

Gavin Gunawardena

INFS 762

9/16/2021

Dr. Liu

Assignment 1

*Note: Added a table of contents since some of the outputs were very long

Table of Contents

2)	1
1.	1
2.	2
3.	2
3)	3
1.	3
2.	3
/* Frequency Table */	3
/* Mean mileage of each Make */	5
/* Mean mileage of each model of each make */	7
3.	19
4.	19
5.	23
6.	24
7.	24
/* Forward Selection Logistic Regression */	24
/* Backward Selection Logistic Regression */	33
/* Stepwise Selection Logistic Regression */	41

2)

1. /* Sort the Data by acctno*/

```
PROC SORT DATA=ASSGN1.DATA1;
```

```
BY ACCTNO;
```

```
RUN;
```

```
PROC SORT DATA=ASSGN1.XTRA1;
```

```
BY ACCTNO;
```

```
RUN;
```

2. Left joins and inner joins are based on the key columns they are joined using, which is usually one but could consist of multiple columns. Left joins include all data from the first table and only data in the second table with a matching value in its key column(s). An inner join does not distinguish between either table and only includes data from either table that has a matching key column value in the table it is being merged with.

3.

```
/* Merge Data1 and xtra1 datasets */
```

```
DATA ASSGN1.COMPLETE;
```

```
    MERGE ASSGN1.DATA1(IN=A) ASSGN1.XTRA1(IN=B);
```

```
    BY ACCTNO;
```

```
    IF A THEN OUTPUT;
```

```
RUN;
```

```
/* Print statement.*/
```

```
PROC PRINT DATA= ASSGN1.COMPLETE(obs=20); /* You need modify this statement – you want to print just first 20 records of ASSGN1.COMPLETE */
```

Obs	Obs	AGEOTD	AGEAVG	TRADES	TOTBAL	HSATRT	CURSAT	BTRDRS	BROPEN	BROLDT	BRHS2X	BRHS3X	BRHS4X	BRHS5X	AUUTIL	CFTRDS	INQ012	PUBREC	BKTIME	CBTYPE
1	1	-5	-5	0	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	1	2	-6	2
2	2	81	46	9	1567	0	0	0	-6	-6	-6	-6	-6	-6	-6	0	1	0	-6	1
3	3	-2	-2	0	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	5
4	4	216	39.52	35	27657	37.14	1	3	1	23	0	1	0	0	-6	6	68	5	-6	1
5	5	107	30.44	13	11467	23.08	0	0	-6	-6	-6	-6	-6	-6	-6	4	0	5	36	1
6	6	-3	-3	0	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	0	-3	-3	3
7	7	62	40	7	34027	0	0	1	0	62	0	0	1	0	-6	1	8	1	-6	1
8	8	79	39.67	16	4920	6.25	1	0	-6	-6	-6	-6	-6	-6	-6	2	2	6	-6	1
9	9	69	13	6	19449	0	0	3	2	22	0	1	1	0	-6	0	7	1	-6	1
10	10	-2	-2	0	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	5
11	11	109	27.6	7	3283	15.29	2	4	2	36	1	2	0	0	-6	0	0	0	-6	1
12	12	93	37.83	6	6222	16.67	0	0	-6	-6	-6	-6	-6	-6	-6	0	0	1	-6	1
13	13	79	46.25	5	78469	0	0	0	-6	-6	-6	-6	-6	-6	-6	2	0	2	-1	1
14	14	246	79.5	8	4199	37.5	2	1	0	79	0	0	0	0	-6	5	1	2	63	1
15	15	80	30.56	9	5834	11.11	0	4	0	43	0	0	0	0	-6	1	1	0	-6	1
16	16	32	-1	2	1175	0	0	0	-6	-6	-6	-6	-6	-6	-6	0	0	1	-6	1
17	17	-5	-5	0	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	0	1	-6	2
18	18	152	50.6	6	2146	0	0	1	0	76	0	0	0	0	-6	1	0	1	-6	1
19	19	67	16.6	7	10922	0	0	1	0	32	0	0	0	0	80.87	1	14	2	-6	1
20	20	115	29.86	10	15779	0	0	1	0	24	0	0	0	0	-6	2	0	0	-6	1

3)

1.

```
/* Replace Missing MNGPAY values with mean of the column */
```

```
proc means data = ASSGN1.COMPLETE MEAN ;
```

```
VAR MNGPAY;
```

```
run;
```

```
/* Result for MNGPAY: 2257.46 */
```

```
/*Imputation step*/
```

```
DATA ASSGN1.COMPLETE;
```

```
SET ASSGN1.COMPLETE;
```

```
IF MNGPAY =. THEN DO;
```

```
    MNGPAY_M=1;
```

```
    MNGPAY=2257.46;
```

```
END;
```

```
ELSE DO;
```

```
    MNGPAY_M=0;
```

```
END;
```

```
RUN;
```

