ФЕДЕРАЛЬНОЕ АГЕНТСТВО ВОЗДУШНОГО ТРАНСПОРТА

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ГРАЖДАНСКОЙ АВИАЦИИ» (МГТУ ГА)

Кафедра аэродинамики, конструкций и прочности
летательных аппаратов

В.В. Трофимов

МЕТОДЫ И АЛГОРИТМЫ ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ

Учебно-методическое пособие
по выполнению практических занятий

для студентов II курса
направления 20.03.01
очной формы обучения

Москва
ИД Академии Жуковского
2018

УДК 519.23(07)
ББК 517.8
Т76

Рецензент:
Кубланов М.С. – д-р техн. наук, проф. каф. АКПЛА

Трофимов В.В.
Т76 Методы и алгоритмы обработки статистических данных [Текст] : учебно-методическое пособие по выполнению практических занятий / В.В. Трофимов. – М.: ИД Академии Жуковского, 2018. – 28 с.

Данное учебно-методическое пособие издается в соответствии с рабочей программой учебной дисциплины «Методы и алгоритмы обработки статистических данных» по учебному плану для студентов II курса направления 20.03.01 очной формы обучения.

Рассмотрено и одобрено на заседании кафедры 06.03.2018 г. и методического совета 14.03.2018 г.

УДК 519.23(07)
ББК 517.8

В авторской редакции

Подписано в печать 31.05.2018 г.
Формат 60x84/16 Печ. л. 1,75 Усл. печ. л. 1,63
Заказ № 300/0514-УМП01 Тираж 40 экз.

Московский государственный технический университет ГА
125993, Москва, Кронштадтский бульвар, д. 20

Издательский дом Академии имени Н. Е. Жуковского
125167, Москва, 8-го Марта 4-я ул., д. 6А
Тел.: (495) 973-45-68
E-mail: zakaz@itsbook.ru

© Московский государственный технический университет гражданской авиации, 2018

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Настоящее пособие содержит задачи и теоретические положения, необходимые для выполнения практических занятий, а также перечень тем семинаров по дисциплине «Методы и алгоритмы обработки статистических данных».

Пособие предназначено для студентов направления образования 20.03.01 – техносферная безопасность.

1. ОЦЕНКА АДЕКВАТНОСТИ РЕЗУЛЬТАТОВ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА ПОВЕДЕНИЮ РЕАЛЬНОГО ОБЪЕКТА

Цель практического занятия: оценка адекватности результатов вычислительного эксперимента поведению реального объекта с помощью оценки непротиворечивости, систематической погрешности и точности.

1.1. Теоретические основы

В теории математического моделирования под адекватностью результатов, полученных с помощью математической модели, понимают их соответствие поведению оригинала. Для выявления этого соответствия необходимо сравнивать отдельные параметры объекта, полученные в расчетах и зарегистрированные при наблюдении за оригиналом в одних и тех же условиях. Очевидно, что сравнивать следует лишь соответствующие друг другу параметры между собой и только в той области функционирования объекта, в которой предполагается его исследовать. Таким образом, необходимо исследовать величину рассогласования результатов контрольного вычислительного эксперимента с результатами натурного наблюдения в тех же условиях: $\Delta v = v_{\text{модели}} - V_{\text{оригинала}}$.

Если моделируется процесс или множество состояний системы, то величина рассогласования принимает множество значений. Поэтому возникает необходимость применения статистических методов.

Для получения этого множества значений рассогласования необходимо иметь:

1) исчерпывающую информацию о поведении оригинала в конкретном случае;

2) исчерпывающие данные результатов контрольного вычислительного эксперимента, воспроизводящего тот же случай поведения объекта.

Для оценки адекватности с точки зрения целей исследования необходимо иметь:

• критерии оценки адекватности.

Цели исследования бывают самыми разнообразными, поэтому возможен выбор различных критериев. Для технических систем и процессов наиболее важными факторами при оценке **АДЕКВАТНОСТИ** необходимо считать **НЕПРОТИВОРЕЧИВОСТЬ** и **ТОЧНОСТЬ**.

Для большинства случаев исследований этих двух составляющих адекватности достаточно, поскольку в технике используются в основном подобные детерминированные математические модели, обладающие общим с оригиналом математическим описанием.

НЕПРОТИВОРЕЧИВОСТЬ подразумевает идентичный характер изменения соответствующих параметров, т.е. идентичный вид основных свойств функциональных зависимостей на отдельных участках траектории, как-то: возрастание, убывание, экстремумы, выпуклость и т.п.

ТОЧНОСТЬ означает, что обобщенная характеристика рассогласования соответствующего параметра модели и оригинала должна быть не больше, чем заранее заданное значение приемлемой погрешности.

В математической статистике известно несколько объектов, которые могут характеризовать непротиворечивость и точность.

НЕПРОТИВОРЕЧИВОСТЬ со статистической точки зрения может означать незначимость каждого отдельного значения рассогласования по сравнению с общим ходом отображаемой зависимости, иными словами, неподверженность рассогласования каким-либо закономерностям, не принципиальность – случайность. Последний термин и служит идеологической основой для построения критерия оценки непротиворечивости. Как известно, нормальный закон распределения характерен для случайной ошибки измерений. Поэтому достаточно проверить статистическую гипотезу о подчиненности рассогласования данных эксперимента и реального поведения объекта НОРМАЛЬНОМУ ЗАКОНУ распределения с нулевым математическим ожиданием $a = 0$ (как у простой ошибки измерений без систематической погрешности). Для подтверждения или отвержения этой гипотезы используется **критерий Пирсона** χ^2 ; при этом сравниваются две величины:

$$\chi^2_{\text{наблюд.}} = \sum_{i=1}^{r} \frac{\left(N_i - Np_i\right)^2}{Np_i} \quad \text{и} \quad \chi^2_{\text{крит.}}\left(\alpha, N\right),$$

где p_i – вероятность попадания в i-й интервал нормально распределенной случайной величины с математическим ожиданием $a = 0$ и среднеквадратическим отклонением $\sigma = s$;

$\chi^2_{\text{крит.}}\left(\alpha, n\right)$ определяется по таблице распределения χ^2 при уровне значимости α (вероятности совершить ошибку первого рода: отвергнуть верную гипотезу) с $n = r–2$ степенями свободы.

