

# What is the significance of the message “glm.fit: fitted probabilities numerically 0 or 1 occurred” in statistical analysis?

Authored by  
**stats writer**

May 5, 2024

## RECOMMENDED CITATION

stats writer (2024). *What is the significance of the message “glm.fit: fitted probabilities numerically 0 or 1 occurred” in statistical analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143094>

The message "glm.fit: fitted probabilities numerically 0 or 1 occurred" is significant in statistical analysis as it indicates that there may be a problem with the data or model being used. This message suggests that the probabilities being calculated are either extremely low (close to 0) or extremely high (close to 1), which can be problematic in statistical analysis. It could be a sign of overfitting or a violation of assumptions in the model, both of which can lead to inaccurate results. Therefore, it is important to carefully examine the data and model when this message appears in order to ensure the validity and reliability of the statistical analysis.

## **Handle: glm.fit: fitted probabilities numerically 0 or 1 occurred**

**One warning message you may encounter in R is:**

**Warning message:**

**glm.fit: fitted probabilities numerically 0 or 1 occurred**

**This warning occurs when you fit a logistic regression model and the predicted probabilities of one or more observations in your data frame are indistinguishable from 0 or 1.**

**It's worth noting that this is a warning message and not an error. Even if you receive this error, your logistic regression model will still be fit, but it may be worth analyzing the original data frame to see if there are any outliers causing this warning message to appear.**

**This tutorial shares how to address this warning**

## message in practice.

### How to Reproduce the Warning

Suppose we fit a logistic regression model to the following data frame in R:

```
#create data frame
```

```
df <- data.frame(y = c(0, 0, 0, 0, 0, 0, 0, 1, 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**

**Our logistic regression model is successfully fit to the data, but we receive a warning message that fitted probabilities numerically 0 or 1 occurred.**

**If we use the fitted logistic regression model to make predictions on the response value of the observations in the original data frame, we can see that nearly all of the predicted probabilities are indistinguishable from 0**

and 1:

**#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**

## How to Handle the Warning

**There are three ways to deal with this warning message:**

**(1) Ignore it.**

In some cases, you can simply ignore this warning message because it doesn't necessarily indicate that something is wrong with the logistic regression model. It simply means that one or more observations in the data frame have predicted values indistinguishable from 0 or 1.

**(2) Increase the sample size.**

In other cases, this warning message appears when you're working with small data frames where there's simply not enough data to provide a reliable model fit. To address this error, simply increase the sample size of observations that you feed into the model.

In other cases, this error occurs when there are outliers in the original data frame and where only a small number of observations have fitted probabilities close to 0 or 1. By removing these outliers, the warning

**message often goes away.**

**The following tutorials explain how to handle other warnings and errors in R:**

ARABPSYCHOLOGY.COM