

Fix in R: there are aliased coefficients in the model

Authored by
Mohammed looti

November 2, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Fix in R: there are aliased coefficients in the model*.
PSYCHOLOGICAL STATISTICS. Retrieved from
<https://statistics.arabpsychology.com/?p=8161>

Decoding the "Aliased Coefficients" Error in Statistical Modeling

The statistical programming environment [R](#) serves as an indispensable tool for developing sophisticated [regression models](#) across various scientific disciplines. Analysts rely on R's robust capabilities to estimate relationships between variables and perform critical post-estimation diagnostics. However, a specific and highly disruptive error can halt this process: the appearance of "aliased coefficients." This issue typically arises during model validation steps, often when attempting to assess the stability of the predictors using tools like the Variance Inflation Factor (VIF).

When attempting to run diagnostics, you might encounter the following precise error message:

Error in vif.default(model) : there are aliased coefficients in the model

This message is not a mere warning about high correlation; it is a definitive signal of **perfect multicollinearity**. This condition indicates a fundamental mathematical flaw in the model specification: two or more predictor variables are perfectly linearly dependent, meaning one variable can be derived exactly from the others. When this occurs, the underlying mathematics required to estimate the parameters fail, rendering the solution indeterminate. Understanding and resolving this core issue is paramount for achieving a valid and interpretable statistical model.

The terminology "aliased coefficients" stems from the fact that, mathematically, the effects of the perfectly correlated variables cannot be separated. In essence, one predictor variable acts as an **alias** for another. The statistical software cannot uniquely attribute the outcome variance to the individual predictors, leading to a singular design matrix. The immediate consequence is the inability to compute unique regression coefficients, necessitating the removal of the redundant variable to restore mathematical stability.

The Mathematical Catastrophe: Perfect Multicollinearity

To appreciate the severity of aliased coefficients, one must understand the role of [multicollinearity](#) in [regression analysis](#). Multicollinearity describes the degree to which independent variables in a multiple regression model are correlated with one another. While some minor correlation among predictors is common and usually tolerable, the presence of **perfect multicollinearity**--where the correlation coefficient is exactly 1 or -1--is statistically fatal.

In the context of linear algebra, fitting a linear regression model requires solving a system of equations, typically involving the inversion of the design matrix ($X'X$). If perfect multicollinearity exists, the columns of the design matrix are linearly dependent. This dependency causes the determinant of the ($X'X$) matrix to be zero, meaning the matrix is singular and therefore non-

invertible. If the matrix cannot be inverted, the standard solution for the Ordinary Least Squares (OLS) coefficients, which involves $(X'X)^{-1}X'Y$, simply cannot be computed.

The failure of the [Variance Inflation Factor](#) (VIF) function is a direct consequence of this singularity. The calculation of the VIF for any predictor involves fitting a secondary regression where that predictor is the outcome and all other predictors are the inputs. If perfect linear dependency exists, this secondary regression also fails due to the same non-invertible matrix problem, immediately generating the "aliased coefficients" error. This error confirms that the model specification violates the basic mathematical assumptions necessary for estimation and diagnostic procedures.

It is crucial to differentiate perfect multicollinearity from high, but imperfect, correlation. High correlation (e.g., $r = 0.95$) inflates the standard errors and makes coefficients unstable, but the model remains mathematically solvable. Perfect correlation ($r = 1.0$) means the model is mathematically insoluble until the redundancy is eliminated. The resolution requires surgical removal of the redundant predictor to ensure the design matrix is full rank, restoring stability for subsequent analysis and interpretation.

Practical Manifestation: Reproducing the Aliasing Error in R

To solidify the concept, let us deliberately construct a scenario in [R](#) that guarantees perfect linear dependence. In this example, we define three predictor variables, `x1`, `x2`, and `x3`, and set `x3` to be an exact multiple of `x2`. By doing so, we ensure that `x2` and `x3` contain identical informational content regarding their relationship with the outcome variable.

We begin by defining our dataset and fitting a standard linear model using the `lm()` function, including all three predictors. Note that some statistical packages, including `lm()` in R, may process this initial step without an error, as they might implicitly drop one of the dependent variables. However, the underlying mathematical instability persists and is instantly revealed when we attempt to run diagnostic checks that require matrix inversion, such as the VIF calculation.

#make this example reproducible

```
set.seed(0)
```

```
#define data where x3 is perfectly dependent on x2
```

```
x1 <- rnorm(100)
```

```
x2 <- rnorm(100)
```

```
x3 <- x2*3
```

```
y <- rnorm(100)
```

```
#fit regression model including the redundant variable
```

```
model <- lm(y~x1+x2+x3)
```

Following the model fit, we utilize the `vif()` function, typically provided by the popular `car` package, to calculate the [VIF](#) values for each predictor. The VIF is designed to quantify the degree of [multicollinearity](#). In the instance of perfect aliasing, the function cannot proceed with its required matrix operations and immediately terminates execution, throwing the error that confirms our setup has resulted in perfect dependency:

library(car)

```
#attempt to calculate VIF values  
vif(model)
```

```
Error in vif.default(model) : there are aliased coefficients in the model
```

This error message explicitly confirms that the computational procedure failed because at least two coefficients in the model are perfectly linked, making them indistinguishable. At this juncture, further meaningful statistical interpretation or inference is impossible until the source of the aliasing is identified and rectified.