2.

```
/* Frequency Table */
```

```
PROC FREQ DATA= ASSGN1.COMPLETE;
```

```
Tables Make / NOPERCENT NOCUM ;
```

```
RUN;
```

MAKE	
MAKE	Frequency
ACURA	67
AMC	1
AUDI	10
BMW	20
BUICK	423
CADILLAC	182
CHEVROLE	1755
CHEVY TR	2
CHRYSLER	355
DAIHATSU	1
DODGE	966
DODGE TR	412
EAGLE	140
FORD	2475
FORD TRU	678
GEO	329
GEO TRUC	74
GMC	102
HONDA	298
HONDA TR	2
HYUNDAI	135
INFINITI	25
ISUZU	5
ISUZU TR	43
JAGUAR	8
JEEP	138
KIA	22
KIA TRUC	1
LEXUS	17
LINCOLN	158

MAKE	
MAKE	Frequency
MAZDA	322
MAZDA TR	47
MERCEDES	11
MERCURY	752
MERKUR	1
MITSUBIS	405
NISSAN	344
NISSAN T	43
OLDSMOBI	674
PLYMOUTH	787
PONTIAC	1126
PORSCHE	2
SAAB	11
SATURN	202
SUBARU	61
SUZUKI	12
SUZUKI T	24
TOYOTA	281
TOYOTA T	26
VOLVO	7
VW	58
Frequency Missing = 2	

```

/* Obtain mean mileage of each Make */
PROC MEANS DATA= ASSGN1.COMPLETE MEAN NONOBS;

CLASS Make;

Var Mileag;

RUN;

```

Analysis Variable : MILEAG MILEAG	
MAKE	Mean
ACURA	99714.19
AMC	73190.00
AUDI	113756.70
BMW	101324.50
BUICK	92437.99
CADILLAC	106010.77
CHEVROL E	93713.24
CHEVY TR	49419.00
CHRYSLER	89254.31
DAIHATSU	79934.00
DODGE	83894.40
DODGE TR	97905.18
EAGLE	87853.48
FORD	85239.30
FORD TRU	94589.53
GEO	79492.80
GEO TRUC	76981.68
GMC	108809.66
HONDA	98146.75
HONDA TR	104925.00
HYUNDAI	68908.31
INFINITI	89005.12
ISUZU	84699.60
ISUZU TR	81960.05
JAGUAR	88383.75
JEEP	102549.41
KIA	43130.32
KIA TRUC	81175.00

Analysis Variable : MILEAG MILEAG	
MAKE	Mean
LEXUS	96426.65
LINCOLN	97932.60
MAZDA	89401.25
MAZDA TR	99357.45
MERCEDE S	135092.82
MERCURY	86677.50
MERKUR	91900.00
MITSUBIS	77214.54
NISSAN	89561.72
NISSAN T	96035.58
OLDSMOB I	95775.97
PLYMOUT H	86373.72
PONTIAC	92796.15
PORSCHE	114599.00
SAAB	114014.00
SATURN	85829.07
SUBARU	92752.26
SUZUKI	40484.25
SUZUKI T	63423.17
TOYOTA	91472.62
TOYOTA T	94988.85
VOLVO	124342.29
VW	93797.93

```

/* Mean mileage of each model of each make */
PROC MEANS DATA= ASSGN1.COMPLETE MEAN NONOBS;

CLASS Make Model;

Var Mileag;

```

RUN;

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
ACURA	INTEGRA	87543.00
	LEGEND	112967.03
	VIGOR	88176.67
AMC	CONCORD	73190.00
AUDI	100	114082.67
	4000	118427.00
	5000	105927.67
	5000CS	88538.00
	80	136746.00
	90	133825.00
BMW	318I	101258.00
	325	117357.50
	325I	90036.17
	325IS	80127.00
	328I	99359.00
	525I	92000.00
	528E	117605.00
	535I	105683.00
	735I	109147.50
BUICK	CENTURY	93271.26
	ELECTRA	115361.50
	LESABRE	97260.71
	PARK AVE	104653.14
	REATA	98824.50
	REGAL	92996.59
	RIVIERA	110340.93
	ROADMAS T	83471.00

