

## Anscombe Data: mydata.anscombe

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.ANSCOMBE	<b>Observations</b>	11
<b>Member Type</b>	DATA	<b>Variables</b>	8
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:20 PM	<b>Observation Length</b>	64
<b>Last Modified</b>	Monday, July 02, 2012 01:41:20 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	2043
<b>Obs in First Data Page</b>	11
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/anscombe.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267647
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
2	X1	Num	8	8.
4	X2	Num	8	8.
6	X3	Num	8	8.
8	X4	Num	8	8.
1	Y1	Num	8	8.2
3	Y2	Num	8	8.2
5	Y3	Num	8	8.2
7	Y4	Num	8	8.2

**Anscombe Data: mydata.anscombe**

Obs	Y1	X1	Y2	X2	Y3	X3	Y4	X4
1	8.04	10	9.14	10	7.46	10	6.58	8
2	6.95	8	8.14	8	6.77	8	5.76	8
3	7.58	13	8.74	13	12.74	13	7.71	8
4	8.81	9	8.77	9	7.11	9	8.84	8
5	8.33	11	9.26	11	7.81	11	8.47	8
6	9.96	14	8.10	14	8.84	14	7.04	8
7	7.24	6	6.13	6	6.08	6	5.25	8
8	4.26	4	3.10	4	5.39	4	12.50	19
9	10.84	12	9.13	12	8.15	12	5.56	8
10	4.82	7	7.26	7	6.42	7	7.91	8
11	5.68	5	4.74	5	5.73	5	6.89	8

Regression Analysis By Example, 5th Edition - p. 29

Anscombe's Quartet is a made up data set designed to illustrate how the correlation coefficient is not adequate by itself for diagnosing a linear relationship between two variables.

## Building Prices Data: mydata.building\_prices

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.BUILDING_PRICES	<b>Observations</b>	24
<b>Member Type</b>	DATA	<b>Variables</b>	10
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	80
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1635
<b>Obs in First Data Page</b>	24
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/building_prices.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267655
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
1	X1	Num	8	8.3
2	X2	Num	8	8.3
3	X3	Num	8	8.3
4	X4	Num	8	8.3
5	X5	Num	8	8.3
6	X6	Num	8	8.3
7	X7	Num	8	8.3
8	X8	Num	8	8.3
9	X9	Num	8	8.3
10	Y	Num	8	8.3

### Building Prices Data: mydata.building\_prices

Obs	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
1	4.918	1.000	3.472	0.998	1.000	7.000	4.000	42.000	0.000	25.900
2	5.021	1.000	3.531	1.500	2.000	7.000	4.000	62.000	0.000	29.500
3	4.543	1.000	2.275	1.175	1.000	6.000	3.000	40.000	0.000	27.900
4	4.557	1.000	4.050	1.232	1.000	6.000	3.000	54.000	0.000	25.900
5	5.060	1.000	4.455	1.121	1.000	6.000	3.000	42.000	0.000	29.900
6	3.891	1.000	4.455	0.988	1.000	6.000	3.000	56.000	0.000	29.900
7	5.898	1.000	5.850	1.240	1.000	7.000	3.000	51.000	1.000	30.900
8	5.604	1.000	9.520	1.501	0.000	6.000	3.000	32.000	0.000	28.900
9	5.828	1.000	6.435	1.225	2.000	6.000	3.000	32.000	0.000	35.900
10	5.300	1.000	4.988	1.552	1.000	6.000	3.000	30.000	0.000	31.500

Regression Analysis By Example, 5th Edition - p. 328-329

Property Valuation: The objective of scientific mass appraisal (or Automated Valuation Models) is to predict the sale price of a home based on selected physical attributes of the property.

Y: sale price of house in thousands of dollars  
 X1: taxes in thousands of dollars  
 X2: number of bathrooms  
 X3: lot size in thousands of square feet  
 X4: living space in thousands of square feet  
 X5: number of garage stalls  
 X6: number of rooms  
 X7: number of bedrooms  
 X8: age of home in years  
 X9: number of fireplaces

## Cigarette Consumption Data: mydata.cigarette\_consumption

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.CIGARETTE_CONSUMPTION	<b>Observations</b>	51
<b>Member Type</b>	DATA	<b>Variables</b>	8
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	64
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	2043
<b>Obs in First Data Page</b>	51
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/cigarette_consumption.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267656
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
2	age	Num	8	8.1
5	black	Num	8	8.1
6	female	Num	8	8.1
3	hs	Num	8	8.1
4	income	Num	8	8.1
7	price	Num	8	8.1
8	sales	Num	8	8.1
1	state	Char	2	\$2.

# Cigarette Consumption Data: mydata.cigarette\_consumption

Obs	state	age	hs	income	black	female	price	sales
1	AL	27.0	41.3	2948.0	26.2	51.7	42.7	89.8
2	AK	22.9	66.7	4644.0	3.0	45.7	41.8	121.3
3	AZ	26.3	58.1	3665.0	3.0	50.8	38.5	115.2
4	AR	29.1	39.9	2878.0	18.3	51.5	38.8	100.3
5	CA	28.1	62.6	4493.0	7.0	50.8	39.7	123.0
6	CO	26.2	63.9	3855.0	3.0	50.7	31.1	124.8
7	CT	29.1	56.0	4917.0	6.0	51.5	45.5	120.0
8	DE	26.8	54.6	4524.0	14.3	51.3	41.3	155.0
9	DC	28.4	55.2	5079.0	71.1	53.5	32.6	200.4
10	FL	32.3	52.6	3738.0	15.3	51.8	43.8	123.6

Regression Analysis By Example, 5th Edition - p. 86-89

A national insurance organization wanted to study the consumption pattern of cigarettes in all 50 states plus the District of Columbia. The data are from 1970 and the response variable of interest is per capita cigarette sales.

