University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Gauss-Markov theorem

Consider the simple regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$. The Gauss-Markov conditions hold, i.e. $E(\epsilon_i) = 0, var(\epsilon_i) = \sigma^2$, and $\epsilon_1, \ldots, \epsilon_n$ are independent.

We have shown in class that the OLS estimators can be expressed as linear combinations of the $Y_i's$. In particular, $\hat{\beta}_1 = \sum_{i=1}^n k_i y_i$ and $\hat{\beta}_0 = \sum_{i=1}^n l_i y_i$. The Gauss-Markov theorem states that these OLS estimates have the smallest variance among all the linear unbiased estimators. We say that the OLS estimates are BLUE.

Proof:
Let $b_1 = \sum_{i=1}^n a_i y_i$ be another linear unbiased estimator of β_1. Since it is unbiased, we have

$$\begin{aligned} Eb_1 &= \beta_1 \\ E\left(\sum_{i=1}^n a_i y_i\right) &= \beta_1 \\ \sum_{i=1}^n a_i E y_i &= \beta_1 \\ \sum_{i=1}^n a_i(\beta_0 + \beta_1 x_i) &= \beta_1 \\ \beta_0 \sum_{i=1}^n a_i + \beta_1 \sum_{i=1}^n a_i x_i &= \beta_1 \end{aligned}$$

This equality holds for ALL β_0 and β_1 if and only if $\sum_{i=1}^n a_i = 0$ and $\sum_{i=1}^n a_i x_i = 1$.

Now for the variance of b_1:

$$\begin{aligned} var(b_1) &= var\left(\sum_{i=1}^n a_i y_i\right) \\ &= \sigma^2 \sum_{i=1}^n a_i^2 \quad \text{(Gauss-Markov condition)} \\ &= \sigma^2 \sum_{i=1}^n (k_i + d_i)^2 \quad (\text{let } a_i = k_i + d_i) \\ &= \sigma^2 \sum_{i=1}^n k_i^2 + \sigma^2 \sum_{i=1}^n d_i^2 + 2\sigma^2 \sum_{i=1}^n k_i d_i \\ &= \sigma^2 \sum_{i=1}^n k_i^2 + \sigma^2 \sum_{i=1}^n d_i^2 \quad \left(\text{because } \sigma^2 \sum_{i=1}^n k_i d_i = 0, \text{ see class notes}\right) \\ &= var(\hat{\beta}_1) + \sigma^2 \sum_{i=1}^n d_i^2 \geq var(\hat{\beta}_1). \end{aligned}$$

Gauss-Markov theorem for the predicted value of y_0 given x_0.

Suppose we wish to predict a new y, say y_0, for a given x_0. The predicted value based on the OLS estimates will be $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{Y} + \hat{\beta}_1(x_0 - \bar{x}) = \sum_{j=1}^{n} \left(\frac{1}{n} + (x_0 - \bar{x})k_j\right) y_j = \sum_{j=1}^{n} r_j y_j$. We can easily show that $E\hat{y}_0 = Ey_0 = \beta_0 + \beta_1 x_0$. Also the variance of the predicted value can be expressed as $var(\hat{y}_0) = \sigma^2 \sum_{j=1}^{n} r_j^2$.

Now consider another unbiased predictor of y_0 of the form $\tilde{y}_0 = \sum_{j=1}^{n} c_j y_j$. Since it is unbiased, we have

$$
\begin{aligned}
E\hat{y}_0 &= \beta_0 + \beta_1 x_0 \\
E\left(\sum_{i=1}^{n} c_j y_j\right) &= \beta_0 + \beta_1 x_0 \\
\sum_{i=1}^{n} c_i E y_i &= \beta_0 + \beta_1 x_0 \\
\sum_{i=1}^{n} c_i(\beta_0 + \beta_1 x_i) &= \beta_0 + \beta_1 x_0 \\
\beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i &= \beta_0 + \beta_1 x_0
\end{aligned}
$$

This equality holds for ALL β_0 and β_1 if and only if $\sum_{i=1}^{n} c_i = 1$ and $\sum_{i=1}^{n} c_i x_i = x_0$.

Now show that $var(\tilde{y}_0) \geq var(\hat{y}_0)$.