How to Fix the “glm.fit: fitted probabilities numerically 0 or 1 occurred” Error in R

Authored by

stats writer

December 4, 2025

The warning message `glm.fit: fitted probabilities numerically 0 or 1 occurred` is a common and often misunderstood notification encountered when performing logistic regression analysis, particularly within the R statistical environment. While many users initially mistake this for a critical failure, it is classified as a warning because the model, utilizing the Iteratively Reweighted Least Squares (IRLS) algorithm, attempts to fit the data but encounters challenges related to the separation of classes. This phenomenon, known as complete or quasi-separation, fundamentally affects the estimation of model coefficients, leading to coefficients that are infinitely large or numerically extreme. Understanding this warning is essential for ensuring the reliability and interpretability of your predictive model.

The core issue arises when a perfect or near-perfect linear dependency exists between one or more predictor variables and the binary outcome variable. When this separation is present, the maximum likelihood estimates for the coefficients do not exist in the traditional sense, resulting in the computational process pushing the fitted probabilities to the numerical extremes of zero or one. To effectively resolve this, we must adopt data preparation techniques, such as identifying and managing highly influential outliers, or employ more robust modeling strategies like penalized maximum likelihood estimation (PMLE) or regularization, which stabilize the coefficient estimates.

Understanding the `glm.fit` Warning

In the process of fitting generalized linear models (GLMs) in R, specifically those using the binomial family (for logistic regression), the statistical engine relies on iterative optimization. When this optimization process encounters an insurmountable situation--where a perfect fit can be achieved by assigning infinite weight to a predictor--the following standard warning message is emitted:

Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred

This warning indicates that for at least one observation in your dataset, the model predicts the outcome with near-certainty (a probability indistinguishable from 0 or 1). Computationally, this means the algorithm is trying to reach an infinite coefficient value to perfectly separate the two outcome classes based on the given predictors, a condition known as complete separation.

It is paramount to recognize that this is strictly a **warning message**, not an error that halts execution. R successfully returns a fitted model, but the accompanying summary output will often show highly inflated standard errors and extreme coefficient estimates, rendering the model untrustworthy for statistical inference. While the model may predict perfectly on the training data, these estimates are unstable and often useless for generalizing to new data.

The Statistical Problem: Complete and Quasi-Separation

The underlying statistical issue driving the numerical instability is the concept of separation. Separation occurs when the predictor variables perfectly predict the outcome variable. This is common in smaller datasets or datasets where a unique combination of predictor values corresponds exclusively to one outcome class.

There are two primary forms of separation:

Complete Separation: This is the most severe case. A linear combination of the predictors perfectly separates the observations into the two outcome categories (0 and 1). If you can draw a hyperplane in the feature space that perfectly divides the positive instances from the negative ones, you have complete separation.
Quasi-Separation (or Partial Separation): This is more subtle. While no perfect separation exists across the entire dataset, a specific range or combination of predictor values perfectly predicts the outcome for a subset of observations. The maximum likelihood estimates still tend toward infinity, but they are technically achievable through specific predictor combinations, often involving outliers.

When separation occurs, the log-likelihood function becomes maximized only as the coefficients approach plus or minus infinity. Since computers cannot represent infinity, the iterative fitting process stops when the change in the coefficients falls below a very small tolerance, resulting in the extreme coefficients and the warning regarding probabilities of 0 or 1.

Reproducing the Warning Message in R

To illustrate how data structure can trigger this numerical issue, consider a highly simplified dataset where the predictor variables almost perfectly partition the outcome variable. We will fit a logistic regression model using the standard `glm()` function in R, which uses the default maximum likelihood estimation method.

The following code block demonstrates the setup and the resultant warning generated by R:

```
#create data frame
df <- data.frame(y = c(0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1),
x1 = c(3, 3, 4, 4, 3, 2, 5, 8, 9, 9, 9, 8, 9, 9, 9),
x2 = c(8, 7, 7, 6, 5, 6, 5, 2, 2, 3, 4, 3, 7, 4, 4))

#fit logistic regression model
model <- glm(y ~ x1 + x2, data=df, family=binomial)

#view model summary
summary(model)
```

Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred

Call:
glm(formula = y ~ x1 + x2, family = binomial, data = df)

Deviance Residuals:
Min 1Q Median 3Q Max
-1.729e-05 -2.110e-08 2.110e-08 2.110e-08 1.515e-05

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -75.205 307338.933 0 1
x1 13.309 28512.818 0 1
x2 -2.793 37342.280 0 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2.0728e+01 on 14 degrees of freedom
Residual deviance: 5.6951e-10 on 12 degrees of freedom
AIC: 6

Number of Fisher Scoring iterations: 24

As demonstrated above, the model successfully completed its fitting process, but the warning indicates the numerical issue. Crucially, examine the coefficients table in the output. The estimated coefficients (e.g., -75.205 for the intercept, 13.309 for x1) are extremely large in magnitude, and their corresponding standard errors are astronomically high (e.g., 307,338.933). This is the hallmark of separation, indicating unreliable parameter estimates.

Interpreting the Extreme Fitted Probabilities

The warning specifically refers to fitted probabilities being numerically 0 or 1. This means the model has assigned near-perfect certainty to the outcomes of the observations in the training data, a sign of severe overfitting caused by the separation issue. We can confirm this by using the fitted model to generate predictions on the original data frame, specifying `type="response"` to obtain probabilities.

