Коэффициенты связи для совокупности номинальных признаков

В.М.Колесов

(Томск)

Вводятся вероятностные меры связи для совокупности номинальных признаков. В мультиномиальной модели данных с использованием многомерной таблицы сопряженности признаков оцениваются направленные и ненаправленные (симметричные) меры связи. Рассматриваются свойства полученных таким образом коэффициентов и их информационные аналоги. Дается пример применения результатов исследования при анализе социологических данных.

Ключевые слова: номинальные признаки, многомерная таблица сопряженности, меры связи, коэффициенты связи, мультиномиальное распределение.

1. Введение

Для измерения зависимости между двумя номинальными признаками предлагается целый ряд коэффициентов связи ([1-3]), что порождает проблему выбора при их практическом ис-

пользовании. В связи с этим Кендалл и Стьюарт подчеркивали, что основной трудностью является то, что коэффициенты связи не имеют простой вероятностной интерпретации [1]. В последующих исследованиях в качестве меры связи рассматривался функционал, характеризующий уменьшение дисперсии наблюдений при наличии зависимости между признаками [4, 5]. Он, в свою очередь, при специальных способах задания функций для подсчета дисперсии позволяет получить так называемые меру концентрации Джини и энтропийную меру Тейла [6]. Их оценивание при мультиномиальной модели данных приводит к коэффициентам связи соответственно Гудмэна-Крускала [7] (известным также под именем Валлиса [8]) и информационным. Своеобразная интерпретация некоторых коэффициентов проводится в работе [8], где определяется вероятностный механизм непосредственно на таблице сопряженности[1].

Рассматриваемый нами подход к измерению связи между признаками основан на интерпретации статистик как оценок соответствующих функционалов (мер связи). В качестве таких функционалов определяются не дисперсионные, а вероятностные меры связи, основанные на вероятностях некоторых событий. Для случая двух признаков их оценивание приводит к коэффициентам Гудмэна-Крускала и информационным. В то же время методика построения указанных мер позволяет естественным образом обобщить ее на случай большего числа признаков. С целью компактности изложения, уменьшения количества индексов далее предполагается, что это число равно 3.

2. Основные предположения при введении мер связи

Пусть изучение некоторой системы связано с наблюдением составляющих ее объектов. Для каждого объекта регистрируются значения трех номинальных признаков π_1, π_2, π_3 со множествами возможных значений соответственно $\{\pi_1(i)\}$, $\{\pi_2(j)\}$, $\{\pi_3(k)\}$; $i=1,\dots,m_1$; $j=1,\dots,m_2$; $k=1,\dots,m_3$.

[1] Показатели связи, меры близости были объектами внимания и многих других исследователей: см. обзоры **Орлов А.И.** Общий взгляд на статистику объектов нечисловой природы //Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1У85; **Раушенбах Г.В.** Меры близости и сходства в социологии //[там же); **Елисеева И.И., Рукаиишников В.О.** Группировка, корреляция, распознавание образов. М.: Статистика, 1977. **(Прим, ред.)**

При определении мер связи между признаками предполагается существование дискретного распределения

Для упрощения записи обозначим I, J, K случайные величины с распределением $\{p_{ijk}\}$. Они определяются на номерах уровней признаков π_1, π_2, π_3 соответственно, и

$p_{ijk}=p(I=i> J=J> K=k)$; $i=1,\dots,m_1$; $j=1,\dots,m_2$; $k=1,\dots,m_3$.

По аналогии введем в рассмотрение еще три случайные величины $\hat{I}$, $\hat{J}$, $\hat{K}$, имеющие то же распределение и статистически независимые с I, J, K. Тогда, подставляя в $P(\hat{I}=i, \hat{J}=j, \hat{K}=k)$ вместо аргументов i, j, k случайные величины I, J, K, получим случайную вероятность $P_S(J=I, J=J, \hat{K}=K)$, которая представляет собой случайную величину со множеством значений $\{p_{ijk}\}$ и распределением $\{p_{ijk}\}$. Для вероятности совпадения всех трех пар случайных величин I, J, K и $\hat{I}$, $\hat{J}$, $\hat{K}$ очевидно представление

$$P(\hat{I}=I, \hat{J}=J, \hat{K}=K)=MP(\hat{I}=I, \hat{J}=J, \hat{K}=K)=\sum\sum\sum p^2_{ijk}, \quad (1)$$

где математическое ожидание M берется по распределению $\{p_{ijk}\}$.

