EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 26. KÖIDE
ÜHISKONNATEADUSED. 1977, NR. 2

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 26
ОБЩЕСТВЕННЫЕ НАУКИ. 1977, № 2

https://doi.org/10.3176/hum.soc.sci.1977.2.03

А. МААМЯГИ

ЭКСПЕРТНАЯ КЛАССИФИКАЦИЯ ОБЪЕКТОВ. РАСПРЕДЕЛЕНИЕ РАССТОЯНИЯ МЕЖДУ КЛАССИФИКАЦИЯМИ НЕКОМПЕТЕНТНЫХ ЭКСПЕРТОВ

1. При решении многих экономических проблем возникает необходимость выделения из основной рассматриваемой совокупности объектов некоторого подмножества «схожих» между собой элементов. Так, при оценке работы предприятий одного профиля целесообразно разделить их на крупные и мелкие и рассматривать эти группы отдельно.

В большинстве случаев «сходство» объектов трудно оценить с помощью какого-то одного количественного показателя. В этом случае классификация объектов производится обычно экспертами без четкой формулировки понятия «сходство», интуитивно. Как правило, из-за различного понимания последнего проведенные разными экспертами разбиения между собой не совпадают. Возникают две проблемы.

Во-первых, неясно, как по нескольким, весьма противоречивым разбиениям найти «истинное». Проблема согласования подробно рассмотрена Б. Г. Миркиным [1].

Во-вторых, при очень большом расхождении мнений экспертов необходимо проверить, не является ли исходная совокупность однородной. Другими словами, не исключено, что «истинного» разбиения не существует и группы, выделенные экспертами, даны случайно.

Если рассматриваемая совокупность объектов разбита на классы, которые упорядочены, для проверки гипотезы о случайном разбиении можно пользоваться коэффициентами конкордации или коэффициентами Спирмэна и Кендэла [2]; практическое применение этих методов приведено в [3, 4].

Часты, однако, случаи, когда упорядочение классов лишено смысла (например, если классификация больных проводится по диагнозу).

При проверке гипотезы о случайном разбиении неупорядоченных классов расстояние Хемминга можно рассматривать как меру «различия» между разбиениями.

Постановка данной задачи принадлежит С. А. Айвазяну.

2. Введем необходимые понятия и обозначения и приведем четкую математическую формулировку данной задачи.

Пусть множество содержит n объектов, которые двумя экспертами разбиваются на классы (матрицы смежности разбиений

$$R=\{r_{ij}\},\quad S=\{s_{ij}\};\quad i=1,\ \ldots\ldots,\ n,\quad j=1,\ \ldots\ldots,\ n).$$

Расстояние Хемминга между двумя разбиениями ($d(R, S)$) вычисляется следующим образом:

$$d(R, S)=\sum_{i=1}^{n}\sum_{j=1}^{n}|r_{ij}-s_{ij}|.$$

Л. Б. Черный [5] доказал, что расстояние Хемминга — единственно возможное расстояние, удовлетворяющее некоторой системе аксиом, выдвинутой ранее [6] для неупорядоченных разбиений.

В случае двух классов это расстояние можно вычислить проще. Если в обоих разбиениях в первом классе находится m_1 одинаковых элементов, а во втором — m_2, то

$$d(R, S)=2(m_1+m_2)(n-m_1-m_2). \tag{1}$$

Постановка I. Эксперты независимо друг от друга разбивают исходное множество точно на k классов (k задается заранее).

Постановка II. Эксперты разбивают исходное множество не более чем на k классов.

Число возможных разбиений n элементов на k классов — $S(n, k)$, где $S(n, k)$ — числа Стирлинга II рода.

Множество всех возможных разбиений будет обозначаться $\{S_i^{(k)}\}_{i=1}^{S(n,k)}$ (постановка I) и $\{S_i^{(k)}\}_{i=1}^{\sum_{j=1}^{k} S(n,j)}$ (постановка II). Любой из элементов этого множества выбирается экспертом с одинаковой вероятностью (предполагается случайная группировка). Тогда для постановки I вероятность выпадения некоторой несовпадающей пары разбиений будет $P_1=\frac{2}{S^2(n,k)}=\frac{1}{H_1}$ и для постановки II — $P_2=\frac{2}{\left(\sum_{j=1}^{k} S(n,j)\right)^2}=\frac{1}{H_2}$.

