# Handout for Lecture 15

Ordinary Least Squares & Goodness of Fit

ECON 340: Economic Research Methods  Instructor: Div Bhagia

Simple Linear Regression Model: $Y = \beta_0 + \beta_1 X + u$

- $Y$: Dependent variable (outcome or response variable)
- $X$: Independent variable (explanatory variable, regressor)
- $\beta_0, \beta_1$: intercept and slope (population parameters)
- $u$: mean zero error term, $E(u) = 0$

Ordinary Least Squares (OLS)

To obtain estimates for the intercept and slope of the line, we minimize the distance between the fitted line and the sample data. Let $X_i$ and $Y_i$ denote the $i$'th observation of $X$ and $Y$ in the sample data.

- $\hat{Y}_i$: predicted value of $Y_i$
- $\hat{\beta}_0, \hat{\beta}_1$: OLS estimators for the intercept and slope
- Residuals/error: $\hat{u}_i = \hat{Y}_i - Y_i$ (Note that we can always write $Y_i = \hat{Y}_i + \hat{u}_i$)

OLS estimators for the intercept $\hat{\beta}_0$ and slope $\hat{\beta}_1$ are obtained by minimizing the sum of squared residuals:

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

A Measure of Goodness of Fit

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

where

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2, \qquad ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2, \qquad RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \hat{u}_i^2$$

Example: Predicting Final Exam Scores

You've collected data on monthly revenue ($Revenue_i$) and advertising spending ($AddSpend_i$) for several months for a small business. You fit the following line using OLS:

$$\widehat{Revenue}_i = 50 + 3 \cdot AddSpend_i, \qquad R^2 = 0.65$$

1. What are the estimated intercept and slope in the given fitted line?

   $\hat{\beta}_0 =$ $\qquad\qquad\qquad$ $\hat{\beta}_1 =$

2. Interpret the intercept and slope coefficient.

3. What is the predicted revenue for a month where the advertising spending was \$50?

4. If in a particular month, the revenue was \$100 and advertising spending was $20$, what would be the residual $\hat{u}$ for this observation?

5. How does the predicted revenue increase due to an increase of \$10 in advertising spending?

6. What percentage of the variability in revenue is explained by advertising spending?

7. (A bit challenging, try at home.) If I tell you, the variance of revenue over months is $125$. Can you tell me what is the variance of advertising spending?