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	SKYHAWK	84358.00
	SKYLARK	83831.68
	SOMERSET	108238.50
CADILLAC	ALLANTE'	68043.00
	BROUGHAM	98100.22
	CIMARRON	135075.33
	DEVILLE	106366.05
	ELDORADO	112785.50
	FLEETWOOD	110157.00
	SEVILLE	98203.33
CHEVROLET	APV	126247.50
	ASTRO	104931.86
	BERETTA	94686.90
	BLAZER	92835.95
	CAMARO	91573.63
	CAPRICE	103983.92
	CAVALIER	87958.00
	CELEBRITY	113884.59
	CHEVETTE	100694.00
	CORSICA	86384.48
	CORVETTE	95206.50
	G-SERIES	105071.43
	IMPALA	92117.00
	LUMINA	96765.14
	LUMINA A	93414.53
	MALIBU	76579.50
	METRO	42050.00
	MONTE CARLO	105072.89

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	NOVA	117994.57
	PICKUP	97236.47
	PRIZM	3229.83
	S10 BLAZ	109257.64
	SPECTRUM	86007.40
	SPORTVAN	103975.63
	SPRINT	91400.00
	SUBURBAN	118408.88
	TAHOE	63000.00
	TRACKER	33850.33
	VENTURE	81525.00
CHEVY TR	TRACKER	49419.00
CHRYSLER	CIRRUS	76054.83
	CONCORDE	83992.43
	CONQUEST	93055.00
	FIFTH AV	92784.82
	IMPERIAL	88910.83
	LEBARON	93069.70
	LHS	73879.00
	NEW YORK	95720.27
	SEBRING	64229.91
	TOWN & C	85804.89
DAIHATSU	CHARADE	79934.00
DODGE	600	95830.50
	ARIES	61510.50
	ARIES AM	112711.67
	AVENGER	60146.00
	CHARGER	73145.00

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	COLT	83867.04
	COLT VIS	101062.50
	DAYTONA	102387.79
	DIPLOMAT	105017.00
	DYNASTY	98091.37
	INTREPID	82674.57
	LANCER	140690.50
	MONACO	69241.43
	NEON	69032.77
	OMNI	86427.75
	OMNI AME	93450.00
	SHADOW	85864.50
	SPIRIT	90559.35
	STEALTH	85599.00
	STRATUS	68848.68
DODGE TR	CARAVAN	97756.98
	GRAND CA	99005.27
	MINI RAM	89879.00
	PICKUP	99558.47
	RAIDER	118164.00
	RAM VAN	87259.67
	RAM WAGO	85316.00
	RAMCHAR G	108942.40
	VAN	116605.00
EAGLE	MEDALLIO	96486.00
	PREMIER	96016.50
	SUMMIT	84756.11
	TALON	87621.03

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	VISION	88039.76
FORD	ASPIRE	57167.25
	CONTOUR	70328.78
	COUNTRY	24759.00
	CROWN VI	98423.43
	ESCORT	81868.58
	FESTIVA	87486.31
	FIESTA	91377.00
	LTD	101304.00
	MUSTANG	82749.73
	PROBE	89053.61
	TAURUS	88918.39
	TEMPO	85215.48
	THUNDER B	91516.95
FORD TRU	AEROSTAR	97468.31
	BRONCO	110898.48
	BRONCO I	110317.78
	CLUB WAG	82383.00
	ECONOLIN	99838.04
	EXPLORER	93928.68
	PICKUP	90776.47
	WINDSTAR	77032.37
GEO	METRO	70741.03
	PRIZM	84390.32
	STORM	86825.79
GEO TRUC	TRACKER	76981.68
GMC	JIMMY	106116.54
	PICKUP	97375.62
	RALLY	114000.00

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	RALLY WA	115325.00
	S15 JIMM	125176.08
	SAFARI	107230.91
	SUBURBA N	176000.00
	VANDURA	131655.14
HONDA	ACCORD	99112.90
	CIVIC	95413.85
	DEL SOL	76366.50
	PRELUDE	117134.67
HONDA TR	PASSPORT	104925.00
HYUNDAI	ACCENT	53071.69
	ELANTRA	65832.50
	EXCEL	74559.00
	SCOUPE	81148.89
	SONATA	75531.93
INFINITI	G20	87187.75
	J30	88238.88
	M30	86200.75
	Q45	95468.25
	QX4	95039.00
ISUZU	I-MARK	77754.00
	IMPULSE	89827.50
	STYLUS	88335.00
ISUZU TR	AMIGO	75404.33
	HOMBRE	10289.00
	PICKUP	86119.78
	RODEO	83175.77
	TROOPER	98718.83
JAGUAR	XJ6	88383.75

