УДК 159.9.072

ОСОБЕННОСТИ ПРИМЕНЕНИЯ КОЭФФИЦИЕНТА ЛИНЕЙНОЙ КОРРЕЛЯЦИИ ПИРСОНА В ПСИХОЛОГИЧЕСКИХ, ПЕДАГОГИЧЕСКИХ И АКМЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ (НА ПРИМЕРЕ ОБУЧЕНИЯ СТУДЕНТОВ ГУМАНИТАРНЫХ СПЕЦИАЛЬНОСТЕЙ)

О. П. Малютина

Воронежский государственный университет

Поступила в редакцию 20 июня 2012 г.

Аннотация: *статья посвящена проблеме применения корреляционного анализа в психологических, педагогических и акмеологических исследованиях (на примере обучения студентов гуманитарных специальностей). Автор предлагает унифицировать формулу коэффициента линейной корреляции Пирсона с целью адаптации изложения основ корреляционного анализа к условиям образовательной среды с учетом наличного уровня математико-статистических компетенций обучающихся.*
Ключевые слова: *статистические методы, корреляционный анализ, коэффициент корреляции, студент.*

Abstract: *the article deals with the problem of correlation analysis in psychological, educational and acmeologic studies (for example, teaching students of humanitarian specialties). The author proposes a formula to unify the Pearson's linear correlation coefficient in order to adapt the presentation framework of correlation analysis to the conditions of the educational environment, taking into account inventory levels of mathematical and statistical skills of students.*
Key words: *statistical methods, correlation analysis, correlation coefficient, student.*

Одной из основных задач, встречающихся в психологических, педагогических и акмеологических исследованиях, является задача выявления линейной связи между изучаемыми признаками, успешно решаемой с помощью корреляционного анализа. Корреляционный анализ как таковой достаточно прост в применении, однако предполагает наличие определенных математико-статистических навыков обработки эмпирического материала у исследователя. Не секрет, что уровень математической подготовки нынешних абитуриентов, поступающих на гуманитарные специальности, а соответственно, и студентов 1–2-х курсов с каждым годом катастрофически снижается, что заставляет адаптировать изложение основ корреляционного анализа к условиям образовательной среды с учетом наличного уровня математико-статистических компетенций обучающихся.

Если базовые знания, умения и навыки студентов предшествующих лет позволяли представлять один и тот же коэффициент корреляции в разных вариациях, в зависимости от контекста его применения, то в настоящее время подобное является «недопустимой роскошью». Поэтому, основываясь на опыте чтения математических дисциплин студентам 1–2-х курсов гуманитарных специальностей (психологических, педагогических, социологических и др.), предлагаем унифицировать формулу коэффициента линейной корреляции Пирсона, чтобы облегчить им, во-первых, процесс освоения соответствующих статистических процедур и, во-вторых, применения в будущем полученных знаний и умений для решения профессиональных задач: избавить от необходимости принятия решения о том, какой именно вариацией коэффициента в том или ином случае можно воспользоваться.

В работах, освещающих результаты психологических, педагогических и акмеологических исследований, а также в известных учебных и учебно-методических пособиях широко применяются следующие варианты формулы:

1. $$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sqrt{s_x s_y}}$$ [1, с. 208],

где x_i – значения, принимаемые переменной x; y_i – значения, принимаемые переменной y; s_x, s_y – выборочные средние квадратические отклонения; n – число пар признаков, между которыми выявляется связь.

2. $$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}}$$

[там же].

3. $$r_{xy} = \frac{M(X-MX)(Y-MY)}{s_x s_y},$$

где MX, MY – математические ожидания соответственно случайных величин X, Y.

4. $$r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{X^2} - (\overline{x})^2}\sqrt{\overline{Y^2} - (\overline{y})^2}},$$

где $\overline{x}$ – средняя по x; $\overline{y}$ – средняя по y ;

5. $$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{(n-1)s_x s_y}$$ (последний коэффициент корреляции часто называют коэффициентом Браве-Пирсона) [2, с. 309; 3, с. 43].

6. $$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$ [1, с. 208; 4, с. 13].

