Understanding the Equivalence of Externally Studentized and Standardized LOOCV Residuals in Linear Models

An Illustration with R

Longhai Li

2025-09-30

1 Introduction

In statistical modeling, identifying data points that don't fit—the **outliers**—is a critical step. The most reliable tool for this job is the **externally studentized residual**. Its power comes from a simple, intuitive idea: to judge a point fairly, you shouldn't use that point when building your model. This is the core principle of **Leave-One-Out Cross-Validation (LOOCV)**.

This article provides a complete walkthrough of this essential concept. We'll start with the basic linear model, introduce the necessary notation, explore the flaws of simpler residuals, and then formally define and prove the equivalence of the conceptual and computational formulas for studentized residuals. Finally, we'll make it all concrete with a simple example.

---

2 The Linear Model

Our discussion is based on the standard multiple linear regression model. In matrix form, the relationship between a response vector $\mathbf{Y}$ and a predictor matrix $\mathbf{X}$ is:

$$\mathbf{Y} = \mathbf{X}\ + \tag{1}$$

Where:

- $\mathbf{Y}$ is an n x 1 vector of the observed outcomes.

1

- **X** is the n x p design matrix of predictor variables (where p is the number of coefficients, including the intercept).
- β is the p x 1 vector of unknown true coefficients we want to estimate.
- ϵ is an n x 1 vector of unobservable random errors, assumed to be independent and identically distributed with a mean of 0 and a variance of σ^2.

---

3 Our Notations

To discuss models fit with all data versus those with one point removed, we need clear notation.

3.1 Full Data Model (Using all *n* observations)

- $\hat{\beta}$: The estimated coefficient vector.
- $\hat{y}_i$: The predicted value for observation i from this model.
- e_i: The **ordinary residual** ($e_i = y_i - \hat{y}_i$).
- $\hat{\sigma}$: The estimated standard deviation of the errors (Residual Standard Error).
- h_{ii}: The **leverage** of observation i, a measure of how much its x-values influence the model.

3.2 Leave-One-Out (LOOCV) Model

- $\hat{\beta}_{-i}$: The coefficient vector estimated after **removing** observation i.
- $\hat{y}_{i,-i}$: The predicted value for observation i, from the model fit **without** observation i.
- $e_{i,-i}$: The **deleted residual** ($e_{i,-i} = y_i - \hat{y}_{i,-i}$).
- $\hat{\sigma}_{-i}$: The standard deviation of the errors estimated from the model fit **without** observation i.

---

4 Non-studentized Residuals

Before getting to the correct solution, it's crucial to understand why simpler methods of standardizing residuals are flawed.

4.1 The Ordinary Residual (e_i): Too Small and x_i Dependent

The most basic residual, e_i, is problematic for two key reasons.

- **Systematically Too Small**: An outlier has an undue influence on the model, pulling the regression line towards itself. This makes its own predicted value, $\hat{y}_i$, artificially close to its actual value, y_i. As a result, its residual, e_i, is **deceptively small** and doesn't reflect the true magnitude of the error.

- x_i-**Dependent Variance**: The variance of an ordinary residual is not constant; it depends on the point's leverage.

$$\text{Var}(e_i) = \sigma^2(1 - h_{ii}) \tag{2}$$

We know

$$\hat{y}_i = \mathbf{x}_i^\top \hat{\beta} = \mathbf{x}_i^\top (\beta + (X^\top X)^{-1} X^\top \epsilon). \tag{3}$$

So the residual is

$$e_i = y_i - \hat{y}_i = \epsilon_i - \mathbf{x}_i^\top (X^\top X)^{-1} X^\top \epsilon. \tag{4}$$

The variance can be derived from the hat matrix $H = X(X^\top X)^{-1} X^\top$. Since $\hat{y} = HY$, we have

$$e = (I - H)\epsilon. \tag{5}$$

Thus,

$$\text{Var}(e) = (I - H)\,\sigma^2\,(I - H)^\top = (I - H)\sigma^2. \tag{6}$$

Therefore, for the ith residual,

$$\text{Var}(e_i) = \sigma^2(1 - h_{ii}). \tag{7}$$

Since leverage (h_{ii}) is always greater than 0, the variance of an ordinary residual is always **less than the true error variance, σ^2**. High-leverage points act as "anchors" for the line and have even smaller variance.