State: State abbreviation  
 Age: median age of person living in a state  
 HS: percentage of people over 25 who completed high school  
 Income: per capita income in dollars  
 Black: percentage of population that is black  
 Female: percentage of population that is female  
 Price: weighted average price (in cents) of a pack of cigarettes  
 Sales: number of packs of cigarettes sold in a state per capita

## Financial Ratios Data: mydata.financial\_ratios

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.FINANCIAL_RATIOS	<b>Observations</b>	66
<b>Member Type</b>	DATA	<b>Variables</b>	4
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	32
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	4078
<b>Obs in First Data Page</b>	66
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/financial_ratios.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267657
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
2	X1	Num	8	8.1
3	X2	Num	8	8.1
4	X3	Num	8	8.1
1	Y	Num	8	8.

# Financial Ratios Data: mydata.financial\_ratios

Obs	Y	X1	X2	X3
1	0	-62.8	-89.5	1.7
2	0	3.3	-3.5	1.1
3	0	-120.8	-103.2	2.5
4	0	-18.1	-28.8	1.1
5	0	-3.8	-50.6	0.9
6	0	-61.2	-56.2	1.7
7	0	-20.3	-17.4	1.0
8	0	-194.5	-25.8	0.5
9	0	20.8	-4.3	1.0
10	0	-106.1	-22.9	1.5

Regression Analysis By Example, 5th Edition - p. 338-341

Analysis of bankruptcy/solvency by financial ratios.

33 firms were observed over a two year period.

If a firm is solvent after two years  $Y=1$ , if the firm went bankrupt  $Y=0$ .

X1: retained earnings / total assets

X2: earnings before interest and taxes / total assets

X3: sales / total assets



## Gasoline Consumption Data: mydata.gasoline\_consumption

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.GASOLINE_CONSUMPTION	<b>Observations</b>	30
<b>Member Type</b>	DATA	<b>Variables</b>	12
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	96
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1363
<b>Obs in First Data Page</b>	30
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/gasoline_consumption.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267660
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
2	X01	Num	8	BEST12.	BEST32.
3	X02	Num	8	BEST12.	BEST32.
4	X03	Num	8	BEST12.	BEST32.
5	X04	Num	8	BEST12.	BEST32.
6	X05	Num	8	BEST12.	BEST32.
7	X06	Num	8	BEST12.	BEST32.
8	X07	Num	8	BEST12.	BEST32.
9	X08	Num	8	BEST12.	BEST32.
10	X09	Num	8	BEST12.	BEST32.
11	X10	Num	8	BEST12.	BEST32.
12	X11	Num	8	BEST12.	BEST32.
1	Y	Num	8	BEST12.	BEST32.

### Gasoline Consumption Data: mydata.gasoline\_consumption

Obs	Y	X01	X02	X03	X04	X05	X06	X07	X08	X09	X10	X11
1	18.9	350	165	260	8	2.56	4	3	200.3	69.9	3910	1
2	17	350	170	275	8.5	2.56	4	3	199.6	72.9	3860	1
3	20	250	105	185	8.25	2.73	1	3	196.7	72.2	3510	1
4	18.3	351	143	255	8	3	2	3	199.9	74	3890	1
5	20.1	225	95	170	8.4	2.76	1	3	194.1	71.8	3365	0
6	11.2	440	215	330	8.2	2.88	4	3	184.5	69	4215	1
7	22.1	231	110	175	8	2.56	2	3	179.3	65.4	3020	1
8	21.5	262	110	200	8.5	2.56	2	3	179.3	65.4	3180	1
9	34.7	89.7	70	81	8.2	3.9	2	4	155.7	64	1905	0
10	30.4	96.9	75	83	9	4.3	2	5	165.2	65	2320	0

Regression Analysis By Example, 5th Edition - p. 255-256

Gasoline consumption (MPG) was collected on 30 models of cars along with physical and mechanical attributes of the cars.

Y: Miles Per Gallon (MPG)  
 X1: displacement in cubic inches  
 X2: horsepower in feet per pound  
 X3: torque in feet per pound  
 X4: compression ratio  
 X5: rear axle ratio  
 X6: carburetor barrels  
 X7: number of transmission speeds  
 X8: overall vehicle length in inches  
 X9: vehicle width in inches  
 X10: vehicle weight in pounds  
 X11: type of transmission, 1=automatic and 0=manual

## Credit Approval Data: mydata.credit\_approval

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.CREDIT_APPROVAL	<b>Observations</b>	690
<b>Member Type</b>	DATA	<b>Variables</b>	16
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	64
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

## Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	2043
<b>Obs in First Data Page</b>	690
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu/i_833463/c_3505/SAS_Data/credit_approval.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267658
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

## Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
1	A1	Char	1	\$1.
2	A2	Num	8	8.3
3	A3	Num	8	8.3
4	A4	Char	1	\$1.
5	A5	Char	2	\$2.
6	A6	Char	2	\$2.
7	A7	Char	2	\$2.
8	A8	Num	8	8.3
9	A9	Char	1	\$1.
10	A10	Char	1	\$1.
11	A11	Num	8	8.3
12	A12	Char	1	\$1.
13	A13	Char	1	\$1.
14	A14	Num	8	8.3
15	A15	Num	8	8.3
16	A16	Char	1	\$1.

