangle$, где $a_i = P(d_i/H_j)$, d_i – решение о принятии гипотезы H_i. Приведена программа, вычисляющая вектор $\langle a_1, a_2, a_3 \rangle$ методом Монте-Карло. Показано, что эти тесты имеют приемлемые вероятности правильных решений.

В главе 4 полученные результаты применяются для анализа эффективности итогов приема в филиалы ННГУ:

- Выделены группы однородных филиалов, для которых выполняется основное предположение разработанной математической модели о пропорциональности дисперсий числа абитуриентов в филиалах одной группы. Показано, что такие группы зависят от того, на какие программы, формы, специальности и их сочетания ведется прием.

- Показано, что самая эффективная приемная кампания на все формы, программы и специальности проводилась в 7Ф, самая неэффективная – в 3Ф. При этом спрос в 1Ф, 2Ф одинаков (с точностью Δ), спрос в 2Ф, 4Ф, 6Ф одинаков (с точностью Δ), однако спрос в 1Ф больше, чем спрос в 4Ф. 6Ф.

- Показано, что при анализе итогов приема на полные программы очного и заочного обучения 3Ф и 7Ф сохраняют свои позиции, а именно, в 7Ф самый высокий спрос, в 3Ф – самый низкий. При этом:

  – Спрос на дневную форму обучения в 2Ф и 6Ф одинаков (с точностью Δ), и уступает спросу в 1Ф. Однако популярность (отношение числа выпускников школ, поступающих на дневное отделение к общему числу абитуриентов – выпускников школ) очной формы обучения в 6Ф выше, чем в 7Ф, 1Ф.

  – Спрос на полную программу обучения по заочной форме в 6Ф является самым низким, т.е. не отличается (с точностью Δ) от спроса в 3Ф. Спрос в 1Ф, 2Ф, 4Ф не отличается между собой.

- Показано, что самая эффективная приемная кампания на сокращенную программу обучения по заочной форме проведена в 4Ф, самая неэффективная — в 3Ф. При этом спрос в 1Ф, 2Ф, 6Ф, 8Ф не отличается (с точностью Δ).

- Показано, что при анализе популярности (отношения числа поступающих на конкретную специальность к общему числу абитуриентов) конкретной специальности соотношение (по сравнению с анализом спроса) изменяется, а именно: самая низкая популярность на специальность "Экономика и управление на предприятии (в машиностроении)" наблюдается в 6Ф; самая высокая — в 5Ф; в то же время самая высокая популярность специальностей "Финансы и кредит" и "Юриспруденция" наблюдается в 6Ф.

Все результаты получены при $\Delta = 10^{-6}$.

Основные результаты диссертации

- Исследована актуальная задача сравнительного анализа итогов приемных кампаний в структурные подразделения территориально распределенной образовательной организации и предложена соответствующая математическая модель.

- Разработан метод многоальтернативного статистического анализа на основе соединения и развития метода построения тестов комбинированной структуры и методов анализа совокупности малых выборок.

- Полученные теоретические результаты применены для анализа реальных данных об итогах приемных кампаний в филиалы Нижегородского государственного университета им. Н.И.Лобачевского.

- Проделанное исследование показало эффективность предложенного подхода.

Список опубликованных работ по теме диссертации

Работы, опубликованные в журналах, рекомендованных ВАК:

[1] Колданов П.А. Статистический анализ совокупности малых выборок //ж. "Нелинейный мир", М., 2007. Т.5. №7-8. С. 531-535.

[2] Колданов П.А. Построение оптимального критерия статистического анализа результатов приема в образовательные учреждения // ж. "Нелинейный мир", М., 2008. Т.6. №11-12. С. 689 - 696.

Другие публикации:

[3] Колданов П.А., Суслова И.В. Информационная поддержка образовательной деятельности ЦДО ННГУ (опыт практической работы) // Материалы Межд. конф. "Прикладная статистика в социально-экономических проблемах". Н.Новгород.: Изд-во Ниж. ун-та, 2003. Т.2. С. 96-99.

[4] Колданов П.А. Анализ однородности наборов ЦДО ННГУ // Материалы Межд. конф. "Прикладная статистика в социально-экономических проблемах". Н.Новгород.: Изд-во Ниж. ун-та, 2003. Т.2. С. 116-121.

[5] Колданов П.А. Оценка эффективности приемной кампании подразделений образовательного учреждения // Материалы Всероссийск. конф. "Математические методы и информационные технологии в экономике, социологии и образовании". Пенза: Приволжский дом знаний, 2005. С. 206-209.

[6] Колданов А.П., Колданов П.А. Применение методов статистического анализа совокупности малых выборок для оценки эффективности деятельности ВУЗа //ж. "Обозрение прикладной и промышленной математики", М., ОПиПМ, 2005. Т.12, выпуск 2. С. 389-390.

[7] Колданов П.А. О повышении эффективности статистического анализа деятельности филиалов ВУЗа //ж. "Обозрение прикладной и промышленной математики", М., ОПиПМ, 2006. Т.14, выпуск 2. С. 316.

[8] Колданов П.А. Применение метода Лемана различения многих гипотез для оценки деятельности филиалов ВУЗа // ж. "Обозрение прикладной и промышленной математики", М., ОПиПМ, 2008. Т.15, выпуск 4. С. 702-703.

[9] Колданов П.А. Применение метода Лемана различения многих гипотез с учетом совокупности малых выборок для оценки эффективности приемных кампаний в филиалы ВУЗа // ж. "Обозрение прикладной и промышленной математики", М., ОПиПМ, 2008. Т.15, выпуск 4. С. 667.