

Logistic Regression Cheat Sheet

The Response – is a categorical variable (we focus here on 2 categories)

The Predictor – can be categorical or numerical

QUESTIONS:

1. Is there a relationship between the response and predictor?

Create a plot of the 2 variables to see

2. How strong is the relationship between the predictor and the response?

Use Residual Deviance. This is a measure of how good the model fits the data. Like RSE in simple linear regression, you compare this to other models to determine the best fit model. Similarly, the smaller the Residual Deviance, the better the fit is.

3. What is the effect of the predictor on the response ?

If Numeric

- As the predictor changes by one unit, the log-odds for being in a category versus the baseline changes by the slope coefficient
 - ◆ *Positive coefficient means log-odds and odds is increasing.* So, it is more likely to be in the category versus the baseline (for the response) as the predictor increases.
 - ◆ *Negative coefficient means log-odds and odds is decreasing.* So, it is less likely to be in the category versus the baseline (for the response) as the predictor increases.

If Categorical

- The slope coefficient refers to the difference in **log-odds for being in the category versus the baseline (for the response)** between the group that is identified by the slope coefficient and its baseline (for the predictor)
 - ◆ *Positive coefficient means that the log-odds is increasing.* This implies that the odds is increasing. So it is more likely to be in the category versus the baseline (for the response) for that group identified by that specific coefficient compared to its baseline (for the predictor).
 - ◆ *Negative coefficient means that the log-odds is decreasing.* This implies that the odds is decreasing so it is less likely to be in the category versus the baseline (for the response) for that group identified by that specific coefficient compared to its baseline (for the predictor)

4. Is the predictor a good predictor of the response?

Here you do a hypothesis test.

Research Hypothesis: The predictor is a good predictor (or there is a difference)

Null Hypothesis: The predictor is NOT a good predictor (or there is no difference)

If $P(|z| > t) < 0.05$, we have evidence to believe our Research hypothesis. We conclude that the predictor is a good predictor of the response.

If $P(|z| > t) > 0.05$, we conclude that the predictor is NOT a good predictor of the response.

5. How good are the predictions based on your model?

Use AIC. Lower AIC implies more accurate predictions. You will use this to compare models to each other, just like you did for simple linear regression with RSE.