Конечной целью исследования, очевидно, является нахождение функции распределения случайной величины $d(R, S)$ или некоторой функции от нее.

Введем

$$N_1=\sum_{i=1}^{S(n,2)} d(U, S_i^{(2)})=\frac{1}{2}\sum_{m=1}^{n-1} C_n^m 2m(n-m)=2^{n-2}n(n-1),$$

$$\begin{aligned} N_2&=\sum_{i=1}^{S(n,2)} d^2(U, S_i^{(2)})=\frac{1}{2}\sum_{m=1}^{n-1} C_n^m 2^2m^2(n-m)^2=\\ &=2^{n-3}n(n-1)(n(n-1)+2), \end{aligned} \tag{2}$$

$$N_3=\sum_{i=1}^{S(n,2)} d^3(U, S_i^{(2)})=2^{n-4}n(n-1)(n^4-2n^3+7n^2-14n+16),$$

где U — матрица, все элементы которой равны 1 (матрица смежности разбиения на один класс).

Пользуясь (1) и (2), можно вычислить первые три момента случайной величины $d(R, S)$ для $k=2$ как для постановки I, так и для постановки II, поскольку

$$\sum_{i=1}^{S(n,2)-1}\sum_{j=i}^{S(n,2)} d^v(S_i^{(2)}, S_j^{(2)})=(2^{n-2}-1)N_v, \quad v=1, 2, 3. \tag{3}$$

Перейдем от расстояния Хемминга к случайной величине

$$z^n_{RT} = \frac{1}{\sqrt{n(n-1)}}\left(2\left[\frac{n}{2}\right]\left(n-\left[\frac{n}{2}\right]\right) - d(R,T)\right) =$$
$$= \frac{1}{\sqrt{n(n-1)}}(n^2-\delta-2d(R,T)), \quad \delta = n-2\left[\frac{n}{2}\right].$$

Для малых значений n закон распределения z^n легко вычислить, учитывая (2), (3).

При $n\to\infty$ моменты случайной величины z^n стремятся к соответствующим моментам χ^2 (1) распределения, а так как моменты последнего определяют единственную функцию распределения, то z^n сходится и по распределению к случайной величине с функцией распределения χ^2 (1).

Доказательство последнего утверждения, а также таблица возможных значений z^n и соответствующих им вероятностей до $n\leqslant 29$ в ближайшее время будут опубликованы.

Из изложенного следует, что если мы имеем более двух классификаций (например, R, T, S), то пользуясь тем фактом, что сумма двух независимых случайных величин, распределенных по закону χ^2 (1), распределена по закону χ^2 (2), можно вместо величины z^n_{RT} (z^n_{ST}, z^n_{RS}) рассмотреть величину $z^n_{RT}+z^n_{TS}$ ($z^n_{RT}+z^n_{RS}$ или $z^n_{TS}+z^n_{RS}$). В таком случае мы имеем величину, которая одновременно учитывает все разбиения. Суммой всех трех величин мы, однако, пользоваться не можем, так как сумма двух слагаемых уже не будет статистически независима от третьего.

3. Разберем теперь случай $k>2$. Рассмотрим группировку n элементов на k классов (R), последний класс которой содержит m элементов, и группировку R' из $n-m$ элементов (из первоначального множества исключаются элементы последнего класса), которая на $n-m$ элементах совпадает с группировкой R. Очевидно, что

$$d(R,U) = d(R',U) + 2m(n-m);$$
$$d^r(R,U) = \sum_{v=0}^{r} C^v_r\, d^v(R',U)\,\{2m(n-m)\}^{r-v}.$$

Обозначим

$$I_r(n,k) = \sum_{i=1}^{S(n,k)} d^r(U, S^{(k)}_i) =$$
$$= \sum_{m=1}^{n-k+1} C^{m-1}_{n-1} \sum_{v=0}^{r} C^v_r \{2m(n-m)\}^{r-v} \sum_{i=1}^{S(n-m,\,k-1)} d^v(U, R^{(k-1)}_i). \qquad (4)$$

