

ИНФОРМАЦИОННАЯ ОЦЕНКА СОВПАДЕНИЯ СТРУКТУР СЛОЖНЫХ СИСТЕМ

к.т.н. В.Ю. Дубницкий, к.ф.-м.н. С.В. Гадецкая
(представил д.т.н., проф. И.В. Кононенко)

На примере анализа структур самолетного парка ВВС РККА и Люфтваффе по состоянию на 1.06.1940 и 22.06.1941 проверены гипотезы об их совпадении и оценена информационная мера расхождения (дивергенция) между ними.

Постановка задачи. Рассмотрим выборки A_1 и A_2 объемом N_i ($i = 1, 2$) соответственно. Каждая из этих выборок разделена на C непересекающихся классов так, что $X_i = (X_{1i}, X_{2i}, \dots, X_{ci})$ и

$$\sum_{j=1}^c X_{ji} = N_i; \quad (i = \overline{1, 2}). \quad (1)$$

Нормируя каждую из выборок по N_i получим структуру выборки A_i в виде C -мерного вектора

$$P_i = (P_{1i}, P_{2i}, \dots, P_{ci}), \quad \sum_{j=1}^c P_{ji} = 1; \quad i = (1, 2). \quad (2)$$

По результатам наблюдений над векторами P_i требуется проверить нулевую гипотезу

$$H_0 : P_1 = P_2; \quad P_j = \frac{X_{j1} + X_{j2}}{N_1 + N_2}, \quad j = 1, 2, \dots, C \quad (3)$$

против сложной альтернативной $H_1 : P_1 \neq P_2$.

Если справедлива гипотеза H_1 , то существует хотя бы одно значение j ($j = \overline{1, C}$), такое, что $P_{j1} \neq P_{j2}$.

В случае справедливости гипотезы H_0 можно говорить о совпадении условий, вызвавших существование выборок A_1 и A_2 . В случае справедливости гипотезы H_1 можно говорить о существенной разнице в этих условиях.

Анализ литературы. Решение поставленной задачи методом многомерного статистического анализа в предположении, что векторы X_i ($i = 1, 2$) имеют нормальное распределение, рассмотрено в работе [1]. В том случае, когда указанное предположение не выполнено или не поддается проверке, в работе [2] предложен способ решения, основанный на анализе информационных свойств соответствующих статистик.

Цель работы. Разработка методики проверки указанных выше статистических гипотез методами теории информации. Объектом исследования выбраны структуры самолетного парка ВВС РККА и ВВС Германии (Люфтваффе) по состоянию на 1.06.1940 и 22.06.1941.

Описание объекта исследования. Количественные данные по структуре самолетного парка Люфтваффе приведены в работе [3] и с достаточной для целей данной работы точностью совпадают с данными других авторов. В этих работах выделяют бомбардировщики, истребители, разведчики и самолеты прочего назначения.

Классификация самолетного парка СССР более сложная. В работе [4] приведена четырехкомпонентная структура, аналогичная германской. В других работах принята пятикомпонентная структура, в которой штурмовая авиация выделена в самостоятельный род ВВС.

В официальных советских учетных документах [5, цит. по 3] штурмовики также выделяли в отдельную строку. В 1940 г. в составе ВВС РККА было 1285 штурмовиков, в т.ч.: И-15 бис – 1105 ед, ДИ-6 – 180 ед. Однако в варианте штурмовика самолетов ДИ-6 было всего 60 машин [7]. В 1941 г. (на 22.06.1941) в составе ВВС было 57 самолетов – штурмовиков [3, 6], хотя известно, что уже было выпущено 249 самолетов Ил-2 [8].

Таким образом, по всей видимости, в учетных документах численность штурмовиков указывалось по тактическому назначению самолета, а не по конструктивному исполнению модели. Этот вывод косвенно подтверждается тем, что с лета 1940 г. в состав советских смешанных авиационных дивизий было введено по одному полку штурмовой авиации [6]. Поэтому для дальнейших расчетов штурмовики были исключены из анализа, а истребители И-15 бис из состава штурмовой авиации учтены в соответствующей строке табл. 1.

Таблица 1
Структура парка боевых самолетов СССР и Германии

Назначение	СССР на 1.06.40 (С40)*		СССР на 22.06.41 (С41)*		Германия на 22.06.41 (Г41)*	
	единиц	доля (P)	единиц	доля (P)	единиц	доля (P)
Бомбардировщики	5234	0,337	6887	0,368	2642	0,462
Разведчики	3348	0,216	1934	0,104	823	0,144
Истребители	6931	0,447	9881	0,528	2249	0,394
Итого	15513	1	18702	1	5714	1

* – условное обозначение выборки.

