

ОЦЕНКИ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ДАННЫХ ЛЕТНЫХ ИСПЫТАНИЙ

Abstract: There are given histogram methods for computing main statistical characteristics of flight-testing data in quasi-real scale of time and their software realization. It is shown the availability of their using for finding quantiles of the empirical distributions, table constructing and joint signs, finding the centre of weight of figures with compound configuration and finding adaptive threshold when recognizing images.

Key words: statistical characteristics, histogram-table methods evaluations, meeting rate, express-processing.

Анотація: Наведені гістограмні методи для обчислення основних статистичних характеристик даних льотних випробувань у квазіреальному часі і їх програмна реалізація. Показана можливість їх використання для знаходження квантилей емпіричних розподілів, представлена побудова таблиць спряженості ознак, знаходження центру ваги фігур зі складною конфігурацією і знаходження адаптивного порога при розпізнанні образів.

Ключові слова: статистичні характеристики, гістограмно-табличні методи, оцінки, частота зустрічності, експрес-обробка.

Аннотация: Приведены гистограммные методы для вычисления основных статистических характеристик данных летных испытаний в квазиреальном масштабе времени и их программная реализация. Показана возможность их использования для нахождения квантилей эмпирических распределений, дано построение таблиц сопряженных признаков, нахождения центра тяжести фигур со сложной конфигурацией и нахождение адаптивного порога при распознавании образов.

Ключевые слова: статистические характеристики, гистограммно-табличные методы, оценки, частота встречаемости, экспресс-обработка.

1. Введение

Испытание изделий новой техники представляет собой один из важнейших этапов технологии их создания. Не является исключением и создание современных самолетов. К числу главных факторов, которые учитываются при летных испытаниях современных самолетов относятся следующие [1]: особенности летной эксплуатации самолетов; особенности летательного аппарата; особенности внешней среды, в которой происходят испытания самолета, и другие факторы, влияющие на ход испытаний.

Основная тенденция при проведении таких испытаний – все увеличивающийся объем экспериментальных данных, которые необходимо собрать, обработать и проанализировать в возможно более короткие сроки.

Одна из причин такого информационного взрыва связана с необходимостью более глубокого проникновения в сущность изучаемых процессов и явлений создаваемого образца летательной аппаратуры, а также всесторонней и комплексной оценки их качественных и количественных характеристик. Другая причина количественного роста информации связана с усложнением самой техники. Существуют и иные причины данного явления.

В результате число датчиков, устанавливаемых на испытываемых образцах, в ряде случаев достигает десятков тысяч. Это приводит к необходимости хранить в памяти компьютера громадное число тарифовочных таблиц либо коэффициентов тарифовочных многочленов. Отметим, что каждому датчику соответствует своя тарифовочная таблица. Помимо этого, существует проблема устранения линейного или нелинейного тренда выбросов и т.д. Это только начальный этап обработки экспериментальных данных.

Сама обработка экспериментальных данных в большей части основывается на использовании методов теории вероятностей и математической статистики. Хотя используют и другие методы. Например, построение математических моделей, построение зависимостей одного параметра от другого.

Наряду с необходимостью повышения точности обработки данных, важным условием является устойчивость (робастность) алгоритмов их обработки, что ведет к повышению уровня достоверности полученных результатов.

Отметим, что сбор, накопление, обработка и анализ измерительной информации являются одним из узких мест в процессе испытаний. При этом необходимо иметь возможность получения результатов испытаний непосредственно на борту испытываемого самолета в квазиреальном времени (возможно, не столь точную и полную), производить паспортизацию магнитных лент и других носителей информации, полученной с датчиков, а также достигнутых результатов на основе использования событийного анализа.

Но и в стационарных условиях существует экспресс-первичная и вторичная обработка экспериментальных данных. Эти типы обработки отличаются как временем обработки, так и глубиной проникновения в исследуемые процессы. Естественно, эти виды обработки отличаются как функциональным, так и алгоритмическим наполнением. Приведенные в настоящей статье алгоритмы, в основном, могут быть эффективно использованы для обработки экспериментальных данных на борту испытываемого самолета, при экспресс-обработке, а частично и при первой обработке.

