Imported Diabetes Dataset Overview

The CONTENTS Procedure

Data Set Name	WORK.DIABETES_DATA	**Observations**	768
Member Type	DATA	**Variables**	9
Engine	V9	**Indexes**	0
Created	12/23/2024 14:28:00	**Observation Length**	72
Last Modified	12/23/2024 14:28:00	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1816
Obs in First Data Page	768
Number of Data Set Repairs	0
Filename	/saswork/SAS_work90F300008524_odaws02-usw2-2.oda.sas.com/SAS_workEC7200008524_odaws02-usw2-2.oda.sas.com/diabetes_data.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2281702010
Access Permission	rw-r--r--
Owner Name	u64112808
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
8	Age	Num	8	BEST12.	BEST32.
6	BMI	Num	8	BEST12.	BEST32.
3	BloodPressure	Num	8	BEST12.	BEST32.
7	DiabetesPedigreeFunction	Num	8	BEST12.	BEST32.
2	Glucose	Num	8	BEST12.	BEST32.
5	Insulin	Num	8	BEST12.	BEST32.
9	Outcome	Num	8	BEST12.	BEST32.
1	Pregnancies	Num	8	BEST12.	BEST32.
4	SkinThickness	Num	8	BEST12.	BEST32.

Metadata of Diabetes Dataset

The CONTENTS Procedure

Data Set Name	WORK.DIABETES_DATA	**Observations**	768
Member Type	DATA	**Variables**	9
Engine	V9	**Indexes**	0
Created	12/23/2024 14:28:00	**Observation Length**	72
Last Modified	12/23/2024 14:28:00	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1816
Obs in First Data Page	768
Number of Data Set Repairs	0
Filename	/saswork/SAS_work90F300008524_odaws02-usw2-2.oda.sas.com/SAS_workEC7200008524_odaws02-usw2-2.oda.sas.com/diabetes_data.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2281702010
Access Permission	rw-r--r--
Owner Name	u64112808
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
8	Age	Num	8	BEST12.	BEST32.
6	BMI	Num	8	BEST12.	BEST32.
3	BloodPressure	Num	8	BEST12.	BEST32.
7	DiabetesPedigreeFunction	Num	8	BEST12.	BEST32.
2	Glucose	Num	8	BEST12.	BEST32.

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
5	Insulin	Num	8	BEST12.	BEST32.
9	Outcome	Num	8	BEST12.	BEST32.
1	Pregnancies	Num	8	BEST12.	BEST32.
4	SkinThickness	Num	8	BEST12.	BEST32.

Sample of the First 10 Rows in the Dataset

Obs	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
6	5	116	74	0	0	25.6	0.201	30	0
7	3	78	50	32	88	31	0.248	26	1
8	10	115	0	0	0	35.3	0.134	29	0
9	2	197	70	45	543	30.5	0.158	53	1
10	8	125	96	0	0	0	0.232	54	1

Summary of Missing Values and Basic Statistics

The MEANS Procedure

Variable	N	N Miss
Pregnancies	768	0
Glucose	768	0
BloodPressure	768	0
SkinThickness	768	0
Insulin	768	0
BMI	768	0
DiabetesPedigreeFunction	768	0
Age	768	0
Outcome	768	0

Frequency Distribution of Outcome Variable

The FREQ Procedure

Outcome	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	500	65.10	500	65.10
1	268	34.90	768	100.00

Checking Missing and Invalid Values in Key Variables

The MEANS Procedure

Variable	N	N Miss	Minimum	Maximum
Glucose	768	0	0	199.0000000
BloodPressure	768	0	0	122.0000000
SkinThickness	768	0	0	99.0000000
Insulin	768	0	0	846.0000000
BMI	768	0	0	67.1000000

Summary After Replacing Invalid Zeros with Missing Values

The MEANS Procedure

Variable	N	N Miss	Minimum	Maximum
Glucose	763	5	44.0000000	199.0000000
BloodPressure	733	35	24.0000000	122.0000000
SkinThickness	541	227	7.0000000	99.0000000
Insulin	394	374	14.0000000	846.0000000
BMI	757	11	18.2000000	67.1000000

Checking Distribution of Missing Values After Cleaning

The MEANS Procedure

Variable	N	N Miss
Pregnancies	768	0
Glucose	763	5
BloodPressure	733	35
SkinThickness	541	227
Insulin	394	374
BMI	757	11
DiabetesPedigreeFunction	768	0
Age	768	0
Outcome	768	0

Summary After Correcting Imputation Logic

The MEANS Procedure

Variable	N	N Miss	Minimum	Maximum
Pregnancies	768	0	0	17.0000000

Variable	N	N Miss	Minimum	Maximum
Glucose	768	0	44.0000000	199.0000000
BloodPressure	768	0	24.0000000	122.0000000
SkinThickness	768	0	7.0000000	99.0000000
Insulin	768	0	14.0000000	846.0000000
BMI	768	0	18.2000000	67.1000000
DiabetesPedigreeFunction	768	0	0.0780000	2.4200000
Age	768	0	21.0000000	81.0000000
Outcome	768	0	0	1.0000000
TYPE	768	0	0	0
FREQ	768	0	768.0000000	768.0000000
Mean_Glucose	768	0	121.6867628	121.6867628
Mean_BP	768	0	72.4051842	72.4051842
Mean_ST	768	0	29.1534196	29.1534196
Mean_Insulin	768	0	155.5482234	155.5482234
Mean_BMI	768	0	32.4574637	32.4574637