## Credit Approval Data: mydata.credit\_approval

Obs	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
1	b	30.830	0.000	u	g	w	v	1.250	t	t	1.000	f	g	202.000	0.000	+
2	a	58.670	4.460	u	g	q	h	3.040	t	t	6.000	f	g	43.000	560.000	+
3	a	24.500	0.500	u	g	q	h	1.500	t	f	0.000	f	g	280.000	824.000	+
4	b	27.830	1.540	u	g	w	v	3.750	t	t	5.000	t	g	100.000	3.000	+
5	b	20.170	5.625	u	g	w	v	1.710	t	f	0.000	f	s	120.000	0.000	+
6	b	32.080	4.000	u	g	m	v	2.500	t	f	0.000	t	g	360.000	0.000	+
7	b	33.170	1.040	u	g	r	h	6.500	t	f	0.000	t	g	164.000	31285.00	+
8	a	22.920	11.585	u	g	cc	v	0.040	t	f	0.000	f	g	80.000	1349.000	+
9	b	54.420	0.500	y	p	k	h	3.960	t	f	0.000	f	g	180.000	314.000	+
10	b	42.500	4.915	y	p	w	v	3.165	t	f	0.000	t	g	52.000	1442.000	+

1. Title: Credit Approval

2. Sources:

(confidential)

Submitted by quinlan@cs.su.oz.au

3. Past Usage:

See Quinlan,

\* "Simplifying decision trees", Int J Man-Machine Studies 27, Dec 1987, pp. 221-234.

\* "C4.5: Programs for Machine Learning", Morgan Kaufmann, Oct 1992

4. Relevant Information:

This file concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data.

This dataset is interesting because there is a good mix of attributes -- continuous, nominal with small numbers of values, and nominal with larger numbers of values. There are also a few missing values.

5. Number of Instances: 690

6. Number of Attributes: 15 + class attribute

**Credit Approval Data: mydata.credit\_approval**

## 7. Attribute Information:

A1: b, a.  
A2: continuous.  
A3: continuous.  
A4: u, y, l, t.  
A5: g, p, gg.  
A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.  
A7: v, h, bb, j, n, z, dd, ff, o.  
A8: continuous.  
A9: t, f.  
A10: t, f.  
A11: continuous.  
A12: t, f.  
A13: g, p, s.  
A14: continuous.  
A15: continuous.  
A16: +,- (class attribute)

## 8. Missing Attribute Values:

37 cases (5%) have one or more missing values. The missing values from particular attributes are:

A1: 12  
A2: 12  
A4: 6  
A5: 6  
A6: 9  
A7: 9  
A14: 13

## 9. Class Distribution

+: 307 (44.5%)  
-: 383 (55.5%)

## Recidivism Data: mydata.recidivism

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.RECIDIVISM	<b>Observations</b>	1445
<b>Member Type</b>	DATA	<b>Variables</b>	18
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	144
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	2
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	909
<b>Obs in First Data Page</b>	881
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/recidivism.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267659
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	270336

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
13	age	Num	8	8.
2	alcohol	Num	8	8.
1	black	Num	8	8.
17	censored	Num	8	8.
3	drugs	Num	8	8.
16	duration	Num	8	8.
11	education	Num	8	8.
6	felony	Num	8	8.
15	follow_up	Num	8	8.
18	log_duration	Num	8	8.6
5	married	Num	8	8.
10	nbr_priors	Num	8	8.
12	nbr_rules	Num	8	8.
9	person	Num	8	8.
8	property	Num	8	8.
4	supervised	Num	8	8.

**Recidivism Data: mydata.recidivism****The CONTENTS Procedure**

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
14	time_served	Num	8	8.
7	workprg	Num	8	8.

### Recidivism Data: mydata.recidivism

Obs	black	alcohol	drugs	supervised	married	felony	workprg	property	person	nbr_priors	education
1	0	1	0	1	1	0	1	0	0	0	7
2	1	0	0	1	0	1	1	1	0	0	12
3	0	0	0	0	0	0	1	1	0	0	9
4	0	0	1	1	0	1	1	1	0	2	9
5	0	0	1	1	0	0	0	0	0	0	9
6	1	0	0	1	0	0	1	0	0	1	12
7	1	1	0	1	0	0	0	0	0	0	12
8	1	0	0	1	0	1	1	1	0	2	10
9	0	0	0	1	1	0	0	0	0	2	10
10	1	0	0	1	1	1	1	0	0	0	11

Obs	nbr_rules	age	time_served	follow_up	duration	censored	log_duration
1	2	441	30	72	72	1	4.276666
2	0	307	19	75	75	1	4.317488
3	5	262	27	81	9	0	2.197225
4	3	253	38	76	25	0	3.218876
5	0	244	4	81	81	1	4.394449
6	0	277	13	79	79	1	4.369448
7	2	327	28	73	73	1	4.290459
8	3	314	34	73	73	1	4.290459
9	0	388	14	73	54	0	3.988984
10	2	430	32	73	60	0	4.094345

### Data Description for Recidivism Data

The dataset considered here is analyzed in Wooldridge (2002) and credited to Chung, Schmidt and Witte (1991). The data pertain to a random sample of convicts released from prison between July 1, 1977 and June 30, 1978. Of interest is the time until they return to prison. The information was collected retrospectively by looking at records in April 1984, so the maximum possible length of observation is 81 months.

Although there are several variables in the data set, the ones of interest are: `duration`, which is the duration in months until return to prison; `censored`, is the censoring indicator variable; `workprg`, an indicator of participation in a work program; `nbr_priors`, the number of previous convictions; `time_served`, the time served rounded to months; `felony`, an indicator of felony sentences; `alcohol`, an indicator of alcohol problems; `drugs`, an indicator of drug use history; `black`, an indicator for African Americans; `married`, an indicator if married when incarcerated; `education`, the number of years of schooling; `age` in months.



