MPG Dataset

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

MPG Dataset

Obs	Auto	ENG_TYPE	CYLINDERS	SIZE	HP	WEIGHT	ACCEL	MPG
37	BMW 320i	0	4	121	110	.	.	21.5
38	VW Rabbit	0	4	89	71	1.925	14	31.9

MPG Dataset

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

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

MPG Dataset

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.

Number of Observations Read	38
Number of Observations Used	18
Number of Observations with Missing Values	20

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	774.27999	129.04667	22.39	<.0001
Error	11	63.40945	5.76450		
Corrected Total	17	837.68944			

Parameter Estimates (Using Listwise Deletion)

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	70.14772	8.03838	8.73	<.0001
CYLINDERS	1	-3.33403	1.56072	-2.14	0.0560
SIZE	1	0.02280	0.03207	0.71	0.4918
HP	1	-0.19546	0.08065	-2.42	0.0338
WEIGHT	1	-0.30623	5.13263	-0.06	0.9535
ENG_TYPE	1	6.59880	3.59008	1.84	0.0932
ACCEL	1	-0.78199	0.58264	-1.34	0.2066

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.

MPG Dataset

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

Group	MPG	CYLINDERS	SIZE	HP	WEIGHT	ENG_TYPE	ACCEL	Freq	Percent
1	X	X	X	X	X	X	X	18	47.37
2	X	X	X	X	X	X	.	1	2.63
3	X	X	X	X	X	.	X	2	5.26
4	X	X	X	X	X	.	.	1	2.63
5	X	X	X	X	.	X	X	3	7.89
6	X	X	X	X	.	X	.	1	2.63
7	X	X	X	.	X	X	X	5	13.16
8	X	X	.	X	X	X	X	2	5.26
9	X	X	.	X	.	X	X	1	2.63
10	X	.	X	X	X	X	X	2	5.26
11	X	.	X	X	X	X	.	1	2.63
12	X	.	X	X	.	X	X	1	2.63

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.

MPG Dataset

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

Model Information	
Data Set	WORK.CARS
Method	MCMC
Multiple Imputation Chain	Single Chain
Initial Estimates for MCMC	EM Posterior Mode
Start	Starting Value
Prior	Jeffreys
Number of Imputations	5
Number of Burn-in Iterations	200
Number of Iterations	100
Seed for random number generator	35399

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

The REG Procedure
Model: MODEL1
Dependent Variable: MPG

Imputation Number=5

Number of Observations Read	38
Number of Observations Used	38

No observations deleted, used the complete data set.

MPG Dataset

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1466.53177	244.42196	63.38	<.0001
Error	31	119.55902	3.85674		
Corrected Total	37	1586.09079			

Original

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	774.27999	129.04667	22.39	<.0001
Error	11	63.40945	5.76450		
Corrected Total	17	837.68944			

Imputation #5

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1466.53177	244.42196	63.38	<.0001
Error	31	119.55902	3.85674		
Corrected Total	37	1586.09079			

Original Dof = 37 versus 37

Parameter Estimates

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	67.82520	4.13231	16.41	<.0001
CYLINDERS	1	-2.91812	0.79886	-3.65	0.0009
SIZE	1	0.02804	0.01877	1.49	0.1452
HP	1	-0.16334	0.03707	-4.41	0.0001
WEIGHT	1	-3.23120	2.84457	-1.14	0.2647
ENG_TYPE	1	6.73878	1.69281	3.98	0.0004
ACCEL	1	-0.53852	0.27385	-1.97	0.0582

MPG Dataset

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;
```

Model Information	
Data Set	WORK.OUTREG
Number of Imputations	5

Parameter Estimates (5 Imputations)										
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	69.852619	4.302167	61.35993	78.34530	169.63	67.825203	71.272311	0	16.24	<.0001
cylinders	-3.146464	0.778468	-4.67432	-1.61861	885.64	-3.360079	-2.918121	0	-4.04	<.0001
size	0.029541	0.019828	-0.00978	0.06886	102.91	0.018318	0.039238	0	1.49	0.1393
hp	-0.158485	0.041582	-0.24122	-0.07575	81.054	-0.176681	-0.134931	0	-3.81	0.0003
weight	-2.690968	3.104658	-8.87031	3.48837	79.269	-4.360640	-1.228726	0	-0.87	0.3887
eng_type	5.980626	1.753324	2.46162	9.49963	51.566	5.080696	6.845293	0	3.41	0.0013
accel	-0.735369	0.311379	-1.36209	-0.10865	46.151	-0.904221	-0.538522	0	-2.36	0.0225

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	70.14772	8.03838	8.73	<.0001
CYLINDERS	1	-3.33403	1.56072	-2.14	0.0560
SIZE	1	0.02280	0.03207	0.71	0.4918
HP	1	-0.19546	0.08065	-2.42	0.0338
WEIGHT	1	-0.30623	5.13263	-0.06	0.9535
ENG_TYPE	1	6.59880	3.59008	1.84	0.0932
ACCEL	1	-0.78199	0.58264	-1.34	0.2066

MPG Dataset

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

MPG Dataset