Аналогично вводятся случайные условные вероятности $P_s(\hat{I}=I/\hat{J}=J, \hat{K}=K)$, $P_s(\hat{J}=J/\hat{I}=I, \hat{K}=K)$ и $P_S(\hat{K}=K/\hat{I}=I, \hat{J}=J)$. Отметим, что для них соотношения типа (1) не выполняются, именно

$$P(\hat{I}=I/\hat{J}=J, \hat{K}=K) \neq MP_S(\hat{I}=I/\hat{J}=J, \hat{K}=K).$$

Другое важное соглашение касается выборки, по которому производится оценивание мер связи. Будем предполагать, что эта выборка повторная, т.е. наблюдения объектов независимы, и распределение $\{p_{ijk}\}$ не изменяется в процессе получения наблюдений. Тогда элементы таблицы сопряженности признаке $T_n(\pi_1,\pi_2,\pi_3)=(n_{ijk})$ $i=1,\dots,m_1$; $j=1,\dots,m_2$; $k=1,\dots,m_3$. имеют мультиномиальное распределение, при котором оценками максимального правдоподобия для вероятностей p_{ijk} являются относительные частоты $p_{ijk}=n_{ijk}/n$.

Для указания того, что по индексу проведено суммирование, он заменяется знаком " + ", например

$$\hat{p}_{+j+}=\sum_i\sum_k p_{ijk}.$$

3. Многомерные аналоги коэффициентов / Гудмэна-Крускала

Проведем анализ для двух номинальных признаков π_1, π_2 с соответствующим распределением $\{p_{ij}\}$ случайных величин I, J. Определим для них направленные меры связи

$$D_{\pi_2/\pi_1}=\frac{MP_s(\hat{J}=J/\hat{I}=I)-MP_s(\hat{J}=J)}{1-MP_s(\hat{J}=J)}, (2)$$

$$D_{\pi_1/\pi_2}=\frac{MP_s(\hat{I}=I/\hat{J}=J)-MP_s(\hat{I}=I)}{1-MP_s(\hat{I}=I)}, (3)$$

Симметричную (ненаправленную) меру связи $D_{\pi1,\pi2}$ зададим как сумму числителей D_{π_2/π_1} и D_{π_1/π_2}, поделенную на сумму их знаменателей. Оценки введенных функционалов приводят к коэффициентам Гудмэна-Крускала для двух номинальных признаков

$$\hat{D}_{\pi_2/\pi_1}=\frac{\sum_i\sum_j(\hat{p}_{ij}-\hat{p}_{i+}\hat{p}_{+j})^2/\hat{p}_{i+}}{1-\sum_j\hat{p}^2_{+j}}, (4)$$

$$\hat{D}_{\pi_1/\pi_2}=\frac{\sum_i\sum_j(\hat{p}_{ij}-\hat{p}_{i+}\hat{p}_{+j})^2/\hat{p}_{+j}}{1-\sum_i\hat{p}^2_{i+}}, (5)$$

$$\hat{D}_{\pi_1/\pi_2}=\frac{\sum_i\sum_j(\hat{p}_{ij}-\hat{p}_{i+}\hat{p}_{+j})^2(1/\hat{p}_{+j}+1/\hat{p}_{i+})}{2-\sum_i\hat{p}^2_{i+}-\sum_j\hat{p}^2_{+j}}, (6)$$

Определение мер связи для коэффициентов (4)-(6) позволяет естественным образом получить их обобщение на случай любого числа признаков r. Каждая из r направленных мер связи при этом характеризует зависимость соответствующего признака от остальных r-1. Так, при r=3, имеем

$$D_{\pi_3/\pi_1\pi_2} = \frac{MP_s(\hat{K}=K/\hat{I}=I,\hat{J}=J) - MP_s(\hat{K}=K)}{1-MP_s(\hat{K}=K)}, (7)$$

$$D_{\pi_2/\pi_1\pi_3} = \frac{MP_s(\hat{J}=J/\hat{I}=I,\hat{K}=K) - MP_s(\hat{J}=J)}{1-MP_s(\hat{J}=J)}, (8)$$

$$D_{\pi_1/\pi_2\pi_3} = \frac{MP_s(\hat{I}=I/\hat{J}=J,\hat{K}=K) - MP_s(\hat{I}=I)}{1-MP_s(\hat{I}=I)}, (9)$$

