

## ПЛАНИРОВАНИЕ В ПАССИВНОМ ЭКСПЕРИМЕНТЕ

\*Национальный технический университет “КПИ”, Киев, Украина

---

**Анотація.** Розглянута проблема застосування теорії планування експерименту до пасивного експерименту. Викладені наукова ідея і алгоритм формування підвибірки з пасивного експерименту з заданими властивостями. Результати впроваджені в Науково-дослідному інституті авіації.

**Ключові слова:** регресійний аналіз, планування експерименту, факторний простір, пасивний експеримент.

**Аннотация.** Рассмотрена проблема применения теории планирования эксперимента к пассивному эксперименту. Описаны научная идея и алгоритм формирования подвыборки из пассивного эксперимента с необходимыми свойствами. Результаты были внедрены в Научно-исследовательском институте авиации.

**Ключевые слова:** регрессионный анализ, планирование эксперимента, факторное пространство, пассивный эксперимент.

**Abstract.** The paper investigates the problem of applying the idea theory design of experiments to matrix of passive experiments. The scientific idea and algorithms of forming subsamples from matrix of passive experiments and completing the construction of samples are described. The work results have been reduced to practice in R&D institute of aviation.

**Keywords:** regression analysis, design of experiment, passive experiment.

### 1. Введение. Проблема и цель работы

В теории планирования эксперимента (ТПЭ) разделяются активный и пассивный эксперименты. К активному относится эксперимент, матрица которого построена в соответствии с ТПЭ, к пассивному – все остальное. Тут, с одной стороны, имеется несоответствие с принятым в науке разделением: наблюдение – эксперимент, с другой, – из рассмотрения исключаются не только наблюдения, но и все эксперименты, в которых полное правление изучаемыми факторами частично или в целом невозможно. Заметим, что в выборочном методе такого разделения не имеется. Дело в том, что хотя изначально ТПЭ возникла в рамках выборочного метода, дальнейшее ее развитие пошло в сторону теоретической математики, что в конечном итоге ухудшило условия применения в практической деятельности и привело к сокращению ее использования и определенной дискредитации [1–3]. Это связано с тем, что критерии оптимальности, хоть и имеют статистический смысл, но представляют трудность в использовании для экспериментаторов [3]. Появление робастного планирования экспериментов явилось поворотом ТПЭ в сторону выборочного метода. Вместе с тем и в робастном планировании осталось деление на активный и пассивный эксперименты. Необходима разработка методов и средств, расширяющих сферу ТПЭ на всю предметную область (экспериментальные исследования и обработка результатов наблюдений). В самом выборочном методе этих средств не имеется, поскольку в нем не рассматривается вопрос получения математических моделей по экспериментальным данным. Этот вопрос относится к регрессионному анализу (РА), вопрос получения матриц для которого, в свою очередь, находится в ведении теории планирования экспериментов. Таким образом, необходима разработка методов и средств формирования выборок из пассивного эксперимента, которые должны иметь свойства, необходимые для получения качественных регрессионных моделей [4].

## 2. Робастные планы – основа для планирования пассивного эксперимента

В выборочном методе основными требованиями являются репрезентативность выборки и точность оценок (прочие свойства оценок зависят от выбранного метода обработки данных). Для обеспечения требований нужны размер выборки, структура выборки, рандомизация и независимость наблюдений (экспериментов) [5–9]. В традиционном планировании эксперимента основой является критерий оптимальности, обеспечивающий только один из необходимых аспектов получения модели [3, 7, 10]. Концепция робастного плана эксперимента [11–13], для которого целью является создание наилучших условий для получения регрессионной модели в целом, позволяет выполнить расширение планирования на пассивный эксперимент. Размещение в факторном пространстве точек робастного плана эксперимента обеспечивает независимость и случайность. Следствием первого являются наилучшие условия для определения структуры уравнения регрессии, применения статистических критериев, получения устойчивых оценок коэффициентов уравнения регрессии.

## 3. Принципиальные ограничения в использовании планирования в пассивном эксперименте

Следует иметь в виду, что предлагаемый подход не может быть распространен на все виды пассивного эксперимента. Есть такие его виды, которые принципиально не могут быть исправлены. Особенно это относится к наблюдениям за установившимися процессами, в которых часть значимо влияющих факторов зафиксирована (не изменяется), часть сильно закоррелирована. Из таких данных, разумеется, может быть сформирована выборка и получена модель, но никакого практического смысла она иметь не будет [3, 7, 11, 12].

## 4. Деформация факторного пространства

Для оценки пригодности выборки к формированию подвыборки плана необходимо проверить ее принципиальную пригодность для таких действий. Это может быть сделано через оценку степени отклонения формы облака рассеивания экспериментальных точек относительно идеального гипершара (гиперкуба).

Для определения степени деформации гиперкуба возможно использование показателя, базирующегося на отношении длин фактических (в экспериментальной выборке) и теоретических (при распределении в виде гиперкуба) диагоналей. В дальнейшем рассматривается нормированное в единичный гиперкуб факторное пространство.