Upon reviewing the prediction output, it becomes clear that the model is making predictions at the numerical limits of machine precision:

#use fitted model to predict response values

df$y_pred = predict(model, df, type="response")

#view updated data frame
df

y x1 x2 y_pred
1 0 3 8 2.220446e-16
2 0 3 7 2.220446e-16
3 0 4 7 2.220446e-16
4 0 4 6 2.220446e-16
5 0 3 5 2.220446e-16
6 0 2 6 2.220446e-16
7 0 5 5 1.494599e-10
8 1 8 2 1.000000e+00
9 1 9 2 1.000000e+00
10 1 9 3 1.000000e+00
11 1 9 4 1.000000e+00
12 1 8 3 1.000000e+00
13 1 9 7 1.000000e+00
14 1 9 4 1.000000e+00
15 1 9 4 1.000000e+00

Observations where the true outcome y is 0 are predicted with probabilities near 2.22e-16, which is numerically zero. Observations where y is 1 are predicted with a probability of exactly 1.0. This demonstrates the model's complete inability to generalize beyond this specific, small dataset, confirming that while the model converges, the resulting parameter estimates are statistically meaningless due to the data's inherent separability.

Strategy 1: Data-Centric Solutions (Increasing Robustness)

When facing the separation warning, the first and often most effective approach is to re-examine the data itself. Issues of separation are frequently symptoms of insufficient data or structural anomalies like severe outliers.

We can adopt several data-centric strategies:

Increase the Sample Size: In many practical scenarios, this warning appears when working with small datasets where accidental separation is highly likely. If the sample size is inadequate, there simply isn't enough variation or overlap in the predictor space to stabilize the maximum likelihood estimation. By gathering more observations, especially those that fall near the separation

boundary, the model gains the necessary data points to estimate non-infinite coefficients, thereby resolving the quasi-separation issue.

Identify and Manage Outliers: Separation can be caused by just a few highly influential data points that pull the regression line toward an extreme position. If only a small number of observations have fitted probabilities close to 0 or 1, those observations might be considered outliers or leverage points. Analyzing the data for influential cases (e.g., using Cook's distance or leverage plots) and either removing them or ensuring they are valid representatives of the population often resolves the warning effectively.

Transform Predictors or Address Collinearity: Although the primary cause is separation, high correlation (collinearity) combined with small sample size can exacerbate the problem, making separation more likely. Transforming continuous predictors (e.g., using log transformations) or creating interaction terms can sometimes redistribute the data points sufficiently to break the linear dependency that causes separation.

The goal of these data-centric solutions is to introduce sufficient overlap between the two outcome classes in the predictor space, allowing the maximum likelihood estimates to converge to finite values.

Strategy 2: Model-Centric Solutions (Implementing Regularization)

If data manipulation is not feasible (e.g., due to limitations on sample size), or if the separation is inherent to the population structure, model modifications are necessary. The most powerful method for dealing with separation without altering the data is the use of penalized maximum likelihood estimation.

The standard maximum likelihood method attempts to maximize the likelihood function without constraints, which fails when separation occurs. Regularization addresses this by adding a penalty term to the log-likelihood function. This penalty discourages the coefficients from growing infinitely large, thereby stabilizing the estimates and ensuring convergence.

Firth Regression (Bias Reduction): A popular technique is Firth's logistic regression, which applies a penalty based on the Jeffreys prior. This method effectively biases the estimates slightly toward zero, solving the problem of infinite coefficients resulting from separation. In R, specialized packages like `logistf` implement this method explicitly for handling separation.

Ridge or Lasso Regularization: Standardized regularization methods (like Ridge or Lasso, often implemented via packages like `glmnet`) also mitigate separation. Ridge regularization (L2 penalty) is particularly effective because it shrinks the coefficients but keeps all predictors in the model, providing finite and stable estimates even under complete separation. This is a robust alternative to manually dealing with high correlation or small sample size issues.

By implementing a form of regularization, we trade the unbiased, but infinite, maximum likelihood

estimates for slightly biased, but stable and finite, penalized estimates that are statistically meaningful.

Why Ignoring the Warning Can Sometimes Be Acceptable

While fixing the separation issue is generally the best practice, there are specific, limited contexts where ignoring the warning might be justified. However, this decision requires careful consideration and understanding of the model's purpose.

If the model is purely intended for descriptive purposes on the training data, and not for statistical inference or generalization, the warning might be less critical. For instance, if you are simply summarizing how perfectly a complex set of features predicts an outcome in a fixed cohort, the fact that the fitted probabilities are 0 or 1 simply reflects the data structure.

However, even in this case, the resulting coefficients and standard errors are invalid. If the goal is prediction on independent data, the model suffering from separation is extremely prone to overfitting. The estimated coefficients are inherently unreliable, and any interpretation of predictor importance based on coefficient magnitude will be flawed. Therefore, while the model output technically exists, relying on it without addressing the separation issue is statistically dubious and generally discouraged.

Conclusion and Best Practices

The warning `glm.fit: fitted probabilities numerically 0 or 1 occurred` is a critical signal that your logistic regression model is suffering from complete or quasi-separation. This invalidates the standard maximum likelihood estimation and leads to unstable, infinite coefficient estimates and meaningless standard errors.

To ensure robust and reliable modeling, always prioritize data quality and sufficient sample representation. Start by investigating potential outliers and ensuring your sample size is adequate. If structural separation persists, the implementation of penalized methods, such as Ridge regression or Firth's correction, offers a mathematically sound way to stabilize the coefficients and produce usable predictive models.

Handling this warning effectively moves your analysis from a simple computational convergence to a statistically valid and interpretable result.