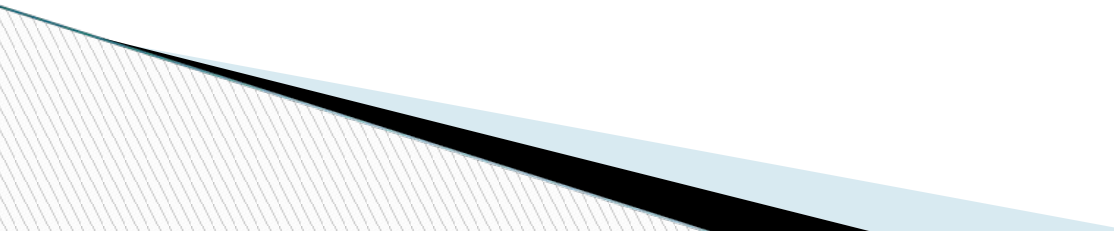


# Математические методы оценки взаимосвязи

или «Как посчитать  
корреляцию?»»



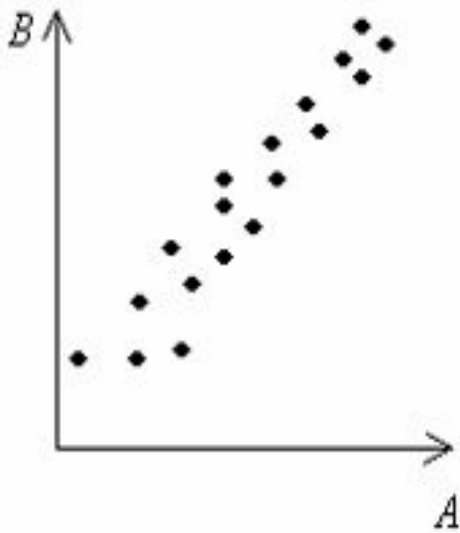
# Корреляционная связь -

это согласованное изменение двух признаков, отражающее, что изменчивость одного признака находится в соответствии с изменчивостью другого

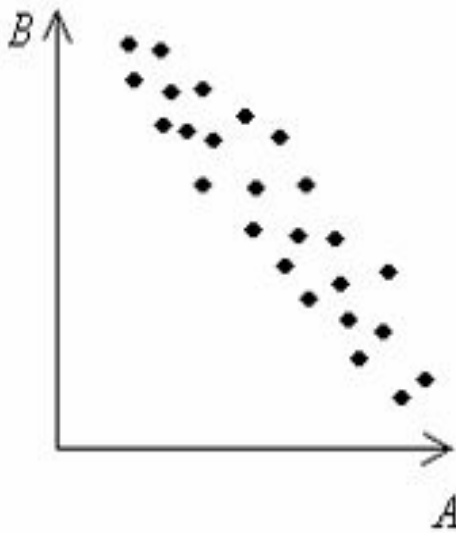
*«Когда изменяется X, то Y тоже меняется»*

Не путать с причинно-следственной связью!

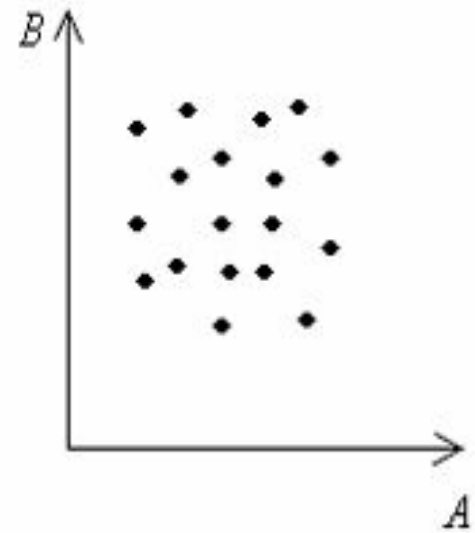
# Виды корреляционной связи



Линейная,  
положительная



Линейная,  
отрицательная



Корреляция  
отсутствует

# Сила корреляционной связи

- Выражается через значение **коэффициента корреляции (R)**
- Бывает:

