

6. Ильева И.А. Информационно-коммуникативная культура личности: Учебное пособие / И.А. Ильева, В.Н. Маркова, Т.С. Одинцова. – Губкин: ИП Уваров В.М., 2009.- 221 с.

Д.П. Денисов
Омская гуманитарная академия

ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ ПОСТРОЕНИЯ ГРАФОВ ЛИНЕЙНОЙ КОРРЕЛЯЦИИ

Корреляционные графы являются одним из наиболее наглядных способов отражения зависимостей между многомерными случайными величинами. В качестве оценки связей, в частности, используются коэффициенты линейной корреляции, удобно интерпретируемые на графе [3]. Вершинам графов соответствуют исследуемые объекты или их признаки, а ребрам - полученные значения коэффициентов связи.

Электронные таблицы MS Excel представляют пользователю богатый инструментарий для выполнения корреляционного анализа и построений [4,5], при этом знание алгоритмических особенностей обработки данных и геометрическая интерпретация расчетов полезны для понимания сущности исследуемых явлений.

Учитывая относительную легкость вычисления линейной корреляции, прием актуален на начальных этапах научного поиска [1]. Коэффициент r_{xy} можно рассчитать по формуле:

$$r_{x,y} = cov(X,Y) / \sigma_x / \sigma_y, (1)$$

где X, Y - пара признаков, $cov(X,Y)$ - выборочная ковариация:

$$cov(X,Y) = \sum_{i=1}^n x_i y_i / (n-1), (2)$$

σ_x, σ_y - среднеквадратические отклонения, в частности:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, (3)$$

и σ_y - аналогично.

Если величины X, Y центрированы и нормированы ($\sigma_x = \sigma_y = 1$; $\bar{x}, \bar{y} = 0$), r_{xy} отражает, исходя из формулы (1), их ковариацию, т.е. $r_{xy} = cov(X,Y)$.

В нормированном пространстве коэффициенту r_{xy} соответствует косинус угла между векторами, описывающими величины X, Y . В координатной форме $\cos\varphi$ (и, соответственно, r_{xy}) выражается через скалярное произведение векторов (числитель):

$$r_{xy} = \cos \varphi = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \sqrt{\sum_{i=1}^n (y_i)^2}}, \quad (4)$$

Если вектора одного направления, или противоположны, $|r_{xy}| = 1$, соответствующие параметры связаны функционально, по мере увеличения угла связь слабеет, и исчезает; если вектора ортогональны (их скалярное произведение, и соответственно, $r_{xy} = 0$). Таким образом, r_{xy} – нелинейная величина, варьирующая в интервале $-1 \div 1$, описываемая тригонометрической функцией $\cos \varphi$.

В таблицах MS Excel r_{xy} определяется статистической функцией **KORREL**, при этом достаточно указать массивы переменных. Алгоритм позволит выполнить расчеты, не вникая в их содержание - автоматизация нивелирует приемы центрирования и нормирования, весьма полезные в плане управления обработкой данных, как и для геометрической интерпретации результатов.

Если переменных несколько, целесообразно обозначить их, как $X_1, X_2 \dots X_j$, соответственно:

$$\mathbf{X}_j = (x_{1j}, x_{2j} \dots x_{ij}). \quad (5)$$

где x_{ij} – результат i -того наблюдения в эксперименте, j - номер изучаемого признака, \mathbf{X}_j - вектор, описывающий признак. Центрирование дат упрощает формулы, смещая начало отсчета в позицию средней арифметической исследуемого параметра, нормирование предполагает сопоставление физически неоднородных величин. Преобразование оправдано, если размерность шкалы и начало отсчета не влияют на сущность явлений. Для стандартизации результата i -го наблюдения j -того признака x_{ij} применяется формула:

$$X_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}} \quad (6)$$

где $\sigma_{x_j} \neq 0$.

Вычисление среднеквадратического отклонения σ_{x_j} (3) осуществляется с помощью функции **СТАНДОТКЛОН**.