Утверждение 1.

$$I_1(n,k) = \{S(n,k) - S(n-1,k)\}\,n(n-1); \qquad (5)$$

$$I_2(n,k) = \{S(n,k) - 2S(n-1,k) + S(n-2,k)\}\,n^2(n-1)^2 +$$
$$+ 2n(n-1)\{S(n-1,k) - S(n-2,k)\}; \qquad (6)$$

$$I_3(n,k)=\{(S(n,k)-S(n-1,k))n^2(n-1)^2-$$
$$-2(n(n-1)-1)(n(n-1)-2)(S(n-1,k)-$$
$$-S(n-2,k))+(n^2(n-1)^2-6n(n-1)+8-$$
$$-8(n-2))(S(n-2,k)-S(n-3,k))\}n(n-1). \qquad (7)$$

Доказательство. Индукция по k. Из (2) получаем утверждение для $k=2$. Для $k+1$ используем (4), учитывая, что

$$S(n,k)=\sum_{j=0}^{n-1} C^j_{n-1}\, S(j,k-1).$$

Утверждение 2.

$$\sum_{v=1}^{S(n,\,i)}\sum_{u=1}^{S(n,\,j)} d(S^{(i)}_v, S^{(j)}_u)=\{S(n-1,j)\,(S(n,i)-S(n-1,i))+$$
$$+S(n-1,i)\,(S(n,j)-S(n-1,j))\}n(n-1); \qquad (8)$$

Доказательство. $d(R,S)=2d(U,RS)-(d(U,R)+d(U,S))$;

$$\sum_v\sum_u d(S^{(i)}_v, S^{(j)}_u)=2\sum_u\sum_v d(U, S^{(i)}_u S^{(j)}_v)-$$
$$-S(n,j)\sum_u d(S^{(i)}_u, U)-S(n,i)\sum_v d(S^{(j)}_v U);$$

$$\sum_u\sum_v d^r(U, S^{(i)}_u S^{(j)}_v)=\sum_{t=j}^{ij} K_t \sum_{q=1}^{S(n,\,t)} d^r(U, R^{(t)}_q), \qquad (9)$$

где K_t — число различных пар разбиений $(S^{(j)}_v, S^{(i)}_u)$ таких, что $S^{(j)}_v \cap S^{(i)}_u = R^{(t)}_q$. Очевидно, что

$$\sum_{t=j}^{ij} K_t S(n,t)=S(n,i)\,S(n,j). \qquad (10)$$

Учитывая последнее равенство и (5), получаем требуемое.

Утверждение 3.

$$\sum_{v=1}^{S(n,\,i)} d(R^{(j)}_u, R^{(j)}_u R^{(i)}_v)=$$
$$=\{S(n,i)-S(n-1,i)\}\{n(n-1)-d(R^{(j)}_u U)\}; \qquad (11)$$

$$\sum_{v=1}^{S(n,\,i)} d^2(U, R^{(j)}_u R^{(i)}_v)=S(n-2,i)\,d^2(U, R^{(j)}_u)+2\{S(n-1,i)-$$
$$-S(n-2,i)\}\{n(n-1)-1\}d(U, R^{(j)}_u)+\sum_{v=1}^{S(n,\,i)} d^2(U, R^{(i)}_v). \qquad (12)$$

$$\sum_{v=1}^{S(n,\,i)} d^2(R_u^{(j)}, R_v^{(i)}) = \{S(n,i) - 4S(n-1,i) +$$

$$+4S(n-2,i)\} d^2(U, R_u^{(j)}) - 2\{S(n,i) - 3S(n-1,i) +$$

$$+2S(n-2,i)\} n(n-1) d(U, R_u^{(j)}) + \sum_{v=1}^{S(n,\,i)} d^2(U, R_v^{(i)}). \tag{13}$$

Доказательство. Докажем (11).

