

Understanding the R Warning: “glm.fit: fitted probabilities numerically 0 or 1 occurred” in Logistic Regression

Authored by
Mohammed loot

November 3, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Understanding the R Warning: “glm.fit: fitted probabilities numerically 0 or 1 occurred” in Logistic Regression*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=8951>

In the field of statistical modeling, particularly when utilizing the [R](#) environment, practitioners frequently encounter various warnings that signal potential issues rather than outright errors. Among the most critical yet frequently misunderstood messages is one that appears during the fitting of a [Generalized Linear Model \(GLM\)](#), especially when conducting [logistic regression](#):

Warning message:

glm.fit: fitted probabilities numerically 0 or 1 occurred

This warning is a fundamental diagnostic tool, indicating a critical structural problem between your data and the statistical model being estimated. Although merely classified as a warning and not a fatal error, ignoring it can severely compromise the reliability, stability, and interpretability of your results. If this issue is left unaddressed, the resulting coefficient estimates will likely be highly biased, unstable, and statistically meaningless.

This expert guide delves into the statistical foundation of this warning, illustrates how to replicate the behavior using [R](#) code, and provides a set of robust, practical methodologies for diagnosing and mitigating the issue of separation in real-world data analysis.

The Statistical Foundation: Understanding Data Separation

The warning `fitted probabilities numerically 0 or 1 occurred` serves as a direct alert to the presence of **separation** within your binary outcome data. Separation is a state where the predictor variables in your model can perfectly, or near-perfectly, predict the binary outcome variable (0 or 1). This deterministic relationship breaks the fundamental assumptions required for standard estimation techniques.

Standard [logistic regression](#) relies on [Maximum Likelihood Estimation \(MLE\)](#) to determine coefficients that maximize the likelihood of observing the specific data points. The successful application of MLE requires the model to produce [fitted probabilities](#) (P) that are strictly bounded between 0 and 1. When separation occurs, the likelihood function continuously increases as the coefficient estimates approach positive or negative infinity, meaning the maximum likelihood is never actually reached within the finite parameter space.

Statisticians typically distinguish between two variants of separation, both of which trigger the problematic `glm.fit` warning, signaling model instability:

Complete Separation: This is the most severe scenario. A linear combination of the predictor variables perfectly classifies the outcome. For instance, if a rule such as "Predictor X > 5" always results in Y = 1, and "Predictor X ≤ 5" always results in Y = 0. To achieve this perfect fit, the model attempts to assign infinite coefficients, which mathematically forces the [fitted probabilities](#) to the exact boundaries of 0 or 1.

Quasi-Complete Separation: This scenario is slightly less absolute but equally disruptive. The outcomes are almost perfectly separated, with the exception of a few overlapping data points. Even in this case, the model's iterative algorithm pushes the coefficients towards extremely large values. This results in predicted values that are numerically indistinguishable from 0 or 1 (e.g., $1e-15$ or $1 - 1e-15$), which is sufficient to trigger the numerical convergence warning.

Crucially, when the model attempts to converge with these infinitely large coefficients, the associated standard errors become astronomical. This renders the estimated coefficients unreliable, making standard statistical inference (such as calculating p-values or confidence intervals) fundamentally impossible. The warning message is therefore an essential notification that the standard MLE routine has failed to produce stable and meaningful estimates.

Demonstration: Reproducing the Warning in R

To fully grasp the practical consequence of separation, we can meticulously construct a small dataset in the [R](#) environment where the binary outcome is visibly separable by the included predictors. This exercise clearly demonstrates how deterministic data patterns cause the [logistic regression](#) model to overfit and produce extreme predicted values.

We define a data frame where observations with low values of `x1` and high values of `x2` consistently align with the outcome `y=0`, while the inverse pattern perfectly aligns with `y=1`. This clear partitioning creates quasi-complete separation, guaranteeing the instability we aim to diagnose.

#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 using the binomial family
```

```
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

Although the code executes successfully and produces a summary table, the immediate appearance of the warning message confirms the underlying separation issue. A closer inspection of the coefficients section reveals the telltale signs of instability: the **Estimate** values are excessively large (e.g., -75.205 for the intercept), and the corresponding **Std. Error** values are astronomical, often reaching into the hundreds of thousands. This drastic inflation of standard errors is the direct result of the model's inability to converge to a finite maximum likelihood solution.

Analysis of Pathological Fitted Probabilities

To visually confirm the numerical failure, we can use this unstable model to generate predicted response values based on the original training data. This step clearly shows how the model is forced against the boundaries, unable to find a nuanced probabilistic relationship.

```
#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
```

The resulting `y_pred` column demonstrates the pathological outcome. For observations where the true outcome \bar{y} is 0, the predicted probability is a vanishingly small number (e.g., $2.220446e-16$), which is numerically interpreted as 0. Conversely, for observations where \bar{y} is 1, the predicted probability is exactly 1.0. While this zero residual deviance might superficially suggest a "perfect" model fit, it is merely evidence of overfitting caused by separation. The model is reflecting a deterministic relationship rather than a true probabilistic one, leading directly to the failure of the underlying [MLE](#) procedure.