## Recidivism Data: mydata.recidivism

### Complete Description of Recidivism Variables

black	alcohol	drugs	supervised	married	felony
workprg	property	person	nbr_priors	education	nbr_rules
age	time_served	follow_up	duration	censored	log_duration

Obs: 1445

1. black	=1 if black
2. alcohol	=1 if alcohol problems
3. drugs	=1 if drug history
4. supervised	=1 if release supervised
5. married	=1 if married when incarcerated.
6. felony	=1 if felony sentence
7. workprg	=1 if in N.C. pris. work prg.
8. property	=1 if property crime
9. person	=1 if crime against person
10. nbr_priors	# prior convictions
11. education	years of schooling
12. nbr_rules	# rules violations in prison
13. age	in months
14. time_served	time served, rounded to months
15. follow_up	length follow period, months
16. duration	max(time until return, follow)
17. censored	=1 if duration right censored
18. log_duration	log(duration)

# Benign Breast Disease 1-3 Matched Case-Control Study: mydata.benign\_breast\_disease

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.BENIGN_BREAST_DISEASE	<b>Observations</b>	200
<b>Member Type</b>	DATA	<b>Variables</b>	14
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	112
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1168
<b>Obs in First Data Page</b>	200
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/benign_breast_disease.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267661
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
13	AGLP	Num	8	8.
9	AGMN	Num	8	8.
3	AGMT	Num	8	8.
8	AGP1	Num	8	8.
7	CHK	Num	8	8.
6	DEG	Num	8	8.
4	FNDX	Num	8	8.
5	HIGD	Num	8	8.
11	LIV	Num	8	8.
14	MST	Num	8	8.
10	NLV	Num	8	8.
2	OBS	Num	8	8.
1	STR	Num	8	8.
12	WT	Num	8	8.

# Benign Breast Disease 1-3 Matched Case-Control Study: mydata.benign\_breast\_disease

Obs	STR	OBS	AGMT	FNDX	HIGD	DEG	CHK	AGP1	AGMN	NLV	LIV	WT	AGLP	MST
1	1	1	39	1	9	0	1	23	13	0	5	118	39	1
2	1	2	39	0	10	0	2	16	11	1	3	175	39	3
3	1	3	39	0	11	0	2	20	12	1	3	135	39	2
4	1	4	39	0	12	1	1	21	11	0	3	125	40	1
5	2	1	38	1	14	2	1	.	14	.	.	118	39	1
6	2	2	38	0	12	1	2	20	15	0	2	183	38	1
7	2	3	38	0	9	0	2	19	11	0	5	218	38	1
8	2	4	38	0	13	1	1	23	13	0	2	192	37	1
9	3	1	38	1	9	0	1	22	15	2	2	125	38	1
10	3	2	38	0	10	0	2	20	14	0	2	123	38	1

## Description of Variables in the Benign Breast Disease 1-3 Matched Case-Control Study Described in Section 7.5 page 244.

Variable	Description	Codes/Values	Name
1	Stratum	1 - 50	STR
2	Observation within Stratum	1 = Case, 2 - 4 = Control	OBS
3	Age at Interview	Years	AGMT
4	Final Diagnosis	1 = Case, 0 = Control	FNDX
5	Highest Grade in School	5 - 20	HIGD
6	Degree	0 = None 1 = High School 2 = Jr. College 3 = College 4 = Masters 5 = Doctoral	DEG
7	Regular Medical Check-ups	1 = Yes, 2 = No	CHK
8	Age at First Pregnancy	Years	AGP1
9	Age at Menarche	Years	AGMN
10	No. of Stillbirths, Miscarriages etc.	0 - 7	NLV
11	Number of Live Births	0 - 11	LIV
12	Weight of the Subject At Interview	Pounds	WT
13	Age at Last Menstrual Period	Years	AGLP
14	Marital Status	1 = Married 2 = Divorced 3 = Separated 4 = Widowed 5 = Never Married	MST

## Child Age and Heart Disease: mydata.child\_age\_heart\_disease

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.CHILD_AGE_HEART_DISEASE	<b>Observations</b>	100
<b>Member Type</b>	DATA	<b>Variables</b>	3
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	24
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	5431
<b>Obs in First Data Page</b>	100
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/child_age_heart_disease.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267662
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
2	AGE	Num	8	8.
3	HEART_DISEASE	Num	8	8.
1	ID	Num	8	8.

**Child Age and Heart Disease: mydata.child\_age\_heart\_disease**

Obs	ID	AGE	HEART_DISEASE
1	1	20	0
2	2	23	0
3	3	24	0
4	5	25	1
5	4	25	0
6	7	26	0
7	6	26	0
8	9	28	0
9	8	28	0
10	10	29	0

**Code Sheet for the Chd-Age data in Table 1.1 page 3 of Applied Logistic Regression: Second Edition**

Variable	Name	Values
1	Identification Code	1-100
2	Age	Years
3	Evidence of Coronary Heart Disease	0 = No, 1 = Yes

## Low Birth Weight Data: mydata.low\_birth\_weight

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.LOW_BIRTH_WEIGHT	<b>Observations</b>	189
<b>Member Type</b>	DATA	<b>Variables</b>	11
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:21 PM	<b>Observation Length</b>	88
<b>Last Modified</b>	Monday, July 02, 2012 01:41:21 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1486
<b>Obs in First Data Page</b>	189
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/low_birth_weight.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267649
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
3	AGE	Num	8	8.
11	BWT	Num	8	8.
10	FTV	Num	8	8.
8	HT	Num	8	8.
1	ID	Num	8	8.
2	LOW	Num	8	8.
4	LWT	Num	8	8.
7	PTL	Num	8	8.
5	RACE	Num	8	8.
6	SMOKE	Num	8	8.
9	UI	Num	8	8.