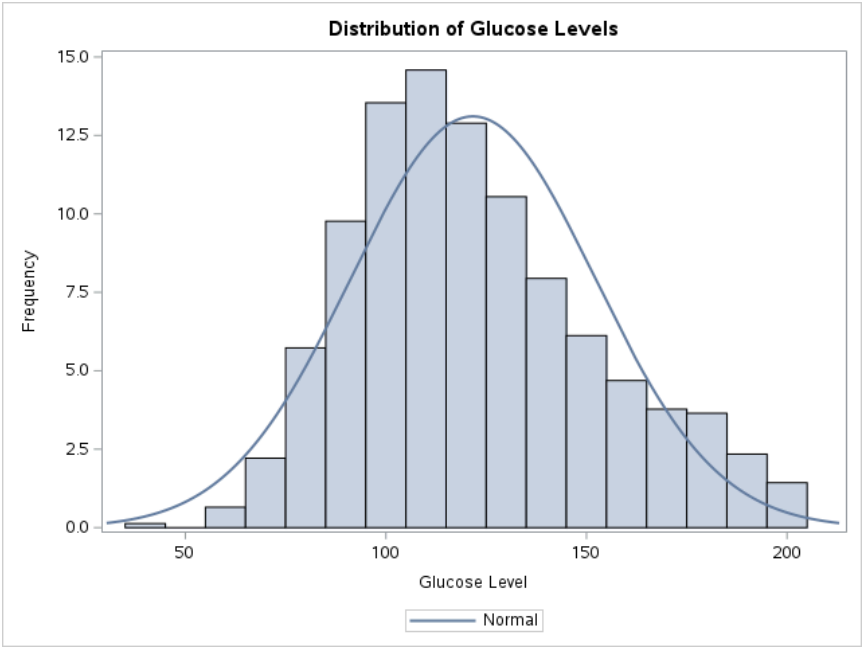
Final Cleaned and Preprocessed Dataset After Correct Imputation

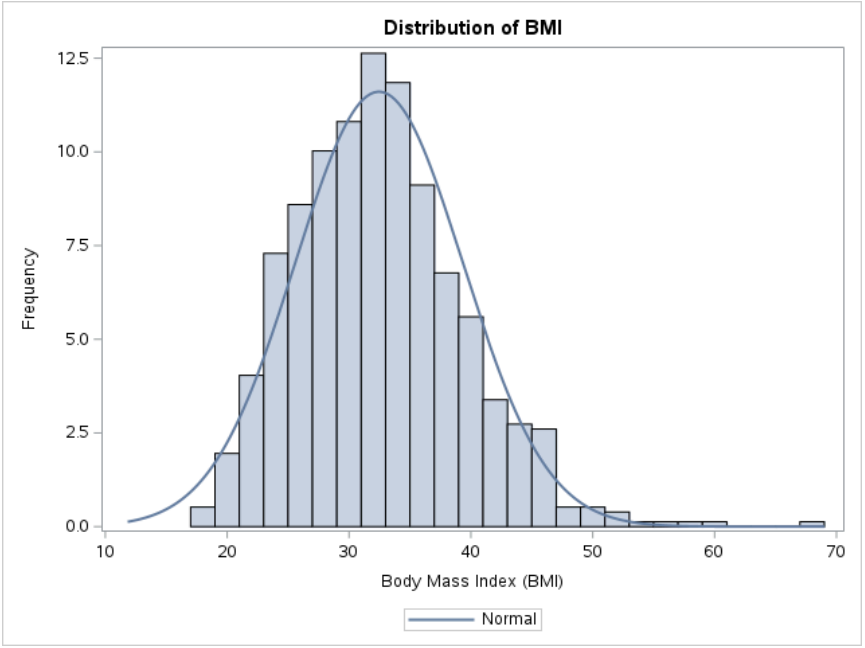
Obs	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome	_TYPE_	_FREQ_	Mean_Glucose	Mean_BP	Mean_ST	Mean_Insulin	Mean_BMI
1	6	148	72	35	155.54822335	33.6	0.627	50	1	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
2	1	85	66	29	155.54822335	26.6	0.351	31	0	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
3	8	183	64	29.153419593	155.54822335	23.3	0.672	32	1	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
4	1	89	66	23	94	28.1	0.167	21	0	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
5	0	137	40	35	168	43.1	2.288	33	1	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
6	5	116	74	29.153419593	155.54822335	25.6	0.201	30	0	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
7	3	78	50	32	88	31	0.248	26	1	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
8	10	115	72.405184175	29.153419593	155.54822335	35.3	0.134	29	0	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
9	2	197	70	45	543	30.5	0.158	53	1	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672
10	8	125	96	29.153419593	155.54822335	32.457463672	0.232	54	1	0	768	121.68676278	72.405184175	29.153419593	155.54822335	32.457463672

