Краевая научно-практическая конференция учебно-исследовательских и проектных работ учащихся 6-11 классов

«Прикладные и фундаментальные вопросы математики»

Прикладные вопросы математики

Зависимость веса школьника от его роста

Шипулина Екатерина Юрьевна,

8 кл., МАОУ « Гимназия 8» г. Перми,

Коньшина Елена Викторовна,

учитель математики

Пермь. 2013.

Оглавление

Введение

Вероятностно-статистические представления, методы, с точки зрения современной науки, являются наиболее эффективными средствами познания и моделирования природных и социальных явлений, процессов, объектов и их характеристик.

Современное естествознание исходит из представлений, согласно которым все явления природы носят статистический характер.

Очень часто встречаются ситуации, когда две различные измеряемые величины связаны друг с другом довольно тесно, хотя и не предельно жестко. Например, толщина ствола и высота дерева, температура воздуха и концентрация в нем вредных веществ и т.д. В математической статистике такие величины называют случайными в том смысле, что при их измерении получают статистический набор значений, несколько отличающихся друг от друга. Многие связи по своей природе, то есть в реальной жизни, либо являются строго линейными, либо их можно привести к линейному виду.

Гипотеза – существует линейная связь между ростом и весом человека.

Предмет исследования - зависимость между ростом и весом человека.

Цель работы: выяснить тесноту связи между случайными величинами: ростом и весом учащихся 7 класса нашей школы.

Задачи:

- познакомиться с основными понятиями теории корреляции;
- провести сбор и обработку данных роста и веса учащихся 7 класса;
- составить уравнение связи между весом и ростом, определить тесноту данной связи.

Методы исследования:

1.Статистический сбор информации.

2.Метод наименьших квадратов.

3.Анализ и сравнение полученных результатов.

Глава I. Теоретическая часть

В математике есть специальный раздел - математическая статистика, посвященный математическим методам сбора, систематизации, обработки и интерпретации статистических данных, а также использование их для научных или практических выводов. Сам термин «статистика» произошёл от латинского слова status, что означает политическое состояние, и первоначально статистикой называли изучение государственных дел, а видных политических деятелей называли statists.Лишь позднее под словом «статистика» стали подразумевать числовые данные, на основе которых государственные деятели делали выводы, а еще позже его стали применять и для числовых данных вообще и постепенно пришли к современному значению.

1.1 Функциональная и корреляционная зависимость

В различных областях медицины, биологии, организации здравоохранения, социально-гигиенических и клинических исследованиях проводятся статистический анализ связей, изучение закономерностей и влияющих факторов.

Существуют два вида проявления количественных взаимосвязанностей между признаками (явлениями, факторами) – функциональные и корреляционные.

При функциональных зависимостях каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной. Такие зависимости наблюдаются в математике и физике. Различные измерительные приборы основаны на функциональной зависимости (высота ртутного столбика дает однозначный ответ о температуре).

Корреляционные или статистические связи, при которых численному значению одной переменной соответствует много значений другой переменной. Пример, между ростом и весом детей существует бесспорная зависимость, но это не значит, что определенному росту строго соответствует определенный вес. В силу участия в формировании веса многих других факторов, каждому значению роста соответствует несколько значений веса, которые могут быть выражены в виде распределения.

Функциональная связь имеет место по отношению к каждому конкретному наблюдению. Корреляционная проявляется в среднем для всей совокупности наблюдений. выявления взаимодействия факторов, определение силы и направленности. Практическое использование корреляционного анализа: выявление взаимодействия факторов, определение силы и направления влияния одних факторов на другие.

Форма связи может быть прямолинейной и криволинейной. Прямолинейная связь – равномерные изменения одного признака соответствуют равномерным изменениям второго признака при незначительных отклонениях. Криволинейная связь – равномерные изменения одного признака соответствуют неравномерным изменениям второго признака.

Направление связи может быть прямое (положительное) или обратное (отрицательное). Если с увеличением одного признака второй также увеличивается или с уменьшением одного другой тоже уменьшается, зависимость прямая, положительная. Если с увеличением одного признака другой уменьшается или с уменьшением первого признака второй увеличивается, зависимость обратная, отрицательная.

По силе связи зависимость может быть сильная (сильно выражена), средняя (умеренно выражена), слабая (слабо выражена).

1.2 Регрессионный анализ

Между переменными **X** и **Y** существует функциональная связь **y = f(x)** когда каждому значению аргумента **X** соответствует единственное значение аргумента **Y**. Регрессия — зависимость среднего значения какой-либо величины **Y** от другой величины **X**. Понятие регрессии в некотором смысле обобщает понятие функциональной зависимости **y = f(x)**. Только в случае регрессии одному и тому же значению **x** в различных случаях соответствуют различные значения **y**.