Если $\chi^2_{\text{наблюд.}} < \chi^2_{\text{крит.}}\left(\alpha, n\right)$, то различие статистического и гипотетического (нормального) законов распределения НЕЗНАЧИМО. Т.е. при заданном уровне значимости α гипотезу о поведении рассогласования между экспериментом и "истиной", как случайной ошибки измерений, можно принять и можно считать результаты вычислительного эксперимента не противоречащими реальности. В случае противоположного неравенства: $\chi^2_{\text{наблюд.}} > \chi^2_{\text{крит.}}\left(\alpha, n\right)$ расхождение ЗНАЧИМО (не может считаться случайным) и гипотезу следует отвергнуть, т.е. результаты вычислительного эксперимента противоречат реальному поведению объекта.

Как известно даже на бытовом уровне, для повышения **ТОЧНОСТИ** измерений проводят не одно измерение, а несколько. Это делается не из-за того, что какое-то из них может оказаться ошибочным, а из-за замечательного свойства дисперсии средней арифметической величины измерений: уменьшаться с ростом числа повторений опытов:

$$D_N = \frac{D}{N}, \qquad \text{или} \qquad \sigma_N = \frac{\sigma}{\sqrt{N}},$$

где D и σ – дисперсия и среднее квадратическое отклонение в одном опыте (измерении), D_N и σ_N – дисперсия и среднее квадратическое отклонение результата осреднения замеров по N опытам. Поэтому с помощью большего числа опытов достигают меньшего РАССЕИВАНИЯ (среднего квадратического отклонения) данных, т.е. большей точности.

Поэтому для оценки точности математической модели по сравнению с данными наблюдения за оригиналом можно использовать величину среднего квадратического отклонения, статистическую оценку s которого можно получить непосредственно из результатов сравнения. Однако такая оценка страдает неполнотой, так как не учитывает, насколько часто встречаются большие и малые, положительные и отрицательные рассогласования. Величина статистического среднего рассогласований $\overline{\Delta v}$ страдает теми же недостатками, но может быть использована в качестве оценки СИСТЕМАТИЧЕСКОЙ погрешности.

СИСТЕМАТИЧЕСКАЯ погрешность свидетельствует о закономерности рассогласования между моделью и оригиналом и не позволяет пользоваться ею. Для оценки систематической погрешности, как указано выше, можно исследовать величину статистического среднего рассогласований $\overline{\Delta v}$. Для этого тоже необходимо знать закон распределения рассогласования. Наличие существенной систематической ошибки, подчиняющейся нормальному закону распределения, проверяется с помощью **критерия Стьюдента**, по которому сравниваются две величины:

$$t = \frac{\overline{\Delta v}}{s}\sqrt{N} \quad \text{и} \quad t_{\text{крит.}}(1-\alpha, N-1).$$

Здесь $t_{\text{крит.}}(1 - \alpha, N - 1)$ определяется по таблице распределения Стьюдента при уровне значимости α (вероятности совершить ошибку первого рода: отвергнуть верную гипотезу) с $N - 1$ степенями свободы.

Если $|t| < t_{\text{крит.}}$, то систематическая ошибка НЕЗНАЧИМА, т.е. несущественна и может быть принята нулевой. В случае противоположного неравенства: $|t| > t_{\text{крит.}}$ – систематическая ошибка ЗНАЧИМА, т.е. не может считаться нулевой.

Так как ТОЧНОСТЬ следует определять единой оценкой всего множества наблюдаемых значений случайной величины рассогласования результатов вычислительного эксперимента и "истинного" значения наблюдаемой величины, то в качестве такой оценки должно выступать a – математическое ожидание рассогласования. Какое истинное значение оно имеет, нам знать не дано, но его можно оценить с помощью ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ РАССОГЛАСОВАНИЯ a оцениваемых параметров – этот подход дает возможность не только учесть все виды рассогласования, но и получить вероятностную характеристику точности. Так, например, может звучать вывод о точности в этом случае: с доверительной вероятностью γ = 0,98 гарантируется рассогласование не более 0,3 м. Критерием оценки точности тогда является соблюдение этой пары значений, приемлемой с точки зрения целей исследования.

Поэтому наиболее полную оценку точности (вернее, погрешности) вычислительного эксперимента дает доверительный интервал для математического ожидания рассогласования: интервал, внутрь которого с заданной доверительной вероятностью попадает "истинное" значение a рассогласования:

$$\overline{\Delta v} - t(\gamma, N) \cdot \frac{s}{\sqrt{N}} < m < \overline{\Delta v} + t(\gamma, N) \cdot \frac{s}{\sqrt{N}},$$

$$\overline{\Delta v} = \frac{1}{N} \cdot \sum_{i=1}^{r} N_i \Delta v_i, \quad s = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{r} N_i \left(\Delta v_i - \overline{\Delta v} \right)^2};$$

где $t(\gamma, N)$ определяется по распределению Стьюдента в случае нормального распределения рассогласования Δv при заданной доверительной вероятности γ и числе степеней свободы N. Здесь N_i – число попаданий в i-й интервал наблюдаемых рассогласований Δv; N – общее число наблюдаемых значений Δv.

Центр этого доверительного интервала определяется значением средней статистической величины рассогласования $\overline{\Delta v}$. РАЗМЕР ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ТЕМ МЕНЬШЕ, ЧЕМ МЕНЬШЕ ДОВЕРИТЕЛЬНАЯ ВЕРОЯТНОСТЬ γ, И ЧЕМ БОЛЬШЕ ЧИСЛО ОПЫТОВ N.

Естественно, при планировании вычислительного эксперимента следует стремиться к тому, чтобы такая оценка погрешности (т.е. доверительный интервал) не выходила за границы требуемой с точки зрения целей исследования погрешности $\pm\ \delta$, чего можно добиться разумным увеличением числа опытов и уменьшением доверительной вероятности. Иными словами, следует стремиться к тому, чтобы доверительный интервал целиком укладывался внутри допустимой погрешности (от $-\delta$ до $+\delta$).