Помимо оценки простейших статистических характеристик, описанных ниже, с помощью гистограммного метода можно решать и другие задачи, например, определение центра тяжести самолета. Теория и практика гистограммных методов приведена в работах [2 – 6].

Другие подходы для оценки статистических характеристик с использованием ЭВМ при обеспечении квазиреального времени приведены в работах [7 – 8], а в работе [9] даны алгоритмы, которые можно использовать для вторичной обработки в стационарных условиях.

2. Вычисление статистических характеристик на основе гистограммно-табличных методов

Ниже приведены гистограммные методы оценки основных статистических характеристик и их использования для вычисления более сложных характеристик при обработке результатов летних испытаний самолетов. Помимо этого приведены результаты вычислений на основе разработанного авторами пакета программ.

2.1. Проверка данных

Строится частотная таблица соотношений значений (имён) переменных и частот их встречаемости в выборке, т.е. значений типа частота.

На основе этой таблицы возможно:

- 1) обнаружить грубые ошибки в выборке за счёт несоответствия отдельных переменных остальным;

2) локализовать выбросы за счет грубого несоответствия частот отдельных переменных (их очень малое значение) от остальных;

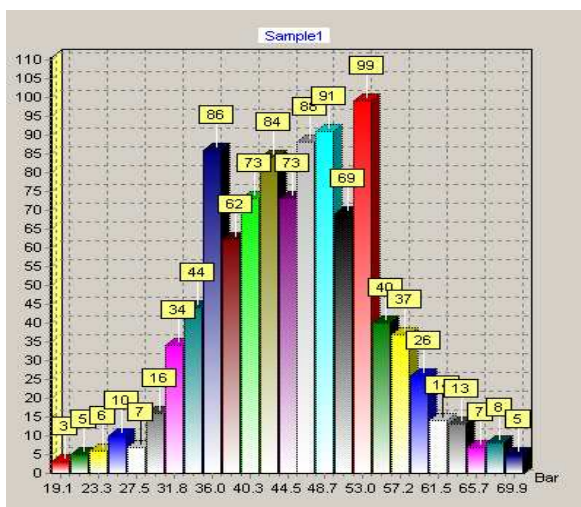


Рис. 1. Гистограмма частот наблюдений

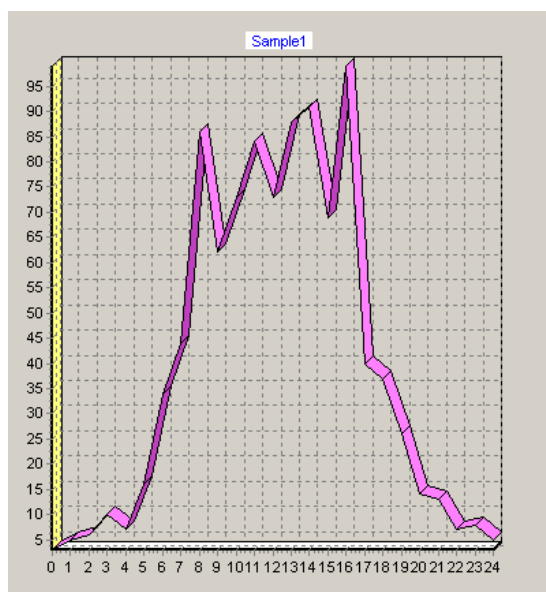


Рис. 2. Полигон частот наблюдений

величины в выборке.

Соединяя средние точки верхних прямоугольников гистограмм прямыми, можно получить *полигон частот наблюдений* (рис. 2). С помощью гистограммы можно получить *оценку выборочной моды*. Этой оценкой может служить середина интервала группировки с наибольшей частотой. Так, например, если интервал $[c_i, c_{i+1}]$ имеет наибольшую частоту f_i , то выборочная мода

3) получить эмпирическое распределение случайной величины X .

Аналогично этому можно поступать с непрерывными переменными за счет введения интервалов группировки.

