

why df of SSreg is k, not n-k-1

Excellent question — this is a subtle but very important point in regression and ANOVA theory. Let's unpack it carefully.

◆ The setup

Suppose we have a multiple linear regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n$$

We partition the total sum of squares (SST) as:

$$\text{SST} = \text{SSR} + \text{SSE}$$

- **SST (Total Sum of Squares):** measures total variation in y
 - **SSR (Regression Sum of Squares):** variation explained by the regression model
 - **SSE (Error or Residual Sum of Squares):** unexplained variation (the residuals)
-

◆ What each term “costs” in degrees of freedom

Component	Formula (conceptual)	DF	Explanation
SST	$\sum(y_i - \bar{y})^2$	$n - 1$	1 d.f. lost for estimating the mean
SSE	$\sum(y_i - \hat{y}_i)^2$	$n - k - 1$	we estimated $k + 1$ parameters ($\beta_0, \beta_1, \dots, \beta_k$)
SSR	$\sum(\hat{y}_i - \bar{y})^2$	k	it measures the improvement in fit due to the k predictors beyond the intercept

◆ Why SSR has k degrees of freedom

Think in terms of how many *independent pieces of information* are used to explain y :

- The fitted values \hat{y}_i lie in a $k + 1$ -dimensional subspace (one dimension per coefficient, including the intercept).
- The mean \bar{y} lies in a 1-dimensional subspace (a constant vector).

- The difference between these two spaces (the “regression space” beyond the mean) therefore has **dimension k** .

So the **regression sum of squares** corresponds to the extra k dimensions gained when moving from just the intercept-only model to the full model with k predictors.

Hence:

$$\text{df}_{\text{reg}} = (k + 1) - 1 = k$$

◆ Intuitive way to see it

- Total variation: $n - 1$ degrees of freedom
- Residual variation: $n - (k + 1)$ degrees of freedom
- Regression variation: the difference between the two

$$\text{df}_{\text{reg}} = (n - 1) - (n - k - 1) = k$$

◆ Summary

Quantity	Definition	Degrees of freedom
Total (SST)	$\sum(y_i - \bar{y})^2$	$n - 1$
Regression (SSR)	$\sum(\hat{y}_i - \bar{y})^2$	k
Error (SSE)	$\sum(y_i - \hat{y}_i)^2$	$n - k - 1$

And:

$$(n - 1) = k + (n - k - 1)$$

balances perfectly.

Would you like me to **illustrate this with a small numerical example (say, n=6, k=2)** so you can see the numbers and df line up concretely?



why df of SSreg is k , not $n-(k+1)-1$

That's a very sharp question — and it shows you're thinking deeply about *what degrees of freedom really mean* in regression. Let's break it down carefully so the logic is crystal clear.

◆ The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

We have:

- n observations
 - k predictors
 - $k + 1$ parameters in total (including the intercept)
-

◆ Three sums of squares

We decompose total variation:

$$\text{SST} = \text{SSR} + \text{SSE}$$

where:

Symbol	Name	Definition	DF
SST	Total	$\sum(y_i - \bar{y})^2$	$n - 1$
SSR	Regression	$\sum(\hat{y}_i - \bar{y})^2$	k
SSE	Error	$\sum(y_i - \hat{y}_i)^2$	$n - k - 1$

and indeed:

$$(n - 1) = k + (n - k - 1)$$

perfectly balances.

◆ Why SSR has only k , not $n - (k + 1) - 1$

The confusion here usually comes from thinking:

"We estimate $k + 1$ parameters, so shouldn't we lose $k + 1$ degrees of freedom again when computing SSR?"

No — because SSR measures the **gain in fit** when we go from a *simpler model* (intercept only) to the *full model* (intercept + k predictors).

Step 1. Start with the simplest model (only intercept)

This model fits:

$$\hat{y}_i = \bar{y}$$

- It estimates only 1 parameter (β_0)
 - It uses up **1 degree of freedom**
 - So residual df for this model = $n - 1$
-

Step 2. Add k predictors

Now the full model estimates $k + 1$ parameters:

- It uses up **$k + 1$ degrees of freedom**
 - So residual df = $n - (k + 1)$
-

Step 3. Compare the two models

When you add k predictors, you go from 1 parameter $\rightarrow k + 1$ parameters.

So you've gained **k extra fitted dimensions**.

That's exactly what the regression sum of squares measures — the improvement in model fit due to those k additional predictors.

Therefore:

$$\text{df}_{\text{reg}} = (k + 1) - 1 = k$$

◆ Geometric interpretation (optional but enlightening)

- All possible \hat{y} vectors lie in a subspace of dimension $k + 1$ (the column space of the design matrix X).
 - The intercept-only model is a 1-dimensional subspace (the line spanned by the constant vector of 1s).
 - The space “added” when including k predictors has dimension k .
 - SSR corresponds to the projection of y onto that k -dimensional added space — hence **k degrees of freedom**.
-

✓ Summary

$$df_{\text{total}} = n - 1$$

$$df_{\text{residual}} = n - (k + 1)$$

$$df_{\text{regression}} = (n - 1) - [n - (k + 1)] = k$$

That's why **SSR has k degrees of freedom**, not $n - (k + 1) - 1$.

It's the *difference in dimensionality* between two nested models — not an additional reduction based on total observations.

Would you like me to show this visually with a simple example (like $n = 6, k = 2$) and a small R table showing how the df are split?