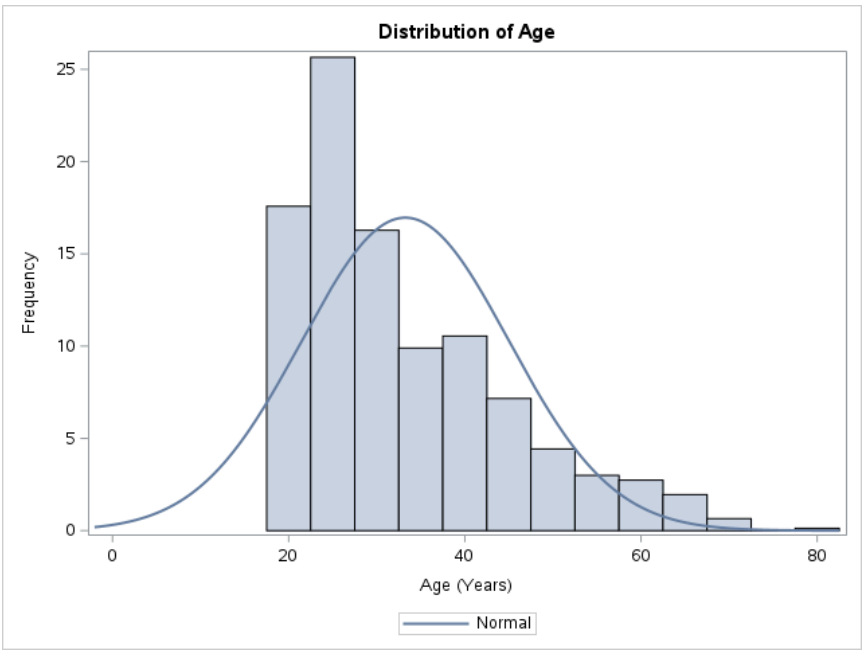
Descriptive Statistics for Key Variables

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Pregnancies	768	3.8450521	3.3695781	0	17.0000000
Glucose	768	121.6867628	30.4359489	44.0000000	199.0000000
BloodPressure	768	72.4051842	12.0963462	24.0000000	122.0000000
SkinThickness	768	29.1534196	8.7909419	7.0000000	99.0000000
Insulin	768	155.5482234	85.0211078	14.0000000	846.0000000
BMI	768	32.4574637	6.8751513	18.2000000	67.1000000
DiabetesPedigreeFunction	768	0.4718763	0.3313286	0.0780000	2.4200000
Age	768	33.2408854	11.7602315	21.0000000	81.0000000







Correlation Analysis of Key Variables

The CORR Procedure

8 Variables:	Pregnancies Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Pregnancies	768	3.84505	3.36958	2953	0	17.00000
Glucose	768	121.68676	30.43595	93455	44.00000	199.00000
BloodPressure	768	72.40518	12.09635	55607	24.00000	122.00000
SkinThickness	768	29.15342	8.79094	22390	7.00000	99.00000
Insulin	768	155.54822	85.02111	119461	14.00000	846.00000
BMI	768	32.45746	6.87515	24927	18.20000	67.10000
DiabetesPedigreeFunction	768	0.47188	0.33133	362.40100	0.07800	2.42000
Age	768	33.24089	11.76023	25529	21.00000	81.00000

Pearson Correlation Coefficients, N = 768
Prob > |r| under H0: Rho=0

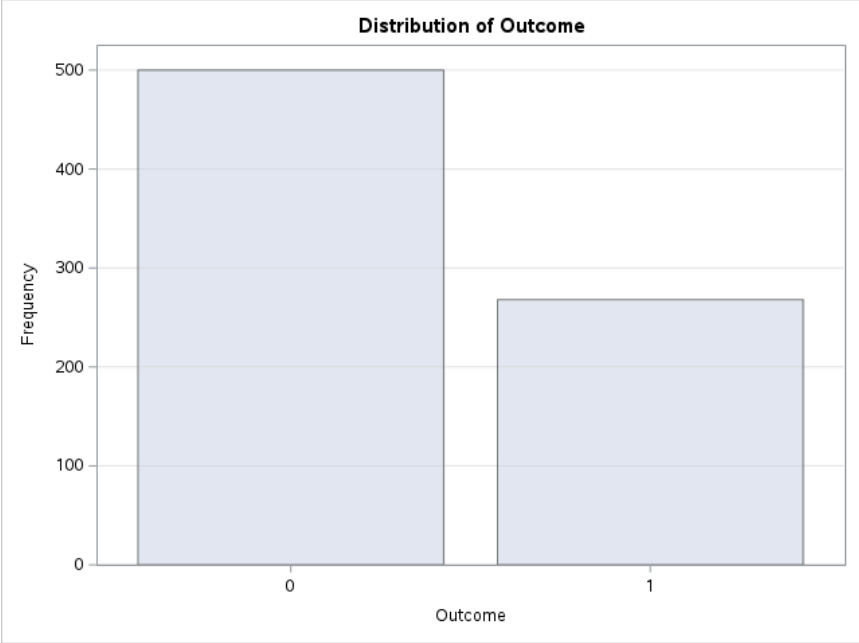
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
Pregnancies	1.00000	0.12791 0.0004	0.20852 <.0001	0.08299 0.0214	0.05603 0.1208	0.02157 0.5507	-0.03352 0.3535	0.54434 <.0001
Glucose	0.12791 0.0004	1.00000	0.21837 <.0001	0.19299 <.0001	0.42016 <.0001	0.23094 <.0001	0.13706 0.0001	0.26653 <.0001
BloodPressure	0.20852 <.0001	0.21837 <.0001	1.00000	0.19282 <.0001	0.07252 0.0445	0.28127 <.0001	-0.00276 0.9391	0.32459 <.0001
SkinThickness	0.08299 0.0214	0.19299 <.0001	0.19282 <.0001	1.00000	0.15814 <.0001	0.54240 <.0001	0.10097 0.0051	0.12787 0.0004
Insulin	0.05603 0.1208	0.42016 <.0001	0.07252 0.0445	0.15814 <.0001	1.00000	0.16659 <.0001	0.09863 0.0062	0.13673 0.0001
BMI	0.02157	0.23094	0.28127	0.54240	0.16659	1.00000	0.15340	0.02552