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
JEEP	CHEROKEE	106651.15
	COMANCHE	102000.00
	GRAND CH	92703.40
	GRAND WA	139761.00
	PICKUP	73442.50
	WAGONEER	138489.00
	WRANGLE R	80908.30
KIA	SEPHIA	43130.32
KIA TRUC	SPORTAGE	81175.00
LEXUS	ES 250	104530.00
	ES 300	87254.29
	LS 400	110520.00
	SC 300	81713.00
LINCOLN	CONTINEN	97428.39
	MARK V	94624.00
	MARK VII	91537.72
	TOWN CAR	101812.16
MAZDA	323	104847.94
	626	87223.07
	929	110479.50
	MIATA	66893.50
	MIATA MX	120920.00
	MILLENNIA	45000.00
	MX-3	78121.16
	MX-6	95319.00
	PROTEGE'	86987.77
	RX-7	110847.50

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
MAZDA TR	B1800	94468.00
	B2000	118715.00
	B2200	93239.30
	B2300	71291.50
	B2600I	109545.00
	B3000	92033.00
	B4000	273204.00
	MPV	99563.00
	NAVAJO	94448.00
	PICKUP	64662.50
MERCEDE S	190E	101528.67
	260E	132781.00
	300CD-T	205540.00
	300E	127538.00
	300SD	270000.00
	500SEL	165417.00
	C-CLASS	25083.00
MERCURY	CAPRI	70495.54
	COUGAR	91074.91
	GRAND MA	93709.88
	LYNX	110918.67
	MARQUIS	106056.00
	MYSTIQUE	66761.90
	SABLE	88479.37
	TOPAZ	84600.85
	TRACER	81556.58
	VILLAGER	79930.56
MERKUR	XR4TI	91900.00
MITSUBIS	3000GT	118318.00
	DIAMANTE	89079.21

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	ECLIPSE	84224.42
	EXPO	90549.00
	GALANT	70332.43
	MIRAGE	74697.10
	MONTERO	82410.50
	PICKUP	77876.33
	PRECIS	118010.67
NISSAN	200SX	99581.50
	240SX	91577.54
	300ZX	130468.00
	ALTIMA	77667.21
	MAXIMA	96788.20
	NX	76297.00
	PULSAR N	100021.50
	SENTRA	84589.36
	STANZA	102095.17
NISSAN T	FRONTIER	54114.00
	PATHFIND	107571.07
	PICKUP	93440.63
	QUEST	88227.67
OLDSMOB I	88	95591.56
	98	111351.69
	ACHIEVA	84653.88
	AURORA	89147.50
	BRAVADA	87918.40
	CALAIS	95390.97
	CIERA	96896.60
	CUSTOM C	57821.00
	CUTLASS	96410.02
	DELTA 88	107011.03

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	FIRENZA	102244.50
	OMEGA	70184.00
	REGENCY	99509.57
	SILHOUET	80216.08
	TORONADO	180546.33
PLYMOUTH	ACCLAIM	89910.86
	BREEZE	39285.64
	CARAVELL	59363.50
	COLT	81052.32
	COLT VIS	85256.00
	FURY	103115.00
	GRAND VO	98220.10
	HORIZON	78431.67
	LASER	90676.00
	NEON	63972.11
	RELIANT	51577.00
	SUNDANCE	88066.01
	VOYAGER	96625.97
PONTIAC	6000	99628.59
	BONNEVIL	102078.90
	FIERO	94579.11
	FIREBIRD	87738.42
	GRAND AM	91274.25
	GRAND PR	94885.39
	LEMANS	79522.29
	PARISIEN	117339.50
	SAFARI	103183.00
	SUNBIRD	94150.83

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	SUNFIRE	65243.26
	TRANS SP	96724.04
PORSCHÉ	928	160000.00
	944	69198.00
SAAB	900	112579.88
	9000	117838.33
SATURN	SC	91785.17
	SC1	77592.22
	SC2	74050.40
	SL	81788.36
	SL1	85775.29
	SL2	90020.17
	SW1	112231.67
	SW2	98067.50
SUBARU	GL	103259.33
	IMPREZA	68188.50
	JUSTY	92926.67
	LEGACY	98363.42
	LOYALE	81646.27
	XT	122218.33
SUZUKI	ESTEEM	38405.00
	SWIFT	43395.20
SUZUKI T	SAMURAI	85792.67
	SIDEKICK	63440.47
	X-90	29704.50
TOYOTA	AVALON	74959.00
	CAMRY	91977.49
	CELICA	100860.29
	COROLLA	89159.82
	CRESSIDA	133811.40