2.2. Статистические выводы о параметрах исследуемого распределения с помощью таблиц частот встречаемости

Строим таблицу либо гистограмму частот встречаемости случайной величины в выборке. По таблице можно построить гистограмму двух других типов:

1) гистограмму относительных частот (нормированная гистограмма), где каждая частота f_i заменяется на относительную частоту $\beta_i = \frac{f_i}{n}$,

$$i = 1, 2, \dots, k, \text{ так что } \sum_{i=1}^k \beta_i = 1;$$

2) гистограмму частот в процентах (процентная гистограмма), где каждая β_i умножается на 100.

Возможно также использование двух шкал, что позволяет сравнивать гистограммы, построенные на одних и тех же интервалах группировки, но для различных выборок из одной и той же генеральной совокупности. На рис. 1 приведен пример гистограммы частот наблюдений (встречаемости) случайной

приблизительно равна $(c_i + c_{i+1})/2$. Оценка процентилей истинного распределения на основе ненормированной эмпирической функции распределения (ФР) (кумулятивная ФР) определяется способом, изложенным ниже.

Пусть $F_1 = 0, F_2 = f_1, F_{k+1} = \sum_{i=1}^k f_i$. Значение F_j представителем со значениями $X \leq c_i$,

$c_i, i = 1, 2, \dots, k$ – называется накопленной частотой. Ненормированной эмпирической ФР называется ломаная, соединяющая точки $(c_1, F_1), (c_2, F_2), (c_3, F_3), \dots, (c_{k+1}, F_{k+1})$. Ломаная, проходящая через точки $(c_1, F_1/n), (c_2, F_2/n), (c_3, F_3/n), \dots, (c_{k+1}, F_{k+1}/n)$, называется нормированной эмпирической ФР, а в случае, когда каждое F_i/n выражено в процентах – *процентной нормированной эмпирической ФР*. На основе процентной нормированной эмпирической ФР можно получить выборочные проценти́ли и выборочные проценти́льные ранги. Эти величины определяются следующим образом: q -я проценти́ль – это число x_q , менее которого принимают значения q % выборки. Обратная по смыслу величина – проценти́льный ранг числа x – равна проценту q части выборки, значения в которой меньше, чем значения x , и равна проценту q части выборки, значения в которой меньше величины x .

Особый интерес имеет 50-я проценти́ль, называемая медианой m , левее которой лежит половина наблюдений. Используются также проценти́ли 25 (1-я квартиль) и 75 (3-я квартиль) и проценти́ли 10, 20, ..., 90, называется соответственно 1-я, 2-я, ..., 9-я децили. Допускаются не только целые проценти́ли.

3. Вычисления основных статистических характеристик на основе частотных таблиц либо гистограмм

На основе частотных таблиц либо гистограмм можно вычислить:

- 1) максимальное и минимальное значения выборки x_{\min}, x_{\max} ;
- 2) размах выборки $x_{\min} - x_{\max}$;
- 3) выборочное среднее $\bar{x} \approx \frac{1}{n} \sum_{i=1}^k f_i \cdot c_i$ на основе частотной таблицы;
- 4) выборочная дисперсия

$$S^2 \approx \frac{1}{n-1} \left(\sum_{i=1}^k f_i \cdot c_i^2 - n \bar{x}^2 \right) \text{ либо } S^2 \approx \frac{1}{n-1} \left(\sum_{i=1}^k f_i \cdot x_i^2 - n \bar{x}^2 \right);$$

- 5) выборочное стандартное отклонение $\sigma = \sqrt{S^2}$;

6) выборочная медиана наблюдения m соответствий $(n+1)/2$ -му значению частоты при суммировании частот (по таблице частот), если n нечетно, или середину из $n/2$ или же $(n/2)+1$, если число наблюдений четно;

7) коэффициент вариации, равный S/\bar{x} , который служит для измерения стандартных отклонений в долях среднего значения. На рис. 3 приведены данные оценки основных статистических характеристик выборки, рассчитанные на компьютере;

```

Frequency :
3 5 6 10 7 16 34 44 86 62 73 84 73 88 91 69 99 40 37
26 14 13 7 8 5
Maximum : 70
Minimum : 18
Frequency Polygon :
19.1 21.2 23.3 25.4 27.5 29.7 31.8 33.9 36.0 38.1 40.3
42.4 44.5 46.6 48.7 50.9 53.0 55.1 57.2 59.3 61.5 63.6
65.7 67.8 69.9
Mathematical expectation : 45.202
Dispersion : 84.595
Standard deviation : 9.198
The third central moment : 103818.400
The fourth central moment : 5231653.000
Asymmetry parameter : 133.432
Coefficient of excess : 731.057
    
```