### Low Birth Weight Data: mydata.low\_birth\_weight

Obs	ID	LOW	AGE	LWT	RACE	SMOKE	PTL	HT	UI	FTV	BWT
1	4	1	28	120	3	1	1	0	1	0	709
2	10	1	29	130	1	0	0	0	1	2	1021
3	11	1	34	187	2	1	0	1	0	0	1135
4	13	1	25	105	3	0	1	1	0	0	1330
5	15	1	25	85	3	0	0	0	1	0	1474
6	16	1	27	150	3	0	0	0	0	0	1588
7	17	1	23	97	3	0	0	0	1	1	1588
8	18	1	24	128	2	0	1	0	0	1	1701
9	19	1	24	132	3	0	0	1	0	0	1729
10	20	1	21	165	1	1	0	1	0	1	1790

#### Code Sheet for the Variables in the Low Birth Weight Study Described in Section 1.6.2 Page 24

Variable	Description	Codes/Values	Name
1	Identification Code	ID Number	ID
2	Low Birth Weight	1 = BWT≤2500g, 0 = BWT>2500g	LOW
3	Age of Mother	Years	AGE
4	Weight of Mother at Last Menstrual Period	Pounds	LWT
5	Race	1 = White, 2 = Black 3 = Other	RACE
6	Smoking Status During Pregnancy	0 = No, 1 = Yes	SMOKE
7	History of Premature Labor	0 = None, 1 = One, 2 = Two, etc.	PTL
8	History of Hypertension	0 = No, 1 = Yes	HT
9	Presence of Uterine Irritability	0 = No, 1 = Yes	UI
10	Number of Physician Visits During the First Trimester	0 = None, 1 = One 2 = Two, etc.	FTV
11	Birth Weight	Grams	BWT

# Matched Pair Low Birth Weight Data: mydata.matched\_pair\_low\_birth\_weight

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.MATCHED_PAIR_LOW_BIRTH_WEIGHT	<b>Observations</b>	112
<b>Member Type</b>	DATA	<b>Variables</b>	9
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	72
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

## Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1816
<b>Obs in First Data Page</b>	112
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/matched_pair_low_birth_weight.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267654
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

## Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
3	AGE	Num	8	8.
8	HT	Num	8	8.
2	LOW	Num	8	8.
4	LWT	Num	8	8.
1	PAIR	Num	8	8.
7	PTD	Num	8	8.
5	RACE	Num	8	8.
6	SMOKE	Num	8	8.
9	UI	Num	8	8.



# Matched Pair Low Birth Weight Data: mydata.matched\_pair\_low\_birth\_weight

Obs	PAIR	LOW	AGE	LWT	RACE	SMOKE	PTD	HT	UI
1	1	0	14	135	1	0	0	0	0
2	1	1	14	101	3	1	1	0	0
3	2	0	15	98	2	0	0	0	0
4	2	1	15	115	3	0	0	0	1
5	3	0	16	95	3	0	0	0	0
6	3	1	16	130	3	0	0	0	0
7	4	0	17	103	3	0	0	0	0
8	4	1	17	130	3	1	1	0	1
9	5	0	17	122	1	1	0	0	0
10	5	1	17	110	1	1	0	0	0

## Code Sheet for the Variables in the 1-1matched Low Birth Weight Study Described in Section 7.3 Page 230

Variable	Description	Codes/Values	Name
1	Pair	1 -50	PAIR
2	Low Birth Weight	1 = BWT<=2500g, 0 = BWT>2500g	LOW
3	Age of Mother	Years	AGE
4	Weight of Mother at Last Menstrual Period	Pounds	LWT
5	Race	1 = White, 2 = Black 3 = Other	RACE
6	Smoking Status During Pregnancy	0 = No, 1 = Yes	SMOKE
7	History of Premature Labor	0 = None, 1 = Yes	PTD
8	History of Hypertension	0 = No, 1 = Yes	HT
9	Presence of Uterine Irritability	0 = No, 1 = Yes	UI

## Longitudinal Low Birth Weight Data: mydata.longitudinal\_low\_birth\_weight

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.LONGITUDINAL_LOW_BIRTH_WEIGHT	<b>Observations</b>	488
<b>Member Type</b>	DATA	<b>Variables</b>	8
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	64
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	2043
<b>Obs in First Data Page</b>	488
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/longitudinal_low_birth_weight.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267653
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
5	AGE	Num	8	8.
2	BIRTH	Num	8	8.
7	BWT	Num	8	8.
1	ID	Num	8	8.
8	LOW	Num	8	8.
6	LWT	Num	8	8.
4	RACE	Num	8	8.
3	SMOKE	Num	8	8.

# Longitudinal Low Birth Weight Data: mydata.longitudinal\_low\_birth\_weight

Obs	ID	BIRTH	SMOKE	RACE	AGE	LWT	BWT	LOW
1	1	1	1	3	28	120	2865	0
2	1	2	1	3	33	141	2609	0
3	2	1	0	1	29	130	2613	0
4	2	2	0	1	34	151	3125	0
5	2	3	0	1	37	144	2481	1
6	3	1	1	2	31	187	1841	1
7	3	2	1	2	35	209	1598	1
8	3	3	1	2	41	217	2015	1
9	4	1	0	3	25	105	3489	0
10	4	2	0	3	30	129	3554	0

## Code Sheet for the Variables in the Longitudinal Low Birth Weight Study Described in Section 8.3

Variable	Description	Codes/Values	Name
1	Identification Code	ID Number	ID
2	Birth Number	1-4	BIRTH
3	Smoking Status During Pregnancy	0 = No, 1 = Yes	SMOKE
4	Race	1 = White, 2 = Black 3 = Other	RACE
5	Age of Mother	Years	AGE
6	Weight of Mother at Last Menstrual Period	Pounds	LWT
7	Birth Weight	Grams	BWT
8	Low Birth Weight	1 = BWT $\leq$ 2500g, 0 = BWT $>$ 2500g	LOW

