"You Do It" Solution

Lecture 11: Multiple Regression Analysis

Prepared as a classroom-ready detailed solution set for ECOM5005

Exercise 1: Explained and Unexplained Variability in Multiple Regression

Question

A two-predictor regression equation using total SAT score (SATSUM) and high-school grades (HSGPA) to predict first-year college grades (FYGPA) is

$$\widehat{\text{FYGPA}} = -0.873 + 0.00144(\text{SATSUM}) + 0.58(\text{HSGPA}).$$

The coefficient of multiple determination is

$$R^2 = 0.358.$$

What percentage of the variability in first-year college grades is **not explained** by SATSUM and HSGPA?

Detailed Solution

Step 1: Recall what R^2 measures.
In a multiple regression model, R^2 is the **proportion of the total variation in the dependent variable** that is explained by all of the independent variables taken together.
Here, the dependent variable is first-year college GPA (FYGPA), while the two explanatory variables are SATSUM and HSGPA. Therefore,

$$R^2 = 0.358$$

means that the model explains a proportion of **0.358** of the observed variability in FYGPA.
Step 2: Convert the explained proportion to a percentage.

$$0.358 \times 100\% = 35.8\%.$$

Thus, **35.8% of the variability** in first-year college grades is explained jointly by SATSUM and HSGPA in this fitted regression model.
Step 3: Find the proportion that is not explained.
The whole of the variation corresponds to 1, or 100%. Therefore,

$$\text{Unexplained proportion} = 1 - R^2.$$

Substitute $R^2 = 0.358$:

$$1 - 0.358 = 0.642.$$

Convert this proportion to a percentage:

$$0.642 \times 100\% = \boxed{64.2\%}.$$

Step 4: State the result in context.
Approximately **64.2% of the variability in first-year college grades is not explained by SATSUM and HSGPA** in this two-predictor linear regression model.
Equivalently,

$$\boxed{100\% - 35.8\% = 64.2\%}.$$

Exercise 1: Interpretation and Final Answer

Interpreting the Coefficient of Multiple Determination

The lecture defines the coefficient of multiple determination as

$$R^2 = \frac{SSR}{SST},$$

where SSR is the regression sum of squares (explained variation) and SST is the total sum of squares (total variation). Hence the complementary share

$$1 - R^2$$

represents the proportion of total variation that is **not explained by the fitted regression model**.

For this model,

$$R^2 = 0.358 \quad \Longrightarrow \quad 1 - R^2 = 0.642.$$

So the model leaves **64.2%** of FYGPA variability unexplained by these two predictors.

What the Regression Equation Contributes

The fitted equation is

$$\widehat{\text{FYGPA}} = -0.873 + 0.00144(\text{SATSUM}) + 0.58(\text{HSGPA}).$$

It is useful for prediction and for interpreting the partial effect of each predictor while holding the other predictor constant. However, the numerical coefficients are **not needed** to answer this particular question. The required information is the reported value of R^2.

A useful distinction:

- $R^2 = 0.358$ means **35.8% of the sample variation in FYGPA is explained** by SATSUM and HSGPA together through the fitted linear model.
- $1 - R^2 = 0.642$ means **64.2% remains unexplained** by those predictors in this model.
- This does **not** mean that 64.2% of students are predicted incorrectly, and it does not mean that each individual prediction has a 64.2% error.

Answer Summary

$$\begin{aligned} \text{Explained variability} &= R^2 \times 100\% \\ &= 0.358 \times 100\% = 35.8\%, \\ \text{Unexplained variability} &= (1 - R^2) \times 100\% \\ &= (1 - 0.358) \times 100\% \\ &= \boxed{64.2\%}. \end{aligned}$$

Final answer: $\boxed{64.2\%}$ of the variability in first-year college grades is not explained by SATSUM and HSGPA in this regression model.

Final Check

For questions involving R^2 in multiple regression, remember:

- R^2 is the proportion of total variation in Y explained by the independent variables **taken together**;
- multiply R^2 by 100% to express explained variation as a percentage;
- use $1 - R^2$ for the proportion of variation that is not explained;
- R^2 lies between 0 and 1;
- do not interpret $1 - R^2$ as the percentage of observations that are "wrong" or as an individual prediction error.