8) значение j-го начального момента:

$$m_j \approx \frac{1}{n} \sum_{i=1}^k f_i \cdot (\bar{c}_i)^j \text{ либо } m_j \approx \frac{1}{n} \sum_{i=1}^k f_i \cdot (\bar{x}_i)^j ;$$

9) значение j-го центрального момента:

$$\mu_j \approx \frac{1}{n} \sum_{i=1}^k f_i \cdot (\bar{c}_i - \bar{x})^j \text{ либо}$$

$$\mu_j \approx \frac{1}{n} \sum_{i=1}^k f_i \cdot (\bar{x}_i - \bar{x})^j ;$$

10) значение коэффициентов асимметрии:

$$\beta_1 = \frac{\mu_3}{\sigma^3},$$

если $\beta_1 = 0$ – то плотность распределения симметрична;

11) значение коэффициентов эксцесса:

$$\beta_2 = \frac{\mu_4}{\sigma^4} \text{ для нормального распределения } \beta_1 = 0 \text{ и } \beta_2 = 3;$$

12) получить другие статистики как выборочные средние от преобразованных наблюдаемых значений. Эти преобразования для выборочных средних следующие:

- среднее гармоничное – $1/\bar{x}$;
- логарифм среднего геометрического – $\log \bar{x}$;
- i - й начальный момент – \bar{x}^i ;
- i - й центральный момент – $(\bar{x} - \bar{x})^i$;
- среднее абсолютное отклонение – $|\bar{x} - \bar{x}|$;

13) построение пробит-графика, где по горизонтальной оси откладывается аргумент эмпирической функции распределения F , а по вертикальной – соответствующее значение аргумента стандартной функции распределения, то $y = \Phi^{-1}(F(x))$.

Для случаев эмпирической ФР, полученной на основе частотной таблицы, пробит-график получается нанесением аргумента эмпирической ФР, который наносится на горизонтальную ось, а нормальная ФР – на вертикальную. Значение аргумента нормального распределения получается путём вычисления $\Phi^{-1}(F_i / n) = z_i, i = 1, 2, \dots, k + 1$;

14) построение таблиц сопряженности признаков. Двухфакторная таблица сопряженности признаков для факторов (критериев) А, имеющих $r \geq 2$ классов(уровней), а фактор В имеет $c \geq 2$ классов, классифицирует выборку. Столбцы таблицы составляют: r классов фактора А, а строки – c классов фактора. В результате получим $r - c$ – таблицу сопряженности.

Таблица. Сопряженность признаков для факторов А и В

А	В					\sum по строкам
	1	2	3	...	С	
1	f_{11}	f_{12}	f_{13}	...	f_{1c}	f_1
2	f_{21}	f_{22}	f_{23}	...	f_{2c}	f_2
3	f_{31}	f_{32}	f_{33}	...	f_{3c}	f_3
...						...
г	f_{r1}	f_{r2}	f_{r3}	...	f_{rc}	f_r
\sum по столбцам	f_1	f_2	f_3	...	f_c	n

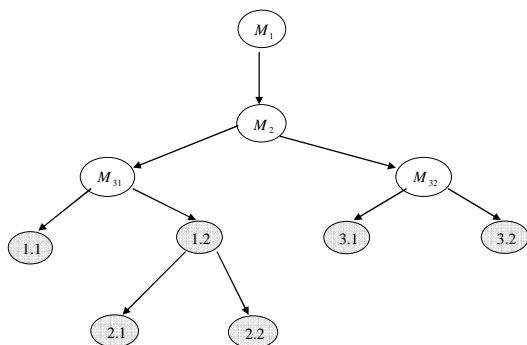


Рис. 4. Связь программных модулей в системе программ статистической обработки информации

Помимо этого, на основе частотных таблиц и гистограмм можно находить квантили эмпирических распределений, осуществлять быструю сортировку данных, находить центр тяжести фигур со сложной конфигурацией, находить адаптивные пороги при распознавании образов и т.д. Заметим, что итоги по строкам и столбцам равны объему выборки, т.е. величине n . На основе таблицы можно находить грубые ошибки в данных, проверять гипотезы о связи между факторами А и В, в том числе гипотезы об однородности, независимости, меры связности

Гудлена-Крускала и т.д.