4.2 The LOOCV Residual ($e_{i,-i}$): Too large and x_i-Dependent Variance

The deleted residual, $e_{i,-i}$, solves the "too small" problem. Because the model isn't influenced by the point it's predicting, the residual is an honest measure of prediction error. However, its variance is still not constant.

x_i-**Dependent Variance**: The variance of a deleted residual also depends on leverage, but in the opposite way.

$$\text{Var}(e_{i,-i}) = \frac{\sigma^2}{1 - h_{ii}} \tag{8}$$

From the key identity,

$$e_{i,-i} = \frac{e_i}{1 - h_{ii}}. \tag{9}$$

Therefore,

$$\text{Var}(e_{i,-i}) = \frac{\text{Var}(e_i)}{(1 - h_{ii})^2} = \frac{\sigma^2(1 - h_{ii})}{(1 - h_{ii})^2} = \frac{\sigma^2}{1 - h_{ii}}. \tag{10}$$

Since $1 - h_{ii}$ is less than 1, the variance of a deleted residual is always **greater than the true error variance, σ^2**. This is because it has two sources of randomness: the error in the point itself (y_i) and the error in the prediction ($\hat{y}_{i,-i}$).

---

5 Studentized LOOCV Residual

The correct solution is to take the honest LOOCV residual and divide it by its true standard error, which properly accounts for its larger, x-dependent variance. This is the **externally studentized residual**, t_i.

Conceptual Definition:

$$t_i = \frac{e_{i,-i}}{\text{SE}(e_{i,-i})} = \frac{e_{i,-i}}{\frac{\hat{\sigma}_{-i}}{\sqrt{1-h_{ii}}}} \tag{11}$$

This final value is a reliable diagnostic. Under the null hypothesis that the observation is not an outlier, it follows a **Student's t-distribution** with *n - p - 1* degrees of freedom.

---

6 Studentized Full Data Residuals (Computational Shortcut Formula)

Calculating the conceptual formula appears to require fitting n different regression models—a computationally expensive task. Fortunately, a mathematical identity allows us to calculate the exact same value using only the results from the single, full data model.

$$t_i = \frac{e_i}{\hat{\sigma}_{-i}\sqrt{1 - h_{ii}}} \tag{12}$$

This is not an approximation; it is an **exact algebraic rearrangement** of the conceptual definition.

6.1 Key Identity Linking $e_{i,-i}$ and e_i

$$e_{i,-i} = \frac{e_i}{1 - h_{ii}} \tag{13}$$

This identity shows that we can find the "pure" leave-one-out residual without actually refitting the model; all we need are the results from the single model fit on the full dataset.

6.2 Proof of Equivalence

The proof that the conceptual and computational formulas are identical hinges on a key algebraic identity that connects the LOOCV residual ($e_{i,-i}$) to the ordinary residual (e_i).

Now, let's start with the conceptual definition of the studentized residual and show how it transforms into the shortcut formula.

Start with the conceptual LOOCV definition:

$$t_i = \frac{e_{i,-i}}{\text{SE}(e_{i,-i})} = \frac{e_{i,-i}}{\frac{\hat{\sigma}_{-i}}{\sqrt{1-h_{ii}}}} \tag{14}$$

Substitute the key identity into the numerator:

$$t_i = \frac{\frac{e_i}{1-h_{ii}}}{\frac{\hat{\sigma}_{-i}}{\sqrt{1-h_{ii}}}} \tag{15}$$

Simplify the complex fraction. We can do this by multiplying the numerator by the reciprocal of the denominator:

$$t_i = \frac{e_i}{1 - h_{ii}} \cdot \frac{\sqrt{1 - h_{ii}}}{\hat{\sigma}_{-i}} \tag{16}$$

Cancel the terms. Since $1 - h_{ii} = (\sqrt{1 - h_{ii}})^2$, one of the $\sqrt{1 - h_{ii}}$ terms in the denominator cancels with the term in the numerator. This leaves us with the computational shortcut formula:

$$t_i = \frac{e_i}{\hat{\sigma}_{-i}\sqrt{1 - h_{ii}}} \tag{17}$$

This proves that the two formulas are mathematically identical. The computational shortcut is simply a clever algebraic rearrangement of the more intuitive LOOCV definition, allowing for efficient and accurate calculation.

