15. Статистические свойства оценок по методу наименьших квадратов параметров парной линейной регрессии.

Рассмотрим задачу "наилучшей" аппроксимации набора наблюдений X_t, Y_t, $t = 1, \dots, n$ линейной функцией $f(X) = a + bX$ в смысле минимизации функционала

$$F = \sum_{t=1}^{n} (Y_t - (a + bX_t))^2 \to min\,, \forall a, b$$

Запишем необходимые условия экстремума:

$$\frac{dF}{da} = -2\sum_{t=1}^{n}(Y_t - a - bX_t) = 0,$$

$$\frac{dF}{db} = -2\sum_{t=1}^{n}X_t(Y_t - a - bX_t) = 0,$$

Или

$$\sum_{t=1}^{n}(Y_t - a - bX_t) = 0,$$

$$\sum_{t=1}^{n}X_t(Y_t - a - bX_t) = 0.$$

Раскроем скобки и получим *стандартную форму нормальных уравнений* (для краткости опустим индексы суммирования у знака суммы):

$$an + b\sum X_t = \sum Y_t,$$

$$a\sum X_t + b\sum {X_t}^2 = \sum X_t Y_t.$$

Решением этой системы будут $\hat{a}, \hat{b}$

$$\hat{a} = \frac{1}{n}\sum Y_t - \frac{1}{n}X_t\hat{b} = \bar{Y} - \bar{X}b,$$

$$\hat{b} = \frac{n\sum X_t Y_t - \sum X_t \sum Y_t}{n\sum X^2{}_t - (\sum X_t)^2} = \frac{Cov\,(X, Y)}{Var\,X}$$

Замечание.

Из первого уравнения следует

$$\bar{Y} = \hat{a} + \hat{b}\bar{X}$$

То есть уравнение прямой линии $Y = \hat{a} + \hat{b}X$, полученное в результате минимизации функционала, проходит через точку $(\bar{X}, \bar{Y})$. Здесь через $\bar{X}, \bar{Y}$ обозначены выборочные средние значения переменных X_t, Y_t:

$$\bar{X} = \frac{1}{n}\sum X_t, \quad \bar{Y} = \frac{1}{n}\sum Y_t,$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \quad Var\,(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Уравнения в отклонениях

Обозначим через $x_t = X_t - \bar{X}, \quad y_t = Y_t - \bar{Y}$ отклонения от средних по выборке значений X_t и Y_t (можно проверить, что ($\bar{x} = \bar{y} = 0$).

Решим теперь ту же задачу: подобрать линейную функцию $f(x) = a + bx$ минимизирующую функционал

$$F = \sum_{t=1}^{n}(y_t - (a + bx_t))^2 \to min.$$

Из геометрических соображений ясно, что решением задачи будет та же прямая на плоскости (x, y), что и для исходных данных X_t и Y_t. В самом деле в силу $\bar{Y} = \hat{a} + \hat{b}\bar{X}$ переход от X, Y к отклонениям x, y означает лишь перенос начала координат в точку $(\bar{X}, \bar{Y})$. Вычисления, которые необходимо проделать для решения задачи, вполне аналогичны предыдущим (заменой X, Y на x, y). Сделав замену X_t, Y_t на x_t, y_t в равенствах для $\hat{a}, \hat{b}$ учитывая, что $\bar{x} = \bar{y} = \frac{1}{n}\sum x_t = \frac{1}{n}\sum y_t = 0$, получим

$$\hat{a} = 0$$

$$\hat{b} = \frac{\sum x_t y_t}{\sum x^2{}_t} = \frac{\sum(X_t - \bar{X})(Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2}.$$

Таким образом, получили другое выражение для углового коэффициента $\hat{b}$.