Analysis Variable : MILEAG MILEAG		
MAKE	MODEL	Mean
	PASEO	91404.40
	SUPRA	109550.00
	TERCEL	89149.39
TOYOTA T	4RUNNER	114765.00
	LAND CRU	91818.50
	PICKUP	89345.73
	PREVIA	92374.67
	TACOMA	94938.00
VOLVO	240	63263.00
	244	158327.00
	740	133987.25
	780	112857.00
VW	CABRIOLE	144046.50
	FOX	109890.67
	GOLF	86857.86
	GTI	107076.00
	JETTA	87614.05
	JETTA II	82861.38
	PASSAT	94961.00
	RABBIT	115982.00

3.

```
/* Create new dataset: REG_DATA */
```

```
DATA ASSGN1.REG_DATA;
```

```
SET ASSGN1.COMPLETE;
```

```
keep AGEOTD AGEAVG TRADES TOTBAL HSATRT CURSAT BRTRDS BROPEN VDDASAV GOOD;
```

```
RUN;
```

4.

```
/* Logistic Regression */
```

```
PROC LOGISTIC DATA= ASSGN1.REG_DATA;
```

```
class VDDASAV;
```

```
model GOOD= AGEOTD AGEAVG TRADES TOTBAL HSATRT CURSAT BTRDRS BROPEN VDDASAV;
```

```
RUN;
```

Model Information		
Data Set	ASSGN1.REG_DAT A	
Response Variable	GOOD	GOOD
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	1404 2
Number of Observations Used	1404 2

Response Profile		
Ordered Value	GOOD	Total Frequency
1	0	4600
2	1	9442

Probability modeled is GOOD='0'.

Class Level Information				
Class	Value	Design Variables		
VDDASAV	BOTH	1	0	0
	DDA	0	1	0
	NONE	0	0	1
	SAV	-1	-1	-1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17366.295
SC	17771.495	17456.893
-2 Log L	17761.945	17342.295

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.6494	11	<.0001
Score	396.4490	11	<.0001
Wald	381.4440	11	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AGEOTD	1	2.7975	0.0944
AGEAVG	1	24.7977	<.0001
TRADES	1	61.8533	<.0001
TOTBAL	1	2.2820	0.1309
HSATRT	1	53.1534	<.0001
CURSAT	1	4.0643	0.0438
BRTRDS	1	0.0406	0.8403
BROPEN	1	27.4814	<.0001
VDDASAV	3	42.6944	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7789	0.0493	249.5410	<.0001
AGEOTD		1	-0.00101	0.000601	2.7975	0.0944
AGEAVG		1	-0.00617	0.00124	24.7977	<.0001
TRADES		1	0.0264	0.00336	61.8533	<.0001
TOTBAL		1	-1.24E-6	8.176E-7	2.2820	0.1309
HSATRT		1	-0.00820	0.00112	53.1534	<.0001
CURSAT		1	-0.0321	0.0159	4.0643	0.0438
BRTRDS		1	0.00277	0.0138	0.0406	0.8403
BROPEN		1	-0.0372	0.00709	27.4814	<.0001
VDDASAV	BOTH	1	-0.1333	0.0639	4.3614	0.0368
VDDASAV	DDA	1	-0.1256	0.0470	7.1556	0.0075
VDDASAV	NONE	1	0.1865	0.0331	31.7225	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGEOTD	0.999	0.998	1.000
AGEAVG	0.994	0.991	0.996
TRADES	1.027	1.020	1.034
TOTBAL	1.000	1.000	1.000
HSATRT	0.992	0.990	0.994
CURSAT	0.968	0.939	0.999
BRTRDS	1.003	0.976	1.030
BROPEN	0.964	0.950	0.977
VDDASAV BOTH vs SAV	0.814	0.664	0.998
VDDASAV DDA vs SAV	0.820	0.695	0.968
VDDASAV NONE vs SAV	1.121	0.978	1.285

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	59.7	Somers' D	0.199
Percent Discordant	39.9	Gamma	0.199
Percent Tied	0.4	Tau-a	0.088
Pairs	4343320	c	0.599

5. Please explain what underfitting means and what overfitting means.

You will lose points if you directly copy materials from Wikipedia and other online sources.

Underfitting occurs when your linear or logistic model is too simple and doesn't fit the dataset you're testing it on. Overfitting is the opposite, where the model fits the training dataset too well, and occurs when your linear or logistic model is too complex. This complexity usually occurs due to utilizing too many independent variables.

6.

Please briefly describe (<= 6 sentences) the OLS linear regression algorithm using your own language.

You will lose points if you directly copy materials from Wikipedia and other online sources.