Если такого условия не удается выполнить на данной серии опытов, то следует или увеличить число опытов N, или уменьшить доверительную вероятность γ. Однако последнее значительно слабее влияет на результат, тем более, что значения доверительной вероятности $\gamma < 0{,}7$ применять не желательно, так как это означает, что почти треть значений рассогласований будет выходить за границы доверительного интервала (и будет трудно уследить за поведением исследуемого параметра).

Единственным практическим недостатком такой оценки может быть лишь необходимость знать закон распределения исследуемого рассогласования.

Только в том случае, когда выполнены условия И НЕПРОТИВОРЕЧИВОСТИ И требуемой ТОЧНОСТИ, можно считать результаты вычислительного эксперимента АДЕКВАТНЫМИ реальности с доверительной вероятностью γ и уровнем значимости α в эксперименте из N опытов.

Условия точности и непротиворечивости можно проверить с помощью статистических критериев по следующему алгоритму, предварительно задав допустимую погрешность δ, уровни значимости и доверительную вероятность, исходя из целей исследования. В данной работе за δ принимается значение 0,1, что соответствует 10 % относительной погрешности. Уровни значимости и доверительную вероятность следует подбирать. В этом алгоритме строго соблюдается последовательность проверки статистических критериев, каждый следующий из которых опирается на вывод предыдущего. Действительно: для построения доверительного интервала и проверки гипотезы о нулевой систематической ошибке необходимо быть уверенным, что рассогласование подчиняется нормальному закону распределения, что может быть проверено по критерию Пирсона вначале алгоритма.

1.2. Алгоритм решения задачи

1. Выбирается один из параметров объекта, для которого есть результаты наблюдения $\{V_k\}$ в N точках, и соответствующий параметр $\{v_k\}$, полученный в контрольном вычислительном эксперименте в тех же условиях в тех же точках.

Вычисляются разности $\Delta v_k = v_k - V_k$.

Вся область значений Δv разбивается на r интервалов таким образом, чтобы в каждый из них попало не менее пяти значений Δv_k.

Производится расчет количества попадания Δv_k в каждый i-й ($1 \le i \le r$) интервал – частот N_i.

Определяются статистические оценки параметров распределения случайной величины Δv: выборочное среднее $\overline{\Delta v}$ и несмещенная оценка дисперсии s^2.

Этот пункт алгоритма выполняется компьютером без участия студента при запуске расчетной части программного обеспечения.

2. Для проверки **непротиворечивости**, т.е. подчиненности рассогласования нормальному закону распределения, применяется критерий согласия Пирсона χ^2. Уровень значимости α достаточно проверить только для двух крайних

(рекомендуемых компьютером) значений. Если $\chi^2_{набл} < \chi^2_{\text{крит}}(\alpha; r-2)$, то распределение $\Delta\nu$ незначимо отличается от нормального, т.е. результаты вычислительного эксперимента можно считать НЕ ПРОТИВОРЕЧАЩИМИ реальному поведению оригинала. Если $\chi^2_{набл} > \chi^2_{\text{крит}}(\alpha; r-2)$, то значимое отличие распределения $\Delta\nu$ от нормального свидетельствует о ПРОТИВОРЕЧИИ результатов вычислительного эксперимента реальному поведению оригинала и исследования адекватности следует прекратить.

3. Для оценки **систематической ошибки** проверяется гипотеза о равенстве нулю математического ожидания ($a = 0$) рассогласования $\Delta\nu$ с помощью критерия Стьюдента. Уровень значимости α достаточно проверить только для двух крайних (рекомендуемых компьютером) значений. Если $|t| > t(1 - \alpha_m; N - 1)$, то дальнейшие исследования адекватности нужно прекратить, так как это означает существование систематической погрешности между результатами вычислительного эксперимента и реальным поведением оригинала. Если $|t| < t(1 - \alpha_m; N - 1)$, то систематическая погрешность отсутствует и можно продолжать исследования.

Замечание. Вывод об отсутствии систематической ошибки ($a = 0$) лишь **подтверждает** возможность исследования непротиворечивости в п. 2, а противоположный вывод – опровергает, т.е. делает его ничтожным.

4. Для оценки **точности** математической модели строится доверительный интервал для математического ожидания рассогласования при заданной доверительной вероятности γ. Если наиболее удаленный от нуля конец доверительного интервала не выходит по модулю за допустимую погрешность $\delta = 0{,}1$, то математическую модель можно считать достаточно точной по отношению к оригиналу. Для выполнения этого условия следует подобрать выгодное (наибольшее) значение доверительной вероятности γ от 0,7 до 0,999.

5. Если по п. 2 можно считать математическую модель не противоречащей оригиналу, а по п. 4 и достаточно точной, то результаты расчетов адекватны реальному поведению оригинала.

Замечание. Если оценка точности математической модели оказывается во много раз лучше допустимой (иными словами, погрешность практически неразличима), то даже в отсутствии непротиворечивости математическую модель можно признать адекватной.

1.3. Программное обеспечение

Задача выполняется с помощью имитатора результатов вычислительного эксперимента на математической модели. Он позволяет сымитировать "точные значения реального объекта" и данные "неточного вычислительного эксперимента" на модели, а также:

– определить выборочные характеристики рассогласования (среднего выборочного и выборочную оценку среднеквадратического отклонения);

– проверить гипотезу о нормальном законе распределения рассогласования результатов вычислительного эксперимента с поведением реального объекта с помощью статистического критерия согласия Пирсона;

– проверить гипотезу о равенстве нулю математического ожидания рассогласования результатов вычислительного эксперимента с поведением реального объекта с помощью статистического критерия Стьюдента;

– построить доверительный интервал для математического ожидания рассогласования (погрешности).

1.4. Порядок выполнения задачи

1. Получить выборочные оценки параметров распределения рассогласования между моделью и оригиналом по данным 120 опытов (результат появляется на экране монитора сразу после входа в режим выполнения расчетов практической работы).

2. Исходя из требуемой обоснованной последовательности действий, провести весь алгоритм оценки адекватности с помощью статистических критериев.