Табличная форма, т.е. основа для построения графа – симметрическая матрица (R) размерности k . Матрица содержит коэффициенты корреляции r_{ij} между парами признаков ($i, j = 1, 2 \dots k$) элементы ее главной диагонали $r_{11} = r_{22} = \dots = r_{kk} = 1$ (автокорреляция с нулевым лагом):

$$R = \begin{array}{|c|c|c|c|} \hline & & r & r_k \\ \hline 1 & & r_{21} & 1 \\ \hline r & & 1 & \cdot \\ \hline 12 & & \cdot & 2 \\ \hline \cdot & & \cdot & \dots \\ \hline r & r & & \\ \hline 1k & 2k & & 1 \\ \hline \end{array} \quad (7)$$

Для вычисления элементов R достаточно рассчитать ковариацию преобразованных (стандартизированных) параметров (1,2,6).

Алгоритм **КОВАР** в MS Excel не учитывает степени свободы варьирующих величин. Поскольку в выборочном исследовании математическое ожидание признаков неизвестно, для расчетов следует использовать формулу (2).

На оценку связи влияет закон распределения переменных: корреляция адекватна, если рассеяние симметрично и описывается кривой Гаусса, в этом случае отклонение от центра $\pm 2\sigma$ обеспечит уровень вероятности $p \approx 0,05$; соответственно, $\pm 3\sigma$, $p \approx 0,01$. По обыкновению исследователь игнорирует правило, т.к. ограничивается выборками меньшего размера, сравнивая t –критерий Стьюдента (функция **СТЮДРАСПОБР**) для $p = 0,05$ с $t_{факт.}$:

$$t_{факт.} = r_{xy} \sqrt{\frac{n-2}{1-r_{xy}^2}} \quad (8).$$

Отметим, что значимость корреляции не доказывает факт взаимовлияния признаков, как и наличия причинно-следственных связи в эксперименте, а лишь констатирует особенности вариации - синхронный (или асинхронный) характер изменения величин. Для обоснования пропорций экспериментатору потребуются более сложные модели, и, соответственно, методы обработки данных (каноническая корреляция, регрессия, факторный анализ и др.).

Рассмотрим пример, иллюстрирующий особенности вычисления r_{ij} , возможности MS Excel и построение графов. В Таблице 1 отражены результаты обработки данных (r_{ij}) конфиденциального анкетирования учащихся различных специальностей (управление, коммерция, финансы) на предмет определения степени автоматизации процессов в условиях прохождения производственной практики.

Анкета отражала ряд количественных показателей:

- 1) число клиентов (x_1), сотрудников в совместном труде на АРМ за день;
- 2) общее число программ (x_2) применяемых на АРМ;
- 3) количество операций (x_3) в одном производственном цикле на АРМ;
- 4) средняя дневная продолжительность компьютерного труда (x_4), часов;
- 5) доля (x_5) автоматизации процессов, %.

Слабо формализуемые характеристики (тип собственности, масштаб деятельности предприятий) указывались в градации 0 и 1, всего обработано 46 анкет (Табл. 1). В эксперименте рассчитывались r_{ij} (элементы $R = \|r_{ij}\|$ представлены в виде столбца), а также определялась максимально возможная для данной выборки теснота корреляционных связей с использованием надстройки "Поиск решения".

Надстройки MS Excel – дополнительные файлы специализированного формата (*.xla; *.xlsx; *.xlsb, *.xlsm), содержащие макросы, расширяющие алгоритмические возможности прикладных задач.

Данные таблицы показывают, что между исследуемыми параметрами существует положительная и достоверная ($p = 0,05$) взаимосвязь. Наиболее тесно

в комплексе взаимодействий обусловлены параметры, определяемые x_2, x_3, x_4, x_5 .

Дифференциация расчетов по группам обнаруживает усиление связей, т.е. выявляет общие тенденции в направлении векторов для соответствующих подпространств. Для всей выборки значимые коэффициенты r_{ij} находятся в интервале $0,35 \div 0,45$ и описывают, следовательно, не более 20,3% вариации показателей; внутри групп отдельные связи близятся к уровню функциональных (в частности, подгруппа "Управление", $r_{45} = 0,90$).

В итоге, игнорируя специализации в опросе, мы теряем (в случае линейной корреляции) удобные проекции наблюдения исследуемых явлений и нивелируем качественные различия групп: "Поиск решения" позволит выбрать наиболее удобные сочетания в этом плане, Таблица 1.

