

Факты, комментарии, заметки (с рабочего стола социолога)

© 1992 г.

Л.Б. КОСОВА

ДОЛЖЕН ЛИ СОЦИОЛОГ ПОМНИТЬ ТАБЛИЦУ УМНОЖЕНИЯ?

КОСОВА Лариса Борисовна — кандидат технических наук, старший научный сотрудник Института социологии РАН. В нашем журнале публикуется впервые.

Зачем все это надо

Слова «... только последовательное использование математических методов...» все чаще встречаются в статьях о предметах традиционно гуманитарных. Это понятно: математика — инструмент универсальный, позволяющий получить достоверные и интересные результаты. Но, как и любой инструмент, она требует умения с ней обращаться. Неправильное использование абсолютно верных методов может привести к серьезным систематическим ошибкам. Понимание этого все чаще приводит к союзу гуманитария и математика — союзу непростому. Каждый, кто работал в таком тандеме, прекрасно помнит сложности взаимных отношений.

Существует ли способ решения данной проблемы? Смею утверждать, что да. Давайте оглянемся на историю развития вычислительной техники. Первые вычислительные комплексы были настолько сложны, что работать с ними приходилось самим разработчикам: так было проще, чем объяснить кому-либо, как это делается. По мере развития компьютеров, совершенствования их архитектуры, процессы поддержания системы и использования ее в различных прикладных целях становились все более и более простыми. Программисты, работавшие на машинах первого поколения, напоминали жрецов, призванных и избранных. Ибо кто, кроме них, мог осуществить ритуал загрузки системы? Сейчас с подобными задачами без труда справляются дети 3—5 лет. Компьютеры не стали проще — нет, появились «дружественные» интерфейсы, ориентированные на неподготовленного пользователя. Они берут на себя те функции, которые раньше вынужден был выполнять программист; прячут от пользователя ненужное ему знание. Данный подход кажется мне заслуживающим внимания не только в сфере computer science.

Является ли социологическим знанием информация о том, что коэффициент Пирсона для переменных V_1 и P_4 равен 13,86 при уровне значимости 0,0027? Социологу важно понять, действительно ли образование влияет на результаты голосования или это только кажется. Методы математической статистики давно формализованы, значит, могут быть спрятаны внутрь программного обеспечения. Я уверена, что социолог должен общаться с компьютером на привычном языке, получая достоверные результаты. Переход от цифри к выводу, представленному в человекоориентированной форме, компьютер (точнее, программное обеспечение) вполне может сделать самостоятельно,

Иными словами, мне кажется, что настала пора перенести методику создания интеллектуальных интерфейсов, широко используемую в computer science в область

статистики, в ту ее сферу, с которой сталкиваются гуманитарии. Опишем небольшую процедуру. Она позволяет анализировать связи между переменными — аналог кросстабуляции пакета SPSS, но прячет цифры от пользователя, общаясь с ним на вполне «человеческом» языке и представляя результаты в форме, более привычной для «статистически» неподготовленного пользователя.

Коротко о методе

Известно, что для двух независимых случайных событий вероятность их совместного появления равна:

$$P(A \cap B) = P(A) \cdot P(B)$$

На «социологический язык» это можно перевести так. Пусть выборка включает N респондентов. N_1 респондентов выбрали позицию анкеты A , N_2 респондентов — позицию B . Тогда, если события независимы, вероятность их одновременного появления равна:

$$\hat{P} = \frac{N_1}{N} \cdot \frac{N_2}{N} = \frac{N_1 \cdot N_2}{N^2}$$

Или, если мы хотим получить легко интерпретируемую величину, можно оценить число респондентов, попадающих в пересечение:

$$\hat{K} = \frac{N_1 \cdot N_2}{N}$$

Но, если опрос проведен, подсчитать реальное число респондентов, выбравших одновременно позиции A и B очень просто. Пусть оно равно K . Еще проще определить разницу между реальным и гипотетическим значениями. Обозначим ее

$$\sigma = K - \hat{K}$$

В каком случае величину отклонения можно считать значимой? Для получения оценки я пользуюсь неравенством Чебышева (правда, необходимо опять перейти к вероятности, но это несложно). Хочу подчеркнуть, что процедура строит интервал, а не хранит статистические таблицы. Мне это кажется ее достоинством.

Как это сделано

Процедура реализована на языке C в операционной среде MS-DOS. Для обеспечения работы с разными форматами хранения данных создан набор интерфейсов. Они позволяют обрабатывать форматы хранения, принятые во Всесоюзном центре изучения общественного мнения (ВЦИОМ), а также ASCII файлы, в которых переменные размещены по столбцам. Именно этот формат продуцирует оператор WRITE пакета SPSS.

Процедура создает два типа выходных файлов. Первый тип, назовем его промежуточным, может быть полезен при обработке ASCII файлов, когда отсутствует какая-либо информация о метках переменных и значений. Промежуточный файл содержит номера связанных переменных и меру связи. Более удобной формой представления результатов мне кажется распечатка непосредственно связанных позиций анкеты. Но в этом случае необходим либо текст анкеты, либо информация о метках переменных и значений.

Как выглядит процесс работы с процедурой для пользователя? Он очень прост — инструкция состоит всего из 6 пунктов, и она не сложнее руководства по эксплуатации электроутюга. Процедура представляет собой «эгзешный» модуль, который вызывается по имени. При вызове процедура распаковывает файл сообщений и высвечивает на экране многооконную панель. Каждое окно подписано. Пользователь задает имена обрабатываемых файлов и номера интересующих его вопросов, после чего он может идти обедать. Процедура проанализирует специфицированные переменные и отыщет

для них все значимые связи. Причем, не пропустит ни одной — компьютеры невероятно пунктуальны! Результат будет представлен в наиболее человеколюбивой форме: не цифры, а текст, который вы сможете читать. Примеры приведены в приложении. Хочу подчеркнуть, что все это процедура сделала сама. Никакой дополнительной аналитической работы не проводилось.

