

## **Лекция 4. Корреляция и ковариация**

# Ковариация

$$\text{cov}(X, Y) = M[(X - M(X))(Y - M(Y))] = M(XY) - M(X)M(Y)$$

## Свойства ковариации

1. Если  $X$  и  $Y$  независимы, то  $\text{cov}(X, Y) = 0$  (*обратное неверно!*)
2. Симметричность:  $\text{cov}(X, Y) = \text{cov}(Y, X)$
3.  $\text{cov}(X, X) = D(X)$
4. Если  $\alpha, \beta = \text{const}$ , то  $\text{cov}(X + \alpha, Y + \beta) = \text{cov}(X, Y)$
5. Для двух выборок  $\text{cov}(X_{(n)}, Y_{(n)}) = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})$

## Неравенство Коши-Буняковского

$$\text{cov}^2(X, Y) \leq D(X)D(Y)$$

**НО! Ковариация – не безразмерная величина**

Функции MS Excel: **КОВАРИАЦИЯ.В**, **КОВАРИАЦИЯ.Г**, **КОВАР**

# Ковариация: функции MS Excel

| Функция                 | Описание   |
|-------------------------|--|
| КОВАРИАЦИЯ . В          | Оценка ковариации для выборок<br>$\frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})$   |
| КОВАР<br>КОВАРИАЦИЯ . Г | Оценка ковариации для генеральных совокупностей<br>$\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})$<br><i>КОВАР - Устаревшая функция из MS Excel 2007</i> |

Сравнить со смещённой и несмещённой оценками дисперсии (см. лекцию 2)

# Корреляция и коэффициент корреляции

Неравенство Коши-Буняковского

$$\text{cov}^2(X, Y) \leq D(X)D(Y)$$

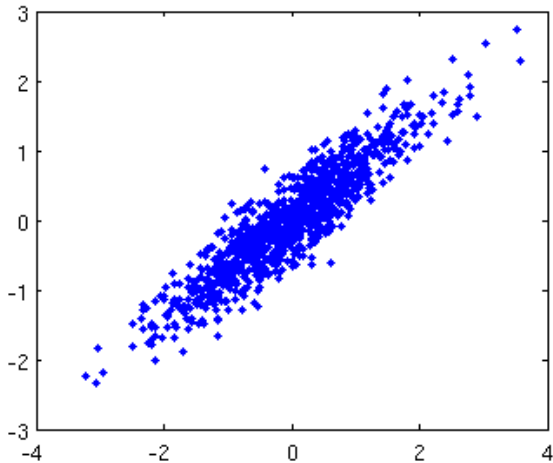
Линейный коэффициент корреляции

$$r_{XY} = \frac{\text{cov}_{XY}}{\sigma_x \sigma_y} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

