

# Элементы корреляционного и регрессионного анализа

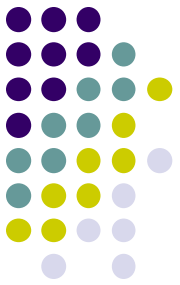
---

**Исследование взаимосвязи  
признаков  
Понятие корреляции**

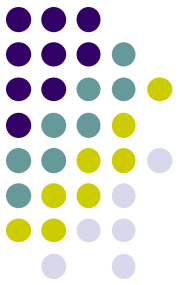




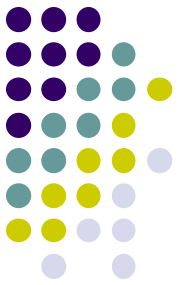
- *Корреляция (корреляционная связь) – это согласованное изменение двух признаков, отражающее тот факт, что изменчивость одного признака находится в соответствии с изменчивостью другого.*
- Если при изменении одной величины изменяется другая, то между показателями этих величин будет наблюдаться корреляция.



Корреляция является **отрицательной**,  
если с увеличением переменной  $X$   
переменная  $Y$  имеет в среднем тенденцию  
к уменьшению.



- Корреляция является положительной, если увеличение одной переменной связано с увеличением другой переменной (чем выше личностная тревожность, тем выше риск заболеть язвой желудка).



- Корреляция, при которой отсутствуют связи между переменными, является нулевой (связь между ростом учеников и их успеваемостью).
- Корреляции также могут быть линейными и нелинейными. Если с увеличением или уменьшением одной переменной вторая переменная в среднем также либо растет, либо убывает, то связь линейна.



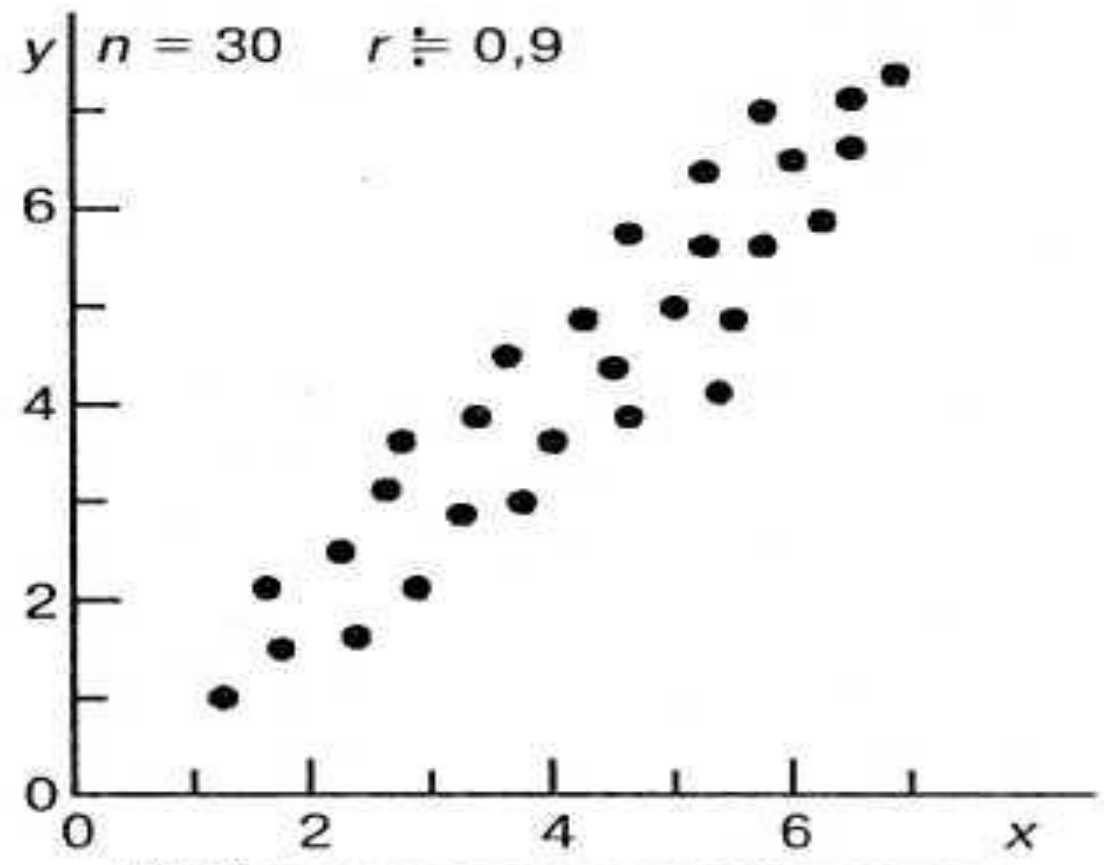
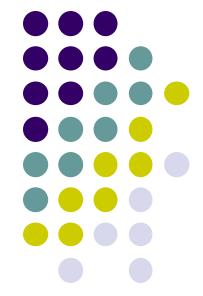
- Наглядное представление о характере корреляционной связи дает диаграмма рассеивания.
- **Диаграмма рассеивания** (точечная диаграмма) используется для наглядного отображения совместного распределения двух переменных.