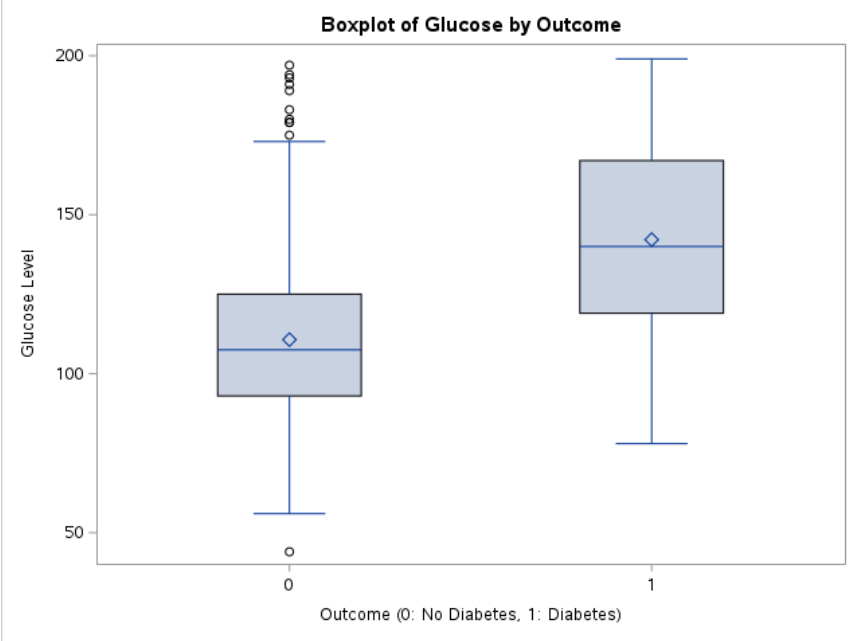
Pearson Correlation Coefficients, N = 768 Prob > |r| under H0: Rho=0								
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
	0.5507	<.0001	<.0001	<.0001	<.0001		<.0001	0.4801
DiabetesPedigreeFunction	-0.03352 0.3535	0.13706 0.0001	-0.00276 0.9391	0.10097 0.0051	0.09863 0.0062	0.15340 <.0001	1.00000	0.03356 0.3530
Age	0.54434 <.0001	0.26653 <.0001	0.32459 <.0001	0.12787 0.0004	0.13673 0.0001	0.02552 0.4801	0.03356 0.3530	1.00000

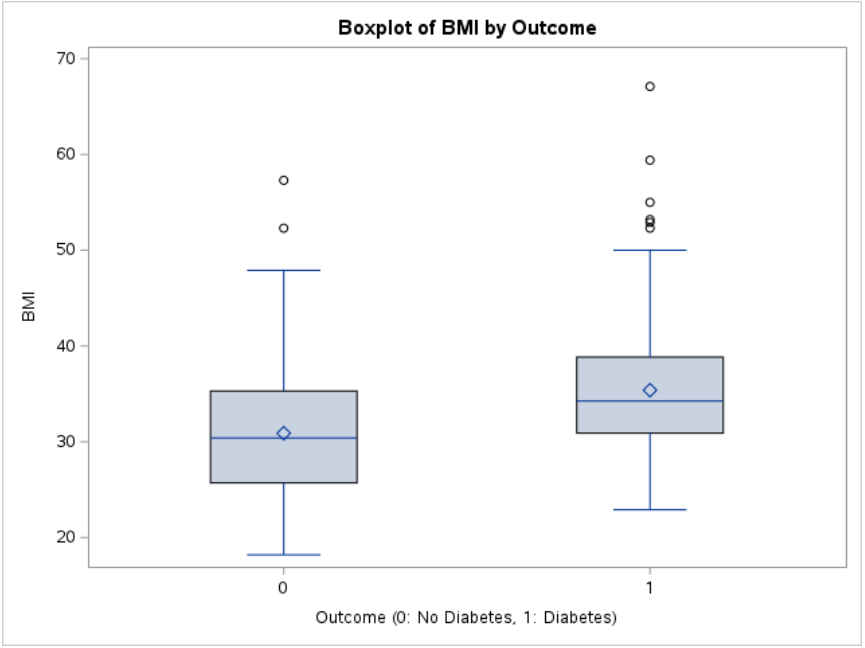
Distribution of Outcome (Diabetes vs. Non-Diabetes)

