

## Practice: Using PROC LOGISTIC to Perform a Multiple Logistic Regression Analysis with Categorical Variables

The insurance company wants to model the relationship between three of a car's characteristics, weight, size, and region of manufacture, and its safety rating. The **stat1.safety** data set contains the data about vehicle safety.

1. Use PROC LOGISTIC to fit a multiple logistic regression model with **Unsafe** as the response variable and **Weight**, **Size**, and **Region** as the predictor variables.
  1. Use the EVENT= option to model the probability of Below Average safety scores.
  2. Specify **Region** and **Size** as classification variables and use reference cell coding. Specify *Asia* as the reference level for **Region**, and 3 (large cars) as the reference level for **Size**.
  3. Request profile likelihood confidence limits, an odds ratio plot, and the effect plot.
  4. Submit the code and view the results.

```
/*st107s03.sas*/
ods graphics on;
proc logistic data=STAT1.safety plots(only)=(effect oddsratio);
  class Region (param=ref ref='Asia')
    Size (param=ref ref='3');
  model Unsafe(event='1')=Weight Region Size / clodds=pl;
  title 'LOGISTIC MODEL (2):Unsafe=Weight Region Size';
run;
```

Here are the [results](#).

2. Do you reject or fail to reject the null hypothesis that all regression coefficients of the model are 0?

The  $p$ -value for the Likelihood Ratio test is  $<.0001$ , and therefore, you reject the null hypothesis.

3. If you reject the global null hypothesis, then which predictors significantly predict safety outcome?

Only **Size** is significantly predictive of **Unsafe**.

4. Interpret the odds ratio for significant predictors.

Only **Size** is significant. The design variables show that Size=1 (Small or Sports) cars have 14.560 times the odds of having a Below Average safety rating compared to the reference category 3 (Large or Sport/Utility). The 95% confidence interval (3.018, 110.732) does not contain 1, implying that the contrast is statistically significant at the 0.05 level.

The contrast from the second design variable is 1.931 (Medium versus Sport/Utility), implying a trend toward greater odds of low safety as size decreases. However, the 95% confidence interval (0.343, 15.182) contains 1, and therefore, the contrast is not statistically significant.