Регрессионный анализ заключается в определении аналитического выражения связи, в котором изменения одной величины (называемой зависимой или результативным признаком) обусловлено влиянием одной или нескольких независимых величин (факторов).

Метод наименьших квадратов (МНК)

Метод наименьших квадратов (МНК), применяется в теории ошибок, для разыскания одной или нескольких величин по результатам измерений, содержащих случайные ошибки. МНК также используется для приближенного представления заданной функции другими (более простыми) функциями и часто оказывается полезным для обработки наблюдений.

Если зависимость линейная, то с помощью метода наименьших квадратов можно составить уравнение прямой линии **y = ax+b** .

Для того чтобы определить параметры **a** и **b** необходимо знать отклонения (точки, находящиеся не на на прямой, а рядом). Суммарное отклонение будет равно:

$$\sum_{i=1}^{n}(Y_{i\exp}-Y_{iteor})^2,$$

где $\mathbf{Y_{iexp}}$ — экспериментальные точки (не обязательно лежащие на прямой), $\mathbf{Y_{iteor}}$ — теоретические точки (лежащие на прямой).

Чтобы все отклонения давали в суммарном отклонении положительные числа, надо возвести в квадрат эти отклонения:

$$\Delta=\sum_{i-1}^{n}(Y_{i\exp}-Y_{iteor})^2=\sum_{i=1}^{n}(Y_{i\exp}-(ax+b))^2=\sum_{i=1}^{n}(Y_i-(ax_i+b))^2,$$

где **Δ** — суммарное квадратичное отклонение, которое зависит от параметров **a** и **b,** $\mathbf{Y_i}$ — экспериментальные значения **Y**, $\mathbf{ax_i + b}$ — теоретические значения **Y**.

Лучшими параметрами **a** и **b** являются такие, которые минимизируют **Δ**, следовательно, среди бесконечного множества прямых, которых дает прямая **y = ax + b**, наилучшей является прямая с такими значениями параметров **a** и **b**, для которых **Δ(a, b)** принимает минимальное значение. Для линейной зависимости значения a и b находятся из решения системы двух уравнений:

$$\begin{cases} a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \end{cases}$$

1.3Корреляционный анализ

Корреляционный анализ - метод, позволяющий обнаружить зависимость между несколькими случайными величинами.

Коэффициент корреляции обозначается r. Он рассчитывается следующим образом:

$r = \frac{\overline{xy} - \bar{x} \times \bar{y}}{s_x \times s_y}$, где

$\overline{xy} = \frac{1}{n}\sum_{i=1}^{m}(x_i \times y_i)$ (произведения случайных величин)

$\bar{x} = \frac{1}{n}\sum x_i$ $\bar{y} = \frac{1}{n}\sum y_i$ (среднее арифметическое значение случайных величин)

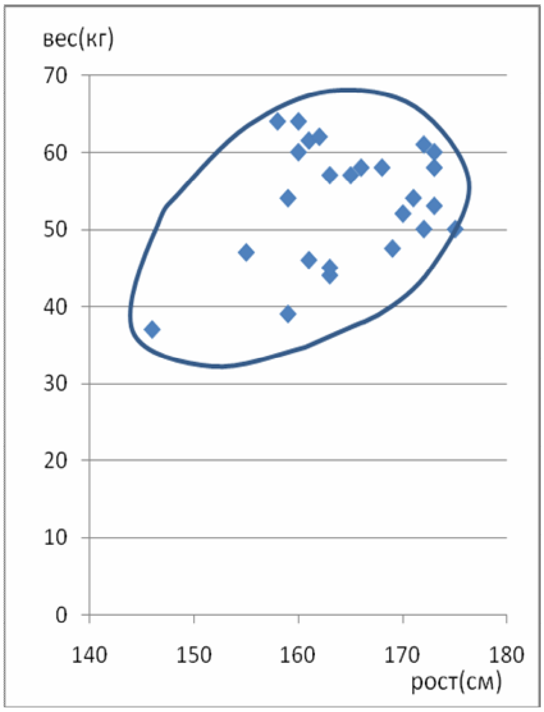
$s_x = \sqrt{\overline{x^2} - \bar{x}^2}$, $s_y = \sqrt{\overline{y^2} - \bar{y}^2}$ (средние квадратичные отклонения).

Коэффициент корреляции r изменяется в пределах от -1 до 1. В данном случае это линейный коэффициент корреляции, он показывает линейную взаимосвязь между x и y : r равен 1 (или -1), если связь линейна (связь 100%-ная), r равен 0 (связи не обнаружено).

Глава II. Определение тесноты связи между случайными величинами: ростом и весом учащихся

В качестве предположительно связанных случайных величин нами были выбраны рост и вес учащихся нашего класса. Наличие связи между ростом и весом представляется заранее очевидным. Из бытовых наблюдений можно предположить , что у высоких людей вес больше, чем у невысоких, и нужно найти численные показатели такой связи.