Количество вершин в гиперкубе равно  $N_{\text{вершин}} = 2^m$ , количество диагоналей  $K = 2^{m-1}$ , при этом расстояние между противоположными вершинами для идеального единичного гиперкуба  $D_{\text{диаг}} = \sqrt{m}$ , где  $m$  – размерность факторного пространства. Степень деформированности факторного пространства может быть оценена по формуле  $Q_{\text{гиперкуб}} = \prod_{i=1}^K \frac{D_i}{D_{\text{диаг}}}$ , где  $K$  – количество диагоналей,  $D_i$  – длина диагонали в реальном (но нормированном) факторном пространстве,  $D_{\text{диаг}}$  – теоретическое значение диагонали. Значение показателя изменяется в интервале  $0 \leq Q \leq 1$ . Значение 1 показатель принимает в том идеальном случае, когда форма факторного пространства образует гипершар или гиперкуб. Если же две или более координатных осей параллельны, то значения показателя приближается к 0. Чем ближе значение показателя к 1, тем форма пространства ближе к идеальному случаю, чем  $Q$  ближе к нулю, тем пространство ближе к вырожденному случаю. Вырожденность (деформированность) факторного пространства легко связать с мульт-

тиколлинеарностью, на оценках которой базируются показатели робастных планов эксперимента. Связь между деформированностью можно видеть в табл. 1.

Таблица 1. Соотношение между закоррелированностью факторов и деформированностью факторного пространства

Коэффициент деформированности для одной оси (двумерный случай)	Коэффициент корреляции между факторами
0,25	>0,8
0,5	>0,45
0,75	>0,25

Для определения граничного коэффициента деформированности для многомерного случая необходимо соответствующие значения возвести в степень, равную числу осей.

Фактическую деформацию факторного пространства в исходной выборке следует оценить до начала работ по формированию из нее подвыборки для построения модели. Связано это с тем, что в случае ее вырожденности (сильной деформированности) работа с ней невозможна.

Для нормированного в единичный гиперкуб пространства координаты вершин можно записать как двоичные числа, которые отвечают номеру вершины минус единица (или нумерация, начиная с нуля). Например, для трехмерного пространства соответствие номеров и координат вершин представлено в табл. 2.

Таблица 2. Соответствие номеров и координат вершин единичного гиперкуба для трехмерного случая

Номер вершины	Координаты вершин			Код вершины
	X	Y	Z	
1	0	0	0	000
2	0	0	1	001
3	0	1	0	010
4	0	1	1	011
5	1	0	0	100
6	1	0	1	101
7	1	1	0	110
8	1	1	1	111

Из табл. 2 легко видеть, что противоположными вершинами являются те, для которых сумма координат равняется  $N_{\text{вершин}} - 1$  (в данном случае 111<sub>2</sub>), например, первая и восьмая, четвертая и пятая и т.д.

Для определения показателя деформированности пространства Q можно воспользоваться следующим алгоритмом.

Алгоритм расчета показателя деформированности факторного пространства.

1. Определение координат всех теоретических вершин в соответствии с их двоичным кодом для фактической размерности факторного пространства.

2. Для каждой теоретической вершины определяются координаты вершины, соответствующей ей в экспериментальной выборке. В качестве таковой принимается точка, ближайшая к координатам теоретической.

3. Для каждой пары противоположных вершин находится расстояние между ними.

4. Рассчитывается частное отношение деформированности  $\frac{D_i}{D_{\text{diag}}}$ .

Если это отношение меньше заданного критического, то координаты соответствующих вершин запоминаются.

6. Вычисляется накопленное значение коэффициента деформированности факторного пространства  $Q$ .

7. В том случае, когда в п.4 все отношения меньше или равны 1, это означает, что пространство имеет форму гипершара, а не гиперкуба. В таком случае окончательное значение показателя деформированности должно быть скорректировано следующим образом:

$$Q_{гипершар} = Q_{гиперкуб} \times (\sqrt{m})^m.$$

В связи с тем, что в многофакторном случае число теоретических диагоналей очень велико при малом размере выборки, каждая точка фактически будет выступать в роли нескольких вершин, то есть использоваться несколько раз.

## 5. Алгоритм формирования подвыборки из пассивного эксперимента. Описание алгоритма формирования подвыборки

1. Найти граничные точки  $\{X_{cp}\} = \{\min X_i; \max X_i\} \forall i \in (1, m)$ .

2. Нормировать исходную выборку к единичному гиперкубу  $x'_{ij} = \frac{x_{ji} - \min x_i}{\max x_i - \min x_i}$ .

3. Задать предельную гипотетическую сложность зависимости отклика от конкретного фактора как степень аппроксимирующего полинома  $f_i$ .

4. Определить нужное количество экспериментов для построения модели, исходя из гипотетической ее сложности:  $N = (1,5 \dots 2) \sum_{i=1}^M (1 + (f_i - 1))$  [5].

5. Если минимальное полученное значение количества экспериментов меньше фактического размера выборки, то принимается решение о невозможности построения модели или происходит переход в п. 3 для изменения сложности модели.

6. Задать критическое значение деформированности факторного пространства  $Q_{кр}$ .