3. Если какие-то результаты проверки критериев Вас не удовлетворяют, повторить исследования с другими значениями уровней значимости,

доверительной вероятности или допустимой погрешности. Следует учесть, что необходимость уменьшения доверительной вероятности вплоть до 0,7 свидетельствует о недопустимо низком качестве вычислительного эксперимента. Естественно, следует стремиться к как можно большему значению доверительной вероятности при соблюдении требуемой погрешности.

4. Основываясь на данных проверки статистических гипотез, сформулировать выводы о наличии систематической погрешности математической модели, о ее непротиворечивости и точности, а также в целом об адекватности математической модели реальному поведению оригинала.

Выполнение задачи завершается написанием:

- значения основных точечных характеристик распределения рассогласования;
- итоговые выбранные значения уровней значимости;
- доверительной вероятности и допустимой погрешности;
- результаты проверки статистических гипотез;
- доверительный интервал для рассогласования;
- выводы о наличии систематической погрешности математической модели, о ее непротиворечивости и точности, а также в целом об адекватности математической модели реальному поведению оригинала.

2. ГЛАДКАЯ АППРОКСИМАЦИЯ ЭКСПЕРИМЕНТАЛЬНОЙ ЗАВИСИМОСТИ

Цель практического занятия: выбор наилучшей аппроксимации экспериментальной поляры самолета Ту-134А полиномами (многочленами) 2-й, 3-й и 4-й степени.

2.1. Теоретические основы

Аппроксимация функции – это приближенная замена заданной сложной функциональной зависимости более простой функцией: алгебраическим полиномом (многочленом), тригонометрическим полиномом, или какой-либо другой функцией, которую можно построить с помощью метода наименьших квадратов. Но в любом случае подбор подходящего класса (вида) функциональной зависимости осуществляется исходя из ФИЗИЧЕСКИХ СВОЙСТВ аппроксимируемой зависимости.

Метод наименьших квадратов основан на отыскании таких значений параметров $\boldsymbol{a}$ функциональной зависимости $\varphi\ (\boldsymbol{x},\ \boldsymbol{a})$ заданного вида, которые минимизируют величину суммы квадратов отклонений вычисленных значений функции от соответствующих наблюдаемых значений:

$$\sum_{i=1}^{n}\left[y_i-\varphi\left(x_i,\boldsymbol{a}\right)\right]^2\to\min .$$

Поэтому все параметры a_j функции $\varphi\ (\boldsymbol{x},\ \boldsymbol{a})$ определяются из условия минимума функции нескольких аргументов, т.е. из системы уравнений:

$$\sum_{i=1}^{n}\left[y_i-\varphi\left(x_i,\boldsymbol{a}\right)\right]\cdot\frac{\partial\varphi}{\partial a_j}=0,\qquad j=0,1,2,\ldots,m .$$

В важном частном случае полиномиальной аппроксимации (многочленами)

$$y=\varphi\left(x,\boldsymbol{a}\right)=a_0+a_1x+a_2x^2+\ldots+a_mx^m=\sum_{j=0}^{m}a_jx^j$$

в результате несложных преобразований получается "система нормальных уравнений метода наименьших квадратов":

$$\sum_{j=0}^{m}a_j\cdot\sum_{i=1}^{n}x_i^{j+k}=\sum_{i=1}^{n}y_ix_i^k,\qquad k=0,1,2,\ldots,m,$$

которая в развернутом виде выглядит следующим образом:

$$\begin{cases}
a_0\cdot n & +a_1\cdot\sum\limits_{i=1}^{n}x_i & +a_2\cdot\sum\limits_{i=1}^{n}x_i^2 & +\ldots & +a_m\cdot\sum\limits_{i=1}^{n}x_i^m & =\sum\limits_{i=1}^{n}y_i\\
a_0\cdot\sum\limits_{i=1}^{n}x_i & +a_1\cdot\sum\limits_{i=1}^{n}x_i^2 & +a_2\cdot\sum\limits_{i=1}^{n}x_i^3 & +\ldots & +a_m\cdot\sum\limits_{i=1}^{n}x_i^{m+1} & =\sum\limits_{i=1}^{n}y_ix_i\\
. & . & . & . & . & .\\
+a_0\cdot\sum\limits_{i=1}^{n}x_i^m & +a_1\cdot\sum\limits_{i=1}^{n}x_i^{m+1} & +a_2\cdot\sum\limits_{i=1}^{n}x_i^{m+2} & +\ldots & +a_m\cdot\sum\limits_{i=1}^{n}x_i^{2m} & =\sum\limits_{i=1}^{n}y_ix_i^m
\end{cases}$$

Таким образом, задача аппроксимации, например, экспериментальной зависимости полиномом m-ой степени сводится к решению системы неоднородных линейных алгебраических уравнений с m+1 неизвестными коэффициентами полинома.

Полярой самолета называется графическое изображение взаимной зависимости аэродинамического коэффициента подъемной силы c_{ya} и аэродинамического коэффициента лобового сопротивления c_{xa} (рис. 2). Поляра характеризует летные качества самолета и имеет некоторые характерные свойства: а) поляра – выпуклая кривая, т.е. у нее нет точек перегиба;

б) у нее есть точка минимального значения c_{xa}.

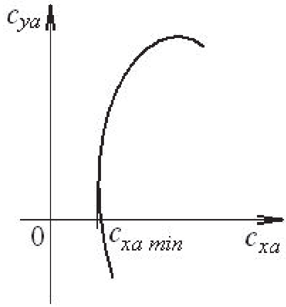


Рис. 2.

Наличие минимального значения c_{xa} означает возможность использовать соответствующую ориентацию самолета относительно воздуха для того, чтобы производить полет на выгодном режиме с минимальным сопротивлением. Обычно такой режим близок к крейсерскому, совершаемому на большой высоте с большой скоростью V. Большая скорость полета соответствует большому значению числа Маха $M = \dfrac{V}{a}$, где a – скорость звука.