Изложение результатов исследований. Решение поставленной задачи состоит из следующих, нижеперечисленных этапов.

1. Выбор одной из альтернативных гипотез H_0 или H_1 .

Предпочтение гипотезы H_0 на содержательном уровне означает, что

различие в структуре ВВС противостоящих сторон было обусловлено случайными причинами, а военно-теоретические взгляды сторон на боевые задачи авиации совпадали. Предпочтение гипотезы H_1 означает, что военно-теоретические взгляды сторон были существенно различными.

2. Оценка расхождения между выборками.

Эта оценка возможна только при использовании понятия дивергенции [2] и дает возможность количественно оценить различие между выборками, чего обычные методы, основанные на критерии χ^2 , не позволяют.

3. Выбор совпадающих частот P_{ji}, P_{ki} , ($j \neq k$, $j = \overline{1, c}$, $k = \overline{1, c}$, $i = 1, 2$), если принимается гипотеза H_1 .

4. Оценка коэффициентов корреляции между частотами P_j, P_k внутри данной выборки.

Для выбора одной из альтернативных гипотез H_0 или H_1 следует, согласно [2], вычислить величину

$$2\hat{I}(H_1, H_0) = 2 \sum_{i=1}^c \left(x_{i1} \cdot \log_2 \frac{(N_1 + N_2)x_{i1}}{N_2(x_{i1} + x_{i2})} + x_{i2} \cdot \log_2 \frac{(N_1 + N_2)x_{i2}}{N_2(x_{i1} + x_{i2})} \right). \quad (4)$$

Если справедлива гипотеза H_0 , то $2\hat{I}(H_1, H_0) \leq \chi_{(c-1); \alpha}^2$. В нашем случае $\chi_{(c-1); \alpha}^2 = \chi_{2; 0,05}^2 = 5,99$.

Результаты расчетов приведены в табл. 2, в которой условные обозначения совпадают с принятыми в табл. 1.

Таблица 2

Проверка гипотез о существенности различия выборок

Обозначения сравниваемых на Р выборок	C40 – C41	C40 – Г41	C41 – Г41
Величина $2\hat{I}(H_1, H_0)$	1200,24	10141,05	468,3

Таким образом, следует считать, что военно-теоретические взгляды командования РККА на использование ВВС в 1940 г. существенно отличались от взглядов, принятых на 1941 г. и от военно-теоретической доктрины Люфтваффе, принятой на 1941 г. Этот вывод требует специального военно-исторического обсуждения, выходящего за рамки данного сообщения.

Так как гипотеза H_0 ранее была отвергнута в пользу H_1 , то расстояние между структурами сравниваемых выборок оценивали, используя понятие дивергенции между выборками A_1 и A_2 [2]:

$$\hat{J}(H_1, H_2) = \frac{N_1 \cdot N_2}{N_1 + N_2} \sum_{j=1}^C \left(\frac{x_{j1}}{N_1} - \frac{x_{j2}}{N_2} \right) \log_2 \frac{N_2 x_{j1}}{N_1 x_{j2}}. \quad (5)$$

Результаты расчетов приведены в табл. 3.

Таблица 3

Оценка дивергенции между выборками

Обозначения сравниваемых пар выборок	C40 – C41	C40 – Г41	C41 – Г41
Величина дивергенции	1212	452,77	470,73
Нормированная величина дивергенции	2,57	0,962	1

Из этих данных можно сделать вывод о том, что структура ВВС РККА в 1940 г. была ближе к структуре Люфтваффе 1941 года.

Поиск совпадающих (отличающихся) частот P_{ji} проводили согласно [9]. Для каждого рода авиации j в стране I определяли его долю

$$P_{ji} = \frac{x_{ji}}{N_i}, \quad j=1, 2, 3, \quad i=1, 2. \quad (6)$$

Величину P_{ji} преобразовали по формуле

$$\varphi_{ji} = \frac{2\pi}{180} \arcsin \sqrt{P_{ji}} \quad (7)$$

и вычисляли величину

$$F_{i\varphi} = (\varphi_1 - \varphi_2)^2 \cdot \frac{N_1 \cdot N_2}{N_1 + N_2}. \quad (8)$$

Если

$$F_{j\varphi} \leq F_{v_1, v_2; \alpha}, \quad v_1 = 1, \quad v_2 = N_1 + N_2 - 2, \quad (9)$$

то принимали решения, что $P_{j1} = P_{j2}$, в противном случае принимали противоположное решение $P_{j1} \neq P_{j2}$.

Так, как объемы выборок велики, то $F_{1, \infty; 0,05} = 3,84$. Результаты вычислений приведены в табл. 4.

Из табл. 4 следует, что сравниваемые структуры различались не только в целом, но и по каждому роду авиации.