OLS Linear Regression, or Ordinary Least Squares, is where a prediction is attempted using a linear function of the inputted independent variables. This is the simplest and most classic linear method for prediction. Its formula is essentially: $\hat{y} = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$. Also, the goal of it is to find the values of w and b that minimize the mean squared error between predictions and the regression targets which are represented by y.

7.

/ Forward Selection Logistic Regression */*

PROC LOGISTIC DATA= ASSGN1.REG_DATA;

class VDDASAV;

model GOOD= AGEOTD AGEAVG TRADES TOTBAL HSATRT CURSAT BRTRDS
BROPEN VDDASAV / Selection=forward;

RUN;

Model Information		
Data Set	ASSGN1.REG_DATA	
Response Variable	GOOD	GOOD
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	14042
Number of Observations Used	14042

Response Profile		
Ordered Value	GOOD	Total Frequency
1	0	4600
2	1	9442

Probability modeled is GOOD='0'.

Forward Selection Procedure

Class Level Information				
Class	Value	Design Variables		
VDDASAV	BOTH	1	0	0
	DDA	0	1	0
	NONE	0	0	1
	SAV	-1	-1	-1

Step 0. Intercept entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L	=	17761.945
----------	---	-----------

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
396.4490	11	<.0001

Step 1. Effect HSATRT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17572.267
SC	17771.495	17587.366
-2 Log L	17761.945	17568.267

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	193.6779	1	<.0001
Score	181.6459	1	<.0001
Wald	176.8417	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
219.1733	10	<.0001

Step 2. Effect AGEAVG entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17482.926
SC	17771.495	17505.575
-2 Log L	17761.945	17476.926

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	285.0192	2	<.0001
Score	271.0157	2	<.0001
Wald	265.3954	2	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
134.7184	9	<.0001

Step 3. Effect VDDASAV entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17441.706
SC	17771.495	17487.005
-2 Log L	17761.945	17429.706

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	332.2391	5	<.0001
Score	316.4653	5	<.0001
Wald	309.0025	5	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
88.7894	6	<.0001

Step 4. Effect TRADES entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17404.425
SC	17771.495	17457.273
-2 Log L	17761.945	17390.425

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	371.5201	6	<.0001
Score	347.2806	6	<.0001
Wald	336.8855	6	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
47.7558	5	<.0001

Step 5. Effect *BROPEN* entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17369.615
SC	17771.495	17430.014
-2 Log L	17761.945	17353.615

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	408.3295	7	<.0001
Score	385.3342	7	<.0001
Wald	372.4776	7	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
11.1095	4	0.0254

Step 6. Effect CURSAT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17365.950
SC	17771.495	17433.898
-2 Log L	17761.945	17347.950

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	413.9947	8	<.0001
Score	391.3689	8	<.0001
Wald	377.6231	8	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
5.5458	3	0.1359

Note No (additional) effects met the 0.05 significance level for entry into the : model.

Summary of Forward Selection						
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	HSATRT	1	1	181.6459	<.0001	HSATRT
2	AGEAVG	1	2	87.4223	<.0001	AGEAVG
3	VDDASAV	3	3	46.2646	<.0001	VDDASAV
4	TRADES	1	4	40.1221	<.0001	TRADES
5	BROPEN	1	5	36.8180	<.0001	BROPEN
6	CURSAT	1	6	5.5929	0.0180	CURSAT

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AGEAVG	1	83.9199	<.0001
TRADES	1	64.1390	<.0001
HSATRT	1	57.1942	<.0001
CURSAT	1	5.5859	0.0181
BROPEN	1	34.6884	<.0001
VDDASAV	3	43.4840	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7923	0.0486	266.0681	<.0001
AGEAVG		1	-0.00774	0.000845	83.9199	<.0001
TRADES		1	0.0234	0.00292	64.1390	<.0001
HSATRT		1	-0.00848	0.00112	57.1942	<.0001
CURSAT		1	-0.0336	0.0142	5.5859	0.0181
BROPEN		1	-0.0375	0.00637	34.6884	<.0001
VDDASAV	BOTH	1	-0.1327	0.0638	4.3197	0.0377

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
VDDASAV	DDA	1	-0.1289	0.0469	7.5520	0.0060
VDDASAV	NONE	1	0.1872	0.0331	31.9711	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGEAVG	0.992	0.991	0.994
TRADES	1.024	1.018	1.030
HSATRT	0.992	0.989	0.994
CURSAT	0.967	0.940	0.994
BROPEN	0.963	0.951	0.975
VDDASAV BOTH vs SAV	0.813	0.663	0.997
VDDASAV DDA vs SAV	0.816	0.692	0.963
VDDASAV NONE vs SAV	1.119	0.976	1.283

