Print

Practice 2.3 (Level 1): Fitting a Lognormal Model

Task

In this practice, you use PROC GLIMMIX to fit a lognormal regression model on the data in **mydata.cars**.

Note: Before you do this practice, you must run the code for practice 2.1 in the same SAS session.

Reminder: Make sure you've defined the **mydata** library.

1. Write a PROC GLIMMIX step to fit a lognormal regression model. Use **Price** as the response variable and **Citympg**, **Citympg^2**, **EngineSize**, **Horsepower**, **Horsepower^2**, and **Weight** as the predictor variables. Use an ODS OUTPUT statement to save the parameter estimates to a data set. Use an OUTPUT statement to create a data set that contains the original predictor variables but not the constructed effects.

   Evaluate the results as follows:

   - Compare the parameter estimates from this lognormal model with the parameter estimates for **LogPrice** that you obtained in practice 2.1.
   - Using the fit statistics produced by PROC GLIMMIX, evaluate the fit of this model compared to the lognormal model that was fit in the last demonstration.

```
ods output ParameterEstimates=params;
proc glimmix data=mydata.cars;
   effect p_City=polynomial(Citympg / degree=2
       standardize(method=moments)=center);
   effect p_hp=polynomial(Horsepower / degree=2
       standardize(method=moments)=center);
   model price = p_City EngineSize p_hp Weight
                 / dist=lognormal
       solution;
   output out=out pred=pred;
   id Manufacturer Model Price Citympg Horsepower
      EngineSize;
run;
```

   The results indicate the following:

   - The parameter estimates for the lognormal model (in the Parameter Estimates table) are identical to those obtained using ordinary least squares regression with the response variable **LogPrice**.
   - The information criteria statistics (in the Fit Statistics table) are all larger for this model than for the previous model with **Hwympg**, **Hwympg^2**, and **Horsepower** as predictor variables. The current model might be more complex than necessary.

2. Generate estimates of the mean of **Price** by back-transforming the predicted values from the lognormal regression model. Store the results in a data set, and print the following variables: **Manufacturer**, **Model**, **Citympg**, **Horsepower**, **EngineSize**, **Price**, **Estimate**, and **Difference** (the difference between **Price** and **Estimate**). Plot the difference between the observed and estimated versus the predicted values. Do these estimates appear to be better than those obtained from the lognormal model developed in the last demonstration?

```
data _null_;
   set params;
```

```
   if Effect='Scale' then call symput('var', Estimate);
run;

data out;
   set out;
   Estimate=exp(pred + &var/2);
   Difference = Price - Estimate;
run;

proc print data=out;
   var Manufacturer Model Citympg Horsepower EngineSize
       Price Estimate Difference;
run;

proc sgplot data=out;
   scatter y=Difference x=Estimate;
   xaxis min=0 max=55;
   yaxis min=-20 max=20;
   refline 0;
run;
```

Based on the results, there do not appear to be any extreme values of **Difference** (in other words, residuals on the original scale) for this model. In that sense, it might be a better model than the one developed in the last demonstration.

Hide Solution

Close