Она позволяет визуально оценить степень связи между изучаемыми признаками.

- Если эллипс минимальной площади, охватывающий все точки на диаграмме, имеет достаточно вытянутую форму, это свидетельствует о наличии связи между случайными величинами  $X$ ,  $Y$ .

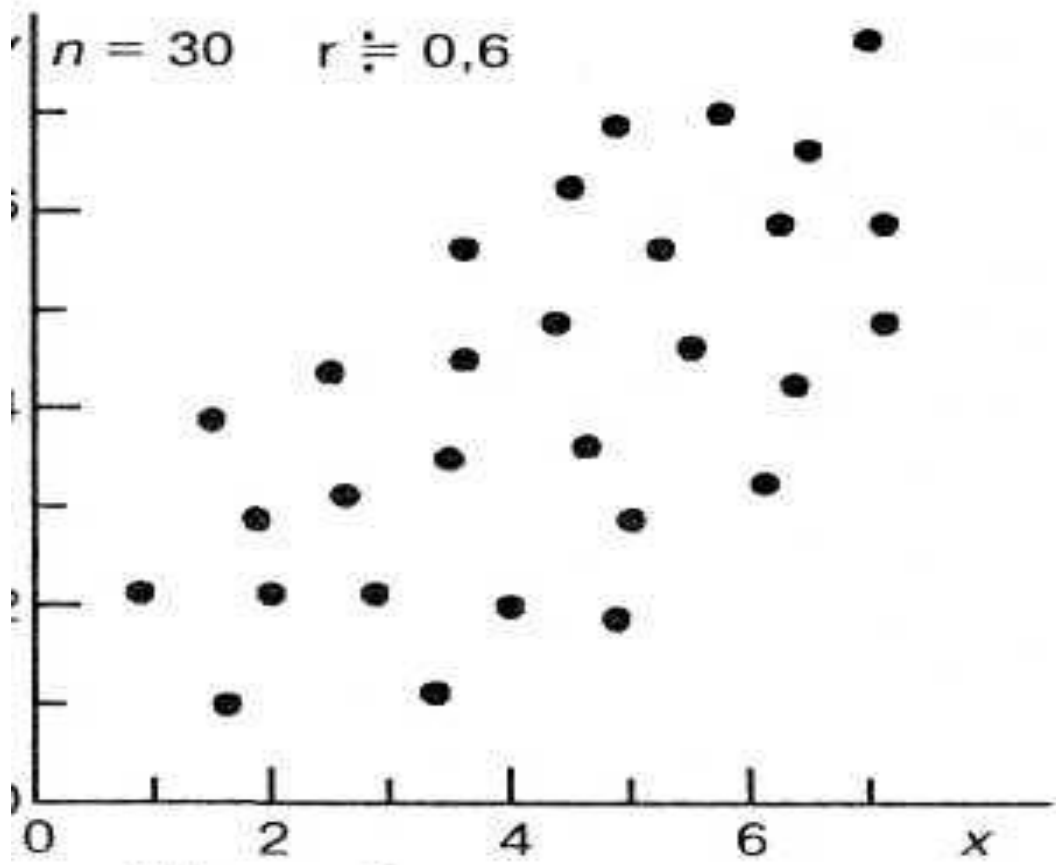
Чем более вытянутая форма у эллипса, тем более выражена связь между признаками.