## ICU Data: mydata.icu

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.ICU	<b>Observations</b>	200
<b>Member Type</b>	DATA	<b>Variables</b>	21
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	168
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	779
<b>Obs in First Data Page</b>	200
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/icu.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267652
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
3	AGE	Num	8	8.
19	BIC	Num	8	8.
7	CAN	Num	8	8.
10	CPR	Num	8	8.
20	CRE	Num	8	8.
8	CRN	Num	8	8.
15	FRA	Num	8	8.
12	HRA	Num	8	8.
1	ID	Num	8	8.
9	INF	Num	8	8.
21	LOC	Num	8	8.
18	PCO	Num	8	8.
17	PH	Num	8	8.
16	PO2	Num	8	8.
13	PRE	Num	8	8.
5	RACE	Num	8	8.

## ICU Data: mydata.icu

### The CONTENTS Procedure

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
6	SER	Num	8	8.
4	SEX	Num	8	8.
2	STA	Num	8	8.
11	SYS	Num	8	8.
14	TYP	Num	8	8.

## ICU Data: mydata.icu

Obs	ID	STA	AGE	SEX	RACE	SER	CAN	CRN	INF	CPR	SYS	HRA	PRE	TYP	FRA	PO2	PH	PCO	BIC	CRE	LOC
1	4	1	87	1	1	1	0	0	1	0	80	96	0	1	1	1	1	1	0	0	0
2	8	0	27	1	1	0	0	0	1	0	142	88	0	1	0	0	0	0	0	0	0
3	12	0	59	0	1	0	0	0	0	0	112	80	1	1	0	0	0	0	0	0	0
4	14	0	77	0	1	1	0	0	0	0	100	70	0	0	0	0	0	0	0	0	0
5	27	1	76	1	1	1	0	0	1	0	128	90	1	1	0	0	0	0	0	0	0
6	28	0	54	0	1	0	0	0	1	0	142	103	0	1	1	0	0	0	0	0	0
7	32	0	87	1	1	1	0	0	1	0	110	154	1	1	0	0	0	0	0	0	0
8	38	0	69	0	1	0	0	0	1	0	110	132	0	1	0	1	0	0	1	0	0
9	40	0	63	0	1	1	0	0	0	0	104	66	0	0	0	0	0	0	0	0	0
10	41	0	30	1	1	0	0	0	0	0	144	110	0	1	0	0	0	0	0	0	0

## Code Sheet for the ICU Study Described in Section 1.6.1 page 22

Variable	Description	Codes/Values	Name
1	Identification Code	ID Number	ID
2	Vital Status	0 = Lived, 1 = Died	STA
3	Age	Years	AGE
4	Sex	0 = Male, 1 = Female	SEX
5	Race	1 = White, 2 = Black, 3 = Other	RACE
6	Service at ICU Admission	0 = Medical, 1 = Surgical	SER
7	Cancer Part of Present Problem	0 = No, 1 = Yes	CAN
8	History of Chronic Renal Failure	0 = No, 1 = Yes	CRN
9	Infection Probable at ICU Admission	0 = No, 1 = Yes	INF
10	CPR Prior to ICU Admission	0 = No, 1 = Yes	CPR
11	Systolic Blood Pressure at ICU Admission	mm Hg	SYS
12	Heart Rate at ICU Admission	Beats/min	HRA
13	Previous Admission to an ICU Within 6 Months	0 = No, 1 = Yes	PRE
14	Type of Admission	0 = Elective, 1 = Emergency	TYP
15	Long Bone, Multiple, Neck, Single Area, or Hip Fracture	0 = No, 1 = Yes	FRA
16	PO2 from Initial Blood Gases	0 = >60, 1 = ≤ 60	PO2
17	PH from Initial Blood Gases	0 = ≥ 7.25, 1 = < 7.25	PH
18	PCO2 from Initial Blood Gases	0 = ≤ 45, 1 = > 45	PCO
19	Bicarbonate from Initial Blood Gases	0 = ≥ 18, 1 = < 18	BIC
20	Creatine from Initial Blood Gases	0 = ≤ 2, 1 = > 2	CRE
21	Level of Consciousness at ICU Admission	0 = No Coma or Deep Stupor, 1 = Deep Stupor, 2 = Coma	LOC

# UMARU Impact Study Data: mydata.umaru\_impact\_study

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.UMARU_IMPACT_STUDY	<b>Observations</b>	575
<b>Member Type</b>	DATA	<b>Variables</b>	9
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	72
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1816
<b>Obs in First Data Page</b>	575
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/umaru_impact_study.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267651
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

## Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
2	AGE	Num	8	8.
3	BECK	Num	8	8.3
9	DFREE	Num	8	8.
1	ID	Num	8	8.
4	IVHX	Num	8	8.
5	NDRUGTX	Num	8	8.
6	RACE	Num	8	8.
8	SITE	Num	8	8.
7	TREAT	Num	8	8.

