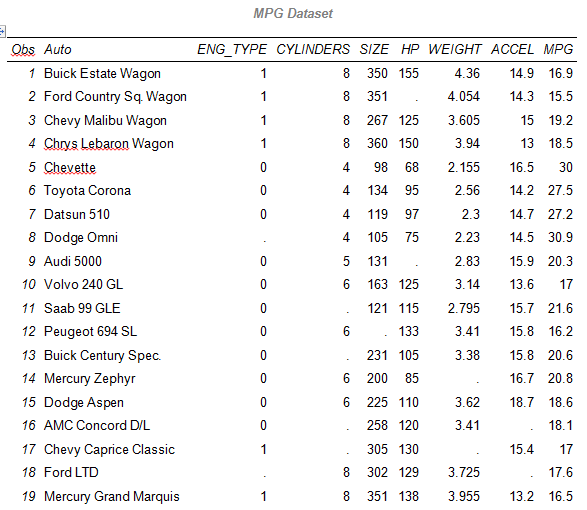
| *Obs* | *Auto* | *ENG\_TYPE* | *CYLINDERS* | *SIZE* | *HP* | *WEIGHT* | *ACCEL* | *MPG* |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *1* | Buick Estate Wagon | 1 | 8 | 350 | 155 | 4.36 | 14.9 | 16.9 |
| *2* | Ford Country Sq. Wagon | 1 | 8 | 351 | . | 4.054 | 14.3 | 15.5 |
| *3* | Chevy Malibu Wagon | 1 | 8 | 267 | 125 | 3.605 | 15 | 19.2 |
| *4* | Chrys Lebaron Wagon | 1 | 8 | 360 | 150 | 3.94 | 13 | 18.5 |
| *5* | Chevette | 0 | 4 | 98 | 68 | 2.155 | 16.5 | 30 |
| *6* | Toyota Corona | 0 | 4 | 134 | 95 | 2.56 | 14.2 | 27.5 |
| *7* | Datsun 510 | 0 | 4 | 119 | 97 | 2.3 | 14.7 | 27.2 |
| *8* | Dodge Omni | . | 4 | 105 | 75 | 2.23 | 14.5 | 30.9 |
| *9* | Audi 5000 | 0 | 5 | 131 | . | 2.83 | 15.9 | 20.3 |
| *10* | Volvo 240 GL | 0 | 6 | 163 | 125 | 3.14 | 13.6 | 17 |
| *11* | Saab 99 GLE | 0 | . | 121 | 115 | 2.795 | 15.7 | 21.6 |
| *12* | Peugeot 694 SL | 0 | 6 | . | 133 | 3.41 | 15.8 | 16.2 |
| *13* | Buick Century Spec. | 0 | . | 231 | 105 | 3.38 | 15.8 | 20.6 |
| *14* | Mercury Zephyr | 0 | 6 | 200 | 85 | . | 16.7 | 20.8 |
| *15* | Dodge Aspen | 0 | 6 | 225 | 110 | 3.62 | 18.7 | 18.6 |
| *16* | AMC Concord D/L | 0 | . | 258 | 120 | 3.41 | . | 18.1 |
| *17* | Chevy Caprice Classic | 1 | . | 305 | 130 | . | 15.4 | 17 |
| *18* | Ford LTD | . | 8 | 302 | 129 | 3.725 | . | 17.6 |
| *19* | Mercury Grand Marquis | 1 | 8 | 351 | 138 | 3.955 | 13.2 | 16.5 |
| *20* | Dodge St Regis | 1 | 8 | 318 | 135 | 3.83 | . | 18.2 |
| *21* | Ford Mustang 4 | 0 | 4 | 140 | . | 2.585 | 14.4 | 26.5 |
| *22* | Ford Mustang Ghia | 1 | 6 | 171 | . | 2.91 | 16.6 | 21.9 |
| *23* | Mazda GLC | 0 | 4 | 86 | 65 | . | 15.2 | 34.1 |
| *24* | Dodge Colt | 0 | 4 | 98 | 80 | 1.915 | 14.4 | 35.1 |
| *25* | AMC Spirit | 0 | 4 | 121 | . | 2.67 | 15 | 27.4 |
| *26* | VW Scirocco | 0 | 4 | 89 | 71 | 1.99 | 14.9 | 31.5 |
| *27* | Honda Accord | 0 | 4 | 98 | 68 | . | 16.6 | 29.5 |
| *28* | Buick Skylark | 0 | 4 | 151 | 90 | 2.67 | 16 | 28.4 |
| *29* | Chevy Citation | 1 | 6 | 173 | 115 | 2.595 | 11.3 | 28.8 |
| *30* | Olds Omega | 1 | 6 | 173 | 115 | 2.7 | 12.9 | 26.8 |
| *31* | Pontiac Phoenix | 0 | 4 | 151 | 90 | 2.556 | 13.2 | 33.5 |
| *32* | Plymouth Horizon | 0 | 4 | 105 | 70 | 2.2 | 13.2 | 34.2 |
| *33* | Datsun 210 | . | 4 | 85 | 65 | 2.02 | 19.2 | 31.8 |
| *34* | Fiat Strada | 0 | 4 | 91 | 69 | 2.13 | 14.7 | 37.3 |
| *35* | VW Dasher | 0 | 4 | . | 78 | . | 14.1 | 30.5 |
| *36* | Datsun 810 | `0 | 6 | . | 97 | 2.815 | 14.5 | 22 |
| *37* | BMW 320i | 0 | 4 | 121 | 110 | . | . | 21.5 |
| *38* | VW Rabbit | 0 | 4 | 89 | 71 | 1.925 | 14 | 31.9 |