- Если большая диагональ эллипса образует с осью абсцисс острый угол, то связь прямая;
- если же угол между большей осью эллипса и осью абсцисс тупой, тогда связь между признаками обратная, т.е. чем больше значение признака  $X$ , тем меньше значение признака  $Y$ .

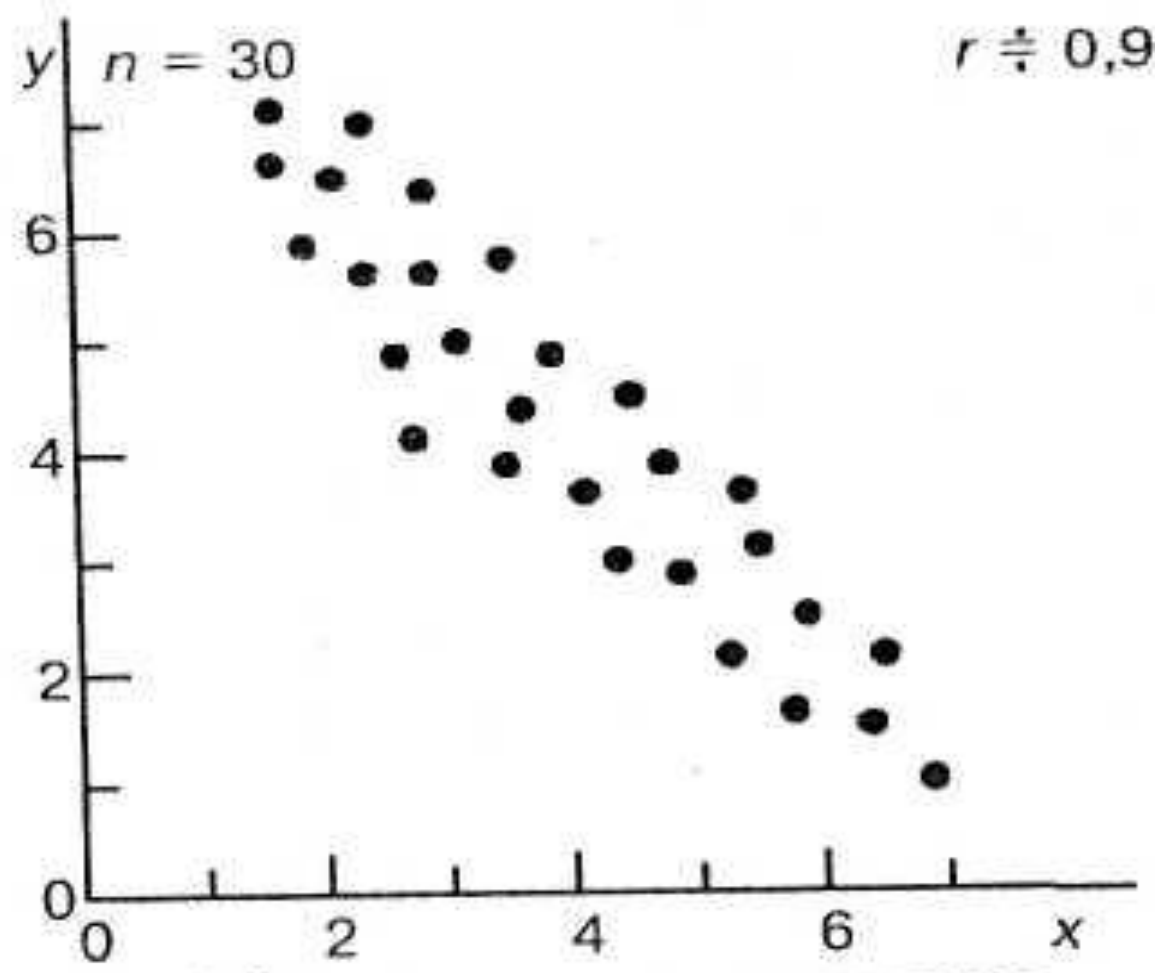


**б) Положительная корреляция**

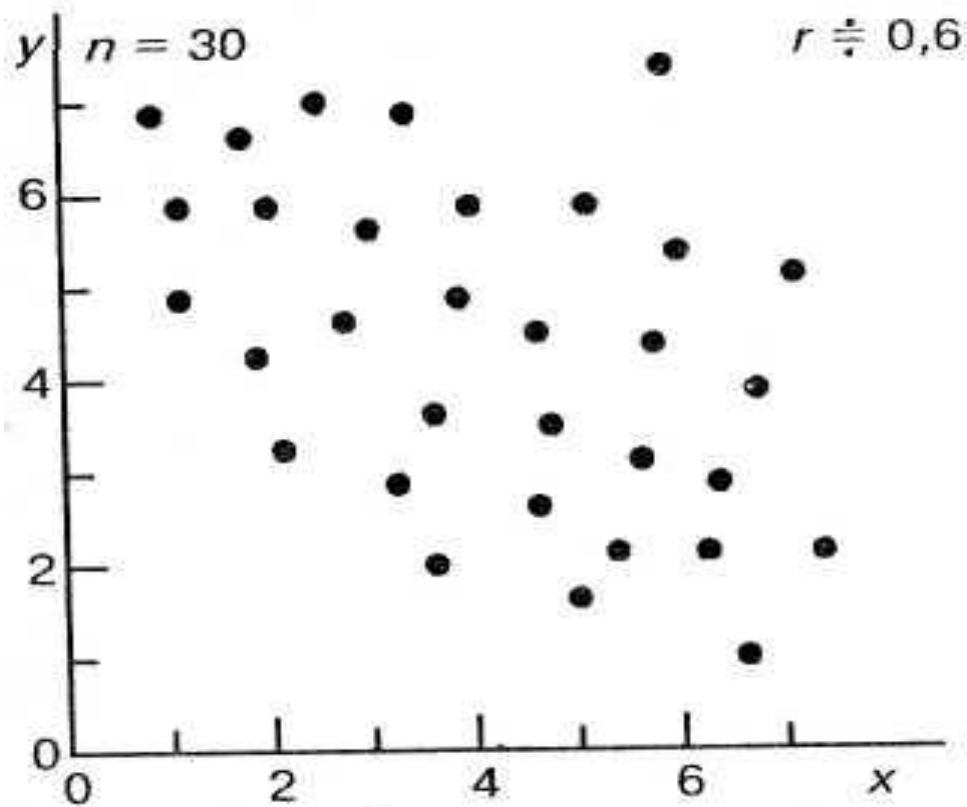




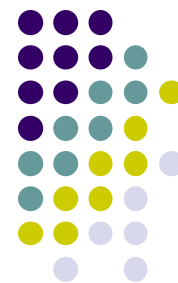
**в) Может быть положительная корреляция**

