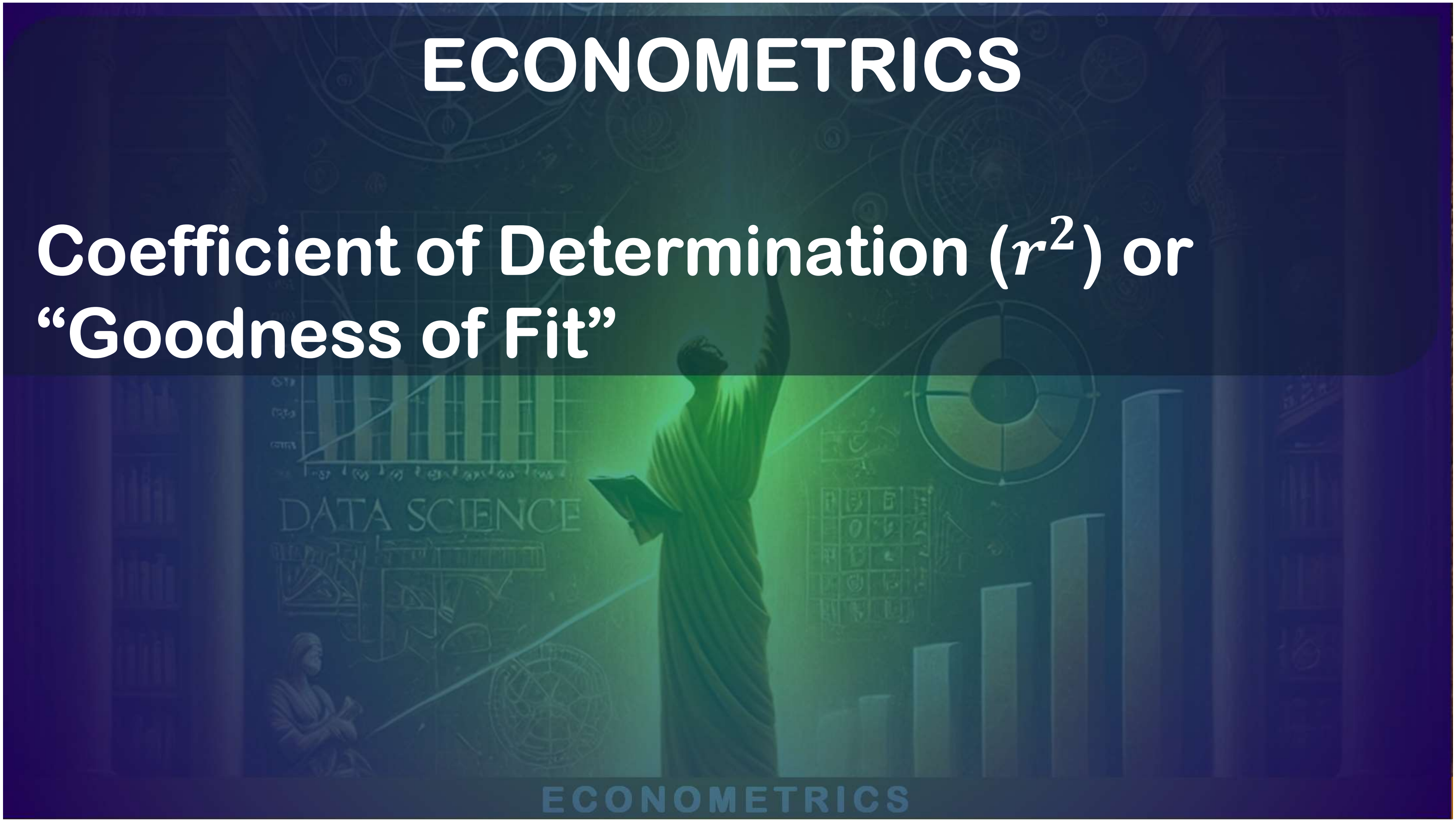
ECONOMETRICS

Coefficient of Determination (r^2) or "Goodness of Fit"