$R(a,b) > 0,70$  - **сильная** связь между a и b

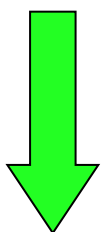
$0,5 < R < 0,7$  - **средняя** сила связи

$0,3 < R < 0,5$  - **умеренная**

$0,2 < R < 0,3$  - **слабая**

$R < 0,2$  - **очень слабая**

# Статистическая значимость корреляции



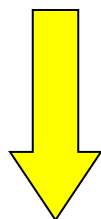
R при  $\alpha < 0,01$

Высоко значимая корреляция



R при  $0,01 < \alpha < 0,05$

Значимая корреляция



R при  $0,05 < \alpha < 0,1$

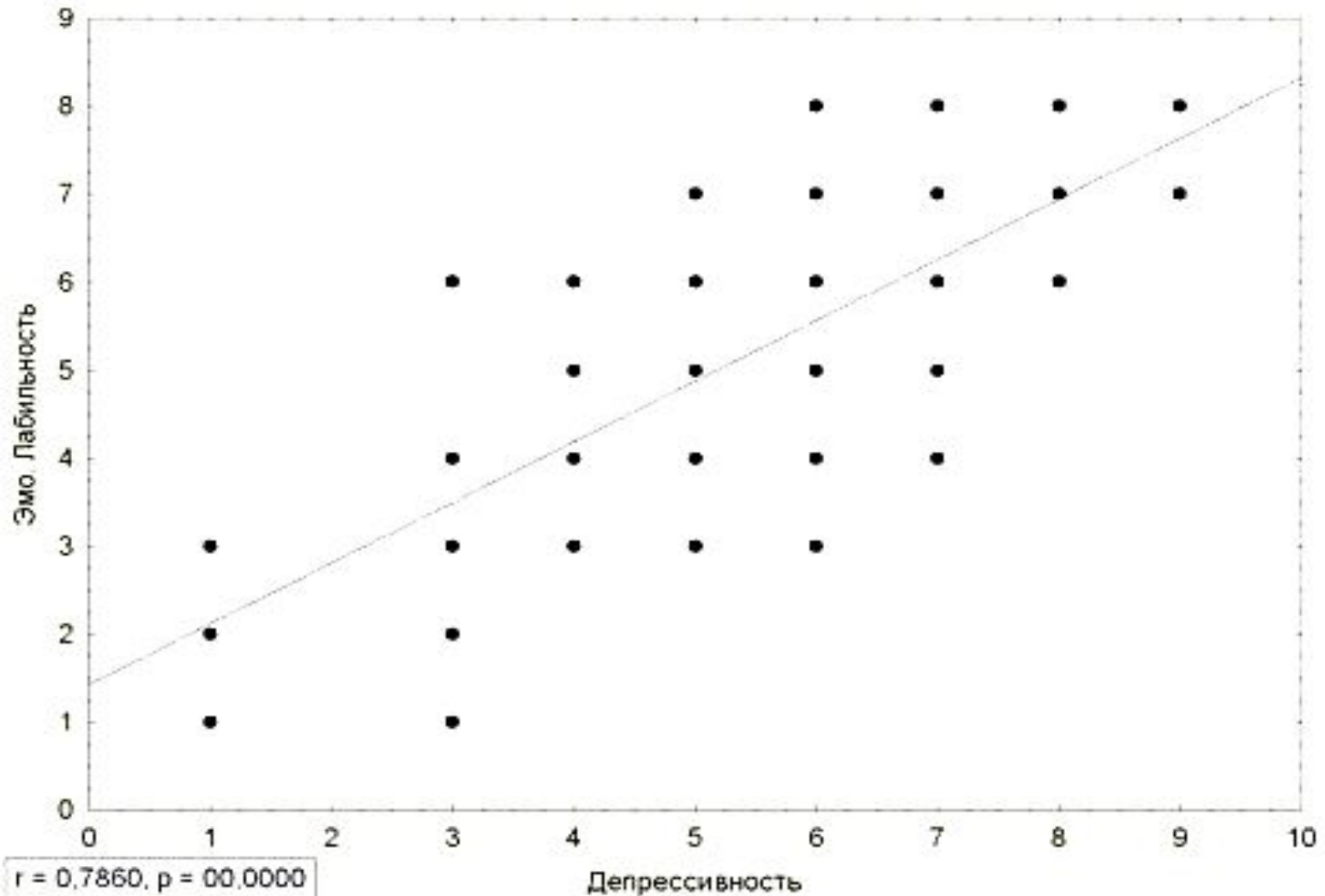
Тенденция достоверной связи



R при  $\alpha > 0,1$

Незначимая

# Пример: связь эмоциональной лабильности и депрессивности



# Коэффициенты корреляции

- для параметрических шкал - коэффициент  **$R_{x,y}$**  Пирсона
- для данных ранговой и параметрических шкал -  **$R_s$**  Спирмена
- для данных ранговых шкал —  $\tau$  (тау) критерий Кендалла
- для дихотомических шкал -  $\phi$  коэффициент ассоциации Пирсона

# Коэффициент корреляции $R_{x,y}$ Пирсона

$$R_{x,y} = \frac{\sum (x_i - \bar{X}) * (y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 * \sum (y_i - \bar{Y})^2}}$$

где  $x_i$  и  $y_i$  — показатели под номером  $i$ ,  
 $\bar{X}$  и  $\bar{Y}$  — средние

**Ограничения:** параметрические данные; нормальное распределение в обеих выборках;  $5 > n_x, n_y > 5000$ ;  $n_x = n_y$

**Возможности:** определяет силу и направление корреляционной связи между двумя признаками (измеренными в одной и той же группе или между двумя рядами значений, полученных в двух группах)



# Коэффициент корреляции $R_s$ Спирмена

$$r = 1 - \frac{6 * \sum d^2}{(n-1) * n * (n+1)}$$