Однако основным условием полета самолета на любых режимах является уравновешивание веса подъемной силой Y_a, рассчитываемой в аэродинамике по следующей формуле:

$$mg = Y_a = c_{ya} \frac{\rho V^2}{2} S,$$

где mg – вес самолета; S – площадь крыла; ρ – плотность воздуха.

Поэтому для того, чтобы летать на выгодных режимах с минимальным c_{xa}, необходимо иметь достаточное для выбранной скорости полета положительное значение c_{ya}. Так как крейсерский полет совершается на большой скорости, то "носик поляры" располагается при небольших значениях коэффициента подъемной силы c_{ya}.

Экспериментальная поляра самолета получается продувками модели в аэродинамической трубе. В каждой продувке после вывода установки на режим заданного числа Маха замеры сил производятся на всех возможных углах атаки модели. В реальном полете далеко не весь этот диапазон может быть реализован. Так, например, при больших числах Маха (при большой скорости полета) нельзя реализовать большие значения коэффициента подъемной силы c_{ya}, так как возникающая при этом большая подъемная сила (больше веса) создает большую перегрузку, угрожающую разрушением конструкции самолета. При малых числах Маха невозможен полет на малых c_{ya}, так как недостаток подъемной силы приведет к падению самолета. Поэтому для расчетов используется лишь та часть полученного в экспериментах диапазона значений коэффициента подъемной силы c_{ya}, которая соответствует числу Маха полета самолета.

Таким образом, аппроксимация экспериментальной поляры, призванная играть роль модели, **ДОЛЖНА ОБЛАДАТЬ ВСЕМИ НЕОБХОДИМЫМИ ХАРАКТЕРНЫМИ СВОЙСТВАМИ ПОЛЯРЫ ЛИШЬ В ТОЙ ЧАСТИ, ДЛЯ КОТОРОЙ ОНА СЛУЖИТ.**

2.2. ПРОГРАМНОЕ ОБЕСПЕЧЕНИЕ

Практическое занятие выполняется с помощью учебной программы, реализующей метод наименьших квадратов для отыскания коэффициентов аппроксимирующего полинома 2-й, 3-й и 4-й степеней по "экспериментальной" зависимости.

2.3. ПОРЯДОК ВЫПОЛНЕНИЯ

1. С помощью программы получить результаты аппроксимации поляры самолета Ту-134 А полиномами 2-й, 3-й и 4-й степени.

2. Проанализировать пригодность каждой из них с точки зрения физичности отображаемой зависимости, свойств аппроксимирующей зависимости и точности приближения. При этом обратить внимание на:

– диапазон эксплуатационных значений c_{ya} (оценивается по числу M и составляет не менее 6 точек из представленной таблицы – сверху, снизу или в середине);

– положение точек перегиба на аппроксимациях;

– положение точек экстремумов на аппроксимациях;

– положение экстремума на экспериментальной поляре;

– погрешности аппроксимации;

– простоту расчетов аппроксимирующего многочлена.

3. Оценить степень достоверности отдельных точек в экспериментальной зависимости (например, если для всех полиномов погрешность высока – более 5 % – и имеет один знак, то доверие к такой экспериментальной точке невелико, и на ее основе делать выводы не следует).

4. Определить наиболее приемлемую по совокупности качеств аппроксимацию и предложить аргументированное обоснование.

<u>Выполнение задачи завершается написанием</u>: итоговой таблицы аппроксимации, аналитического выражения выбранного полинома и аргументированных выводов по результатам анализа приближения.

3. ДИСПЕРСИОННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

<u>Цель практического занятия</u>: проведение дисперсионного анализа результатов дефектоскопии размера трещины с целью выявления ее роста в течение срока эксплуатации.

3.1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

После обнаружения трещины в некотором силовом элементе планера самолета проводятся плановая ее дефектоскопия после каждых 2 посадок. В данной работе таких плановых работ проводится 9 (m = 9). Каждый раз замеры трещины проводятся 5-кратно (все $N_i = n = 5$) для определения среднего значения. Результаты приводятся к начальному размеру трещины, т.е. фиксируются относительные значения роста трещины.

Для эксплуатации ЛА важно знать, растет ли со временем трещина. Если роста сверх допустимого размера нет, то эксплуатация безопасна. Если рост заметен, то при достижении трещиной определенного размера необходимо проводить ремонтные работы. Непосредственно по результатам замеров установить факт роста трещины на практике чрезвычайно трудно. Это объясняется значительной погрешностью приборов и слабой скоростью развития трещины. Поэтому необходимо проведение статистического анализа.

Дисперсия некоторого объема данных характеризует разброс, "размазанность" значений вокруг среднего. Поэтому выявление факта зависимости результатов такого однофакторного эксперимента от исследуемого фактора (числа посадок) осуществляется с помощью дисперсионного анализа. Для этого результаты замеров располагают в виде матрицы y_{ij} где каждая строка соответствует определенному циклу замеров (после 2, после 4 и т.д. посадок), а номер позиции в строке соответствует порядковому номеру единичного замера из 5 в цикле.

По элементам этой матрицы рассчитываются частные дисперсии: МЕЖГРУППОВАЯ, отражающая разброс средних (по циклам замеров) экспериментальных данных между собой из-за **влияния исследуемого фактора**:

$$s_A^2 = \frac{n}{m-1}\sum_{i=1}^{m}(\overline{y}_i - \overline{\overline{y}})^2,$$

и ОСТАТОЧНАЯ, отражающая разброс результатов единичных опытов вокруг средних по циклам экспериментальных данных каждого цикла, обусловленная **неконтролируемой погрешностью эксперимента**:

$$s_0^2 = \frac{1}{m(n-1)} \sum_{j=1}^{n} \sum_{i=1}^{m} (y_{ij} - \overline{y}_i)^2.$$

Оценку значимости исследуемого фактора (времени эксплуатации) получают с помощью критерия Фишера для сравнения двух дисперсий при заданном уровне значимости α:

– если межгрупповая дисперсия ЗНАЧИМО БОЛЬШЕ остаточной:

$$\frac{s_A^2}{s_0^2} > F_{1-\alpha}[m-1,\ m(n-1)],$$

то влияние фактора существенно и его необходимо учитывать;