4. Программная реализация системы программ статистической обработки информации на ПК

На рис. 4 изображены связи программных модулей в системе обработки информации на основе гистограммно-табличных методов, обеспечивающих работу в квазиреальном времени.

Программные модули здесь предназначены для выполнения следующих функций:

M_1 – генерация случайных чисел;

M_2 – построение частотной таблицы;

M_{31} – вычисление статистических характеристик;

M_{32} – построение гистограмм;

1.1 – определение начальных моментов, в том числе и математического ожидания;

1.2 – определение центральных моментов, в том числе и дисперсии;

2.1 – определение коэффициента асимметрии;

2.2 – определение коэффициента эксцесса;

3.1 – построение полигона частот;

3.2 – определение моды, процентилей, в том числе и медианы.

5. Дальнейшие исследования в этом направлении

1. Скользящее перемещение выборки по генеральной совокупности (пересекающиеся выборки).

2. Непрерывное оценивание статистических характеристик за счет добавления нового значения и вычитания первого значения.

3. Проверка стационарности на основе скользящего перемещения выборки (1-й вариант дальнейших исследований) и т.п.

4. Оценка точности либо достоверности полученных статистических характеристик в зависимости от величины интервала $[x_i, x_{i+1}]$ в гистограммном варианте и вычисления на её основе основных статистических характеристик (сравнение с "точными" методами, основанными на таблице частот встречаемости случайной величины).

5. Получение и использование статистических норм оценивания, основанных на моментах.

6. Выводы

В статье отмечается, что для экспресс- и первичной обработки данных летных испытаний целесообразно использовать гистограммные методы оценки статистических характеристик. Приведены алгоритмы для реализации этих методов, которые демонстрируются соответствующими вычислениями на компьютере. Приведены графы взаимосвязи программных модулей пакета программ, реализующих эти методы и сформированы дальнейшие направления исследований в этом направлении.

СПИСОК ЛИТЕРАТУРЫ

1. Пашковский И.М., Леонов В.А., Поплавский Б.К. Летные испытания самолетов и обработка результатов испытаний. – М.: Машиностроение, 1985. – 426 с.
2. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ: Пер. с англ. – Мир, 1982. – 488 с.
3. Гальчук В.Я., Соловьев А.П. Техника научного эксперимента. – Ленинград: Судостроение, 1982. – 256 с.
4. Теслер Г.С. Знаходження середнього арифметичного, асиметрії та ексцесу вибірки у реальному часі на ЕОМ // Обробка дослідних даних: Збірник. – Львів: ФМІ АН УРСР, 1980. – С. 33 – 38.
5. Теслер Г.С. Оцінка статистичних характеристик на основі дискретних або інтервальних варіаційних рядів // Обробка дослідних даних: Збірник. – Львів: ФМІ АН УРСР, 1980. – С. 39 – 41.
6. Теслер Г.С. Обчислення автокореляційної функції у реальному часі // Обробка дослідних даних: Збірник. – Львів: ФМІ АН УРСР, 1980. – С. 42 – 44.
7. Петренко П.А., Теслер Г.С. Обработка данных в вычислительных системах и сетях. – Киев: Техника, 1980. – 232 с.
8. Иванов В.В. Методы вычислений на ЭВМ: Справочное пособие. – Киев: Наукова думка, 1986. – 584 с.
9. Серенко И.В., Парасюк И.Н., Верева О.В. Пакет прикладных программ для обработки и анализа данных методами теории вероятности и математической статистики на ЕС ЭВМ (ДЕЛЬТАСТАТ): Препр. / АН Украины. Ин-т кибернетики им. В.М. Глушкова. – К.: 1985. – 50 с.