**UMARU Impact Study Data: mydata.umaru\_impact\_study**

Obs	ID	AGE	BECK	IVHX	NDRUGTX	RACE	TREAT	SITE	DFREE
1	1	39	9.000	3	1	0	1	0	0
2	2	33	34.000	2	8	0	1	0	0
3	3	33	10.000	3	3	0	1	0	0
4	4	32	20.000	3	1	0	0	0	0
5	5	24	5.000	1	5	1	1	0	1
6	6	30	32.550	3	1	0	1	0	0
7	7	39	19.000	3	34	0	1	0	1
8	8	27	10.000	3	2	0	1	0	0
9	9	40	29.000	3	3	0	1	0	0
10	10	36	25.000	3	7	0	1	0	0

**Description of Variables in the UMARU IMPACT Study**  
**Described in Section 1.6.4 page 26**

Variable	Description	Codes/Values	Name
1	Identification Code	1-575	ID
2	Age at Enrollment	Years	AGE
3	Beck Depression Score at Admission	0.000-54.000	BECK
4	IV Drug Use History at Admission	1 = Never, 2 = Previous 3 = Recent	IVHX
5	Number of Prior Drug Treatments	0-40	NDRUGTX
6	Subject's Race	0 = White, 1 = Other	RACE
7	Treatment Randomization Assignment	0 = Short, 1 = Long	TREAT
8	Treatment Site	0 = A, 1 = B	SITE
9	Remained Drug Free For 12 Months	1 = Remained Drug Free 0 = Otherwise	DFREE



# Prostate Cancer Study Data: mydata.prostate\_cancer\_study

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.PROSTATE_CANCER_STUDY	<b>Observations</b>	380
<b>Member Type</b>	DATA	<b>Variables</b>	9
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	72
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

## Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1816
<b>Obs in First Data Page</b>	380
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/prostate_cancer_study.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267648
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

## Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
3	AGE	Num	8	8.
2	CAPSULE	Num	8	8.
6	DCAPS	Num	8	8.
5	DPROS	Num	8	8.
9	GLEASON	Num	8	8.
1	ID	Num	8	8.
7	PSA	Num	8	8.
4	RACE	Num	8	8.
8	VOL	Num	8	8.

## Prostate Cancer Study Data: mydata.prostate\_cancer\_study

Obs	ID	CAPSULE	AGE	RACE	DPROS	DCAPS	PSA	VOL	GLEASON
1	1	0	65	1	2	1	1	0	6
2	2	0	72	1	3	2	7	0	7
3	3	0	70	1	1	2	5	0	6
4	4	0	76	2	2	1	51	20	7
5	5	0	69	1	1	1	12	56	6
6	6	1	71	1	3	2	3	0	8
7	7	0	68	2	4	2	32	0	7
8	8	0	61	2	4	2	67	27	7
9	9	0	69	1	1	1	4	24	7
10	10	0	68	2	1	2	13	0	6

## Code Sheet for the Prostate Cancer Study Described in Section 1.6.3 page 25

Variable	Description	Codes/Values	Name
1	Identification Code	1 - 380	ID
2	Tumor Penetration of Prostatic Capsule	0 = No Penetration, 1 = Penetration	CAPSULE
3	Age	Years	AGE
4	Race	1= White, 2 = Black	RACE
5	Results of the Digital Rectal Exam	1 = No Nodule 2 = Unilobar Nodule (Left) 3 = Unilobar Nodule (Right) 4 = Bilobar Nodule	DPROS
6	Detection of Capsular Involvement in Rectal Exam	1 = No, 2 = Yes	DCAPS
7	Prostatic Specific Antigen Value	mg/ml	PSA
8	Tumor Volume Obtained from Ultrasound	cm3	VOL
9	Total Gleason Score	0 - 10	GLEASON

## Mammography Experience Study Data: mydata.mammography\_experience

### The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.MAMMOGRAPHY_EXPERIENCE	<b>Observations</b>	412
<b>Member Type</b>	DATA	<b>Variables</b>	7
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	56
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

### Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	2334
<b>Obs in First Data Page</b>	412
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/mammography_experience.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267650
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	139264

### Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format
6	BSE	Num	8	8.
7	DETC	Num	8	8.
5	HIST	Num	8	8.
2	ME	Num	8	8.
1	OBS	Num	8	8.
4	PB	Num	8	8.
3	SYMPT	Num	8	8.

# Mammography Experience Study Data: mydata.mammography\_experience

Obs	OBS	ME	SYMPT	PB	HIST	BSE	DETC
1	1	0	3	7	0	1	2
2	2	0	2	11	0	1	3
3	3	0	3	8	1	1	3
4	4	1	3	11	0	1	3
5	5	2	4	7	0	1	3
6	6	0	3	7	0	1	3
7	7	2	4	6	0	1	2
8	8	0	4	6	0	1	3
9	9	0	2	6	0	1	3
10	10	1	4	6	0	1	3

## Code Sheet for the Variables in the Mammography Experience Study Described in Section 8.1.2 page 264.

Variable	Description	Codes/Values	Name
1	Identification Code	1-412	OBS
2	Mammograph Experience	0 = Never 1 = Within One Yea 2 = Over One Year Ago	ME
3	"You do not need a mamogram unless you develop symptoms"	1 = Strongly Agree 2 = Agree 3 = Disagree 4 = Stongly Disagree	SYMPT
4	Perveived benefit of mammography*	5 - 20	PB
5	Mother or Sister with a history of breast cancer	0 = No, 1 = Yes	HIST
6	"Has anyone taught you how to examine your own breasts: that is BSE"	0 = No, 1 = Yes	BSE
7	"How likely is it that a mamogram could find a new case of breast cancer"	1= Not likely 2 = Somewhat likely 3 = Very likely	DETC

\*The variable PB is the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

## NHANES III Data: mydata.nhanes4

## The CONTENTS Procedure

<b>Data Set Name</b>	MYDATA.NHANES3	<b>Observations</b>	17031
<b>Member Type</b>	DATA	<b>Variables</b>	16
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	Monday, July 02, 2012 01:41:22 PM	<b>Observation Length</b>	128
<b>Last Modified</b>	Monday, July 02, 2012 01:41:22 PM	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	17
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1022
<b>Obs in First Data Page</b>	993
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data/nhanes3.sas7bdat
<b>Release Created</b>	9.0202M3
<b>Host Created</b>	SunOS
<b>Inode Number</b>	267645
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	2236416