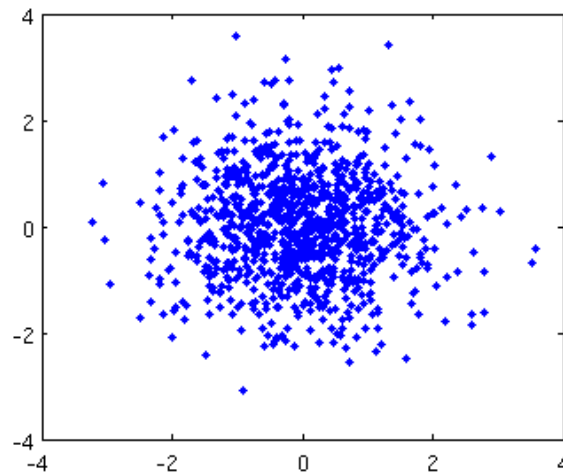
Функции MS Excel: **КОРРЕЛ**

$$-1 \leq r_{XY} \leq 1$$

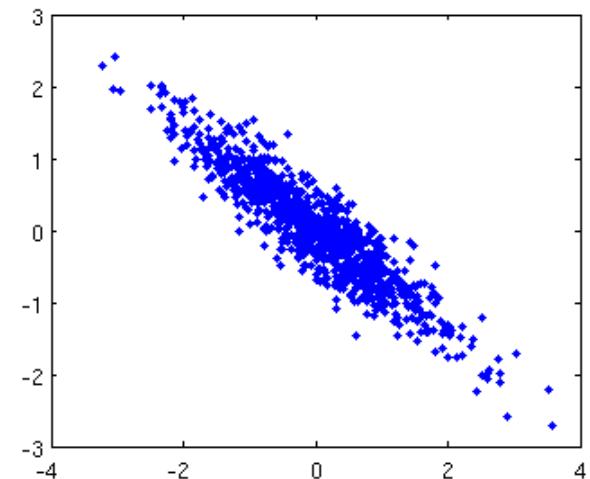
**Положительная  
корреляция**



**Отсутствие  
корреляции**



**Отрицательная  
корреляция**



# Корреляция и коэффициент корреляции

Статистическая значимость коэффициента корреляции

$$t_{emp} = r_{emp} \frac{\sqrt{n-2}}{\sqrt{1-r_{emp}^2}} \sim t_{\alpha, n-2}$$

Доверительный интервал коэффициента корреляции

Преобразование Фишера

$$z = 0.5 \ln \left( \frac{1+r}{1-r} \right)$$

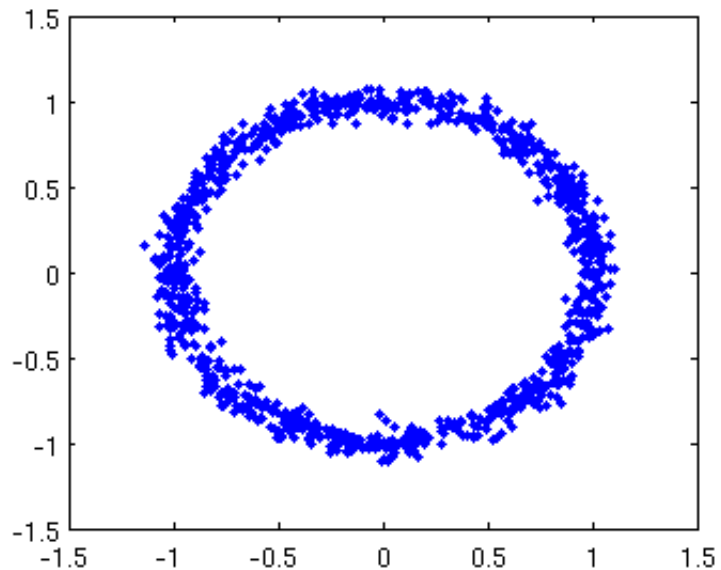
Величина  $z$  – распределена нормально,  $S_E = \frac{1}{\sqrt{n-3}}$

