

Контрольные вопросы к практическому занятию № 10

(ответы записываются в тетрадь)

1. Виды связей между признаками, примеры.
2. Функциональная зависимость (определение).
3. Какие явления могут быть смоделированы функциональной зависимостью?
4. Корреляционная зависимость (определение).
5. Способы обнаружения корреляционной связи.
6. Оценка силы и направления связи по величине коэффициента корреляции.
7. Методика расчета коэффициента линейной корреляции.
8. Как оценивается достоверность коэффициента корреляции?
9. Как выполняется корреляционный анализ с помощью программы MS Excel?
10. Методика расчета рангового коэффициента корреляции.
11. Дайте определение термина «Регрессия».
12. Значение коэффициента регрессии и методика его расчета.

Теория Метод корреляции

При проведении исследования в биологии или медицине, как правило, регистрируются множество учетных признаков и возникает вопрос об их взаимном изменении, т.е. обнаружение зависимостей между ними. Выявление наличия таких взаимосвязей является одной из важнейших задач любой науки, в том числе и медицины.

Различают две формы количественных связей между явлениями или процессами: функциональную и корреляционную. Под ФУНКЦИОНАЛЬНОЙ понимают такую связь, при которой любому значению одного из признаков соответствует строго определенное значение другого. Например, зависимость площади круга от длины окружности в геометрии, или в физике путь, пройденный телом от времени. Наиболее известным видом функциональной зависимости является линейная, которая выражается математической формулой: $y = ax + b$.

В биологии и медицине установить функциональную зависимость, как правило, не удастся. Объекты этих исследований имеют большую изменчивость и зависят от огромного числа факторов, учесть которые просто невозможно. В этом случае определяется наличие КОРРЕЛЯЦИОННОЙ связи, при которой значению каждой средней величины одного признака соответствует несколько значений другого взаимосвязанного с ним признака.

Например: связь между ростом и массой тела человека. У группы людей с одинаковым ростом наблюдается различная масса тела, однако она варьирует в определенных пределах вокруг средней величины. Поэтому такую зависимость нужно оценивать с использованием понятия случайной величины с привлечением теории вероятности. Такую форму зависимостей называют «Корреляционной».

При поиске зависимости между признаками может быть обнаружена взаимосвязь, различная по направлению и силе:

- Прямая (при увеличении одного признака увеличивается второй);
- Обратная (при увеличении одного признака второй уменьшается).

Степень взаимосвязи признаков по силе (тесноте) принято обозначать как:

- Отсутствие;
- Слабая;
- Средняя;
- Сильная;
- Полная.

Способами выявления корреляционной взаимосвязи между признаками являются:

- Визуальные (таблицы и графики).
- Статистические (корреляция и регрессия).

Обнаружение корреляции между двумя признаками еще не говорит о существовании причинной связи между ними, а лишь указывает на

ВОЗМОЖНОСТЬ таковой или на наличие фактора, определяющего изменение обеих переменных совместно.

Приёмы визуализации данных позволяют обнаружить корреляционную зависимость лишь при небольшом числе наблюдений и только приблизительно. Для обнаружения

корреляционной взаимосвязи с помощью таблицы в ней располагают ранжированные вариационные ряды и затем определяют совместное изменение признаков. График более наглядно демонстрирует такую зависимость и позволяет оценить ее форму: линейная, параболическая, и др.

Наиболее точным способом обнаружения взаимосвязи между признаками является вычисление коэффициента корреляции. В зависимости от природы обрабатываемых данных применяются параметрические или непараметрические методы вычисления этого коэффициента.

При вычислении коэффициента корреляции исследователь получает возможность судить о силе связи (степени сопряженности) и ее направлении, а также с требуемой долей вероятности делать вывод о проявлении этой связи в генеральной совокупности. Чем больше коэффициент корреляции, тем с большей степенью уверенности можно говорить о наличии корреляционной зависимости между признаками. Если каждому заданному значению одного признака соответствуют близкие друг к другу, тесно расположенные около средней величины значения другого признака, то связь является более тесной. Когда эти значения сильно варьируют, связь менее тесная. Таким образом, мера корреляции указывает, насколько тесно связаны между собой параметры.

Коэффициент корреляции может принимать значения от -1 до +1. Направление обнаруженной взаимосвязи определяют по знаку коэффициента корреляции. При его положительном значении обнаруженная связь является прямой, при отрицательном – обратной. Теснота связи оценивается по модулю этого коэффициента. Условно выделяют следующие уровни корреляционной связи:

Таблица 1

$ r_{xy} $	Вид связи между признаками x и y
0	отсутствует
от 0 до 0,3	слабая
от 0,3 до 0,7	средняя
от 0,7 до 0,9	сильная
От 0,9 до 0,95	
1	Функциональная (полная)

Однако говорить о наличии корреляции можно только в тех случаях, когда она статистически значима ($p < 0,05$). Поэтому после вычисления коэффициента корреляции производится вычисление его ошибки репрезентативности и критерия достоверности.