$$\sum_{v=1}^{S(n,\,i)} d(U, R_v^{(i)}) = \sum_{v=1}^{S(n,\,i)} \sum_{p=1}^{n} \sum_{q=1}^{n} |1 - (r_v^{(i)})_{pq}|.$$

Согласно (5),

$$\sum_{v=1}^{S(n,\,i)} |1 - (r_v^{(i)})_{pq}| = (S(n,i) - S(n-1,i)). \tag{14}$$

Рассмотрим $\sum_{v=1}^{S(n,\,i)} d(R_u^{(j)}, R_u^{(j)} R_v^{(i)})$. Из $n(n-1)$ элементов $((r_u^{(j)})_{pq} - (r_u^{(j)} r_v^{(i)})_{pq})$ нулевое значение для любого разбиения $R_v^{(i)}$ дают $d(U, R_u^{(j)})$ элементов, на оставшихся $\{n(n-1) - d(U, R_u^{(j)})\}$ элементах, учитывая (14), очевидно

$$\sum_{v=1}^{S(n,\,i)} |(r_u^{(j)})_{pq} - (r_u^{(j)} r_v^{(i)})_{pq}| = \{S(n,i) - S(n-1,i)\}.$$

Докажем (12).

$$d^2(U, R_v^{(i)} R_u^{(j)}) = 4\{\sum_{p=1}^{n-1} \sum_{q=p+1}^{n} |1 - (r_u^{(j)} r_v^{(i)})_{pq}|\}^2 =$$

$$= 4\{\sum_{t=1}^{\frac{n(n-1)}{2}} (b_{u\,v}^{(i)(j)})_t\}^2,$$

где

$$(b_{u\,v}^{(i)(j)})_t = |1 - (r_u^{(j)} r_v^{(i)})_{pq}|, \quad t = (q-p) + \sum_{l=1}^{p-1} (n-l).$$

$$\frac{1}{4} d^2(U, R_u^{(j)} R_v^{(i)}) = \sum_{t=1}^{\frac{n(n-1)}{2}} (b_{u\,v}^{(i)(j)})_t +$$

$$+2 \sum_{t=1}^{\frac{n(n-1)}{2}-1} \sum_{s=t+1}^{\frac{n(n-1)}{2}} (b_{u\,v}^{(i)(j)})_t (b_{u\,v}^{(i)(j)})_s. \tag{15}$$

По (11) и, учитывая, что

$$d(R_u^{(j)}, R_u^{(j)}R_v^{(i)}) = d(U, R_u^{(j)}R_v^{(i)}) - d(U, R_u^{(j)}),$$

$$\sum_{v=1}^{S(n,\,i)} \sum_{t=1}^{\frac{n(n-1)}{2}} (b_{u\,v}^{(i)(j)})_t = \frac{1}{2}\{(S(n,i) -$$

$$- S(n-1,i))\,n(n-1) + S(n-1,i)\,d(U, R_u^{(j)})\}.$$

Рассмотрим

$$(b_{u\,v}^{(i)(j)})_t\,(b_{u\,v}^{(i)(j)})_s = 1.$$

Тогда при

$$(r_u^{(j)})_{p_1q_1} = 0, \quad t = (q_1 - p_1) + \sum_{v=1}^{p_1-1}(n-v),$$

$$(r_u^{(j)})_{p_2q_2} = 0, \quad s = (q_2 - p_2) + \sum_{v=1}^{p_2-1}(n-v),$$

$(r_v^{(i)})_{p_1q_1}$, $(r_v^{(i)})_{p_2q_2}$ могут иметь любые значения и, следовательно, при таких $(b_{u\,v}^{(i)(j)})_t\,(b_{u\,v}^{(i)(j)})_s$

$$\sum_{v=1}^{S(n,\,i)} (b_{u\,v}^{(i)(j)})_t\,(b_{u\,v}^{(i)(j)})_s = S(n,i).$$

Рассмотрим теперь случай $\begin{cases} (r_u^{(j)})_{p_1q_1} = 0 \\ (r_u^{(j)})_{p_2q_2} = 1 \\ (r_v^{(i)})_{p_2q_2} = 0. \end{cases}$