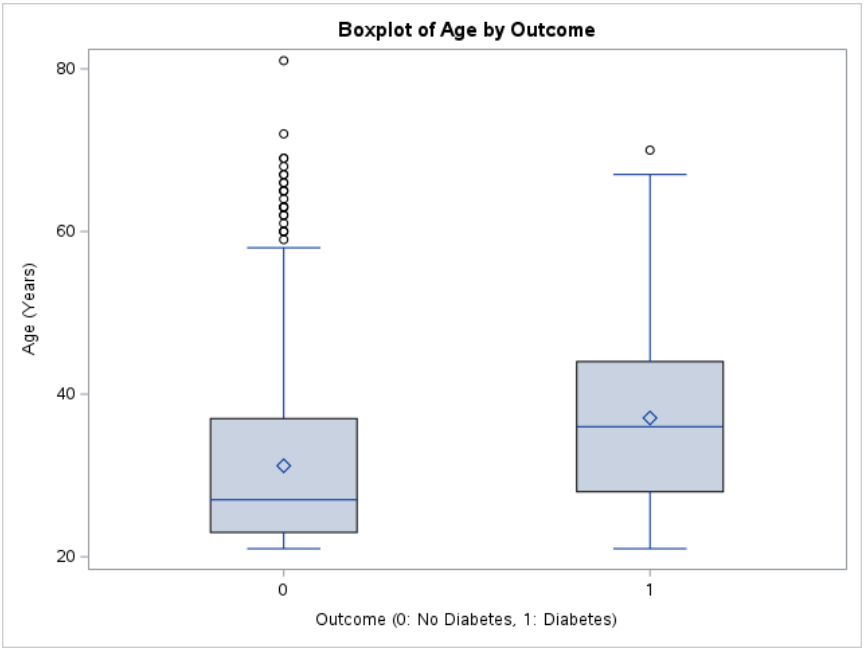
The FREQ Procedure

Outcome	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	500	65.10	500	65.10
1	268	34.90	768	100.00









Summary of Scaled Features (Z_Glucose, Z_BMI, Z_Age)

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Z_Glucose	768	1.16371E-16	1.0000000	-2.5524673	2.5401947
Z_BMI	768	3.420658E-15	1.0000000	-2.0737673	5.0388035
Z_Age	768	2.216109E-16	1.0000000	-1.0408711	4.0610692

Summary of Engineered Features (Interaction_Glucose_BMI and BMI_Category)

The MEANS Procedure

Analysis Variable : Interaction_Glucose_BMI				
N	Mean	Std Dev	Minimum	Maximum
768	0.2306405	0.9857566	-3.6290500	7.5082119

Summary of Engineered Features (Interaction_Glucose_BMI and BMI_Category)

The FREQ Procedure

BMI_Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Normal	102	13.28	102	13.28
Obese	483	62.89	585	76.17
Overweight	179	23.31	764	99.48
Underweight	4	0.52	768	100.00

Logistic Regression Model: Predicting Diabetes Outcome

The LOGISTIC Procedure

Model Information	
Data Set	WORK.DIABETES_ENGINEERED
Response Variable	Outcome
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	768
Number of Observations Used	768

Response Profile		
Ordered Value	Outcome	Total Frequency
1	1	268
2	0	500

Probability modeled is Outcome='1'.

Stepwise Selection Procedure

Class Level Information				
Class	Value	Design Variables		
BMI_Category	Normal	-1	-1	-1
	Obese	1	0	0
	Overweight	0	1	0
	Underweight	0	0	1

Step 0. Intercept entered:

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

-2 Log L	=	993.484

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
237.3265	7	<.0001

Step 1. Effect Z_Glucose entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	797.444
SC	1000.128	806.731
-2 Log L	993.484	793.444

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	200.0402	1	<.0001
Score	186.6069	1	<.0001
Wald	144.3272	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
63.0684	6	<.0001

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

Step 2. Effect Z_BMI entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	758.148
SC	1000.128	772.080
-2 Log L	993.484	752.148

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	241.3357	2	<.0001
Score	218.4398	2	<.0001
Wald	160.7062	2	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
27.0851	5	<.0001

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

Step 3. Effect Z_Age entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	745.683
SC	1000.128	764.258
-2 Log L	993.484	737.683

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	255.8010	3	<.0001
Score	229.2972	3	<.0001
Wald	167.5878	3	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
13.0131	4	0.0112

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

Step 4. Effect BMI_Category entered:

Model Convergence Status
Quasi-complete separation of data points detected.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	737.914
SC	1000.128	770.421
-2 Log L	993.484	723.914

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	269.5695	6	<.0001
Score	234.3783	6	<.0001
Wald	163.6561	6	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.0179	1	0.8936

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

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

Summary of Stepwise Selection							
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Z_Glucose		1	1	186.6069		<.0001
2	Z_BMI		1	2	40.4664		<.0001
3	Z_Age		1	3	14.7914		0.0001
4	BMI_Category		3	4	13.0100		0.0046

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Z_Glucose	1	102.2856	<.0001
Z_BMI	1	4.0634	0.0438
Z_Age	1	14.2635	0.0002
BMI_Category	3	12.2097	0.0067