Таблица 1

Взаимосвязь исследуемых параметров в эксперименте

Пере- менные	Коэффициент корреляции по группам, r_{ij}/r_{ij}^z									
	Все ответы		Управление		Коммерция		Финансы-1		Финансы-2	
x_1x_2	0,17	0,49	-0,14	-0,25	0,25	0,24	0,07	0,46	0,15	0,38
x_1x_3	0,07	0,65	-0,60	-0,88	0,00	-0,04	-0,02	0,17	0,37	0,81
x_1x_4	0,02	0,36	-0,44	-0,66	0,10	0,84	0,20	0,55	0,20	0,37
x_1x_5	0,22	0,29	-0,47	-0,63	0,60	0,75	0,28	0,38	0,43	0,34
x_2x_3	0,43	0,74	0,33	0,44	0,82	0,93	0,25	0,88	0,54	0,62
x_2x_4	0,45	0,85	0,85	0,78	0,03	0,26	0,66	0,72	0,27	0,68
x_2x_5	0,41	0,76	0,80	0,84	0,58	0,80	0,25	0,85	0,20	0,70
x_3x_4	0,22	0,65	0,63	0,89	0,17	0,00	-0,05	0,68	0,25	0,53
x_3x_5	0,35	0,54	0,82	0,85	0,27	0,57	0,33	0,81	0,21	0,73
x_4x_5	0,36	0,87	0,90	0,99	-0,19	0,75	0,40	0,95	0,63	0,75
Z / Z_{max}	0,27	0,62	0,60	0,72	0,30	0,52	0,25	0,65	0,32	0,59
n/n_z	46	15	6	5	12	7	12	7	15	10

Целевую функцию $Z \rightarrow \max$ определяем, как среднее значение модуля $|r_{ij}|$ при $i \neq j$, дифференцированно по подгруппам. В качестве ограничений указываем, в частности, пределы изменения σ_{x_j} , и допустимые отклонения средних значений x_j от центра рассеяния.

Стандартизация x_{ij} , и суммирование произведений аналогичны калибровке и взвешиванию регистрируемых событий: алгоритмическая структура функционирует подобно нейронной сети [2], эталону соответствует максимум функции Z в интервале от 0 до 1; "Поиск решения", как активный компонент системы расчетов, игнорирует сочетания, обнаруживающие низкий уровень корреляций. Чувствительность сети регулируется нижним пределом целевой функции Z , ограничениями для σ_{x_j} и центрирования x_j .

На Рис.1 отражены значимые корреляции r_{ij} , а также сопоставимые по величине коэффициенты r_{ij}^z для оптимизированной выборки на примере подгруппы "Финансы-2" из Таблицы 1. Оптимизация обнаруживает новые связи, на оценку значимости которых влияет количество ограничений Z_{max} .

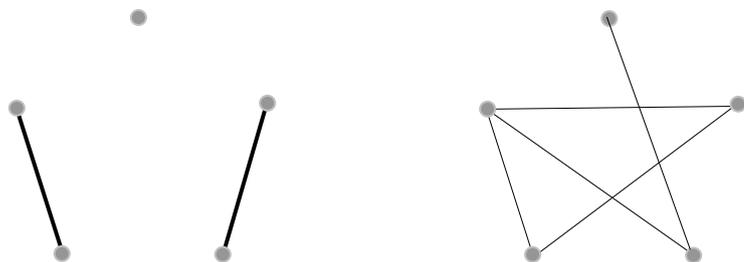


Рис. 1. Графы линейной корреляции показателей, а) при $n=15$; б) $n=10$.

Если алгоритм обнаруживает новый признак классификации объектов, т.е. наблюдения, соответствующие Z_{max} , действительно характеризуются некоторым особым свойством или образуют кластер относительно других дат (координаты рассеяния определяются пределами центрирования и ограничениями для σ_{xj}), постановка нулевой гипотезы обоснована.

В нашем примере слабо формализуемые признаки для $n = 46$ распределились в следующей пропорции (Рис. 2):

а) тип собственности (государственная, ИЧП, смешанная) - 6,5; 32,6; 60,9 (% от выборки);

б) масштаб деятельности предприятия (малое, среднее, крупное - 37,0; 45,7; 17,4 (%)) соответственно.

Для 15 наблюдений (Z_{max}) распределение составило 6,7; 20,0; 73,3 и 26,7; 60,0; 13,3 (%), т.е. в кластер попали различные по условиям автоматизации процессов объекты. Таким образом, некоторое усиление связей в эксперименте с указанными признаками не связано и, по-видимому, объясняется причинами случайного характера.