Pinpointing the Redundancy using the Correlation Matrix

Once the "aliased coefficients" error has confirmed the existence of perfect dependency, the immediate diagnostic task is to identify precisely which predictor variables are responsible. While the error confirms the problem, it does not specify the culprits. The most reliable and straightforward method for identifying perfect linear relationships is to examine the [correlation matrix](#) of all independent variables.

To generate this matrix in [R](#), we must first consolidate our independent variables (`x1`, `x2`, and `x3`) into a single data structure, along with the dependent variable (`y`). We then apply the `cor()` function to this data frame. The key objective is to scan the resulting matrix for correlation coefficients that possess a magnitude of exactly 1.0 (or -1.0).

#place variables in data frame

```
df <- data.frame(x1, x2, x3, y)
```

```
#create correlation matrix for data frame  
cor(df)
```

```
x1 x2 x3 y  
x1 1.00000000 0.126886263 0.126886263 0.065047543  
x2 0.12688626 1.000000000 1.000000000 -0.009107573  
x3 0.12688626 1.000000000 1.000000000 -0.009107573
```

```
y 0.06504754 -0.009107573 -0.009107573 1.000000000
```

Careful examination of the output confirms the perfect dependency. Specifically, the intersection of row `x2` and column `x3` (and their symmetrical counterparts) shows a correlation coefficient of exactly 1.000000000. This numerical confirmation proves that `x2` and `x3` are perfectly collinear. Since one is merely a linear transformation of the other, they are statistically redundant, confirming they are the source of the aliasing error. This diagnostic step is critical because it moves the analysis from confirming the error exists to identifying the exact structural flaw in the data.

Resolving the Model Instability: Variable Removal

The resolution for aliased coefficients is definitive and straightforward: one of the perfectly correlated variables must be removed from the [regression model](#) specification. This corrective action is justified because, by definition, the information contained in the redundant variable is completely encompassed by the remaining variable. Therefore, removing one predictor results in absolutely no loss of explanatory power or information content for the model.

In our illustrative example, having confirmed that `x2` and `x3` are perfectly dependent, we choose to eliminate `x3`. We then refit the linear model, using only `x1` and `x2` as predictors. The success of this intervention is immediately verifiable by rerunning the VIF diagnostic check.

library(car)

```
#make this example reproducible
```

```
set.seed(0)
```

```
#define data (same as before)
```

```
x1 <- rnorm(100)
```

```
x2 <- rnorm(100)
```

```
x3 <- x2*3
```

```
y <- rnorm(100)
```

```
#fit regression model, removing the redundant variable x3
```

```
model <- lm(y~x1+x2)
```

```
#calculate VIF values for predictor variables in the corrected model
```

```
vif(model)
```

```
x1 x2
```

```
1.016364 1.016364
```

As evidenced by the successful execution of the `vif()` function, the removal of the aliased coefficient has stabilized the model. The output now provides valid [VIF](#) values for the remaining predictors. Since these VIF values are extremely close to 1.0 (indicating minimal correlation), we can confidently conclude that the perfect [multicollinearity](#) has been resolved. The model is now mathematically sound and ready for reliable estimation and interpretation.

Best Practices for Preemptively Avoiding Aliasing

While the example above demonstrated a clear-cut, intentional case of perfect correlation, aliasing often manifests in more subtle ways when dealing with complex data structures, especially those involving categorical variables, transformations, or interaction terms. Establishing preventative practices can save significant diagnostic time later.

Common structural mistakes that inadvertently lead to perfect collinearity and aliased coefficients include:

The Dummy Variable Trap: When modeling a nominal categorical variable with N categories, $N-1$ dummy variables are required. Including all N dummy variables along with the model intercept creates perfect dependency, as the N th category is perfectly defined by the absence of the other $N-1$ categories. One category must always be excluded as the reference group.

Including Redundant Transformations: Adding both the raw scores of a predictor and a simple linear transformation (like standardized scores, or calculating a variable in both inches and centimeters) guarantees perfect aliasing. Only one representation is necessary.

Perfectly Collinear Interaction Terms: This occurs when an interaction term (e.g., $A * B$) is perfectly dependent on its constituent parts due to restrictions in the data. For example, if variable A only ever has a value greater than zero when variable B is exactly 1, the interaction term $A * B$ will be perfectly collinear with A .

A robust preventative measure, especially before fitting intricate models in [R](#), is to preemptively inspect the [correlation matrix](#) of all planned predictor variables. Identifying coefficients of exactly 1.0 or -1.0 before model fitting allows the researcher to correct the model specification immediately, ensuring the stability and invertibility of the design matrix. By addressing perfect dependencies early, analysts can avoid the runtime failure signaled by the "aliased coefficients" error and proceed confidently with stable statistical inference.