– если межгрупповая дисперсия ЗНАЧИМО МЕНЬШЕ остаточной:

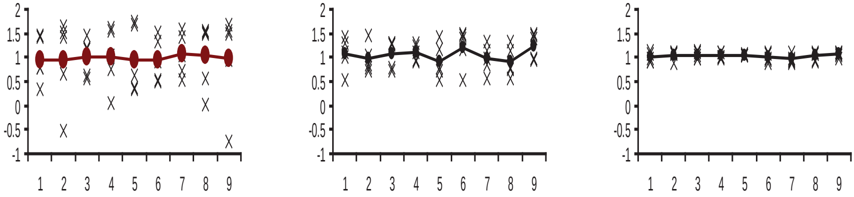
$$\frac{s_0^2}{s_A^2} > F_{1-\alpha}[m(n-1),\ m-1],$$

то влияние фактора несущественно и им можно пренебречь;

– в остальных случаях, когда нельзя говорить о ЗНАЧИМОМ превосходстве одной из дисперсий над другой, влияние исследуемого фактора сравнимо с погрешностью эксперимента или влиянием неучтенных факторов, поэтому конкретный вывод невозможен.

На рис. 3 показаны примеры различных случаев экспериментальных данных, характеризующихся различными соотношениями дисперсий. По оси абсцисс отложены лишь **номера** уровней исследуемого входного фактора, но не его физическая величина – так обычно строятся исследования в дисперсионном анализе, чтобы не привносить лишней информации.

Основываясь лишь на зрительном восприятии этого рисунка, нельзя сказать, существует ли зависимость функции, отложенной по ординате, от параметра, отложенного по абсциссе. Этого нельзя сказать даже в том случае, если расположить очередность уровней исследуемого входного фактора в порядке возрастания частных средних, соответствующих этим уровням, которые на рисунке обозначены кружочками и соединены сплошной линией.



Несмотря на это дисперсионный анализ позволяет сделать достаточно

а б в

Рис. 3.

уверенный вывод о влиянии исследуемого входного фактора на выходной.

В случае "а" бóльшая дисперсия – остаточная (внутренняя): $\frac{s_0^2}{s_A^2} = 8,07 > F_{1-\alpha}(N-k, k-1) = 5,15$, что свидетельствует о значительном влиянии **неучтенных** факторов, которые "забивают" возможную зависимость от исследуемого входного фактора. В этих условиях естественно считать эту зависимость несущественной.

В случае "б" больше уже межгрупповая дисперсия, но отношение дисперсий не достигает критического значения по критерию Фишера: $\frac{s_A^2}{s_0^2} = 1,21 < F_{1-\alpha}(k-1, N-k) = 3,04$, следовательно, сделать уверенный вывод о влиянии или невлиянии исследуемого входного фактора нельзя.

В случае "в" межгрупповая дисперсия не только больше, но и **значимо** больше остаточной: $\frac{s_A^2}{s_0^2} = 9,02 > F_{1-\alpha}(k-1, N-k) = 3,04$, поэтому необходимо сделать вывод о существенности влиянии исследуемого входного фактора.

Значимое превосходство одной из дисперсий определяется с помощью таблицы распределения Фишера, в которой приведены критические значения ОТНОШЕНИЯ БОЛЬШЕЙ ДИСПЕРСИИ К МЕНЬШЕЙ для уровня значимости 0,01 и двух чисел степеней свободы f_1 и f_2. Для межгрупповой дисперсии число степеней свободы определяется величиной (m – 1), а для остаточной – величиной $m \cdot (n$ – 1). Таким образом, о значимом превосходстве одной из дисперсий можно говорить, когда соответствующее их расчетное отношение превышает критическое, определенное по таблице распределения Фишера.

Таблица критических значений распределения Фишера

f_2 – число степеней свободы для меньшей дисперсии	f_1 – число степеней свободы для большей дисперсии	
	8	36
8	6,03	5,15
36	3,04	2,21

3.2. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Задача выполняется с помощью имитатора результатов дефектоскопии. Он позволяет:

– сымитировать данные замеров трещины;

– вычислить средние значения замеров в цикле дефектоскопии;

– вычислить межгрупповую и остаточную дисперсии;

– вычислить их отношения друг к другу.

3.3. ПОРЯДОК ВЫПОЛНЕНИЯ

1. Получить с помощью расчетной части программы результаты замеров и их обработки.

2. Выявить бо́льшую дисперсию и выписать соответствующее отношение дисперсий (большее 1).

3. Определить числа степеней свободы для каждой из полученных дисперсий.

4. По таблице, приведенной в описании работы, найти критическое значение критерия Фишера.

5. Обосновать вывод о наличии или отсутствии роста трещины.

Итоги работы: формулы для межгрупповой и остаточной дисперсий; числа степеней свободы межгрупповой и остаточной дисперсий; выражение критерия Фишера и числовые значения его частей; вывод о наличии или отсутствии роста трещины.

4. СОСТАВЛЕНИЕ КОНТРОЛЬНЫХ КАРТ СРЕДНИХ ЗНАЧЕНИЙ ПАРАМЕТРА

Требуется разработать контрольную карту средних значений контролируемого параметра $\overline{x}$, определяемых по данным N замеров, для текущего контроля качества технологического процесса. Нижнюю и верхнюю контрольные границы определить по значениям соответствующих уровней значимости (вероятностей ошибок I рода в нижнюю и верхнюю стороны): α^- и α^+. Номинальное значение контролируемого параметра составляет a, а дисперсия σ.

Исходные данные для решения составлять следующим образом:

N=4;

a = 0, последние 3 цифры № зачетки;

σ = 0,0 последняя 1 значащая цифра № зачетки;

α^-= 0,0 последние 2 цифры № зачетки (00 заменять на 10);

α^+ = 0,000 последние 2 цифры № зачетки (00 заменять на 10).

Перед выполнением задачи необходимо ознакомиться с материалами учебного пособия ([2] § 5.8).

Пример: Разработать контрольную карту средних значений контролируемого параметра $\overline{x}$, определяемых по данным N замеров, для текущего контроля качества технологического процесса.