We have a quite a few of missing values throughout the CARMPG data set.

SAS Code: Regression 1

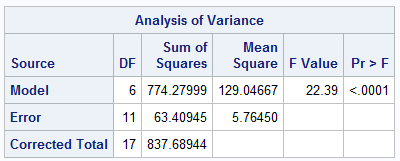
The model of interest is linear regression

First use PROC REG for an analysis that uses (by default) list-wise deletion

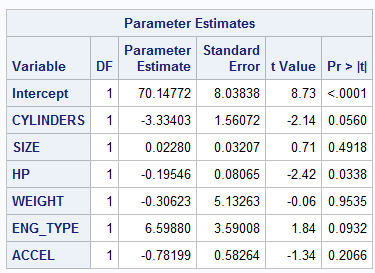
|  |
| --- |
| \* what data is missing from dataset?;  \* use PROC REG with listwise deletion;  title 'Predicting MPG (initial)';  **proc** **reg** data=cars;  model mpg = cylinders size hp weight eng\_type accel;  **run**;  **quit**; |

Based on PROC REG, out of 38 observations, 20 observations were deleted because of missing values, and only 18 observations were kept. This should cause us some concern.





Parameter Estimates (Using Listwise Deletion)



The question is how are are the parameter estimates for the linear regression without creating an imputation.

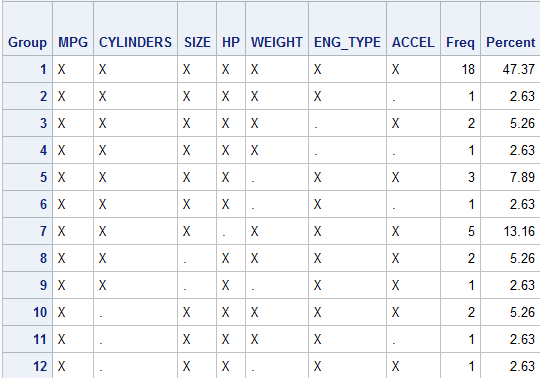
Maybe we can do a better job if we use imputation to fill in the missing values and rerun the analysis with “completed” data.

Examine Missing Pattern

The code displays the patterns of missing data, so you can determine if patterns are monotone or non-monotone (arbitrary).

|  |
| --- |
| \* is the missing data monotone or non-monotone?;  \* the data is non-monotone;  title 'MI Pattern';  ods select misspattern;  **proc** **mi** data=cars nimpute=**0**;  var mpg cylinders size hp weight eng\_type accel;  **run**;  **quit**; |

SAS Reports Missing Data Patterns



Is this monotone or non-monotone?

Based on the output of missing data, there is no pattern to displayed, values are missing everywhere

This is a non-monotone of missingness.

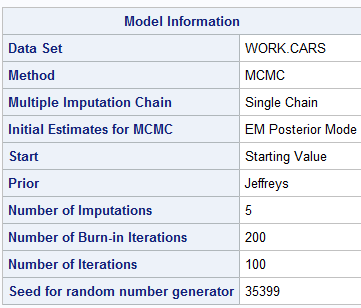
Using SAS PROC MI, Step 1

Use the default method (MCMC) since this missing pattern is arbitrary.

