15. Функциональная, корреляционная и статистическая зависимости. Ковариация и корреляция (определения, свойства – с доказательством)

В статистике выделяют три основных вида зависимости между случайными величинами X и Y, различающихся по своей природе и строгости связи.

1. Виды зависимостей между переменными

- **Функциональная зависимость:** Самый строгий вид связи, при котором каждому значению переменной X соответствует ровно одно определенное значение переменной Y. В общем виде записывается уравнением $Y = f(X)$. Примером может служить зависимость площади круга от его радиуса.
- **Статистическая зависимость:** Более общая форма связи, при которой изменение одной величины (X) влечет за собой изменение **закона распределения** другой величины (Y). Это означает, что при фиксированном x величина Y остается случайной, но ее возможные значения и их вероятности меняются в зависимости от x.
- **Корреляционная зависимость:** Частный случай статистической связи, при которой изменение значения X приводит к изменению **среднего значения** (математического ожидания) величины Y. Математически это выражается через функции регрессии: $M(Y|x) = f(x)$.

2. Ковариация (корреляционный момент)

Определение: Ковариацией $cov(X, Y)$ называется математическое ожидание произведения отклонений случайных величин от их средних значений:

$$cov(X,Y) = M[(X - M(X))(Y - M(Y))]$$

Свойства ковариации и доказательства:

1. **Симметричность:** $cov(X, Y) = cov(Y, X)$.
2. **Связь с дисперсией:** $cov(X, X) = D(X)$.
   - *Доказательство:* $cov(X, X) = M[(X - M(X))(X - M(X))] = M[(X - M(X))^2] = D(X)$.

3. **Вынос константы:** $cov(kX, Y) = k \cdot cov(X, Y)$.
4. **Ковариация с константой:** $cov(X, C) = 0$.
5. **Расчетная формула:** $cov(X, Y) = M(XY) - M(X)M(Y)$.
   - *Доказательство:*

     $$M[(X - M(X))(Y - M(Y))] = M[XY - X \cdot M(Y) - Y \cdot M(X) + M(X)M(Y)] =$$

     $$= M(XY) - M(X)M(Y) - M(Y)M(X) + M(X)M(Y) = M(XY) - M(X)M(Y)$$

     .

Ковариация характеризует как силу связи, так и её направление: если $cov(X, Y) > 0$, величины имеют тенденцию возрастать одновременно; если $cov(X, Y) < 0$ — при возрастании одной величины другая убывает.

3. Коэффициент корреляции

Определение: Коэффициент корреляции r_{xy} — это безразмерная величина, характеризующая тесноту **линейной** связи между величинами:

$$r_{xy} = \frac{cov(X, Y)}{\sigma_x \sigma_y}$$

где σ_x, σ_y — средние квадратические отклонения величин X и Y.

Свойства коэффициента корреляции:

1. **Ограниченность:** $-1 \leq r_{xy} \leq 1$.
   - Если $|r_{xy}| = 1$, то между переменными существует строгая **линейная функциональная зависимость**.
2. **Независимость и корреляция:** Если X и Y независимы, то $r_{xy} = 0$.
   - *Важно:* Если $r_{xy} = 0$, величины называются **некоррелированными**. Это гарантирует отсутствие линейной связи, но между ними всё ещё может существовать сильная нелинейная зависимость.
3. **Инвариантность:** Значение r_{xy} не меняется при изменении единиц измерения или масштаба переменных, так как оно является нормированным значением ковариации.

Аналогия:

Ковариация — это как направление ветра: она говорит, дуют ли переменные в одну сторону. Корреляция — это "спидометр", который не просто показывает направление, но и измеряет силу этого ветра по шкале от -1 до 1, при этом не обращая внимания на то, измеряете вы скорость в узлах или километрах в час.