Lesson Goal

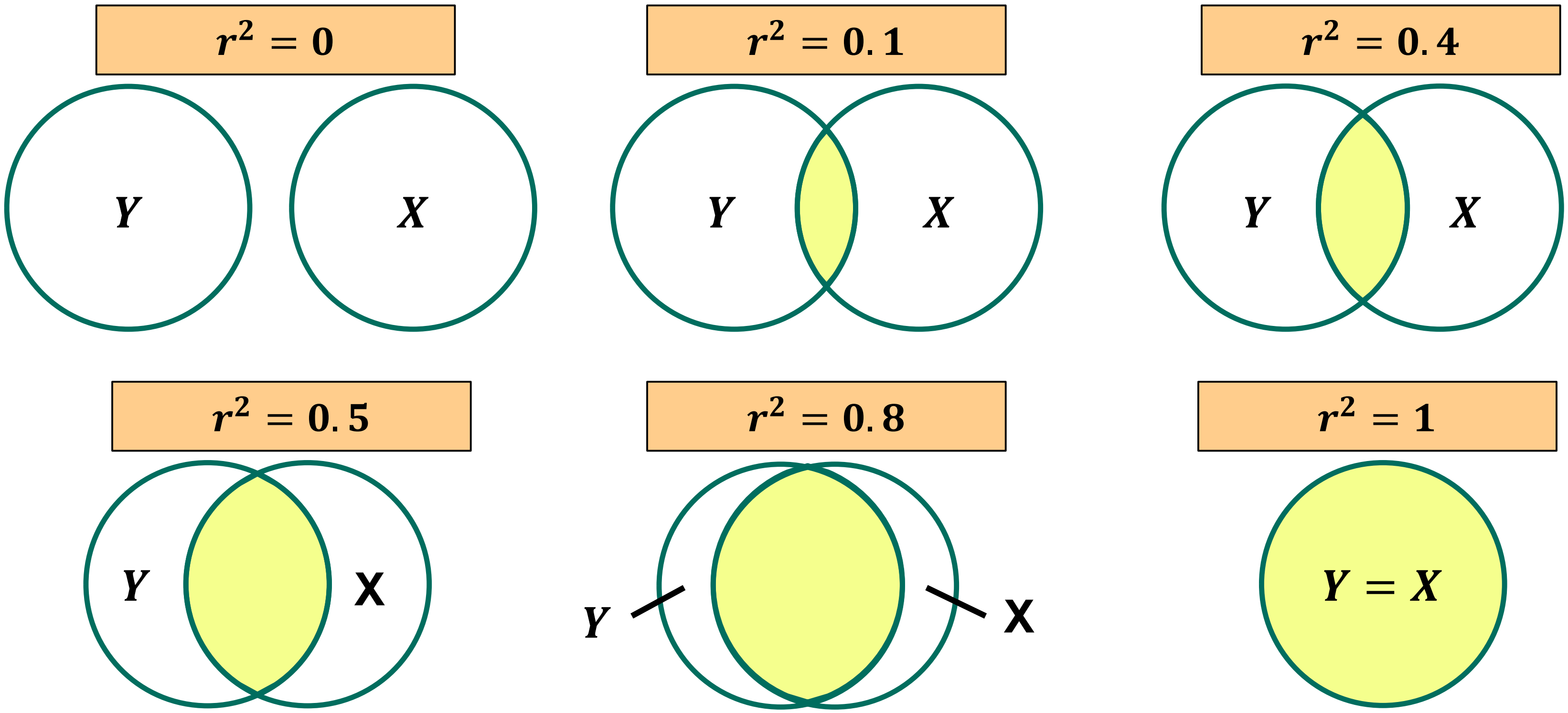
- Assess the **goodness of fit** of regression models.

Coefficient of Determination r^2

- We look at the goodness of fit of the fitted regression line to a set of data.

- The coefficient of determination r^2 (two-variable case) or R^2 (multiple regression) is a summary measure of how well the sample regression line fits the data.

Coefficient of Determination r^2 - Venn Diagram/Ballentine



Source: Peter Kennedy, "Ballentine: A Graphical Aid for Econometrics," Australian Economics Papers, vol. 20, 1981, pp. 414–416.

Coefficient of Determination r^2 - Proof

Recall, $Y_i = \hat{Y}_i + \hat{U}_i$

Deviation form, $y_i = \hat{y}_i + \hat{u}_i$

where $y_i = Y_i - \bar{Y}$, $\hat{y}_i = \hat{Y}_i - \bar{Y}$, $\hat{u}_i = \hat{U}_i - 0$

$$\sum y_i^2 = \sum(\hat{y}_i + \hat{u}_i)^2$$
$$\sum y_i^2 = \sum(\hat{y}_i^2 + 2\hat{y}_i\hat{u}_i + \hat{u}_i^2)$$
$$\sum y_i^2 = \sum \hat{y}_i^2 + 2\sum \hat{y}_i\hat{u}_i + \sum \hat{u}_i^2$$

But, $\sum \hat{y}_i \hat{u}_i = 0$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

TSS $\Leftarrow \sum y_i^2 = \sum(Y_i - \bar{Y})^2$ → the total variation of the actual Y values about their sample mean.