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.8835	97.5504	0.0016	0.9682
Z_Glucose		1	1.0829	0.1071	102.2856	<.0001
Z_BMI		1	0.2809	0.1394	4.0634	0.0438
Z_Age		1	0.3514	0.0931	14.2635	0.0002
BMI_Category	Obese	1	3.3565	97.5505	0.0012	0.9726
BMI_Category	Overweight	1	2.7169	97.5504	0.0008	0.9778
BMI_Category	Underweight	1	-7.6757	292.7	0.0007	0.9791

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Z_Glucose	2.953	2.394	3.643
Z_BMI	1.324	1.008	1.740
Z_Age	1.421	1.184	1.705

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
BMI_Category Obese vs Normal	5.779	2.089	15.988
BMI_Category Overweight vs Normal	3.049	1.187	7.827
BMI_Category Underweight vs Normal	<0.001	<0.001	>999.999

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.4	Somers' D	0.668
Percent Discordant	16.6	Gamma	0.668
Percent Tied	0.0	Tau-a	0.304
Pairs	134000	c	0.834

Confusion Matrix for Logistic Regression Predictions

The FREQ Procedure

Frequency

Table of Outcome by Predicted_Class			
	Predicted_Class		
Outcome	0	1	Total
0	438	62	500
1	118	150	268
Total	556	212	768

Statistics for Table of Outcome by Predicted_Class

Statistic	DF	Value	Prob
Chi-Square	1	165.7423	<.0001
Likelihood Ratio Chi-Square	1	162.4576	<.0001
Continuity Adj. Chi-Square	1	163.5693	<.0001
Mantel-Haenszel Chi-Square	1	165.5265	<.0001
Phi Coefficient		0.4646	
Contingency Coefficient		0.4213	
Cramer's V		0.4646	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	438
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 768

TP	TN	FP	FN
150	438	62	118

Performance Metrics for Logistic Regression Model

Obs	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F1_Score
1	150	438	62	118	0.766	0.708	0.560	0.876	0.625

ROC Curve and AUC for Logistic Regression

The LOGISTIC Procedure

Model Information	
Data Set	WORK.DIABETES_ENGINEERED
Response Variable	Outcome
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	768
Number of Observations Used	768

Response Profile		
Ordered Value	Outcome	Total Frequency
1	0	500
2	1	268

Probability modeled is Outcome='0'.

Class Level Information				
Class	Value	Design Variables		
BMI_Category	Normal	-1	-1	-1
	Obese	1	0	0
	Overweight	0	1	0

Class Level Information				
Class	Value	Design Variables		
	Underweight	0	0	1

Model Convergence Status
Quasi-complete separation of data points detected.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	739.897
SC	1000.128	777.047
-2 Log L	993.484	723.897

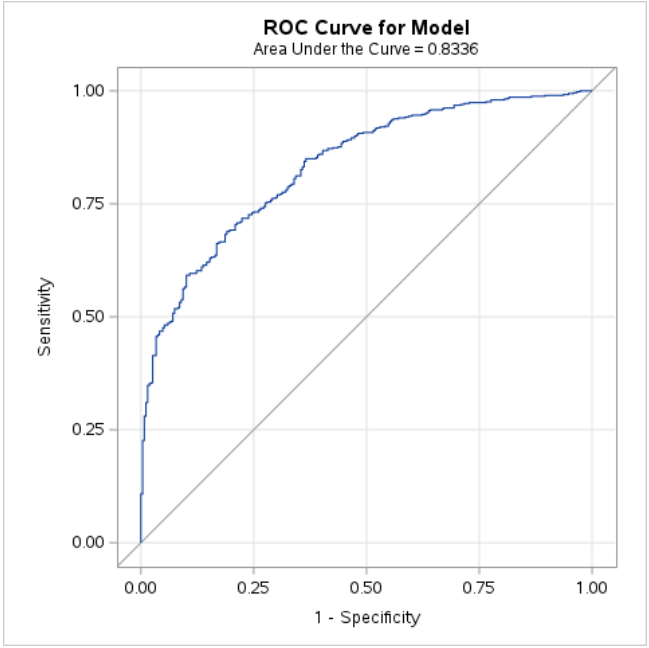
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	269.5874	7	<.0001
Score	237.3265	7	<.0001
Wald	163.3526	7	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Z_Glucose	1	95.5254	<.0001
Z_BMI	1	4.0329	0.0446
Z_Age	1	13.9333	0.0002
Interaction_Glucose_	1	0.0179	0.8936
BMI_Category	3	12.1903	0.0068

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	3.8762	97.6895	0.0016	0.9683
Z_Glucose		1	-1.0868	0.1112	95.5254	<.0001
Z_BMI		1	-0.2846	0.1417	4.0329	0.0446
Z_Age		1	-0.3499	0.0937	13.9333	0.0002
Interaction_Glucose_		1	0.0156	0.1164	0.0179	0.8936
BMI_Category	Obese	1	-3.3483	97.6897	0.0012	0.9727
BMI_Category	Overweight	1	-2.7095	97.6895	0.0008	0.9779
BMI_Category	Underweight	1	7.6503	293.1	0.0007	0.9792