$n$  – количество пар рангов,  
 $d$  – разность между рангами по  $X$  и по  $Y$

**Возможности:** измеряет силу и направление корреляционной связи; работает со всеми количественными данными; непараметрический.

**Ограничения:**  $5 < n < 40$  (таблицы); при увеличении  $n$  становится менее мощным; требует поправок для связанных рангов

# Алгоритм расчета $R_s$ Спирмена

1. В таблице 1,2 столбцы заполнить значениями  $X$  и  $Y$
2. В столбце 3 присвоить ранги значениям  $X$  по Правилам ранжирования. Аналогично для значений  $Y$  (в столбце 4)
3. **Разности** между каждой парой рангов (по строка) занести в 5-й столбец как  $d$
4. Возвести каждую разность **в квадрат**, записать в 6-й столбец как  $d^2$
5. Посчитать сумму всех  $d^2$ , записать ее как  $\sum(d^2)=$

# Алгоритм расчета $R_s$ Спирмена

6. Считать значение коэффициента:

$$r = 1 - \frac{6 * \sum d^2}{(n-1) * n * (n+1)}$$

7. При наличии связанных рангов

в числитель дроби сделать поправки **+ $T_x$**  и **+ $T_y$** :

$$T_x = \frac{\sum (a_x^3 - a_x)}{12}$$

$$T_y = \frac{\sum (a_y^3 - a_y)}{12}$$

где  $a_x$  и  $a_y$  — объем каждой группы одинаковых рангов в соответствующем ранговом ряду

# Алгоритм расчета $R_s$ Спирмена

8. Определить по таблицам критические значения; сопоставить с  $R_s$
9. Если  $R_s$  попадает в интервал  $p < 0,01$  признать имеющуюся корреляцию статистически высоко значимой
10. Сделать статистический вывод

# The End

Спасибо за внимание!