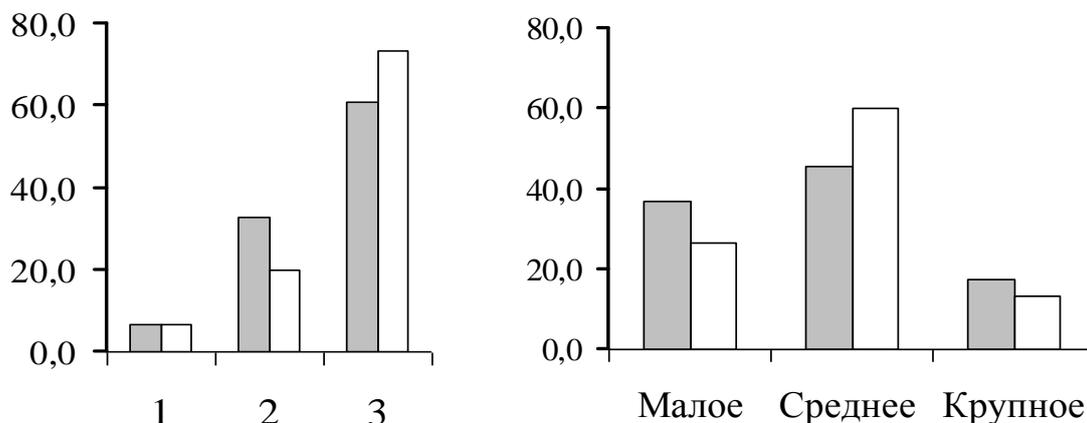


Рис. 2. Влияние оптимизации на распределение дат в выборке,

а) форма собственности: 1 – государственная, 2- частная, 3 – смешанная; б) размер предприятия; - все ответы, $n = 46$; - $n = 15$.

Поскольку любое научное наблюдение является элементом системы исследований, расчет коэффициентов линейной корреляции и построение графов с учетом возможностей оптимизации представляется полезным в плане корректировки структуры планируемых экспериментов, как и обоснования наиболее эффективных направлений в поиске.

Библиографический список:

1. Горшков М. Экологический мониторинг. Учеб. Пособие/ Горшков М. - Владивосток: Изд-во ТГЭУ, 2010.- С. 213-226.

2. Денисов, Д.П. Проектирование нейронной сети в табличном процессоре MS EXCEL / Д.П. Денисов, О.К. Касимова // Наука о человеке: гуманитарные исследования: научный журнал / гл. ред. А.Э. Еремеев. - Омск: НОУ ВПО "Омская гуманитарная академия", 2010. - № 1. - С.230-234.

3. Денисов Д.П. Эффективный алгоритм определения структуры интеллекта и его реализация в системе "Дедуктор"/ Д.П. Денисов. О.К. Касимова// Наука и общество: проблемы современных исследований: сб. научных статей: в 3 ч. – Ч.3. Проблемы современных исследований в гуманитарных науках / под ред. А.Э. Еремеева. – Омск: Изд-во НОУ ВПО "ОмГА", 2010. – С. 222 – 230.

4. Остапенко Р. И. Многомерный анализ данных для психологов: учебно-методическое пособие / Р. И. Остапенко. – Воронеж: ВГПУ, 2012. – 72 с.: ил.

5. Aravind H. A Simple Approach to Clustering in Excel/ Aravind H., Rajgopal S., Soman K P.// International Journal of Computer Applications, Vol. 11,№ .7, 2010.- P.19-25.

Б. Е. Батанасова

ГККП «Педагогический колледж имени Ж. Мусина

(с казахским языком обучения)»

г. Кокшетау, Республика Казахстан

ПЕДАГОГИЧЕСКАЯ ЭТИКА УЧИТЕЛЯ В ФОРМИРОВАНИИ ЛИЧНОСТИ УЧЕНИКА

Происходящие в мире социально - экономические и производственно- технические изменения предъявляют новые требования в систему образования. В гуманизации обучения важную роль имеет личность учителя. Компетентность учителей приобретает в последние годы все большую актуальность в связи с тем , что постоянно трансформируется сфера образования, появляются всевозможные разновидности техники и технологии, возрастает уровень запроса общества к учителям. Каким же должен быть хороший педагог?... Классическая педагогика считает: он должен любить детей, знать детскую психологию, владеть методикой преподавания. А современная педагогика требует от учителя педагогической этики. Педагогическая этика - это соблюдение педагогом