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	59.6	Somers' D	0.197
Percent Discordant	39.9	Gamma	0.198
Percent Tied	0.4	Tau-a	0.087
Pairs	4343320	c	0.598

```
/* Backward Selection Logistic Regression */
PROC LOGISTIC DATA= ASSGN1.REG_DATA;
```

```

class VDDASAV;
model GOOD= AGEOTD AGEAVG TRADES TOTBAL HSATRT CURSAT BRTRDS
BROPEN VDDASAV / Selection=backward;
RUN;

```

Model Information		
Data Set	ASSGN1.REG_DAT A	
Response Variable	GOOD	GOOD
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	1404 2
Number of Observations Used	1404 2

Response Profile		
Ordered Value	GOOD	Total Frequency
1	0	4600
2	1	9442

Probability modeled is GOOD='0'.

Backward Elimination Procedure

Class Level Information				
Class	Value	Design Variables		
VDDASAV	BOTH	1	0	0
	DDA	0	1	0
	NONE	0	0	1
	SAV	-1	-1	-1

Step 0. The following effects were entered:

Intercept AGEOTD AGEAVG TRADES TOTBAL HSATRT CURSAT BRTRDS BROPEN VDDASAV

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17366.295
SC	17771.495	17456.893
-2 Log L	17761.945	17342.295

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.6494	11	<.0001
Score	396.4490	11	<.0001
Wald	381.4440	11	<.0001

Step 1. Effect BRTRDS is removed:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17364.336
SC	17771.495	17447.384
-2 Log L	17761.945	17342.336

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.6088	10	<.0001
Score	396.4093	10	<.0001
Wald	381.5208	10	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.0406	1	0.8402

Step 2. Effect TOTBAL is removed:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17364.663
SC	17771.495	17440.161
-2 Log L	17761.945	17344.663

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	417.2816	9	<.0001
Score	394.2356	9	<.0001
Wald	379.7758	9	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
2.3049	2	0.3159

Step 3. Effect AGEOTD is removed:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17365.950
SC	17771.495	17433.898
-2 Log L	17761.945	17347.950

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	413.9947	8	<.0001
Score	391.3689	8	<.0001
Wald	377.6231	8	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
5.5458	3	0.1359

Note No (additional) effects met the 0.05 significance level for removal from the : model.

Summary of Backward Elimination						
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Variable Label
1	BRTRDS	1	8	0.0406	0.8403	BRTRD S
2	TOTBAL	1	7	2.2584	0.1329	TOTBA L
3	AGEOTD	1	6	3.2485	0.0715	AGEOT D

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AGEAVG	1	83.9199	<.0001
TRADES	1	64.1390	<.0001
HSATRT	1	57.1942	<.0001
CURSAT	1	5.5859	0.0181
BROPEN	1	34.6884	<.0001
VDDASAV	3	43.4840	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7923	0.0486	266.0681	<.0001
AGEAVG		1	-0.00774	0.000845	83.9199	<.0001
TRADES		1	0.0234	0.00292	64.1390	<.0001
HSATRT		1	-0.00848	0.00112	57.1942	<.0001
CURSAT		1	-0.0336	0.0142	5.5859	0.0181
BROPEN		1	-0.0375	0.00637	34.6884	<.0001
VDDASAV	BOTH	1	-0.1327	0.0638	4.3197	0.0377
VDDASAV	DDA	1	-0.1289	0.0469	7.5520	0.0060
VDDASAV	NONE	1	0.1872	0.0331	31.9711	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGEAVG	0.992	0.991	0.994
TRADES	1.024	1.018	1.030
HSATRT	0.992	0.989	0.994
CURSAT	0.967	0.940	0.994
BROPEN	0.963	0.951	0.975
VDDASAV BOTH vs SAV	0.813	0.663	0.997
VDDASAV DDA vs SAV	0.816	0.692	0.963
VDDASAV NONE vs SAV	1.119	0.976	1.283

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	59.6	Somers' D	0.197
Percent Discordant	39.9	Gamma	0.198
Percent Tied	0.4	Tau-a	0.087
Pairs	4343320 0	c	0.598

/*Stepwise Selection Logistic Regression */

PROC LOGISTIC DATA= ASSGN1.REG_DATA;

class VDDASAV;

model GOOD= AGEOTD AGEAVG TRADES TOTBAL HSATRT CURSAT BRTRDS
BROPEN VDDASAV / Selection=stepwise;

RUN;

Model Information		
Data Set	ASSGN1.REG_DATA	
Response Variable	GOOD	GOOD
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	1404 2
Number of Observations Used	1404 2

Response Profile		
Ordered Value	GOOD	Total Frequency
1	0	4600
2	1	9442

Probability modeled is GOOD='0'.