7 Example

```r
# Libraries
library(dplyr)
library(tidyr)
library(ggplot2)

# -------------------------------
# 1) Data and full-model fit
# -------------------------------
set.seed(123)
n <- 20
x <- 1:n
y <- 2 + 3 * x + rnorm(n, mean = 0, sd = 5)
y[6] <- y[6] + 25  # add an outlier at x = 6
full_data  <- data.frame(x = x, y = y)

full_model <- lm(y ~ x, data = full_data)
leverage   <- hatvalues(full_model)
e_full     <- resid(full_model)          # e_i
y_hat      <- fitted(full_model)         # yhat_i
sigma_hat  <- summary(full_model)$sigma       # sigma_hat

# -------------------------------
# 2) LOOCV quantities (refit n times)
# -------------------------------
e_del         <- numeric(n)   # e_{i,-i}
sigma_minus_i <- numeric(n)   # sigma_{-i}
ns_loocv      <- numeric(n)   # non-studentized LOOCV: e_{i,-i} / sigma_{-i}
stud_loocv    <- numeric(n)   # studentized LOOCV: e_{i,-i} / (sigma_{-i} / sqrt(1 - h_ii))

for (i in 1:n) {
  loocv_model <- lm(y ~ x, data = full_data[-i, ])
  yhat_minus  <- predict(loocv_model, newdata = full_data[i, , drop = FALSE])
  e_del[i]    <- full_data$y[i] - yhat_minus
  sigma_minus_i[i] <- summary(loocv_model)$sigma

  # non-studentized LOOCV residual
  ns_loocv[i] <- e_del[i] / sigma_minus_i[i]

  # studentized LOOCV residual (conceptual definition)
```

```r
    stud_loocv[i] <- e_del[i] / (sigma_minus_i[i] / sqrt(1 - leverage[i]))
}

# ------------------------------
# 3) Shortcut studentized residual
# ------------------------------
stud_shortcut <- rstudent(full_model)

# ------------------------------
# 4) Non-studentized full-data residual
# ------------------------------
ns_full <- e_full / sigma_hat  # e_i / sigma_hat

# ------------------------------
# 5) Assemble ASCII-named results
# ------------------------------
residuals_df <- data.frame(
  x = full_data$x,
  leverage = as.numeric(leverage),
  e_full = as.numeric(e_full),
  sigma_hat = as.numeric(sigma_hat),
  e_del = as.numeric(e_del),
  sigma_minus_i = as.numeric(sigma_minus_i),
  ns_residual_e_over_sigma = as.numeric(ns_full),
  ns_loocv_e_del_over_sigma_minus_i = as.numeric(ns_loocv),
  studentized_loocv = as.numeric(stud_loocv),
  studentized_shortcut_rstudent = as.numeric(stud_shortcut)
)

residuals_df_print <- residuals_df[,-c(1,3:6)]
num_cols <- vapply(residuals_df_print, is.numeric, logical(1))
residuals_df_print[num_cols] <- lapply(residuals_df_print[num_cols], function(z) round(z, 3)

brk <- if (knitr::is_html_output()) "<br/>" else "\\\\"

display_names <- c(
  "$h_{ii}$",
  paste0("Non-studentized Full", brk, "$\\frac{e_i}{\\hat{\\sigma}}$"),
  paste0("Non-studentized LOOCV", brk, "$\\frac{e_{i,-i}}{\\hat{\\sigma}_{-i}}$"),
  paste0("Studentized LOOCV", brk, "$\\frac{e_{i,-i}}{\\hat{\\sigma}_{-i}/\\sqrt{1-h_{ii}}}$"
  paste0("Studentized Full (rstudent)", brk, "$\\frac{e_i}{\\hat{\\sigma}_{-i}\\sqrt{1-h_{ii}}
)
```

```
knitr::kable(
  residuals_df_print,
  caption = "Residual variants and related quantities.",
  col.names = display_names,
  escape = FALSE,
  align = rep("r", length(display_names))
)
```

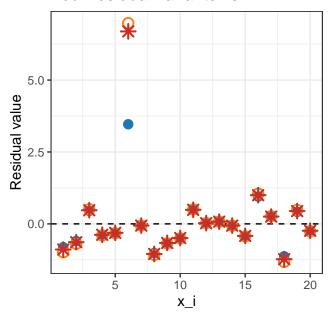Table 1: Residual variants and related quantities.