Исходные данные для решения:

N=5;

a = 0,0450;

σ = 0,02;

α^-= 0,110;

$\alpha^{+} = 0{,}105$.

Решение. В основе разработки контрольных карт для $\overline{x}$ лежат доверительные интервалы с опорной точкой в a и определяемые заданной доверительной вероятностью γ:

$$P\{a - \delta^{-} < \overline{x} < a + \delta^{+}\} = \gamma.$$

Преобразуем это выражение, вычтя из всех трех частей неравенства величину a и поделив полученное на $\sigma / \sqrt{N}$:

$$P\left\{-\frac{\delta^{-}}{\sigma}\sqrt{N} < \frac{\overline{x} - a}{\sigma}\sqrt{N} < \frac{\delta^{+}}{\sigma}\sqrt{N}\right\} = \gamma.$$

В нашем несимметричном случае, когда заданы вероятности ошибок в каждую сторону, следует строить доверительный интервал из двух неравных частей:

$$P\left\{-\frac{\delta^{-}}{\sigma}\sqrt{N} < \frac{\overline{x} - a}{\sigma}\sqrt{N} < 0\right\} = \gamma^{-} \text{ и } P\left\{0 < \frac{\overline{x} - a}{\sigma}\sqrt{N} < \frac{\delta +}{\sigma}\sqrt{N}\right\} = \gamma^{+},$$

где $\gamma^{-} + \gamma^{+} = \gamma$. Вероятность ошибки I рода в нижнюю сторону представляет собой вероятность **не**попадания в левый доверительный полуинтервал $\alpha^{-} = 0{,}5 - \gamma^{-}$, откуда $\gamma^{-} = 0{,}5 - \alpha^{-} = 0{,}39$. Аналогично: $\gamma^{+} = 0{,}5 - \alpha^{+} = 0{,}395$.

Как известно ([2] § 5.4), выборочная функция $\dfrac{\overline{x} - a}{\sigma}\sqrt{N}$ распределена по нормированному закону, что позволяет для вычисления вышеуказанных вероятностей попадания в интервал использовать функцию Лапласа, таблица которой приведена в приложении 1. С помощью этой таблицы по известным значениям функции (вероятностям $\Phi(u^{-}) = \gamma^{-} = 0{,}39$ и $\Phi(u^{+}) = \gamma^{+} = 0{,}395$) найдем значения аргумента (крайних значений интервала):

$$u^{-} = -\frac{\delta^{-}}{\sigma}\sqrt{N} = -1{,}23 \quad \text{и} \quad \frac{\delta^{+}}{\sigma}\sqrt{N} = 1{,}25.$$

Отсюда, вычисляя δ^{-} и δ^{+}, определяем контрольные границы:

$$a-\delta^{-}=a-1,23\frac{\sigma}{\sqrt{N}}=0,045-1,23\frac{0,02}{\sqrt{5}}=0,0340,$$

$$a+\delta^{+}=a+1,25\frac{\sigma}{\sqrt{N}}=0,045+1,25\frac{0,02}{\sqrt{5}}=0,0562.$$

Таким образом, средняя величина 5 замеров контролируемого параметра должна удовлетворять условию:

$$0,0340<\overline{x}<0,0562.$$

5. ОПРЕДЕЛЕНИЕ НЕОБХОДИМОГО ОБЪЕМА ИСПЫТАНИЙ

Двумя способами (с помощью одностороннего доверительного интервала и альтернативной гипотезы вида: $H1$: $a = a_1$) определить необходимый объем летных испытаний для решения вопроса о возможности эксплуатации самолета нового типа на аэродроме с располагаемой посадочной дистанцией $L_{а/д}$. Допускаются погрешность δ и вероятности: ошибочного отвержения возможности эксплуатации до α и ошибочного принятия – до β. В качестве значения среднеквадратического отклонения величины единичной посадочной дистанции принять σ.

Исходные данные для решения выбирать следующим образом:

δ = последняя 1 значащая цифра № зачетной книжки **и 0** (десятки м);

σ = **пред**последняя 1 значащая цифра № зачетной книжки **и 0** (десятки м);

α = 0, последние 2 цифры № зачетной книжки (в **%**) (00 заменять на 10);

β = 0,0 последние 2 цифры № зачетной книжки (в **%**, в 10 раз меньше α) (00 заменять на 01).

Перед выполнением задачи необходимо ознакомиться с материалами учебного пособия ([2] § 7.3).

Пример: Требуется определить необходимый объем летных испытаний для решения вопроса о возможности эксплуатации самолета на аэродроме с располагаемой посадочной дистанцией $L_{а/д}$.

Исходные данные для решения:

δ = 50 м;

σ = 50 м;

α = 0,1;

β = 0,01.

I способ – с помощью одностороннего доверительного интервала ([2] § 7.3). Так как отклонения значений посадочной дистанции в меньшую сторону для целей нашей практической задачи несущественны, то достаточно построить **односторонний** доверительный интервал: от $-\infty$ до $L_{\text{а/д}}$. Не входящая в него часть числовой оси правее $L_{\text{а/д}}$ представляет собой область риска принять неверное решение. Т.е. вероятность попадания истинного значения посадочной дистанции в эту область не должна превышать α = 0,1 %. В предположении, что центр распределения находится левее $L_{\text{а/д}}$ на 50 м: $a_0 = L_{\text{пос}} = L_{\text{а/д}} - \delta$, среднее арифметическое значение $\bar{L}$ посадочных дистанций при N посадках должно удовлетворять условию: $P\left\{\bar{L} > a_0 + \delta = L_{\text{à/ä}}\right\}$.

После простых преобразований дополнительная к этой вероятность:

$$P\left\{\frac{\bar{L}-a_0}{\sigma}\sqrt{N} < \frac{\delta}{\sigma}\sqrt{N}\right\} = 1-\alpha = 0,999 =$$

$$= P\left\{\frac{\bar{L}-a_0}{\sigma}\sqrt{N} < 0\right\} + P\left\{0 < \frac{\bar{L}-a_0}{\sigma}\sqrt{N} < \frac{\delta}{\sigma}\sqrt{N}\right\} = 0,5 + 0,499$$

с помощью таблицы функции Лапласа (Приложение 1) по значению функции $\Phi\,(u)$ = 0,499 дает значение аргумента:

$$u = \frac{\delta}{\sigma}\sqrt{N} > 3,09 \quad \text{откуда} \quad N > \frac{u\cdot\sigma}{\delta} = \frac{3,09^2\cdot 50^2}{50^2} = 9,55.$$

Таким образом, при этом способе оценки объема эксперимента необходимо произвести 10 посадок.

