A Proof of the Decomposition Identity: TSS = ESS + RSS

Shayan Shahrabi

May 2025

Abstract

In the context of linear regression, the decomposition of variance is a fundamental result that relates total variation in the response variable to the variation explained by the model and the residual (unexplained) variation. This document provides both an algebraic and geometric proof of the identity $\text{TSS} = \text{ESS} + \text{RSS}$, grounded in the properties of the Ordinary Least Squares (OLS) estimator. Emphasis is placed on the orthogonality of residuals and the projection geometry underlying OLS. The result not only enhances interpretability of regression but also underpins measures such as R^2.

1. Introduction

In linear regression, understanding how variation in the response variable is partitioned is essential for evaluating model performance. This partitioning involves three key quantities:

- **Total Sum of Squares (TSS)**: Measures total variation around the mean:

$$\text{TSS} = \sum (Y_i - \bar{Y})^2$$

- **Residual Sum of Squares (RSS)**: Captures unexplained variation (errors):

$$\text{RSS} = \sum (Y_i - \hat{Y}_i)^2$$

- **Explained Sum of Squares (ESS)**: Reflects variation explained by the model:

$$\text{ESS} = \sum (\hat{Y}_i - \bar{Y})^2$$

OLS Regression Review

The Ordinary Least Squares (OLS) estimator fits the model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

by minimizing the residual sum of squares (RSS). It satisfies key properties:

1. Orthogonality of Residuals:

- The residual vector $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ is orthogonal to the predictor space.
- Mathematically: $\mathbf{X}^T\mathbf{e} = \mathbf{0}$

2. Vanishing Cross-Term:

- The vector $\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}$ lies in $\text{col}(\mathbf{X})$
- Hence, $\mathbf{e}^T(\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) = 0$

Implication: These properties justify the decomposition:

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

2. Algebraic Derivation

2.1. Variance Decomposition Identity

Starting from:

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$$

Squaring both sides and summing over all i gives:

$$\text{TSS} = \text{RSS} + \text{ESS} + 2\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$$

Due to the orthogonality between residuals and fitted values, the cross-term vanishes:

$$\sum(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 0$$

Thus, we confirm the decomposition:

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

2.2. Geometric Interpretation

- The fitted vector $\hat{\mathbf{Y}}$ is the orthogonal projection of $\mathbf{Y}$ onto $\text{col}(\mathbf{X})$

- The residual $\mathbf{e}$ is orthogonal to this subspace
- By the Pythagorean theorem:

$$\|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2 = \|\mathbf{e}\|^2 + \|\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}\|^2$$

This geometric view reinforces the algebraic decomposition.

3. Conclusion

The properties of the least squares estimator lead to a clean and intuitive decomposition of variance:

- The total variation (TSS) is split into explained (ESS) and residual (RSS) components
- The residual sum of squares is minimized: $\text{RSS} \leq \text{TSS}$

$$\boxed{\text{TSS} = \text{ESS} + \text{RSS} \quad \text{with} \quad \text{RSS} \leq \text{TSS}}$$