→ Total Sum of Squares

ESS $\Leftarrow \sum \hat{y}_i^2 = \sum(\hat{Y}_i - \bar{Y})^2$ → variation of the estimated Y values about their mean.

→ Explained Sum of Squares

(sum of squares due to regression / sum of squares explained by regression)

RSS $\Leftarrow \sum \hat{u}_i^2 = \sum(Y_i - \hat{Y}_i)^2$ → residual or unexplained variation of the Y values about regression line

→ Residual sum of Squares

Coefficient of Determination r^2 - Proof

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$TSS = ESS + RSS$$

$$\frac{TSS}{TSS} = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

where $r^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$

$$1 = r^2 + \frac{RSS}{TSS}$$

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

$$= 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$$

$$= 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

Coefficient of Determination r^2 - Definition

- r^2 measures the proportion or percentage of the total variation in Y explained by the variation in X in the regression model.

$$r^2 = \frac{ESS}{TSS} = \frac{\sum\left(\widehat{Y_i} - \bar{Y}\right)^2}{\sum(Y_i - \bar{Y})^2}$$

$$r^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum\left(Y_i - \widehat{Y_i}\right)^2}{\sum(Y_i - \bar{Y})^2}$$

- Simply put, r^2 is the ratio of explained variation to the total variation.
  - r^2 lies between 0 and 1.

Correlation Coefficient (r)

- Another formula for r^2 is:

$$r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \cdot \sum y_i^2}$$

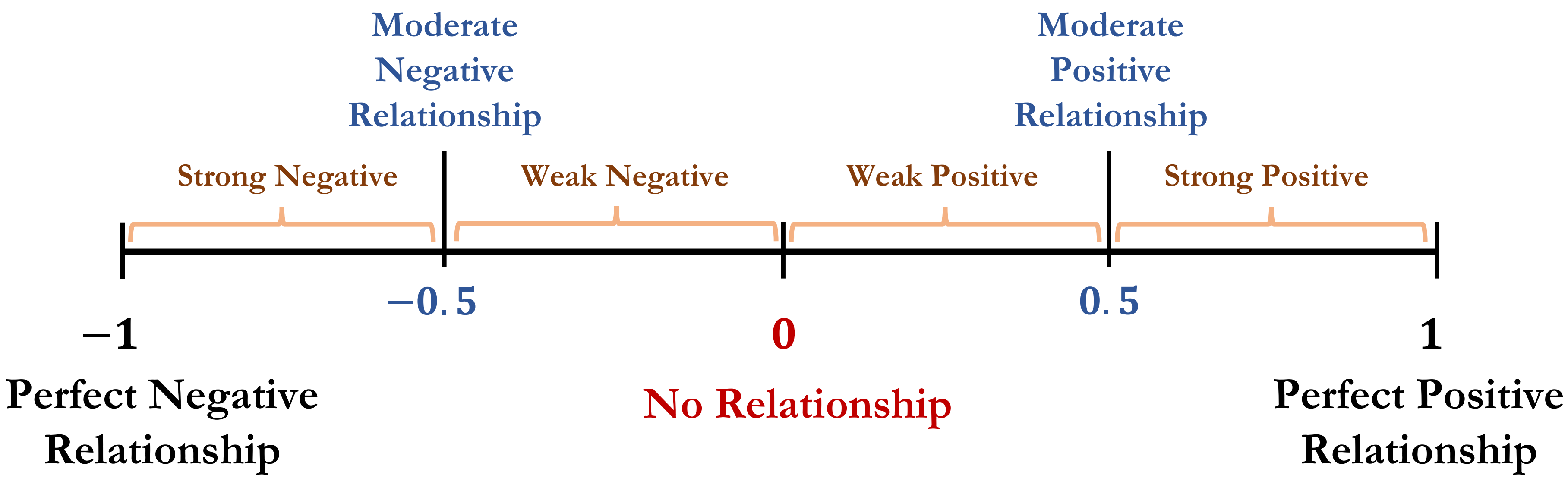
- The square root of r^2 is the correlation coefficient r.

$$r = \pm\sqrt{r^2}$$

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \cdot \sum(Y_i - \bar{Y})^2}}$$

Correlation Coefficient (r)

- Correlation coefficient r measures the **strength** and **direction** of the relationship between two variables. It ranges from -1 to 1.



THANK YOU!

Next Lesson: Numerical Example of Regression Analysis