Что же в результате?

Как мне кажется, процедура является удобным для работы инструментом. Одним из возможных инструментов, но разве этого мало? Процедура реализует принцип «черного ящика» — пользователь не должен понимать, что у нее внутри. Он вообще может не знать имени Чебышева. В данном случае на результатах это не отразится: вся статистика спрятана от пользователя.

Процедура полезна не только тем, кто сомневается в своей статистической подготовке. Любая аналитическая работа начинается с выявления связей между переменными. Процедура сделает это за вас. Скажите ей — какой опрос вас интересует, и она сообщит вам какие связи существуют между переменными. Она более «дружелюбна», чем SPSS, и позволяет обрабатывать весь опрос одновременно. Сейчас процедура ничего не говорит о характере связей. Но она допускает развитие. (Я бы сказала, что она ждет его. Или развитие ждет ее?) Анализ природы связей формализован, следовательно может быть запрограммирован. Программирование потребует некоторого времени и сил, но вполне возможно.

Хочется еще раз подчеркнуть: пришло время перенести хорошо зарекомендовавшие себя методы интеллектуализации программного обеспечения в области статистической обработки. Пользователь не должен выполнять рутинную, легко формализуемую работу. Ее необходимо переложить на компьютер. Творческое — человеку, остальное — компьютеру.

В заключение — несколько примеров, иллюстрирующих работу процедуры. Это фрагменты результатов обработки опроса, проведенного ВЦИОМ в 1990 г.

Примеры

Как отмечалось выше, процедура позволяет проанализировать зависимость между номинальными переменными. Это же можно сделать с помощью других программных средств — пакета SPSS, например. Сравним алгоритм работы социолога с пакетом SPSS и описываемой процедурой. В приложении 1 приведен пример работы оператора CROSSTABS. Проанализировано две пары дихотомических переменных. Социолог должен проанализировать подсчитанные статистики и в зависимости от уровня значимости решить, есть между ними связь или нет. Причем, надо построить и проанализировать столько таблиц, сколько пар переменных надо проверить.

В приложении 2 представлены результаты работы описанной в статье процедуры. Она прячет цифры от пользователя: сама анализирует статистики, объем выборки и уровень значимости. Проанализировано 20 дихотомических переменных, распечатаны только значимо связанные. Процедура указывает число респондентов в подвыборке и меру связи. Еще раз напомним — процедура сама продуцирует данный формат выдачи. Никакой дополнительной аналитической работы не проводилось.

A16 by A7

Count Col Pet	A7		Row Total
	0	1	
0	450 66.3	360 66.2	810 66.2
1	229 33.7	184 33.8	413 33.8
Column Total	679 55.5	544 44.5	1223 100.0

Page 1 of 1

Chi-Square	Value	DF	Significance
Pearson	00128	1	97143
Continuity Correction	00000	1	100000
Likelihood Ratio	00123	1	97199
Mantel-Haenszel test for linear association	00128	1	97144

Minimum Expected Frequency — 183.706

A20 by A7

Count Col Pet	A7		Row Total
	0	1	
0	647 95.3	442 81.3	1089 89.0
1	32 4.7	102 18.8	134 11.0
Column Total	679 55.5	544 44,5	1223 100.0

Page 1 of 1

Chi-Square	Value	DF	Significance
Pearson	60.99899	1	00000
Continuity Correction	59.56867	1	00000
Likelihood Ratio	62.33332	1	00000
Mantel-Haenszel test for linear association	60.94911	1	00000

Minimum Expected Frequency — 59.604

- Евреи есть среди моих близких родственников (27)
Евреи есть среди моих близких друзей 1.94
- Евреи есть среди моих отдаленных родственников (52)
Евреи уезжают — хотят обеспечить лучшее будущее своим детям 2.82
- Евреи есть среди моих близких друзей (149)
Евреи есть среди моих близких родственников 1.94
Евреи уезжают — талантливые люди не имеют здесь перспектив 2.04
Евреи уезжают — хотят обеспечить лучшее будущее своим детям 2.18
Некоторые из моих соседей — евреи 2.76
Евреи есть среди моих коллег и соучеников 2.78
Евреи уезжают из-за угроз, оскорблений 4.68
- Евреи есть среди моих коллег и соучеников (312)
Евреи уезжают из-за экономического кризиса 2.06
Евреи уезжают, т.к. не верят, что здесь возможен прогресс 2.24
Евреи есть среди моих близких друзей 2.78
Евреи есть среди моих знакомых 3.20
Евреи уезжают — хотят обеспечить лучшее будущее своим детям 3.90
Некоторые из моих соседей — евреи 4.11
Евреи уезжают — талантливые люди не имеют здесь перспектив 4.32
- Некоторые из моих соседей — евреи (183)
Евреи есть среди моих знакомых 2.74
Евреи есть среди моих близких друзей 2.76
Евреи есть среди моих коллег и соучеников 4.11
Евреи уезжают из-за угроз, оскорблений 4.57
Евреи уезжают — талантливые люди не имеют здесь перспектив 7.33
- Евреи есть среди моих знакомых (364)
Евреи уезжают — хотят обеспечить лучшее будущее своим детям 2.54
Некоторые из моих соседей — евреи 2.74
Евреи есть среди моих коллег и соучеников 3.20
Евреи уезжают — талантливые люди не имеют здесь перспектив 14.58
- В моем окружении нет евреев (544)
Я не знаю почему евреи уезжают 3.47
Евреи уезжают, т.к. поверили пропаганде 21.97
Евреи уезжают, т.к. не любят нашу страну 23.35