Построение симметричных мер очевидно, поэтому выражения для них далее не приводятся. Свойства функционалов (7)-(9) аналогичны свойствам мер связи при r=2. А именно, их значения принадлежат отрезку [0,1]. Направленная мера связи равна нулю при статистической независимости между соответствующим ей признаком и остальными г-1 признаками. Ее единичное значение указывает на наличие детерминированной зависимости, т.е. по значениям уровней г-1 признаков можно без ошибочно определить уровень соответствующего данной мер признака. Оценки функционалов (7)-(9), представляющие направленные коэффициенты связи, имеют вид

$$\hat{D}_{\pi_3/\pi_1\pi_2} = \frac{\sum_i\sum_j\sum_k(\hat{p}_{ijk} - \hat{p}_{ij+}\hat{p}_{++k})^2/\hat{p}_{ij+}}{1-\sum_k \hat{p}^2_{++k}}, (10)$$

$$\hat{D}_{\pi_2/\pi_1\pi_3} = \frac{\sum_i\sum_j\sum_k(\hat{p}_{ijk} - \hat{p}_{i+k}\hat{p}_{+j+})^2/\hat{p}_{i+k}}{1-\sum_j \hat{p}^2_{+j+}}, (11)$$

$$\hat{D}_{\pi_1/\pi_2\pi_3} = \frac{\sum_i\sum_j\sum_k(\hat{p}_{ijk} - \hat{p}_{+jk}\hat{p}_{i++})^2/\hat{p}_{+jk}}{1-\sum_i \hat{p}^2_{i++}}, (12)$$

Направленные коэффициенты связи не убывают при уве личении числа анализируемых признаков. При этом направ ленные меры связи для совокупности признаков остаются неиз -

менными только в том случае, если добавляемый признак статистически независим от исходной совокупности.

Данное утверждение следует из соотношений между математическими ожиданиями условных случайных вероятностей при увеличении числа признаков. Например, при r=3 из выражений

$$MP_s(\hat{J}=J/\hat{I}=I,\hat{K}=K) - MP_s(\hat{J}=J) = \sum_i\sum_j\sum_k(p_{ijk}p_{i++} - p_{ij+}p_{i+k})^2/(p_{i+k}p^2{}_{i++}), (13)$$

$$MP_s(\hat{I}=I/\hat{J}=J,\hat{K}=K) - MP_s(\hat{I}=I) = \sum_i\sum_j\sum_k(p_{ijk}p_{+j+} - p_{ij+}p_{+jk})^2/(p_{+jk}p^2{}_{+j+}), (14)$$

следует $D_{\pi_2/\pi_1\pi_3} \geq D_{\pi_2/\pi_1}, D_{\pi_1/\pi_2\pi_3} \geq D_{\pi_1/\pi_2}$

Указанные свойства направленных мер и коэффициентов связи для совокупности номинальных признаков аналогичны свойствам множественных и выборочных множественных коэффициентов корреляции в числовой статистике [1]. В смысле практического использования первые предназначены для решения той же задачи отсеивания признаков, что и вторые, но в номинальной шкале измерения признаков. Действительно, малость коэффициентов связи Гудмэна-Крускала D_{π_2/π_1} и D_{π_1/π_2}, не гарантирует малости направленных коэффициентов связи более высокого порядка ($D_{\pi 1/\pi 2\pi 3}, D_{\pi 2/\pi 1\pi 3}$ и т.д.). Их исследование позволяет обнаружить «скрытые» зависимости между признаками, не проявляющиеся при анализе парных коэффициентов связи.

Асимптотическая нормальность распределения направленных коэффициентов при r>2 доказывается так же, как и для коэффициентов Гудмэна-Крускала [9]. Пусть символ $\hat{=>}$ означает сходимость по распределению. Представляя {p_{ijk}} и { p_{iik} } в виде столбцов $\hat{p}$ и p, имеем: $\sqrt{n}(p-p) \Rightarrow N(0, \sum_{,}), где \sum_{,} = diag(p) - p * p^T$, где £ = diag(p)-p-p^T. Обозначим v числитель, а δ знаменатель направленной меры связи.