Seed is a positive integer to start the psuedo-random number generator.

|  |
| --- |
| \* create mi data using default MCMC for non-monotone;  title 'MI with MCMC';  **proc** **mi** data=cars out=miout seed=**35399** nimpute=**5**;  var mpg cylinders size hp weight eng\_type accel;  **run**;  **quit**; |

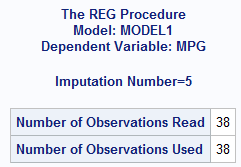
From SAS Output



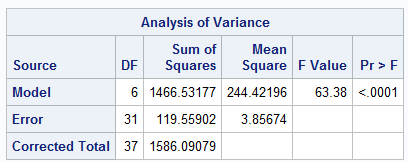
Run Analysis Using Imputed Data

|  |
| --- |
| \* run reg with mi data;  title 'Predicting MPG with MI (final)';  **proc** **reg** data=miout outest=outreg covout;  model mpg = cylinders size hp weight eng\_type accel;  by \_Imputation\_;  **run**;  **quit**; |

Output: Imputation #5



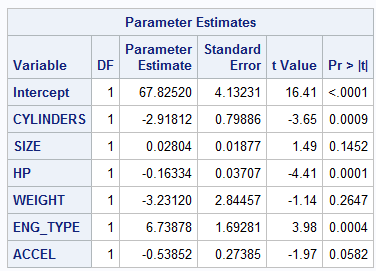
No observations deleted, used the complete data set.



|  |  |
| --- | --- |
| Original | Imputation #5 |
|  |  |

Original Dof = 37 versus 37

Parameter Estimates



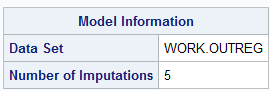
Our output results only gave Imputation #5. We were never able to compare the estimates of Imputation #1 to Imputation #5.

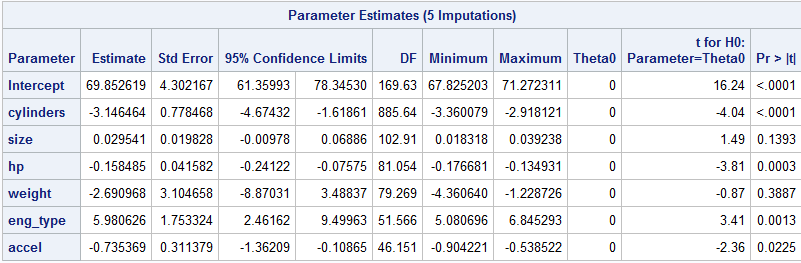
Summary of Five Analyses”

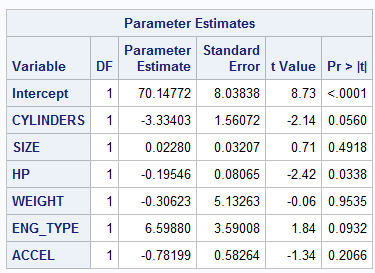
SAS Data Set OUTREG

Code:

|  |
| --- |
| \* combine results;  title 'Predicting MPG (combined)';  **proc** **mianalyze** data=outreg;  modeleffects Intercept cylinders size hp weight eng\_type accel;  **run**; |







|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Compare Parameter Estimates to Original | | | | |
| Variable | **Original Estimate** | **Original Std Error** | **Combined Estimate** | **Combined Std Error** |
| Intercept | 70.14772 | 8.03838 | 69.852619 | 4.302167 |
| Cylinders | -3.33403 | 1.56072 | -3.146464 | 0.778468 |
| Size | 0.0280 | 0.03207 | 0.029541 | 0.019828 |
| HP | -0.19546 | 0.08065 | -0.158485 | 0.041582 |
| Weight | -0.30623 | 5.13263 | -2.690968 | 3.104658 |
| ENG\_TYPE | 6.59880 | 3.59008 | 5.980626 | 1.753324 |
| ACCEL | -0.78199 | 0.58264 | -0.735369 | 0.311379 |

We don’t expect the combined estimates to be close to the original estimates, but we do some have confidence that they are better estimates of the parameters.

we appreciated the natural variance within original data set. And we used that in order to fill in the data sets. We did this five (5) different times to get five (5) different ideas of how this data set may work. Then we were able to combine those data into a single analysis using MIANALYZE. So the estimates that we see on the right

We have some confidence in that we have a good set of estimates based on the natural variabiliy within the original data set.

Source 2.3 PROC MI Example II, MSDS 7333, Quantifying the World

Summary:

Multiple imputation:

* Provides an analysis that reflects the uncertainty due to missing values
* Creates a representative random sample of the missing values
* Is typically better than single imputation methods because it results in valid statistical inferences that reflect the uncertainty due to missing values