# Задачи на коэффициент корреляции

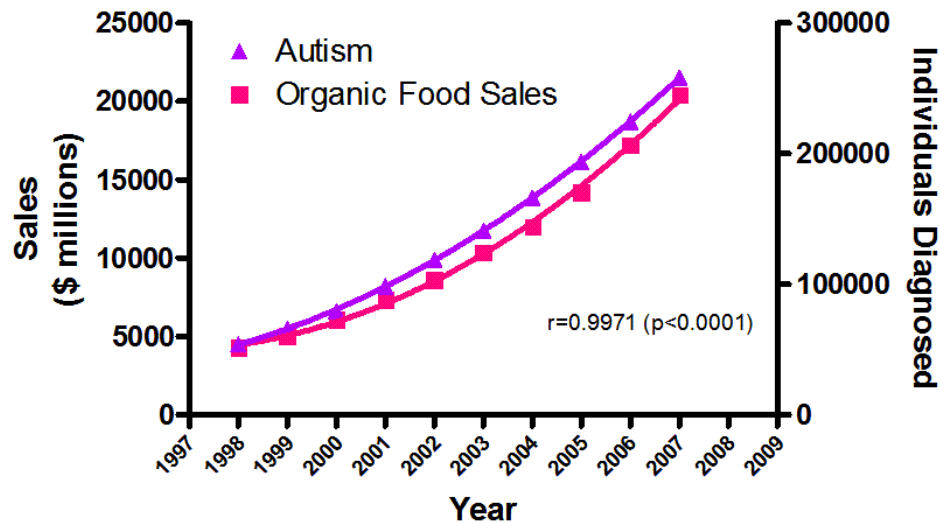
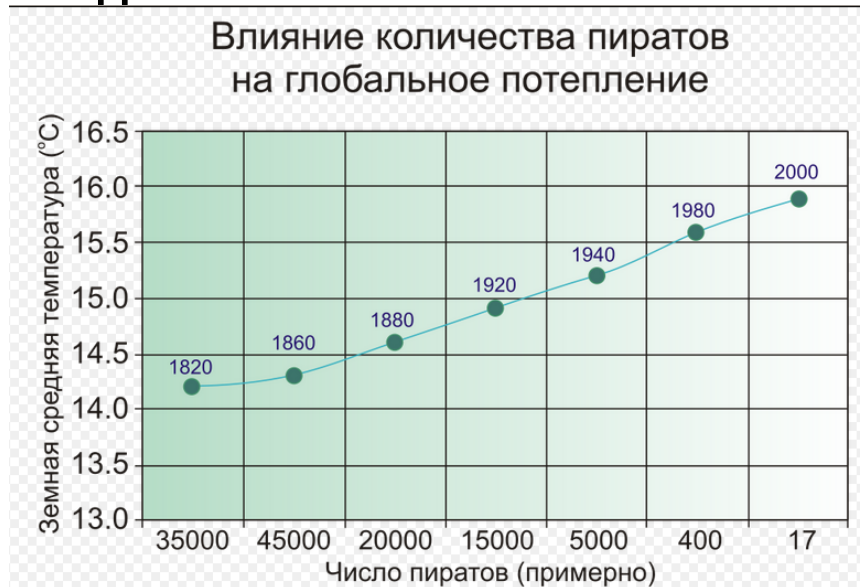
1. Построить две независимые выборки из 100 нормально распределённых случайных чисел ( $\mu = 0$ ,  $\sigma^2 = 1$ ), рассчитать коэффициент корреляции как с помощью функции **КОРРЕЛ**, так и с использованием **КОВАРИАЦИЯ.В**, **ДИСП.В** и **СТАНДОТКЛОН.В**. Построить X-Y диаграмму
2. Построить две независимые выборки из 100 нормально распределённых случайных чисел ( $N(0;1)$ ). Рассчитать коэффициенты корреляции и построить X-Y диаграммы для следующих величин:
  - $N_1(0;1)$  и  $2N_1(0;1)$
  - $N_1(0;1)$  и  $-2N_1(0;1)$
  - $N_1(0;1)$  и  $2N_1(0;1) + N_2(0;1)$
  - $N_1(0;1)$  и  $-2N_1(0;1) + N_2(0;1)$
3. Сгенерировать числа от 0 до 6.5 с интервалом 0.1 (далее – величина  $t$ ). Рассчитать величины  $x = \sin t$  и  $y = \cos t$ , построить X-Y диаграмму и рассчитать коэффициент корреляции. Объяснить полученный результат.

# Корреляция и коэффициент корреляции

Помните о нелинейных зависимостях



Корреляция – не причинно-следственная связь

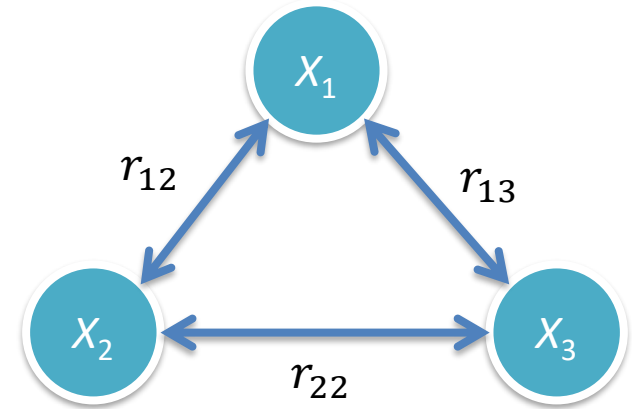


# Корреляционная матрица

Пусть  $X_1, X_2, X_3$  – случайные величины. Тогда матрица вида

$$(r) = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ \dots & 1 & r_{23} \\ \dots & \dots & 1 \end{pmatrix}, \text{ где } r_{ij} -$$

коэф. корреляции между  $X_i$  и  $X_j$   
- корреляционная матрица



**Найти корреляционную матрицу**

|       |    |   |   |   |   |   |    |   |   |    |
|-------|----|---|---|---|---|---|----|---|---|----|
| $X_1$ | 1  | 2 | 3 | 4 | 5 | 6 | 7  | 8 | 9 | 10 |
| $X_2$ | 2  | 1 | 3 | 4 | 9 | 8 | 10 | 5 | 7 | 6  |
| $X_3$ | 10 | 9 | 8 | 7 | 6 | 5 | 4  | 3 | 2 | 1  |