Наиболее часто применяемыми в настоящее время методами обнаружения корреляции являются параметрический анализ по Пирсону и

непараметрический анализ по Спирмену. Этими методами проверяется нулевая гипотеза (H_0) об отсутствии связи между параметрами. Если такая гипотеза отклоняется при заданном уровне значимости (p), можно говорить о наличии взаимосвязи между параметрами.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ПО ПИРСОНУ используется при решении задачи исследования линейной связи двух нормально распределенных параметров. Кроме проверки на нормальность распределения каждого параметра, до проведения корреляционного анализа рекомендуется строить график в координатах оцениваемых параметров, чтобы визуально определить характер зависимости.

Коэффициент корреляции Пирсона (r_{xy}) или коэффициент линейной корреляции, был разработан в 90-х годах XIX века Карлом Пирсоном, Фрэнсисом Эджуортом и Рафаэлем Уэлдоном в Англии. Он рассчитывается по формуле:

$$r_{xy} = \frac{COV_{XY}}{\sigma_X \sigma_Y} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

где: r_{xy} – коэффициент линейной корреляции Пирсона;
 COV_{XY} – ковариация признаков X и Y ;
 σ_X – среднее квадратическое отклонение признака X ;
 σ_Y – среднее квадратическое отклонение признака Y ;
 \bar{X} – средняя арифметическая признака X ;
 \bar{Y} – средняя арифметическая признака Y .

В медицинской литературе встречается упрощенная запись этой формулы:

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$$

где: r_{xy} – коэффициент линейной корреляции Пирсона;
 d_x – отклонение каждой варианты признака x от средней этого признака: $d_x = x - M_x$;
 d_y – отклонение каждой варианты признака y от средней этого признака: $d_y = y - M_y$.

В программе Microsoft Excel значение коэффициента линейной корреляции Пирсона может быть вычислено функцией

=КОРРЕЛ(Диапазон ячеек 1-го ряда; Диапазон ячеек 2-го ряда).

Для прогнозирования уровня корреляции в генеральной совокупности определяют ошибку репрезентативности этого коэффициента m_r . Она вычисляется по формуле:

$$m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}}$$

где: m_r – ошибка репрезентативности коэффициента корреляции;
 r_{xy} – коэффициент линейной корреляции Пирсона;
 n – число парных вариант (пар наблюдений).

Актив.

Достоверность коэффициента линейной корреляции оценивается с помощью коэффициента Стьюдента (t_r), который вычисляется с использованием его ошибки:

$$t_r = \frac{r_{xy}}{m_r}$$

где: t_r – коэффициент достоверности Стьюдента;
 r_{xy} – коэффициент линейной корреляции Пирсона;
 m_r – ошибка репрезентативности коэффициента корреляции.

Если число парных вариантов $n > 30$, то при $t_r > 2$ связь считается достоверной при уровне значимости $p < 0,05$. Если число парных вариантов $n < 30$, то критическое значение $t_{r\text{Крит}}$ находят по таблице критических значений Стьюдента при степени свободы $df = n - 2$. В программе Excel это значение вычисляется функцией

=СТЮДРАСПОБР(Уровень значимости p ; Степени свободы df).

С целью уменьшения объема вычислений может применяться функция =КОРРЕЛ(Диапазон1; Диапазон2) или надстройка «Анализ данных» и ее модуль «Корреляционный анализ».

Отсутствие линейной корреляции еще не означает, что параметры полностью независимы. Связь между ними может быть нелинейной, или признаки, используемые в вычислениях, могут не подчиняться нормальному закону распределения. Поэтому, помимо вычисления коэффициента линейной корреляции, прибегают к использованию непараметрических коэффициентов корреляции. К ним относятся:

- Коэффициент ранговой корреляции Спирмена;
- Коэффициент ранговой корреляции Кендалла;
- Коэффициент корреляции знаков Фехнера;

Активна
Чтобы ак

- Коэффициент множественной ранговой корреляции (конкордации).

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ПО СПИРМЕНУ применяется для обнаружения взаимосвязи двух параметров, если распределение хотя бы одного из них отлично от нормального.

Каждому показателю x и y присваивается ранг. На основе полученных рангов рассчитываются их разности d . Затем вычисляется коэффициент корреляции (ρ) по формуле:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)},$$

где: ρ – коэффициент корреляции Спирмена;

d – разность рангов;

n – число парных вариантов (пар наблюдений).

Ошибка репрезентативности коэффициента корреляции Спирмена определяется по формуле:

$$m_\rho = \sqrt{\frac{1 - \rho^2}{n - 2}},$$

а коэффициент достоверности Стьюдента:

$$t_\rho = \frac{\rho}{m_\rho},$$

где: t_ρ – коэффициент достоверности Стьюдента;

ρ – коэффициент корреляции Спирмена;

m_ρ – ошибка репрезентативности коэффициента корреляции Спирмена.

Оценка коэффициента корреляции Спирмена и его достоверности выполняется так же, как и коэффициента линейной корреляции Пирсона.

Активна

Рекомендуемая литература:

1. Гельман В.Я. Медицинская информатика. Практикум. СПб: Питер, 2001. – 420 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика : Учебное пособие – 12-е изд., перераб. и доп. – М. : Юрайт, 2011. – 478 с. :ил.
3. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
4. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. -М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>