h_{ii}	Non-studentized Full\\$\frac{e_i}{\hat{\sigma}}$	Non-studentized LOOCV\\$\frac{e_{i,-i}}{\hat{\sigma}_{-i}}$	Studentized LOOCV\\$\frac{e_{i,-i}}{\hat{\sigma}_{-i}/\sqrt{1-h_{ii}}}$	Studentized Full (rstudent)\\$\frac{e_i}{\hat{\sigma}_{-i}\sqrt{1-h_{ii}}}$
0.186	-0.810	-0.989	-0.892	-0.892
0.159	-0.594	-0.694	-0.637	-0.637
0.135	0.450	0.509	0.474	0.474
0.114	-0.367	-0.403	-0.380	-0.380
0.095	-0.305	-0.328	-0.312	-0.312
0.080	3.464	6.979	6.693	6.693
0.068	-0.060	-0.062	-0.060	-0.060
0.059	-1.012	-1.078	-1.046	-1.046
0.053	-0.655	-0.681	-0.663	-0.663
0.050	-0.490	-0.505	-0.492	-0.492
0.050	0.487	0.502	0.489	0.489
0.053	0.024	0.025	0.024	0.024
0.059	0.076	0.079	0.076	0.076
0.068	-0.060	-0.063	-0.061	-0.061
0.080	-0.411	-0.436	-0.418	-0.418
0.095	0.948	1.048	0.997	0.997
0.114	0.244	0.269	0.253	0.253
0.135	-1.127	-1.320	-1.228	-1.228
0.159	0.417	0.484	0.444	0.444
0.186	-0.222	-0.265	-0.239	-0.239

```
# One-panel plot of all four residual variants vs x_i with distinct shapes
plot_df <- residuals_df %>%
  select(
    x,
    ns_residual_e_over_sigma,
    ns_loocv_e_del_over_sigma_minus_i,
```

```r
    studentized_loocv,
    studentized_shortcut_rstudent
  ) %>%
  pivot_longer(
    cols = -x,
    names_to = "residual_type",
    values_to = "residual_value"
  )

# Shapes (ASCII only)
# - ns_loocv_e_del_over_sigma_minus_i: hollow circle (1)
# - studentized_loocv: asterisk (8)
# - studentized_shortcut_rstudent: asterisk (8)
# - ns_residual_e_over_sigma: solid circle (16)
shape_map <- c(
  ns_residual_e_over_sigma = 16,                 # solid circle
  ns_loocv_e_del_over_sigma_minus_i = 1,         # hollow circle
  studentized_loocv = 8,                         # asterisk
  studentized_shortcut_rstudent = 8              # asterisk
)

# Readable legend labels
labels_map <- c(
  ns_residual_e_over_sigma = "Non-studentized Full Data",
  ns_loocv_e_del_over_sigma_minus_i = "Non-studentized LOOCV",
  studentized_loocv = "Studentized LOOCV",
  studentized_shortcut_rstudent = "Studentized Full Data"
)

# Four distinct colors
color_map <- c(
  ns_residual_e_over_sigma = "#1f77b4",          # blue
  ns_loocv_e_del_over_sigma_minus_i = "#ff7f0e", # orange
  studentized_loocv = "#2ca02c",                 # green
  studentized_shortcut_rstudent = "#d62728"      # red
)

ggplot(
  plot_df,
  aes(x = x, y = residual_value,
      shape = residual_type, color = residual_type, group = residual_type)
) +
```

```r
geom_hline(yintercept = 0, linetype = "dashed") +
geom_point(size = 3, stroke = 1) +
scale_shape_manual(
  values = shape_map,
  breaks = names(labels_map),
  labels = unname(labels_map),
  name = "Residual Type"
) +
scale_color_manual(
  values = color_map,
  breaks = names(labels_map),
  labels = unname(labels_map),
  name = "Residual Type"
) +
labs(
  title = "Four residual variants vs x",
  x = "x_i",
  y = "Residual value"
) +
theme_bw() +
theme(
  legend.position = "right",
  legend.title = element_text(face = "bold")
)
```

Four residual variants vs x