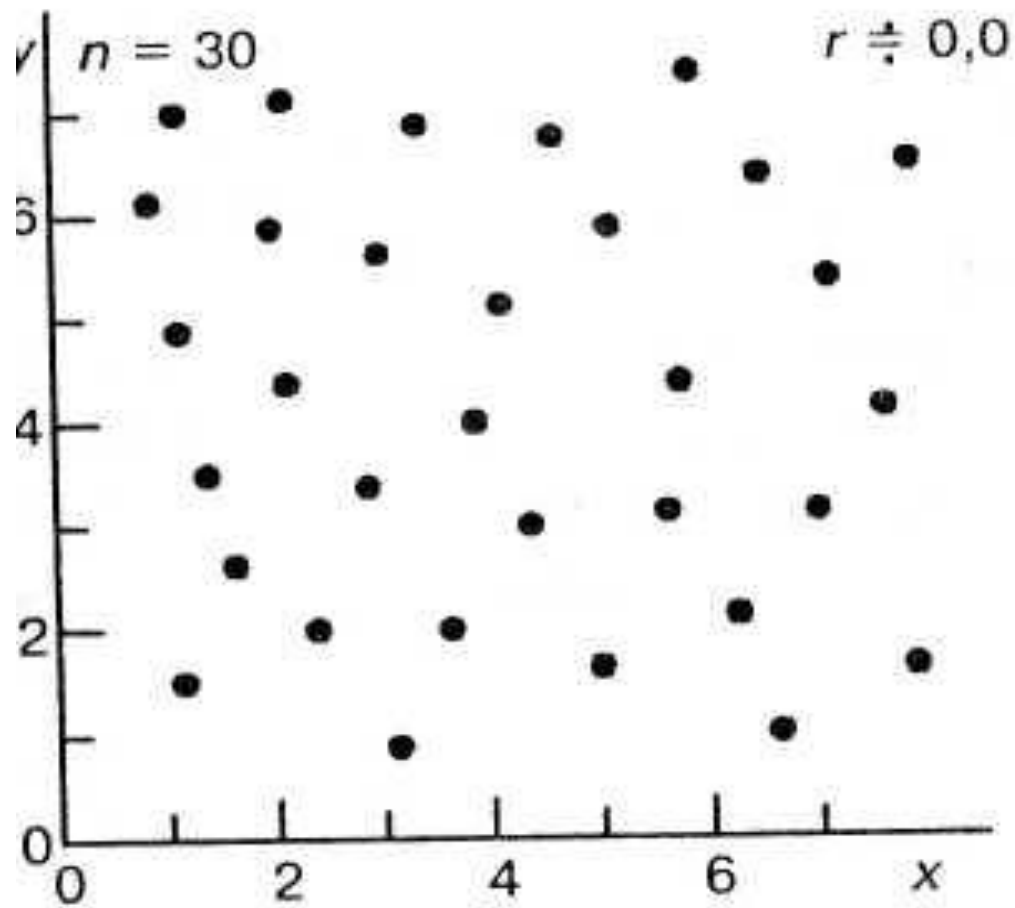


**д) Отрицательная корреляция**

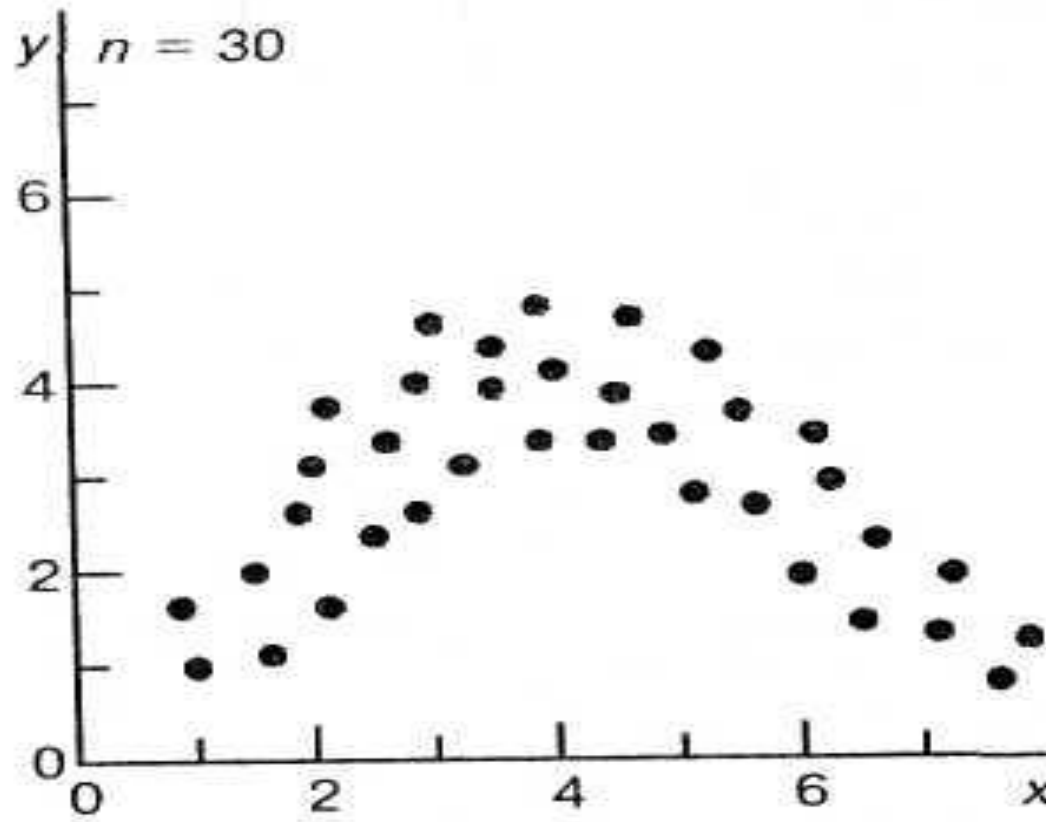


**е) Может быть отрицательная корреляция**

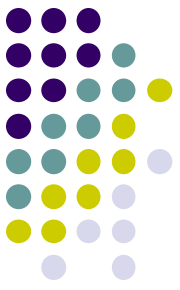




г) Нет корреляции

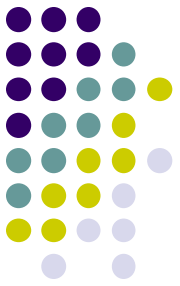


**ж) Нет линейной, но возможна криволинейная корреляция**



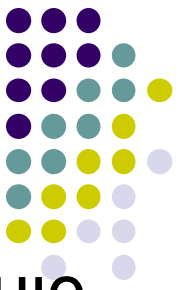
- В качестве числовой характеристики вероятностной связи величин используются **коэффициенты корреляции**.
- Степень, сила или теснота корреляционной связи определяются по величине коэффициента корреляции.
- Значения коэффициента могут находиться в диапазоне от  $-1$  до  $+1$ .
- **Сила связи** не зависит от ее направленности и определяется по **абсолютному значению** коэффициента корреляции.
- Показателем направления связи является **знак** коэффициента корреляции.

# Общая классификация корреляционных связей:



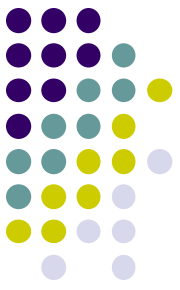
- 1) сильная, или тесная -  $|r| \geq 0,7$ ;
- 2) средняя -  $0,5 \leq |r| < 0,7$  ;
- 3) умеренная -  $0,3 \leq |r| < 0,5$  ;
- 4) слабая -  $0,2 \leq |r| < 0,3$  ;
- 5) очень слабая  $|r| < 0,2$ .

# ***Частная классификация*** **корреляционных связей:**



- 1) высокая значимая корреляция - при  $r$ , уровню статистической значимости  $P \leq 0,01$ ;
- 2) значимая корреляция - при  $r$ , соответствующем уровню статистической значимости  $P \leq 0,05$  ;