7. Определить критическое расстояние между точками в выборке  $d_{кр} = \sqrt{\sum_{i=1}^m \Delta_{x_i}^2}$ , где

$$\Delta_{x_i} = \frac{(x_{i \max} - x_{i \min})}{(f_i + 1)}.$$

8. Найти матрицу расстояний между точками  $d_{ij} = \sqrt{\sum_{l=1}^m (x_{il} - x_{jl})^2}$ .

9. Взяв за основу точки, которыми создается ось наибольшей длины, оставить в выборке только те точки, расстояние между которыми больше критического.

10. В том случае, когда в сформированной подвыборке после выполнения предыдущего пункта окажется точек меньше, чем рассчитано в п. 4, то пошагово добавляются новые точки. Выбираются точки, максимально отдаленные от уже имеющихся. После каждого включения характеристики матрицы пересчитываются и принимается решение о продолжении или прекращении процесса.

В тех ситуациях, когда физически возможно проведение дополнительных экспериментов, предлагается алгоритм определения точек, которыми необходимо дополнить выборку для придания ей необходимых свойств.

Эти точки определяются следующим образом.

1. Выбирается диагональ с деформацией больше критической.

2. Для данной диагонали рассчитываются координаты теоретических вершин.

3. Рассчитывается возможный допуск на изменение координат.
4. Координаты и допуск пересчитываются к натуральным значениям.
5. Пп.1–4 выполняются для всех деформированных диагоналей.

После того, как определенные таким образом точки будут присоединены к отобранной ранее подматрице, она приобретет необходимые свойства для построения регрессионной модели.

## 6. Выводы

В работе рассмотрен вопрос распространения методов планирования эксперимента на так называемый пассивный эксперимент. Предложенный подход базируется на идеях робастного планирования экспериментов и позволяет разделить выборку пассивного эксперимента на обучающую, с необходимыми для построения регрессионной модели статистическими свойствами, и контрольную. Предложены алгоритмы для формирования этой подвыборки, показатели степени деформированности пространства и рекомендации по принятию решений.

По данным алгоритмам были написаны макросы для табличного редактора Excel на языке VBA. Разработанные методы и программы использовались в Научно-исследовательском институте авиации при построении математических моделей по данным накопленной статистики.

## СПИСОК ЛИТЕРАТУРЫ

1. Лапач С.Н. Проблемы построения математических моделей экспериментально-статистическими методами / С.Н. Лапач // Прогресивна техніка і технологія машинобудування, приладобудування і зварювального виробництва. Праці НТУУ “КПІ”. – К.: НТУУ “КПІ”, 1998. – Т. 2. – С. 25 – 29.
2. Лапач С.М. Регресійний аналіз. Проблеми росту / Лапач С.М. // 13 міжнар. наук. конф. імені академіка М. Кравчука, (Київ, 13–15 травня 2010 р.). – К.: НТУУ, 2010. – Т. 3. – С. 71.
3. Налимов В.В. Логические основания планирования эксперимента / В.В. Налимов, Т.И. Голикова. – М.: Металлургия, 1981. – [2-е изд. перераб. и доп.]. – 152 с.
4. Лапач С.Н. Основные проблемы построения регрессионных моделей / С.Н. Лапач, С.Г. Радченко // Математичні машини і системи. – 2012. – № 4. – С. 125 – 133.
5. Дружинин Н.К. Выборочное наблюдение и эксперимент / Дружинин Н.К. – М.: Статистика, 1977. – 176 с.
6. Кокрен У. Методы выборочного исследования / Кокрен У. – М.: Статистика, 1976. – 440 с.
7. Налимов В.В. Теория эксперимента / Налимов В.В. – М.: Наука. ГРФМЛ, 1971. – 208 с.
8. Финни Д. Введение в теорию планирования эксперимента / Финни Д. – М.: Наука, ГРФМЛ, 1970. – 288 с.
9. Хикс Ч. Основные принципы планирования эксперимента / Хикс Ч. – М.: Мир, 1967. – 406 с.
10. Математическая теория планирования эксперимента / Под ред. С.М. Ермакова. – М.: Наука, ГРФМЛ, 1983. – 392 с.
11. Лапач С.Н. Статистические методы в фармакологии и маркетинге фармацевтического рынка / Лапач С.Н., Пасечник М.Ф., Чубенко А.В. – К.: ЗАТ “Укрспецмонтаж”, 1999. – 312 с.
12. Лапач С.Н. Статистические методы в медико-биологических исследованиях с использованием Excel / Лапач С.Н., Чубенко А.В., Бабич П.Н. – К.: Морион, 2001. – [2 изд. перераб. и доп.]. – 408 с.
13. Лапач С.М. Забезпечення необхідних властивостей вибірки для побудови регресійної моделі / С.М. Лапач // Физические и компьютерные технологии. Труды 15-й междунар. научн.-техн. конф., (Харьков, 2–3 декабря 2009 г.). – Харьков: ХНПК «ФЭД», 2009. – С. 179 – 182.

*Стаття надійшла до редакції 14.06.2013*