Асимптотическая дисперсия коэффициента связи (например, $\hat{D}_{\pi 3/\pi 1\pi 2}$) находится δ-методом [9] из соотношении

$$\gamma = \sum_i \sum_j \sum_k p_{ijk}^2 / p_{ij+} - \sum_k p_{++k}^2 ,$$

$$\delta = 1 - \sum_k p_{++k}^2$$

$$\gamma'_{ijk} = \frac{\partial \gamma}{\partial p_{ijk}} = \frac{2 p_{ij+} p_{ijk} - \sum_u p_{iju}^2}{p_{ij+}^2} - 2 p_{++k} ,$$

$$\delta'_{ijk} = \frac{\partial \delta}{\partial p_{ijk}} = -2 p_{++k} ,$$

$$d_{ijk} = \frac{\delta \gamma'_{ijk} - \gamma \delta'_{ijk}}{\delta^2} .$$

Тогда, если $d^T=(d_{111}...d_{m1m2m3})$ и $D_{\pi 3/\pi 1 \pi 2} \in [0,1]$, то

$\sqrt{n}(D_{\pi 3/\pi 1 \pi 2} - D_{\pi 3/\pi 1 \pi 2}) \Rightarrow N(0, \sigma^2{}_{\pi 3/\pi 1 \pi 2})$, где

$\sigma^2{}_{\pi 3/\pi 1 \pi 2} = d^T = \sum d$

Аналогичные результаты можно получить и для других коэффициентов связи при соответствующей замене выражении для v и δ.

4. Информационные коэффициенты связи для совокупности признаков

Введение D-функционалов связи основано на нормировке (т.е. приведении значений на отрезок [0, 1]) математических ожиданий случайных условных вероятностей $P_s(\bullet)$.

При синтезе информационных мер связи нами будет ис пользована та же вероятностная модель признаков с заменой линейной функции $P_s(\bullet)$ на $\log P_s(\bullet)$ и соответствующим приве-дением значений мер связи на отрезок [0, 1]. Полученные таким образом меры будем называть U-функционалами связи. При г =2 имеем

где количество информации

$$U_{\pi_2/\pi_1} = \frac{M \log P_s(\hat{J} = J / \hat{I} = I) - M \log P_s(\hat{J} = J)}{- M \log P_s(\hat{J} = J)} = \frac{I(\pi_1, \pi_2)}{H(\pi_2)} . (15)$$

$$U_{\pi_1/\pi_2} = \frac{M \log P_s(\hat{I} = I / \hat{J} = J) - M \log P_s(\hat{I} = I)}{- M \log P_s(\hat{I} = I)} = \frac{I(\pi_1, \pi_2)}{H(\pi_1)} , (16)$$

$$U_{\pi_1, \pi_2} =$$

$$\frac{M \log P_s(\hat{J} = J / \hat{I} = I) - M \log P_s(\hat{J} = J) + M \log P_s(\hat{I} = I / \hat{J} = J) - M \log P_s(\hat{I} = I)}{- M \log P_s(\hat{J} = J) - M \log P_s(\hat{I} = I)} =$$

$$= \frac{2 I(\pi_1, \pi_2)}{H(\pi_1) + H(\pi_{21})} , (17)$$

$$I(\pi_1, \pi_2) = \sum_i \sum_j p_{ij} \log(p_{ij} / p_i p_j),$$

энтропии признаков

$$H(\pi_1) = -\sum_i p_{i+} \log p_{i+} , H(\pi_2) = -\sum_i p_{+j} \log p_{+j} .$$

Оценки функционалов (15), (16) представляют направлен-ные информационные коэффициенты для двух признаков, а оценка (17) — соответствующий им симметричный коэффициент связи.

Как и в случае D-функционалов, методика построения информационных мер связи позволяет получить их обобщение для r>*2.* Например, направленные информационные меры связи

для трех признаков

$$U_{\pi_3/\pi_1\pi_2} = \frac{M \log P_s(\hat{K} = K / \hat{I} = I, \hat{J} = J) - M \log P_s(\hat{K} = K)}{- M \log P_s(\hat{K} = K)} = \frac{I((\pi_1, \pi_2), \pi_3)}{H(\pi_3)} . (1$$

$$U_{\pi_2/\pi_1\pi_3} = \frac{M \log P_s(\hat{J} = J / \hat{I} = I, \hat{K} = K) - M \log P_s(\hat{J} = J)}{- M \log P_s(\hat{J} = J)} = \frac{I((\pi_1, \pi_3), \pi_2)}{H(\pi_2)} , (1$$

$$U_{\pi_1/\pi_2\pi_3} = \frac{M \log P_s(\hat{I} = I / \hat{J} = J, \hat{K} = K) - M \log P_s(\hat{I} = I)}{- M \log P_s(\hat{I} = I)} = \frac{I((\pi_2, \pi_3), \pi_1)}{H(\pi_1)} , (20$$