Как и в (14), $\displaystyle\sum_{v=1}^{S(n,\,i)} (b_{u\,v}^{(i)(j)})_s\,(b_{u\,v}^{(i)(j)})_t = \{S(n,i) - S(n-1,i)\}.$

Для случая $\begin{cases} (r_u^{(j)})_{p_1q_1} = 1 \\ (r_u^{(j)})_{p_2q_2} = 1 \\ (r_v^{(i)})_{p_1q_1} = 0 \\ (r_v^{(i)})_{p_2q_2} = 0 \end{cases}$

$$\sum_{v=1}^{S(n,\,i)} (b_{u\,v}^{(i)(j)})_t\,(b_{u\,v}^{(i)(j)})_s = \{S(n,i) - 2S(n-1,i) + S(n-2,i)\}.$$

Последнее получим, если из числа всех возможных способов разбиений исключим те, в которых объекты p_1 и q_1 или p_2 и q_2 принадлежат к одному классу.

И окончательно:

$$\frac{1}{4}\sum_{v=1}^{S(n,\,i)} d^2(U, R_u^{(j)}R_v^{(i)}) = \frac{n(n-1)}{2}\,(S(n,i) - S(n-1,i)) +$$

$$+\frac{1}{2}S(n-1,i)\,d(U,R_u^{(j)})+\frac{d(U,R_u^{(j)})}{2}\left(\frac{d(U,R_u^{(j)})}{2}-1\right)S(n,i)+$$

$$+d(R_u^{(j)},U)\left\{\frac{n(n-1)}{2}-\frac{d(R_u^{(j)},U)}{2}\right\}\{S(n,i)-S(n-1,i)\}+$$

$$+\left\{\frac{n(n-1)}{2}-\frac{d(R_u^{(j)},U)}{2}\right\}\left\{\frac{n(n-1)}{2}-\right.$$

$$\left.-\frac{d(R_u^{(j)},U)}{2}-1\right\}\{S(n,i)-2S(n-1,i)+S(n-2,i)\},$$

что и доказывает (12).

Аналогично доказывается (13). Рассмотрим тоже три случая:

а) $(r_u^{(j)})_{p_1q_1}=0$	б) $(r_u^{(j)})_{p_1q_1}=1$	в) $(r_u^{(j)})_{p_1q_1}=1$
$(r_u^{(j)})_{p_2q_2}=0$	$(r_u^{(j)})_{p_2q_2}=0$	$(r_u^{(j)})_{p_2q_2}=1$
$(r_v^{(i)})_{p_1q_1}=1$	$(r_v^{(i)})_{p_1q_1}=0$	$(r_v^{(i)})_{p_1q_1}=0$
$(r_v^{(i)})_{p_2q_2}=1$	$(r_v^{(i)})_{p_2q_2}=1$	$(r_v^{(i)})_{p_2q_2}=0$

Случай в) уже рассмотрен. В случае а)

$$\sum_{v=1}^{S(n,i)}(b_{u\,v}^{(i)(j)})_s(b_{u\,v}^{(i)(j)})_t=S(n-2,i)$$

(число таких разбиений, где объекты p_1 и q_1 попадают в один класс; p_2 и q_2 также в одном классе). В случае б) аналогичные рассуждения дают

$$\sum_{v=1}^{S(n,i)}(b_{u\,v}^{(i)(j)})_s(b_{u\,v}^{(i)(j)})_t=\{S(n-1,i)-S(n-2,i)\}.$$

Суммируя все три случая и учитывая, что

$$\sum_{v=1}^{S(n,i)}d(R_u^{(j)},R_v^{(i)})=n(n-1)\{S(n,i)-S(n-1,i)\}-$$

$$-\{S(n,i)-2S(n-1,i)\}d(U,R_u^{(j)}),$$

получаем (13). Последнее вытекает из (8) и из того, что

$$d(R_u^{(j)},R_v^{(i)})=2d(U,R_u^{(j)}R_v^{(i)})-d(U,R_u^{(j)})-d(U,R_v^{(i)}).$$

Утверждение 4.