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
9	BMPHTIN	Num	8	8.2
8	BMPWTLBS	Num	8	8.2
7	DMARACER	Num	8	8.
12	HAR1	Num	8	8.
13	HAR3	Num	8	8.
16	HBP	Num	8	8.
5	HSAGEIR	Num	8	8.
6	HSSEX	Num	8	8.
10	PEPMNK1R	Num	8	8.
11	PEPMNK5R	Num	8	8.
2	SDPPSU6	Num	8	8.
3	SDPSTRA6	Num	8	8.
1	SEQN	Num	8	8.
14	SMOKE	Num	8	8.
15	TCP	Num	8	8.
4	WTPFH6	Num	8	8.2

## NHANES III Data: mydata.nhanes4

Obs	SEQN	SDPPSU6	SDPSTRA6	WTPFH6	HSAGEIR	HSSEX	DMARACER	BMPWTLBS	BMPHTIN
1	3	1	44	1735.14	21	1	1	179.50	70.40
2	4	1	43	1725.01	32	0	1	135.60	63.90
3	9	2	43	19451.83	48	0	1	149.70	61.80
4	10	1	6	27769.56	35	1	1	203.50	69.80
5	11	2	40	1245.52	48	1	1	155.30	66.20
6	19	1	35	3860.97	44	1	2	189.60	70.20
7	34	1	13	5031.70	42	0	2	125.80	62.60
8	44	1	8	28148.69	24	0	1	122.80	64.40
9	45	1	22	4582.03	67	0	2	149.60	64.30
10	48	1	24	26919.29	56	0	1	239.90	67.60

Obs	PEPMNK1R	PEPMNK5R	HAR1	HAR3	SMOKE	TCP	HBP
1	120	67	2	.	1	268	0
2	126	86	2	.	1	160	0
3	131	73	1	2	2	236	0
4	130	82	2	.	1	225	0
5	120	70	1	2	2	260	0
6	133	85	1	1	3	187	0
7	100	67	1	1	3	216	0
8	.	.	1	1	3	137	0
9	.	.	2	.	1	.	0
10	128	73	1	2	2	156	0

## Variables in the NHANES III Data Set Described in Table 6.3 page 215

Variable	Description	Codes/Values	Name
1	Respondent Identification	Number	SEQN
2	Pseudo-PSU	1,2	SDPPSU6
3	Pseudo-stratum	01 - 49	SDPSTRA6
4	Statistical weight	225.93 - 139744.9	WTPFH6
5	Age	years	HSAGEIR
6	Sex	0 = Female, 1 = Male	HSSEX
7	Race	1 = White, 2 = Black, 3 = Other	DMARACER
8	Body Weight	pounds	BMPWTLBS
9	Standing Height	inches	BMPHTIN
10	Average Systolic BP	mm Hg	PEPMNK1R
11	Average Diastolic BP	mm Hg	PEPMNK5R
12	Has respondent smoked > 100 cigarettes in life	1 = Yes, 2 = No	HAR1
13	Does respondent smoke cigarettes now?	1 = Yes, 2 = No	HAR3
14	Smoking	1 = if HAR1 = 2 2 = if HAR1=1 & HAR3=2 3 = if HAR1=1 & HAR3=1	SMOKE
15	Serum Cholesterol	mg/100ml	TCP
16	High Blood Pressure	0 if PEPMNK1R <= 140 1 if PEPMNK1R > 140	HBP

Directory	
<b>Libref</b>	MYDATA
<b>Engine</b>	V9
<b>Physical Name</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data
<b>Filename</b>	/courses/u_northwestern.edu1/i_833463/c_3505/SAS_Data
<b>Inode Number</b>	3332789
<b>Access Permission</b>	rw-rw-r-x
<b>Owner Name</b>	chadrbhatti
<b>File Size (bytes)</b>	4096

#	Name	Member Type	File Size	Last Modified
1	ANSCOMBE	DATA	139264	02Jul12:13:41:20
2	BENIGN_BREAST_DISEASE	DATA	139264	02Jul12:13:41:21
3	BUILDING_PRICES	DATA	139264	02Jul12:13:41:21
4	CHILD_AGE_HEART_DISEASE	DATA	139264	02Jul12:13:41:21
5	CIGARETTE_CONSUMPTION	DATA	139264	02Jul12:13:41:21
6	CREDIT_APPROVAL	DATA	139264	02Jul12:13:41:21
7	FINANCIAL_RATIOS	DATA	139264	02Jul12:13:41:21
8	GASOLINE_CONSUMPTION	DATA	139264	02Jul12:13:41:21
9	ICU	DATA	139264	02Jul12:13:41:22
10	LONGITUDINAL_LOW_BIRTH_WEIGHT	DATA	139264	02Jul12:13:41:22
11	LOW_BIRTH_WEIGHT	DATA	139264	02Jul12:13:41:21
12	MAMMOGRAPHY_EXPERIENCE	DATA	139264	02Jul12:13:41:22
13	MATCHED_PAIR_LOW_BIRTH_WEIGHT	DATA	139264	02Jul12:13:41:22
14	NHANES3	DATA	2236416	02Jul12:13:41:22
15	PROSTATE_CANCER_STUDY	DATA	139264	02Jul12:13:41:22
16	RECIDIVISM	DATA	270336	02Jul12:13:41:21
17	UMARU_IMPACT_STUDY	DATA	139264	02Jul12:13:41:22