- 3) тенденция достоверной связи - при  $r$  , соответствующем уровню статистической значимости  $P \leq 0,10$  ;
- 4) незначимая корреляция - при  $r$ , не достигающем уровня статистической значимости.





- Коэффициент корреляции  $r_{xy}$ -Пирсона характеризует наличие только линейной связи между признаками.
- Формула для подсчета коэффициента корреляции Пирсона:

$$r_{xy} = \frac{n \cdot \sum (x_i \cdot y_i) - \left( \sum x_i \cdot \sum y_i \right)}{\sqrt{\left[ n \cdot \sum x_i^2 - \left( \sum x_i \right)^2 \right] \cdot \left[ n \cdot \sum y_i^2 - \left( \sum y_i \right)^2 \right]}}$$

где  $x_i$  - значения, принимаемые переменной  $X$ ,  
 $y_i$  - значения, принимаемые переменной  $Y$ ;  
 $n$  – объем выборок.

# Условия применения коэффициента корреляции Пирсона:

1. Сравнимые переменные должны быть получены в интервальной шкале или шкале отношений.
2. Распределения переменных  $X$  и  $Y$  должны быть близки к нормальному.
3. Число варьирующих признаков в сравниваемых переменных  $X$  и  $Y$  должно быть одинаковым.
4. Таблицы уровней значимости для коэффициента корреляции Пирсона (таблица 20 Приложения) рассчитаны от  $n = 5$  до  $n = 1000$ .

5. Оценка уровня значимости по таблицам осуществляется при числе степеней свободы  $k = n - 2$ .
6. При расчете коэффициента корреляции нельзя произвольно переставлять элементы в коррелируемых столбцах.