Lecture Summary: Jan. 30, 2023

- Coefficient of correlation:

$$r = \pm\sqrt{R^2} = \begin{cases} +\sqrt{R^2}, & \text{if } \hat{\beta}_1 > 0, \\ -\sqrt{R^2}, & \text{if } \hat{\beta}_1 < 0. \end{cases}$$

Alternative expression:

$$r = \frac{\sum_i (Y_i - \bar{Y})(x_i - \bar{x})}{\sqrt{\sum_i (Y_i - \bar{Y})^2 \sum_i (x_i - \bar{x})^2}}.$$

- Properties of r:

1. $-1 \le r \le 1$.
2. $r \approx \pm 1$, if there is a strong linear association between x and Y.
3. $r \approx 0$, if there is a weak or no linear association between x and Y.
4. r is a measure of linear association only.

- Covariance and correlation between two random variables

$$\text{cov}(X, Y) = \text{E}\{(X - \mu_X)(Y - \mu_Y)\} = \text{E}(XY) - \text{E}(X)\text{E}(Y),$$

$$\rho = \text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)},$$

where $\mu_X = \text{E}(X)$, $\text{sd}(X) = \sqrt{\text{var}(X)}$, etc.

- Special case: (X, Y) has a bivariate normal distribution.

- Testing for hypothesis about ρ: Assuming that the bivariate normal distribution holds for (X, Y).

Consider $\text{H}_0 : \rho = 0$ vs $\text{H}_1 : \rho \neq 0$ (or $\rho > 0$, or $\rho < 0$). The test-statistic is

$$t^* = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}},$$

which has a t_{n-2} distribution under H_0.