7. $$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{s_x\, s_y}.$$

Ограничимся лишь семью вариациями. Как видим, спектр модификаций формулы коэффициента линейной корреляции Пирсона достаточно широк, а учитывая особенности восприятия студентами математических формул, – просто пугает. Предлагаем рассматривать на занятиях со студентами-гуманитариями только одну вариацию, а именно последнюю. Она удобна тем, что содержит в знаменателе s_x, s_y – выборочные средние квадратические отклонения. Они вычисляются всегда с целью определения «нормальности» распределения изучаемых признаков. Заметим, что корреляционный анализ может быть применен только в случае нормального распределения. В числителе седьмого варианта формулы коэффициента линейной корреляции Пирсона фигурирует сумма произведений отклонений каждого индивидуального значения признака от среднего, что, в свою очередь, является промежуточным результатом для вычисления среднего квадратического отклонения. Это значительно облегчает вычислительную задачу, минимизируя производимые студентами арифметические действия.

В рамках данной статьи продемонстрируем пошаговый переход от шестого варианта формулы к седьмому, иными словами,

от $$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

к $$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{s_x\, s_y}.$$

Проведем следующие несложные преобразования.

1. Умножим и разделим знаменатель на n, причем n, на которое делим, представим в виде $(\sqrt{n})^2$.

Получаем $$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\frac{n}{(\sqrt{n})^2}\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}.$$

2. Внесем под знак большого корня в знаменателе $(\sqrt{n})^2$, но уже в виде $\sqrt{n}\sqrt{n} = \sqrt{n\,n}$. Подчеркнем, что преобразованная формула примет вид:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n\sqrt{\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}.$$

3. Мы видим, что коэффициент $\frac{1}{n}$ не участвует в суммировании, поэтому без ограничения общности можем внести оба эти коэффициента по одному под знак каждой из сумм, получив следующую модификацию формулы:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}}}.$$

В силу того что $D_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $D_y = \frac{1}{n}\sum_{i=1}^{n}(y_i - y)^2$, заметим, что под знаком большого корня фигурируют не что иное, как формулы выборочных дисперсий. Однако обычно выборочные дисперсии определяются через выборочные средние квадратические отклонения, т.е. $D_x = s_x^2$, $D_y = s_y^2$. Таким образом, мы имеем:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n\sqrt{D_x D_y}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n\sqrt{s_x^2\ s_y^2}} =$$

$$= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n\ \mathrm{s_x}\ s_y}.$$

4. Поднимая *n* в числитель, окончательно получаем:

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{s_x\ s_y}.$$

Важно сказать, что переходы других вариантов формулы (с первого по пятый) в седьмой аналогичны рассмотренному выше.

Таким образом, оптимизация учебного процесса в вузе, в частности адаптация программного материала к уровню математической, а с 2004 г. (с возвращением комбинаторики, теории вероятностей и математической статистики в обязательный школьный курс) и статистической подготовки студентов-гуманитариев младших курсов, не может ограничиться рамками корреляционного анализа. В перспективе мы видим возможность минимизировать подобные сложности в преподавании и других видах статистического анализа – с целью более структурированного изложения учебных дисциплин «Математика в социально-гуманитарной сфере», «Математическая статистика», «Математические методы в психологии», «Экспериментальная психология», «Математические модели в социологии».

ЛИТЕРАТУРА

1. *Ермолаев О. Ю*. Математическая статистика для психологов : учебник / О. Ю. Ермолаев. – М. : Моск. психол.-соц. ин-т : Флинта, 2002. – 336 с.

2. *Годфруа Ж*. Что такое психология : в 2 т. / Ж. Годфруа. – М. : Мир, 1996. – Т. 2. – 376 с.

3. *Малютина О. П*. Математические методы в психологии : учеб.-метод. пособие / О. П. Малютина. – Воронеж : ЦНТИ, 2010. – 48 с.

4. *Харченко М. А*. Корреляционный анализ : учеб.-метод. пособие / М. А. Харченко. – Воронеж : ВГУ, 2008. – 31 с.

Воронежский государственный университет

Малютина О. П., кандидат физико-математических наук, доцент кафедры общей и социальной психологии

E-mail: maloxdub@land.ru

Тел.: 8-903-852-93-98

Voronezh State University

Malutina O. P., Candidate of Physical and Mathematical Sciences, Associate Professor of the General and Social Psychology Department

E-mail: maloxdub@land.ru

Tel.: 8-903-852-93-98