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Z_Glucose	0.337	0.271	0.419
Z_BMI	0.752	0.570	0.993
Z_Age	0.705	0.586	0.847
Interaction_Glucose_	1.016	0.808	1.276
BMI_Category Obese vs Normal	0.173	0.062	0.479
BMI_Category Overweight vs Normal	0.327	0.127	0.843
BMI_Category Underweight vs Normal	>999.999	<0.001	>999.999

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.4	Somers' D	0.667
Percent Discordant	16.6	Gamma	0.667
Percent Tied	0.0	Tau-a	0.304
Pairs	134000	c	0.834



Refined Logistic Regression Model: Excluding Non-Significant Interaction Term

The LOGISTIC Procedure

Model Information	
Data Set	WORK.DIABETES_ENGINEERED
Response Variable	Outcome
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	768
Number of Observations Used	768

Response Profile		
Ordered Value	Outcome	Total Frequency
1	1	268
2	0	500

Probability modeled is Outcome='1'.

Stepwise Selection Procedure

Class Level Information				
Class	Value	Design Variables		
BMI_Category	Normal	-1	-1	-1
	Obese	1	0	0
	Overweight	0	1	0
	Underweight	0	0	1

Step 0. Intercept entered:

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

-2 Log L	=	993.484

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
234.3783	6	<.0001

Step 1. Effect Z_Glucose entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	797.444
SC	1000.128	806.731
-2 Log L	993.484	793.444

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	200.0402	1	<.0001
Score	186.6069	1	<.0001
Wald	144.3272	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
62.9865	5	<.0001

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

Step 2. Effect Z_BMI entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	758.148
SC	1000.128	772.080
-2 Log L	993.484	752.148

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	241.3357	2	<.0001
Score	218.4398	2	<.0001
Wald	160.7062	2	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
27.0788	4	<.0001

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

Step 3. Effect Z_Age entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	745.683
SC	1000.128	764.258
-2 Log L	993.484	737.683

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	255.8010	3	<.0001
Score	229.2972	3	<.0001
Wald	167.5878	3	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
13.0100	3	0.0046

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

Step 4. Effect BMI_Category entered:

Model Convergence Status
Quasi-complete separation of data points detected.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	995.484	737.914
SC	1000.128	770.421
-2 Log L	993.484	723.914

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	269.5695	6	<.0001
Score	234.3783	6	<.0001
Wald	163.6561	6	<.0001

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

Note: All effects have been entered into the model.

Summary of Stepwise Selection							
	Effect			Number	Score	Wald	
Step	Entered	Removed	DF	In	Chi-Square	Chi-Square	Pr > ChiSq
1	Z_Glucose		1	1	186.6069		<.0001
2	Z_BMI		1	2	40.4664		<.0001
3	Z_Age		1	3	14.7914		0.0001
4	BMI_Category		3	4	13.0100		0.0046

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Z_Glucose	1	102.2856	<.0001
Z_BMI	1	4.0634	0.0438
Z_Age	1	14.2635	0.0002
BMI_Category	3	12.2097	0.0067

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.8835	97.5504	0.0016	0.9682
Z_Glucose		1	1.0829	0.1071	102.2856	<.0001
Z_BMI		1	0.2809	0.1394	4.0634	0.0438
Z_Age		1	0.3514	0.0931	14.2635	0.0002
BMI_Category	Obese	1	3.3565	97.5505	0.0012	0.9726
BMI_Category	Overweight	1	2.7169	97.5504	0.0008	0.9778
BMI_Category	Underweight	1	-7.6757	292.7	0.0007	0.9791

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Z_Glucose	2.953	2.394	3.643
Z_BMI	1.324	1.008	1.740
Z_Age	1.421	1.184	1.705
BMI_Category Obese vs Normal	5.779	2.089	15.988
BMI_Category Overweight vs Normal	3.049	1.187	7.827
BMI_Category Underweight vs Normal	<0.001	<0.001	>999.999

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.4	Somers' D	0.668
Percent Discordant	16.6	Gamma	0.668
Percent Tied	0.0	Tau-a	0.304
Pairs	134000	c	0.834

Confusion Matrix for Refined Logistic Regression Predictions

The FREQ Procedure

Frequency

Table of Outcome by Predicted_Class			
	Predicted_Class		
Outcome	0	1	Total
0	438	62	500
1	118	150	268
Total	556	212	768

Statistics for Table of Outcome by Predicted_Class

Statistic	DF	Value	Prob
Chi-Square	1	165.7423	<.0001
Likelihood Ratio Chi-Square	1	162.4576	<.0001
Continuity Adj. Chi-Square	1	163.5693	<.0001
Mantel-Haenszel Chi-Square	1	165.5265	<.0001
Phi Coefficient		0.4646	
Contingency Coefficient		0.4213	
Cramer's V		0.4646	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	438
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 768

TP	TN	FP	FN
150	438	62	118

Structure of the Imported Test Dataset

The CONTENTS Procedure

Data Set Name	WORK.NEW_DATA	Observations	20
Member Type	DATA	Variables	9
Engine	V9	Indexes	0
Created	12/23/2024 14:28:02	Observation Length	72
Last Modified	12/23/2024 14:28:02	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO

Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1816
Obs in First Data Page	20
Number of Data Set Repairs	0
Filename	/saswork/SAS_work90F300008524_odaws02-usw2-2.oda.sas.com/SAS_workEC7200008524_odaws02-usw2-2.oda.sas.com/new_data.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2281702013
Access Permission	rw-r--r--
Owner Name	u64112808
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes					
#	**Variable**	**Type**	**Len**	**Format**	**Informat**
8	Age	Num	8	BEST12.	BEST32.
6	BMI	Num	8	BEST12.	BEST32.
3	BloodPressure	Num	8	BEST12.	BEST32.
7	DiabetesPedigreeFunction	Num	8	BEST12.	BEST32.
2	Glucose	Num	8	BEST12.	BEST32.
5	Insulin	Num	8	BEST12.	BEST32.
9	Outcome	Num	8	BEST12.	BEST32.
1	Pregnancies	Num	8	BEST12.	BEST32.
4	SkinThickness	Num	8	BEST12.	BEST32.

Processed Test Dataset

Obs	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome	Z_Glucose	Z_BMI	Z_Age	BMI_Category
1	6	101	91	35	33	26.6873934	2.4321090289	71	0	-0.67970	-0.83904	3.21088	Overweight
2	3	132	109	11	114	24.456739844	2.1373931782	42	0	0.33870	-1.16326	0.74490	Normal
3	12	81	64	29	67	18.3097561	1.8321508508	48	0	-1.33673	-2.05672	1.25510	Underweight
4	14	167	111	37	34	22.37453289	0.6663638074	22	0	1.48850	-1.46591	-0.95578	Normal
5	10	187	111	16	59	33.64952296	0.7145639746	62	1	2.14553	0.17290	2.44558	Obese
6	7	109	96	17	101	35.383861892	0.1970406149	65	1	-0.41689	0.42498	2.70068	Obese
7	12	117	111	44	130	31.331119445	1.8055909352	77	0	-0.15407	-0.16408	3.72109	Obese
8	4	81	100	23	72	38.378619327	0.3661379699	73	0	-1.33673	0.86026	3.38095	Obese
9	6	143	104	26	43	32.323694561	1.1544076045	26	0	0.70007	-0.01981	-0.61565	Obese
10	9	139	113	45	173	38.129112862	0.5841260856	48	0	0.56866	0.82400	1.25510	Obese

Contents of Scored Data

The CONTENTS Procedure

Data Set Name	WORK.NEW_DATA_PREDICTIONS	**Observations**	20
Member Type	DATA	**Variables**	17
Engine	V9	**Indexes**	0
Created	12/23/2024 14:28:02	**Observation Length**	152
Last Modified	12/23/2024 14:28:02	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label	Posterior Probabilities for DATA=WORK.NEW_DATA_PROCESSED.		
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	861
Obs in First Data Page	20
Number of Data Set Repairs	0
Filename	/saswork/SAS_work90F300008524_odaws02-usw2-2.oda.sas.com/SAS_workEC7200008524_odaws02-usw2-2.oda.sas.com/new_data_predictions.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2281702014
Access Permission	rw-r--r--
Owner Name	u64112808
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes						
#	**Variable**	**Type**	**Len**	**Format**	**Informat**	**Label**

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
8	Age	Num	8	BEST12.	BEST32.	
6	BMI	Num	8	BEST12.	BEST32.	
13	BMI_Category	Char	11			
3	BloodPressure	Num	8	BEST12.	BEST32.	
7	DiabetesPedigreeFunction	Num	8	BEST12.	BEST32.	
14	F_Outcome	Char	12			From: Outcome
2	Glucose	Num	8	BEST12.	BEST32.	
15	I_Outcome	Char	12			Into: Outcome
5	Insulin	Num	8	BEST12.	BEST32.	
9	Outcome	Num	8	BEST12.	BEST32.	
17	P_0	Num	8			Predicted Probability: Outcome=0
16	P_1	Num	8			Predicted Probability: Outcome=1
1	Pregnancies	Num	8	BEST12.	BEST32.	
4	SkinThickness	Num	8	BEST12.	BEST32.	
12	Z_Age	Num	8			
11	Z_BMI	Num	8			
10	Z_Glucose	Num	8			

Predictions for Processed Test Dataset

Obs	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome	Predicted_Prob	Predicted_Class
1	6	101	91	35	33	26.6873934	2.4321090289	71	0	0.26699	0
2	3	132	109	11	114	24.456739844	2.1373931782	42	0	0.12137	0
3	12	81	64	29	67	18.3097561	1.8321508508	48	0	0.00000	0
4	14	167	111	37	34	22.37453289	0.6663638074	22	0	0.19510	0
5	10	187	111	16	59	33.64952296	0.7145639746	62	1	0.93728	1
6	7	109	96	17	101	35.383861892	0.1970406149	65	1	0.52249	1
7	12	117	111	44	130	31.331119445	1.8055909352	77	0	0.63824	1
8	4	81	100	23	72	38.378619327	0.3661379699	73	0	0.36711	0
9	6	143	104	26	43	32.323694561	1.1544076045	26	0	0.50227	1
10	9	139	113	45	173	38.129112862	0.5841260856	48	0	0.68164	1