Вероятностной моделью для структур с ограничениями (1), (2) может служить модель мультиномиального распределения вероятностей. Доказательство этого приведено в работе [10]. Связь между отдельными составляющими структуры внутри данной выборки может быть измерена с помощью коэффициента корреляции [11]:

$$r_{jk} = \sqrt{\frac{P_j P_k}{(1 - P_j)(1 - P_k)}}, j \neq k, j = \overline{1, C}, k = \overline{1, C}. \quad (6)$$

Результаты расчетов приведены в табл. 5.

Таблица 4

Проверка гипотез о совпадении (различии) структуры ВВС РККА и Германии по каждому из родов авиации

Род авиации	Сравниваемые пары	Расчетные величины				
		φ_1	φ_2	η_1	η_2	F_{ji}
Бомбардировочная	C40-C41	1,2303	1,3016	5234	6887	15,11
	C41-Г41	1,3016	1,4967	6887	2642	72,68
Разведывательная	C40-C41	0,9594	0,6534	3348	1934	1141,7
	C41-Г41	0,6534	1934	0,7785	823	9,035
Истребительная	C40-C41	1,1545	1,6248	6931	9881	901,0
	C41-Г41	1,6248	1,3551	9881	2249	133,2

Таблица 5

Оценка статистической связи между отдельными родами ВВС РККА и Германии

Дата сравнения	Страна	Условные обозначения сравниваемых пар	Коэффициент корреляции
1.06.40	СССР	Б – И	0,641
		Б – Р	0,374
		Р – И	0,275
22.06.41	СССР	Б – И	0,807
		Б – Р	0,260
		Р – И	0,360
22.06.41	Германия	Б – И	0,747
		Б – Р	0,380
		Р – И	0,331

*) Б – бомбардировщики, И – истребители, Р – разведчики.

Следует отметить, что в 1941 г. в ВВС РККА резко снизилась, в сравнении с Люфтваффе, доля разведывательной авиации и ее корреляционная связь с бомбардировочной. Принимая во внимание низкий уровень связи в ВВС РККА в 1941 г., что отмечено в работе [12], можно предположить, что это обстоятельство было одной из причин плохого

взаимодействия ВВС и сухопутных войск в первые месяцы войны.

Выводы.

1. Описана методика сравнения выборок из мультиномиальных совокупностей.

2. Приведен пример применения методики для сравнения пары организационных структур специального назначения.

3. По результатам статистического анализа структур ВВС РККА и Люфтваффе можно предположить, что они формировались по различным военно-теоретическим принципам.

ЛИТЕРАТУРА

1. *Андерсон Т. Введение в многомерный статистический анализ.* – М.: Физматгиз, 1963. – 406 с.
2. *Кульбак С. Теория информации и статистика.* – М.: Физматгиз, 1967. – 407 с.
3. *Мельтюхов М.И. Упущенный шанс Сталина. Советский Союз и борьба за Европу 1939 – 1941 г.г.* – М.: Вече, 2002. – 542 с.
4. *Советская военная энциклопедия. Т.2.* – М.: Воениздат, 1976. – 203 с.
5. *РГВА (Российский государственный архив военной истории). Ф. 29. Оп. 73. Д. 21. Л. 14-18.*
6. *Дрогозов И.Г. Воздушный щит Страны Советов.* – Минск.: Харвест, 2003. – 544 с.
7. *Шавров В.Б. История конструкций самолетов, в СССР до 1938 года.* – М.: Машиностроение, 1978. – С. 462, 479 – 482.
8. *Авиация Второй мировой войны / Пер. с англ. В.В.Найденова.* – Смоленск: Русич, 2001. – 128 с.
9. *Плохинский Н.А. Биометрия.* – М.: Московский университет, 1970. – 358 с.
10. *Кендалл М., Стьюарт А. Теория распределений.* – М.: ГРФМЛ, 1966. – С. 198 – 199.
11. *Поллард Дж. Справочник по вычислительным методам статистики.* – М.: Финансы и статистика, 1982. – С. 111 – 112.
12. *Война и мы. Кн. 1. Человеческий фактор / Мухин Ю.И. Совещание.* – М. Б.М., 2000. – С. 75 – 126.

Поступила 5.08.2004

***ДУБНИЦКИЙ Валерий Юрьевич**, кандидат технических наук, доцент Харьковского банковского института Украинской академии банковского дела. В 1975 г. окончил Харьковский институт радиоэлектроники. Область научных интересов – исследование операций.*

***ГАДЕЦКАЯ Светлана Викторовна**, кандидат физико-математических наук, доцент, заведующая кафедрой Харьковского банковского института Украинской академии банковского дела. В 1986 г. окончила Харьковский государственный университет. Область научных интересов – исследование операций.*