$$\mathrm{II}(i,j)=\sum_{u=1}^{S(n,j)}\sum_{v=1}^{S(n,i)}d^2(R_u^{(j)},R_v^{(i)})=$$

$$=n^2(n-1)^2\{S(n-2,i)\,(S(n,j)-2S(n-1,j)+S(n-2,j))+$$

$$+S(n-2,j)\,(S(n,i)-2S(n-1,i)+S(n-2,i))+$$

$$+2(S(n-1,i)-S(n-2,i))(S(n-1,j)-S(n-2,j))\}+$$
$$+2n(n-1)\{S(n,i)(S(n-1,j)-S(n-2,j))+$$
$$+S(n,j)(S(n-1,i)-S(n-2,i))-$$
$$-4(S(n-1,i)-S(n-2,i))(S(n-1,j)-S(n-2,j))\}.$$

Для доказательства просуммируем (13) вторично.

Чтобы найти сумму кубов всех возможных расстояний вычислим $\sum\limits_{v=1}^{S(n,\,i)} d(U, R_v^{(i)})\, d(R_u^{(j)}R_v^{(i)}, U)$ из (13), учитывая, что

$$d^2(R_u^{(j)}, R_v^{(i)})=4d^2(U, R_u^{(j)}R_v^{(i)})+(d(U, R_u^{(j)})+d(U, R_v^{(i)}))^2-$$
$$-4d(U, R_u^{(j)}R_v^{(i)})(d(U, R_v^{(i)})+d(U, R_u^{(j)})),$$

и просуммируем дважды (по u и v)

$$d^3(R_u^{(j)}, R_v^{(i)})=8d^3(U, R_u^{(j)}R_v^{(i)})-(d(U, R_u^{(j)})+d(U, R_v^{(i)}))^3+$$
$$+6d(U, R_u^{(j)}R_v^{(i)})(d^2(U, R_u^{(j)})+d^2(U, R_v^{(i)}))-$$
$$-12d^2(U, R_v^{(i)}R_u^{(j)})(d(U, R_u^{(j)})+d(U, R_v^{(i)}))+$$
$$+12d(U, R_v^{(i)}R_u^{(j)})\,d(U, R_u^{(j)})\,d(U, R_v^{(i)}),$$

откуда и получим требуемое.

Иным, более простым способом вычислены и суммы расстояний четвертой и пятой степени. Из-за громоздкости формулы здесь не приводятся. Предельные значения (при $n\to\infty$) для соответствующих моментов некоторого линейного преобразования расстояния Хемминга будут приведены несколько ниже.

Введем величину

$$y_{RT}^n=\frac{d(R,T)-\frac{2}{k}\left(1-\frac{1}{k}\right)n(n-1)}{\sqrt{n(n-1)}},$$

и вычислим ее первые моменты, учитывая (2) и (4), а также значения моментов для расстояния Хемминга третьего, четвертого и пятого порядков. Поскольку $S(n,k)=\frac{1}{k!}\sum\limits_{i=0}^{k}(-1)^{k-i}C_k^i\, i^n$, то $S(n,k)\sim\frac{k_n}{k!}$ и для первых пяти моментов случайной величины y_{RT}^n получаем:

$$\lim_{n\to\infty} Ey^n=0;$$

$$\lim_{n\to\infty} E(y^n)^2=\frac{4}{k^4}(k-1)(k^2-2k+2);$$

$$\lim_{n\to\infty} E(y^n)^3=\frac{16}{k^6}(k-1)(k^3-6k^2+8k-4);$$

$$\lim_{n\to\infty} E(y^n)^4 = \frac{48}{k^8}(k-1)(k^5-3k^4-4k^3+32k^2-36k+12);$$

$$\lim_{n\to\infty} E(y^n)^5 = \frac{64}{k^{10}}(k-1)(10k^6-78k^5-12k^4+516k^3-784k^2+$$
$$+448k-112).$$

Приведенные результаты можно, очевидно, получить и из следующей общей теоремы.