II способ – с помощью альтернативной гипотезы вида: H_1: $a = a_1$.

Проанализируем требование точности. Заданная погрешность δ = 50 м может интерпретироваться, как величина уверенного (с некоторой вероятностью) **различения** двух значений $L_{\text{пос}}$. Тогда в качестве a_1 и a_0 следует рассматривать значения, различающиеся на эту величину: $a_1 = a_0 + \delta$.

Погрешность $\delta = 50$ м, рассматриваемая как расстояние между центрами двух однотипных распределений, выразится суммой расстояний по оси абсцисс до границы критической области x^* от центров a_1 и a_0, т.е. суммой величин:

$$u_{0,5-\alpha} = \frac{x^* - a_0}{\sigma\sqrt{N}} \quad \text{и} \quad u_{0,5-\alpha} = \frac{a_1 - x^*}{\sigma\sqrt{N}}.$$

Тогда, исключая x^*, можно получить выражения для необходимого объема эксперимента N: $N > (u_{0,5-\beta} + u_{0,5-\alpha})^2 \times \dfrac{\sigma^2}{(a_1 - a_0)^2}$.

Расчеты для рассматриваемого примера дают по таблице функции Лапласа (Приложение 1) $u_{0,5-\alpha} = 3{,}09$, $u_{0,5-\beta} = 3{,}72$, откуда $N > 46{,}38$. Т.е. на практике следовало бы по результатам 47 посадок вычислить среднюю величину посадочной дистанции $\overline{L}$ и принять ее в качестве a_0, выдвинув тем самым гипотезу H_0: $a = a_0$. Если вычисленная после этого граница критической области x^* окажется правее полученной величины $\overline{L}$ и левее $L_{\text{а/д}}$, то не будет оснований отвергать гипотезу H_0, т.е. можно разрешить эксплуатацию нового типа самолета на данном аэродроме.

6. ПОСТРОЕНИЕ ПЛАНА ЭКСПЕРИМЕНТА

Построить дробный план с числом опытов $N = 2^{f-k}$ f-факторного двухуровневого эксперимента и показать конструкцию возможных сочетаний основных факторов.

Исходные данные для выполнения работы выбирать по таблице.

Последняя цифра № зачетки	f	k	Последняя цифра № зачетки	f	k
0	10	6	5	5	1
1	11	7	6	6	2
2	5	2	7	7	3
3	6	3	8	8	4
4	7	4	9	9	5

Перед выполнением задачи необходимо ознакомиться с материалами учебного пособия ([2] § 7.6 и 7.7).

Пример: Требуется построить дробный план с числом опытов $N = 2^{f-k}$ f - факторного двухуровневого эксперимента.

Исходные данные: $f = 4$ – количество основных факторов, $k = 1$ – количество возможных сочетаний основных факторов.

Для построения дробного плана $2^{f–k}$ f-факторного эксперимента можно взять полный план ($f–k$)-факторного эксперимента и добавить k столбцов любых взаимодействий (парных, тройных и т.д.). Такая конструкция обеспечивает ортогональность, т.е. возможность определения всех коэффициентов линейной регрессии.

Полный план можно составлять, руководствуясь следующим правилом, обеспечивающим полный перебор всевозможных комбинаций двух уровней основных факторов (кроме фиктивного x_0, который присутствует во всех опытах). В первом столбце знаки меняют через один. Во втором знаки встречаются парами, т.е. чередуются через 2. В третьем – четверками, чередуясь через 4. Далее, если необходимо – через следующие степени 2. Построенный по этому правилу полный план всегда обладает свойствами симметричности и ортогональности, что проверяется непосредственно.

В таблице 1 показан составленный по этому правилу полный план 2^3 трехфакторного двухуровневого эксперимента.

Для построения дробного плана $2^{4–1}$ четырехфакторного эксперимента достаточно добавить к этому плану один столбец взаимодействия, например, $x_4 = x_1x_2$, составив его из произведений соответствующих уровней факторов x_1 и x_2. Итоговый дробный план $2^{4–1}$ четырехфакторного двухуровневого эксперимента приведен в таблице 2.

Таблица 2.

№ опыта	Факторы			
	x_1	x_2	x_3	x_4
1	+1	+1	+1	+1
2	+1	–1	+1	+1
3	+1	+1	–1	+1
4	+1	–1	–1	+1
5	+1	+1	+1	–1
6	+1	–1	+1	–1
7	+1	+1	–1	–1
8	+1	–1	–1	–1
		план		

Таблица 3.

№ опыта	x_0	x_1	x_2	x_3	x_4
1	+1	+1	+1	+1	+1
2	+1	–1	+1	+1	–1
3	+1	+1	–1	+1	–1
4	+1	–1	–1	+1	+1
5	+1	+1	+1	–1	+1
6	+1	–1	+1	–1	–1
7	+1	+1	–1	–1	–1
8	+1	–1	–1	–1	+1

7. ОЗНАКОМЛЕНИЕ С КОМПЬЮТЕРНОЙ ПРОГРАММОЙ GARLINA ДЛЯ ПРИЕМА ЗАЧЕТА

СПИСОК ЛИТЕРАТУРЫ

1. Математическое моделирование. Методология и методы разработки математических моделей механических систем и процессов: Учебное пособие. Часть II. Третье издание. – М.: МГТУ ГА, 2004. – 125 с.

2. Трофимов В. В. Методы и алгоритмы обработки статистических данных: Пособие к изучению дисциплины, выполнению лабораторных работ, практических занятий и домашних заданий для студентов II курса направления образования 280700 дневного обучения. – М.: МГТУ ГА, 2012. – 32 с.