Нами были созданы данные по парам «рост- вес» 7 класса учащихся нашей школы (n= 25). Если отложить эти значения на плоскости, то каждой паре соответствует одна точка, а все точки составляют слегка наклонное облако точек. Это означает, что чем больше рост, тем больше вес (хотя конечно, существуют и такие пары, которые сильно выпадают из общей тенденции- такова статистика).



Расположив в два столбца полученные пары значений, вычислим далее их квадраты и произведения, которые требуются для вычисления коэффициента корреляции (см.приложение 1).

Поделим суммы во 2-ом и 3-м столбцах на n , получим средние арифметические значения роста и веса соответственно. Рассчитаем коэффициент корреляции:

$$r = \frac{\overline{xy} - \bar{x} \times \bar{y}}{s_x \times s_y} = \frac{8835{,}16 - 164{,}68 \times 53{,}56}{6{,}81011 \times 7{,}455629} \approx 0{,}30$$

Полученное значение r=0,30 попадает в интервал связи слабой силы (т.к. сильной обычно считают связь, начиная с **r= 0,7).** Таким образом, можно сделать вывод, что сила связи между случайными величинами- ростом и весом человека в подростковом возрасте довольно незначительная , это объясняется тем, что в возрасте 13-14ти лет подросток растет, его рост и вес в этом возрасте постоянно меняются, и точную связь установить невозможно.

Тем самым выдвинутая нами гипотеза подтвердилась: существует определенная связь между случайными величинами: ростом и весом человека.

Подставив все полученные нами экспериментальные значения в формулы, получим окончательно для нашей задачи:

$$\begin{cases} a \times 4117 + b \times 25 = 1339 \\ a \times 679147 + b \times 4117 = 220879 \end{cases}$$

Решая систему методом подстановки, получим $a \approx 0{,}32$, $b \approx 0{,}9$. Уравнение зависимости веса от роста : у= 0,32х+0,9 (у - вес х – рост). График зависимости представлен в приложении 2.

Этой формулой могут воспользоваться те, кто хочет определить соответствует ли его рост и вес среднестатическому параметру ученика 7 класса.

Заключение

В ходе работы корреляционная связь между весом и ростом была исследована в следующей последовательности:

1. Выявление связи между весом и ростом.
2. Описание связи в табличной, графической и аналитической форме.
3. Измерение тесноты связи.

По исходным данным было построено корреляционное поле, и установлен факт наличия линейной связи между весом и ростом. На основе анализа графически представлена корреляционная модель $y = 0{,}32x + 0{,}9$.

Мы выяснили тесноту связи между весом и ростом учащихся нашего класса, рассчитав коэффициент корреляции ($r=0{,}30$) . Связь между весом и ростом получилась слабая, это объясняется тем, что в нашем возрасте организм растет. Можно предположить, что если собрать статистические данные по парам «рост- вес» взрослых людей, то формулы получатся другие. Таким образом, наша гипотеза подтвердилась: между весом и ростом существует зависимость, которую можно описать уравнением линейной функции.

Список литературы

1. Коршунова Н., Плясунов В. Математика в экономике. И: Вита-Пресс, 1996.
2. Шалабанов А.К., Роганов Д.А.Эконометрика: Учебно-методическое пособие.Казань: ТИСБИ, 2008.

Приложение 1

Расчетная таблица для определения параметров корреляционной модели вида y=ax+в

№	X(рост)	Y(вес)	xy	x^2	y^2
1	146	37	5402	21316	1369
2	155	47	7285	24025	2209
3	158	64	10112	24964	4096
4	159	39	6201	25281	1521
5	159	54	8586	25281	2916
6	160	60	9600	25600	3600
7	160	64	10240	25600	4096
8	161	46	7406	25921	2116
9	161	61,5	9901,5	25921	3782,25
10	162	62	10044	26244	3844
11	163	44	7172	26569	1936
12	163	45	7335	26569	2025
13	163	57	9291	26569	3249
14	165	57	9405	27225	3249
15	166	58	9628	27556	3364
16	168	58	9744	28224	3364
17	169	47,5	8027,5	28561	2256,25
18	170	52	8840	28900	2704
19	171	54	9234	29241	2916
20	172	50	8600	29584	2500
21	172	61	10492	29584	3721
22	173	53	9169	29929	2809
23	173	58	10034	29929	3364
24	173	60	10380	29929	3600
25	175	50	8750	30625	2500
сумма	4117	1339	220879	679147	73106,5
Среднее арифметическое	164,68	53,56	8835,16	27165,88	2924,26
s	6,81011	7,455629			
s^2	46,3776	55,5864			

Приложение 2

График корреляционной модели