Теорема. *Случайная величина*

$$w^n_{RT} = \frac{1}{n} d(R, T) - \frac{2}{k}\left(1-\frac{1}{k}\right) n$$

при $n \to \infty$ сходится по распределению к случайной величине, характеристическая функция которой имеет вид

$$\frac{1}{\left(1-2\frac{k-2}{k^2}it\right)^{k-1}\left(1+\frac{4}{k^2}it\right)^{\frac{(k-1)^2}{2}}}.$$

Доказательство теоремы мы опускаем, оно будет опубликовано отдельно.

4. Пример. Рассмотрим совокупность из 100 элементов, которая разбивается тремя экспертами на два класса. Разбиение первого эксперта: в первом классе элементы 1, 2, ... , 30; во втором — 31, 32, ... , 100 (разбиение R). Разбиение второго эксперта: в первом классе объекты 1, 2, ... , 40 (разбиение T). Разбиение третьего эксперта: первый класс состоит из элементов 1, 2, ... , 35, 96, 97, ... , 100 (S).

Тогда

$$d(R, T) = d(S, T) = d(R, S) = 2\cdot 10\cdot 90 = 1800;$$

$$z^{100}_{RT} = z^{100}_{TS} = z^{100}_{RS} = \frac{(100)^2 - 2\cdot 1800}{\sqrt{100\cdot 99}} \approx 64{,}3.$$

Воспользуемся таблицами $\chi^2(t)$ распределения [7]. Для $\alpha=0{,}05$ (0,01) и $t=1$ получаем

$$z_1 = 0{,}98\cdot 10^{-3};\quad z_2 = 5{,}02\quad (z_1 = 0{,}39\cdot 10^{-6};\quad z_2 = 7{,}88).$$

Поскольку $64{,}3 > z_2$, гипотезу о случайном разбиении придется отбросить.

ЛИТЕРАТУРА

1. Б. Г. Миркин, Проблема группового выбора. М., 1974.
2. М. Кендэл, Ранговые корреляции. М., 1975.
3. С. А. Айвазян, З. И. Бежаева, О. В. Староверов, Классификация многомерных наблюдений. М., 1974.
4. Многомерный статистический анализ в социально-экономических исследованиях. Ученые записки по статистике, **XXVI**. М., 1974.

5. Л. Б. Черный, Метод пространства разбиений в анализе качественных данных. Автореферат на соискание ученой степени кандидата технических наук. М., 1973.
6. Д. Г. Кемени, Д. Л. Снелл, Кибернетическое моделирование. М., 1972.
7. Г. Г. Абезгауз, А. П. Тронь, Ю. Н. Копенкин, И. А. Коровина, Справочник по вероятностным расчетам. М., 1970.

Институт экономики
Академии наук Эстонской ССР

Поступила в редакцию
28/VI 1976

A. MAAMÄGI

OBJEKTIDE EKSPERTKLASSIFIKATSIOON. EBAKOMPETENTSETE EKSPERTIDE KLASSIFIKATSIOONIDE VAHELISTE KAUGUSTE JAOTUSSEADUS

Resümee

Artiklis vaadeldakse kahe järjestamata klassifikatsiooni olulise erinevuse hüpoteesi statistilist kontrolli. Esitatakse klassifikatsioonide sarnasust iseloomustava statistiku piirjaotus objektide arvu piiramatu kasvamise korral ja meetod, kuidas leida jaotusseadus väikese arvu objektide puhul ühe- ja kaheklassilistele klassifikatsioonidele. Statistiku viis esimest momenti on leitud mistahes arvu klasside jaoks.

Eesti NSV Teaduste Akadeemia
Majanduse Instituut

Toimetusse saabunud
28. VI 1976

A. MAAMÄGI

AN EXPERT CLASSIFICATION OF OBJECTS. DISTRIBUTION OF DISTANCES BETWEEN CLASSIFICATIONS GIVEN BY INCOMPETENT EXPERTS

Summary

The paper discusses a statistical test of a hypothesis about a substantial difference between two unranged classifications. A limiting distribution for a statistic characterization of the resemblance of the classifications is presented for a large number of objects; besides, a technique is suggested for computing the distribution in case of a small number of objects for one- and two-class classifications. The first five moments of the statistics have been found for any number of classes.

Academy of Sciences of the Estonian SSR,
Institute of Economics

Received
June 28, 1976