где $I((\pi_u,\pi_v),\pi_w)$ — количество информации в признаках $(\pi_u,/\pi_v)$ относительно признака π_w, т.е.

$$I((\pi_1,\pi_2),\pi_3 = \sum_i\sum_j\sum_k p_{ijk}\log(p_{ijk}/p_{ij+}p_{++k},$$

$$I((\pi_1,\pi_3),\pi_2 = \sum_i\sum_j\sum_k p_{ijk}\log(p_{ijk}/p_{i+k}p_{+j+},$$

$$I((\pi_2,\pi_3),\pi_1 = \sum_i\sum_j\sum_k p_{ijk}\log(p_{ijk}/p_{+jk}p_{i++}.$$

Непосредственно из свойств количества информации [10 следует, что при увеличении числа признаков информационные меры связи не убывают

$$U_{\pi 1/\pi 2} \le U_{\pi 1/\pi 2\pi 3} \le \ldots \le U_{\pi 1/\pi 2\ldots\pi m}$$

Соотношения типа (21) справедливы для направленной меры связи любого признака (не только π_1), входящего в исходную совокупность. Статистические свойства оценок U-функционалов аналогичны тем же для D-функционалов. Для нахождения дисперсии асимптотического распределения использует ранее указанная методика. Например, для направленного коэффициента $U_{\pi 3/\pi 1\pi 2}$ получаем:

$$\sqrt{n\left(\hat{U}_{\pi 3/\pi 1\pi 2} - U_{\pi 3/\pi 1\pi 2}\right)} \Rightarrow N(0,\sigma^2_{\pi 3/\pi 1\pi 2}\_,$$ где $\sigma^2_{\pi 3/\pi 1\pi 2}$ вычисляется при

$$\nu = \sum_i\sum_j\sum_k p_{ijk}\log(p_{ijk}/p_{ij+}p_{++k}),$$

$$\delta = \sum_k p_{++k}\log(p_{++k}),$$

$$\gamma'_{ijk} = 1 - p_{ijk}/p_{ij+}p_{++k}),$$

$$\delta'_{ijk} = 1 + \log(p_{++k}).$$

5. Иллюстративный пример

Предложенный подход к анализу взаимосвязей для совокупности номинальных признаков предоставляет возможное! более углубленного исследования данных, поскольку информация, полученная по парным коэффициентам связи, отражает наличие лишь локальных зависимостей. В этом смысле введен*.

ные меры связи, представленные D и U-функциона обработке номинальных признаков так же важны, ка фициенты множественной корреляции в числовой ста качестве иллюстративного примера приведем резул работки трех признаков с использованием D-функцио

Расчеты проводились по материалам анкетного вания студентов, выполненного в 1989 году по прогр щественное мнение» Гособразования СССР. Признакам соответствовали следующие вопросы анкеты и альте ответы на них:

(π_1) Если бы Вам лично пришлось выбирать путь обществ развития, какой бы Вы выбрали?

1. Дальнейшую демократизацию;
2. Сегодняшний уровень развития демократии
3. Твердую власть популярного в народе руков

(π_2) Изменилось ли Ваше отношение к перестройке за год реализации?

1. Не изменилось;
2. Изменилось в худшую сторону;
3. Изменилось в лучшую сторону;
4. Трудно сказать.

(π_3) Ощущаете ли Вы лично потребность в активизации с общественной деятельности? 1.Да; 2. Нет.

Объем выборки, по которой рассчитывались к енты связи, равен 1101 наблюдению. Ниже приводятс сопряженности признаков, значения статистик x^2, соответ им достигнутые уровни значимости Pr и значения коэф Гудмэна-Крускала. В таблицах сопряженности пре абсолютные частоты, уровни признаков кодиру номерами, небольшой процент пропущенных значен работке игнорировался.