Strategic Solutions for Handling Separation Warnings

Successfully addressing the separation warning necessitates a methodical approach, involving careful inspection of the data, consideration of sample size constraints, and, if necessary, the implementation of specialized statistical methodologies. The chosen strategy must align with the nature of the data and the ultimate analytical goals.

1. Data Diagnostics and Feature Engineering

The initial and most crucial step is to pinpoint the exact source of the separation within the dataset. Often, separation is caused by specific data artifacts rather than intrinsic population properties.

Investigate [Outliers](#) and Edge Cases: Sometimes, one or two extreme observations or [outliers](#) create the boundary condition necessary for separation. If these points are deemed erroneous or unrepresentative, their removal may resolve the issue, allowing the MLE procedure to converge successfully. However, data removal must always be rigorously justified.

Check for Collinearity: High correlation among independent variables can exacerbate separation issues. Analysts should check for strong collinearity; if redundant predictors are identified, removing or combining them (e.g., via principal components) can stabilize the model.

Modify Categorical Predictors: If a categorical variable has a level that perfectly predicts one outcome (e.g., a rarely occurring demographic category always results in $Y=0$), separation will occur. Solutions include collapsing that category into a broader group or, if the variable is non-essential, removing it entirely.

2. Increasing Sample Size

The issue of separation is frequently a symptom of insufficient data, particularly in studies involving rare events or small populations. Small datasets are far more prone to complete separation because there are simply not enough observations to create the necessary overlap between the two outcome classes.

Where logistically possible, expanding the sample size is the most statistically robust intervention. Collecting more data, especially observations that fall into the ambiguous or overlapping regions between $Y=0$ and $Y=1$, introduces the required uncertainty. This ambiguity prevents the model coefficients from diverging to infinity, thus stabilizing the estimates and significantly reducing the magnitude of the standard errors.

3. Employing Penalized Regression Techniques

When separation is inherent to the population, or when collecting additional data is infeasible, standard MLE must be replaced by advanced penalized estimation methods. The gold standard technique for stabilizing [logistic regression](#) under conditions of separation is **Firth's Correction** (or penalized maximum likelihood).

Firth regression resolves the issue by introducing a penalty term (specifically, the Jeffreys prior) into the likelihood function. This penalty subtly biases the coefficient estimates toward zero, effectively preventing them from diverging toward infinity. This crucial modification ensures that finite, stable standard errors are produced, thereby solving the convergence problem inherent in separation.

In the [R](#) environment, implementing Firth regression is straightforward using specialized packages like `logistf`. For persistent or intrinsic separation issues, this robust statistical technique is generally the recommended course of action.

4. Model Re-evaluation and Transformation

If the aforementioned methods do not yield a stable model, the analyst should consider fundamentally altering the model structure or the input features.

Remove Problematic Predictors: If diagnostic analysis isolates the separation to a single, highly influential variable, removing it may be necessary to achieve model convergence. This involves

balancing predictive power against model stability.

Transform Continuous Variables: For continuous features causing quasi-complete separation, transformations--such as logarithmic scaling or grouping the variable into meaningful discrete bins--can introduce the necessary non-linearity or overlap to stabilize the coefficient estimates.

Consider Alternative Classifiers: If the data exhibits true complete separation (i.e., the relationship is fundamentally deterministic), a probabilistic model like [logistic regression](#) based on [MLE](#) may be inappropriate. In such cases, non-probabilistic methods--such as Classification and Regression Trees (CART) or Random Forests--which are designed to handle perfect separation without convergence issues, should be explored.

Conclusion: The Imperative of Addressing the Warning

The warning message `glm.fit: fitted probabilities numerically 0 or 1 occurred` is a critical indicator that your statistical model is suffering from **complete or quasi-complete separation**. While the model may technically produce an output, the resulting coefficients and their standard errors are mathematically unstable and should be considered completely unreliable for scientific inference or prediction.

Data analysts must treat this warning not as a nuisance, but as a mandatory prompt for detailed diagnosis. By systematically identifying the root cause--be it small sample size, extreme [outliers](#), or intrinsic data structure--and applying appropriate mitigation strategies like Firth regression or data refinement, you ensure that your statistical analysis is both robust and statistically valid. Ignoring this warning guarantees that any conclusions drawn from the model will be fundamentally flawed.

Additional Resources for R Diagnostics

For further resources on diagnosing and resolving advanced modeling issues in [R](#), consult the following specialized materials:

How to interpret large standard errors in GLMs.

Using the `logistf` package for penalized regression.

Detailed documentation on the `family=binomial` parameter in the `glm()` function.