Stepwise Selection Procedure

Class Level Information				
Class	Value	Design Variables		
VDDASAV	BOTH	1	0	0
	DDA	0	1	0
	NONE	0	0	1
	SAV	-1	-1	-1

Step 0. Intercept entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L	=	17761.945
----------	---	-----------

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
396.4490	11	<.0001

Step 1. Effect HSATRT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17572.267
SC	17771.495	17587.366
-2 Log L	17761.945	17568.267

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	193.6779	1	<.0001
Score	181.6459	1	<.0001
Wald	176.8417	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
219.1733	10	<.0001

Note No effects for the model in Step 1 are
: removed.

Step 2. Effect AGEAVG entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.94 5	17482.926
SC	17771.49 5	17505.575
-2 Log L	17761.94 5	17476.926

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	285.0192	2	<.0001
Score	271.0157	2	<.0001
Wald	265.3954	2	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
134.7184	9	<.0001

Note No effects for the model in Step 2 are
: removed.

Step 3. Effect VDDASAV entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17441.706
SC	17771.495	17487.005
-2 Log L	17761.945	17429.706

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	332.2391	5	<.0001
Score	316.4653	5	<.0001
Wald	309.0025	5	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
88.7894	6	<.0001

Note No effects for the model in Step 3 are
: removed.

Step 4. Effect TRADES entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17404.425
SC	17771.495	17457.273
-2 Log L	17761.945	17390.425

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	371.5201	6	<.0001
Score	347.2806	6	<.0001
Wald	336.8855	6	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
47.7558	5	<.0001

Note No effects for the model in Step 4 are
: removed.

Step 5. Effect BROPEN entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17369.615
SC	17771.495	17430.014
-2 Log L	17761.945	17353.615

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	408.3295	7	<.0001
Score	385.3342	7	<.0001
Wald	372.4776	7	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
11.1095	4	0.0254

Note No effects for the model in Step 5 are
: removed.

Step 6. Effect CURSAT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	17763.945	17365.950
SC	17771.495	17433.898
-2 Log L	17761.945	17347.950

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	413.9947	8	<.0001
Score	391.3689	8	<.0001
Wald	377.6231	8	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
5.5458	3	0.1359

Note No effects for the model in Step 6 are
: removed.

Note No (additional) effects met the 0.05 significance level for entry into the
: model.

Summary of Stepwise Selection								
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed						
1	HSATRT		1	1	181.6459		<.0001	HSATRT
2	AGEAVG		1	2	87.4223		<.0001	AGEAVG
3	VDDASAV		3	3	46.2646		<.0001	VDDASAV
4	TRADES		1	4	40.1221		<.0001	TRADES
5	BROPEN		1	5	36.8180		<.0001	BROPEN
6	CURSAT		1	6	5.5929		0.0180	CURSAT

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AGEAVG	1	83.9199	<.0001
TRADES	1	64.1390	<.0001
HSATRT	1	57.1942	<.0001
CURSAT	1	5.5859	0.0181
BROPEN	1	34.6884	<.0001
VDDASAV	3	43.4840	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7923	0.0486	266.0681	<.0001
AGEAVG		1	-0.00774	0.000845	83.9199	<.0001
TRADES		1	0.0234	0.00292	64.1390	<.0001
HSATRT		1	-0.00848	0.00112	57.1942	<.0001
CURSAT		1	-0.0336	0.0142	5.5859	0.0181
BROPEN		1	-0.0375	0.00637	34.6884	<.0001
VDDASAV	BOTH	1	-0.1327	0.0638	4.3197	0.0377
VDDASAV	DDA	1	-0.1289	0.0469	7.5520	0.0060
VDDASAV	NONE	1	0.1872	0.0331	31.9711	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGEAVG	0.992	0.991	0.994
TRADES	1.024	1.018	1.030
HSATRT	0.992	0.989	0.994
CURSAT	0.967	0.940	0.994
BROPEN	0.963	0.951	0.975
VDDASAV BOTH vs SAV	0.813	0.663	0.997
VDDASAV DDA vs SAV	0.816	0.692	0.963
VDDASAV NONE vs SAV	1.119	0.976	1.283

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	59.6	Somers' D	0.197
Percent Discordant	39.9	Gamma	0.198
Percent Tied	0.4	Tau-a	0.087
Pairs	4343320 0	c	0.598