Таблица сопряженности признаков π_1 и π_2

π_1	π_2			
	1	2	3	4
1	144	276	94	332
2	3	5	4	23
3	24	65	6	64

X^2= 19,014; Рг=0,004; $\hat{D}_{\pi 1/\pi 2}$ =0,07; $\hat{D}_{\pi 2/\pi 1}$ =0,07

Таблица сопряженности признаков π_3 и π_2

π_1	π_2			
	1	2	3	4
1	80	129	70	210
2	94	229	39	217

X^2= 30,624; Рг<0,001;

$\hat{D}_{\pi 3/\pi 2}$ =0,029; $\hat{D}_{\pi 2/\pi 3}$ =0,007.

Таблица сопряженности признаков π_3 и π_1

π_1	π_2		
	1	2	3
1	80	129	70
2	94	229	39

X^2= 11,075; Рг=0,004; $\hat{D}_{\pi 3/\pi 1}$ =0,011; $\hat{D}_{\pi 1/\pi 3}$ =0,002

Из таблиц следует вывод о наличии зависимости между признаками. Однако коэффициенты связи между ними малый что, казалось бы, говорит о ее «слабости». Ситуация вполне аналогична уже известным. Одну из них приводит Г.Крамер [11]. По I выборке в 25263 супружеских пар он исследовал зависимость | между двумя социально-демографическими характеристиками:

годовым доходом и количеством детей в семье. При двенадцати степенях свободы статистика x^2 оказалась равной 568,5, что показывает чрезвычайно высокую значимость отклонения от гипотезы о независимости признаков (на обычно используемом уровне значимости 0,05 критическое значение статистики равно 21,026). Рассчитанный по той же таблице сопряженности коэффициент связи Крамера равен всего лишь 0,0075. На основании этого им было сделано заключение о «слабой зависимости» между этими характеристиками. Можно также встретить достаточно категоричные утверждения общего характера, основанные на парных коэффициентах связи, о «слабой зависимости» социально-демографических и других социологических показателей, например: «Коэффициент сопряженности Крамера, как правило, не превышает значения 0,3, а это достаточно слабая связь» [12, с. 119].

В то же время имеются замечания о том, что аргументация заключений с использованием значений коэффициентов связи между парами признаков не является столь очевидной [13].

Какие при таких обобщениях могут подстерегать социолога неожиданные результаты, можно показать, вычислив коэффициент связи между признаком т, и совокупностью признаков (π_2, π_3). Хотя значения коэффициентов между любыми парами очень малы, $\hat{D}_{\pi 1/\pi 2\pi 3}$ =0,467. Таким образом, связи между интересующим исследователя признаком и другими показателями не могут характеризоваться лишь на основе изучения парных коэффициентов. Для этого необходимо использование множественных коэффициентов.

Литература

1. *Кендолл М.Дж., СтьюартА.* Статистические выводы и связи. М.: Наука, 1973.
2. *Миркин Б.Г.* Группировки в социально-экономических исследованиях. М.: Финансы и статистика, 1985.
3. *ЕнюковИ.С.* Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА. М.: Финансы и статистика, 1986.

4. *Magidson J.* Qualitative Variance, Entropy and Correlation Ratios for Nominal Dependent Variables //Social Sci. Research. 1982 V 77 P. 177-194.
5. *Habennan S.J.* Analysis of Dispersion of Multinomial Responses //J. Amer. Stat. Ass. 1982. V.77. P.568-580.
6. *Tlieil H.* On the Estimation of Relationships Involving Qualitative Variables //Amer. J. of Sociology. 1970. V.76. P.103-154.
7. *Goodman LA., Kruskal W.H.* Measures of Association for Cross Classification //J. Amer. Stat. Ass. 1954. V.49. P.732-764.
8. *Миркин Б.Г.* Анализ качественных признаков и структур. М.: Статистика, 1980.
9. *Bishop Y.M.M., Fienberg S.E., Holland P. W.* Discrete Multivariate Analysis. Cambridge, 1985.
10. *ТaраceiiКО Ф.П.* Введение в курс теории информации. Томск: Изд-во Томского ун-та, 1963.
11. *П.Крамер Г.* Математические методы статистики. М.: Мир, 1975.™
12. *Давыдов АЛ.* Репрезентативность выборки //Социол исслед ч 1990. №1.
13. *Толстова Ю.Н.* Методология математического анализа данных //Социол. исслед. 1990. №6.: