

FALSEPOS: A Macro for Sensitivity, Specificity, Positive and Negative Predictive Values and Receiver Operating Characteristic (ROC) Curves.

Alfred J. Anderson, Aurora Health Care

The decision that a patient has a certain disease often depends upon a test or series of tests. The utility of a diagnostic test depends upon the tests ability to detect patients with disease (sensitivity) and exclude patients without disease (specificity). Further, when a patient tests positive, what is the probability that the disease is present (positive predictive value) and when the test is negative, what is the probability that the disease is not present (negative predictive value).

		DISEASE			
		Present	Absent		
TEST	Positive	a True Pos	b False Pos	Positive Predictive Value	$\frac{a}{a + b}$
	Negative	c False Neg	d True Neg	Negative Predictive Value	$\frac{d}{c + d}$
		$\frac{a}{a + c}$	$\frac{d}{b + d}$		
		Sensitivity	Specificity		

When the diagnostic test is a continuous variable, a decision needs to be made concerning the optimal cutoff point to achieve the desired sensitivity and specificity. By changing the cutoff point for a variable measured in patients known to have the disease and in patients without the disease, one can determine the best cutoff point to use for future diagnoses. As the cutoff point changes over a range of the data, sensitivity and specificity would alternately increase and decrease with similar changes in the positive and negative predictive values. Plots of sensitivity versus specificity and the positive versus the negative predictive values over a range of the diagnostic variable are useful in selecting the best cutoff point. A ROC curve, which is a plot of sensitivity versus the false positive fraction (one minus specificity), over a range of cutoff points provides another visual aid in the selection of the best cutoff point. The best cutoff point often depends on the costs of not missing the disease (both monetary and medical) and not on a point selected by a statistical procedure. The use of an appropriate statistical procedure, though, can aid in the decision. The Logistic procedure (SAS/STAT*) is included in order to calculate the probability of the disease or event over the range of the diagnostic variable and to assess statistical significance.

The macro FALSEPOS requires a variable that is included in a SAS* Data Set called 'TEST' which is equal to 1 for disease present (or event) and 2 for disease absent (no event). Parameters that are in the macro call are the SAS Data Set name, the name of the continuous variable, the range of the variable for the analysis in the form of a 'from' 'to' 'by' and if the disease state is associated with high levels or low levels of the continuous variable.

```
%MACRO FALSEPOS(DATA=,FROMV=,TOV=,BYV=,LOHI=,VARBL=,);
```

```
DATA=    Name of the SAS data set.
FROMV=   Lowest value for the analysis.
TOV=     Highest value for the analysis.
BYV=     Increment or by variable.
LOHI=    Event associated with lower=0 or higher=1 levels.
VARBL=   Name of the continuous variable.
```

```
DATA MMMM;
```

```
SET AAAA;
```

```
--data step statements--
IF GROUP ='P' THEN TEST=1;    /* event or disease */
IF GROUP ='C' THEN TEST=2;    /* no event or no disease */
IF TEST =. THEN DELETE;
IF CONTVA=. OR CONTVB=. THEN DELETE;
KEEP TEST CONTVA CONTVB;
RUN;
```

```
%FALSEPOS(DATA=MMMM, FROMV=120, TOV=160, BYV=2, LOHI=1, VARBL=CONTVA);
%FALSEPOS(DATA=MMMM, FROMV=12, TOV=50, BYV=2, LOHI=0, VARBL=CONTVB);
```

The macro initiates arrays, up to size 24 for both levels of 'TEST', that will contain the counts of the number of patients that are above and below the cutoff points. The cutoff points are calculated using the FROMV, TOV and BYV macro variables. The difference between FROMV minus TOV must be evenly divisible by BYV. These counts are used to calculate sensitivity, specificity and the positive and negative predictive values. Footnotes for the cutoff points are derived by concatenating a letter code, with an equal sign and with the cutoff points. These codes are combined to form three footnotes. Plots are included for sensitivity and specificity versus VARBL, positive and negative predictive values versus VARBL and the ROC curve with the footnotes. Proc Logistic, with the Output option, creates a data set containing predicted probabilities and upper and lower 95% confidence limits. Duplicates and the extreme probabilities are deleted to reduce the length of a print of this data and to limit the scaling of a plot of the probability of an event versus the continuous variable.

The macro was run in SAS version 6.06 under MVS.

```

XMACRO FALSEPOS(DATA=,FROMV=,TOV=,BYV=,LOHI=,VARBL=);
DATA INAA;

```

```

SET ADATA END=LASTREC;
IF VARBL= . THEN DELETE;
XLET N= ZEVAL(1-((STOV-SFROMV)/SBYV));
ARRAY PPP(8H) PPP1-PPPH8;
ARRAY PPN(8H) PPN1-PPNH8;
ARRAY NNN(8H) NNN1-NNNH8;
ARRAY XXX(8H) XXX1-XXXH8;
ARRAY FOT(24) FOT1-FOT24;
ARRAY ALP(24) ALP1-ALP24;

```

```

IF N=1 THEN DO;
DO I=1 TO 8H;
PPN(I)=0;
PPP(I)=0;
NNN(I)=0;
NNP(I)=0;
XXX(I)=0;
END;
NUNH=0; NUNP=0; END;
RETAIN PPN1-PPNH8 PPP1-PPPH8 NNN1-NNNH8 NNP1-NNPNH
XXX1-XXXH8 NUNH NUNP;
IF TEST=2 THEN NUNH=NUNH+1;
ELSE IF TEST=1 THEN NUNP=NUNP+1;
I=0;
DO K= SFROMV TO STOV BY SBYV;
I=I+1;
XXX(I)=K; END;

```

```

/*FOR RISK IN LOWER TAIL */
IF SLOHI='0' THEN DO;

```

```

DO I=1 TO 8H;
IF TEST=2 THEN DO;
IF SVARBL<XXX(I) THEN NNN(I)=NNN(I)+1;
IF SVARBL<XXX(I) THEN PPN(I)=PPN(I)+1; END;
ELSE IF TEST=1 THEN DO;
IF SVARBL<XXX(I) THEN NNP(I)=NNP(I)+1;
IF SVARBL<XXX(I) THEN PPP(I)=PPP(I)+1; END;
END;

```

```

/*FOR RISK IN UPPER TAIL */
IF SLOHI='1' THEN DO;

```

```

DO I=1 TO 8H;
IF TEST=2 THEN DO;
IF SVARBL<XXX(I) THEN NNN(I)=NNN(I)+1;
IF SVARBL<XXX(I) THEN PPN(I)=PPN(I)+1; END;
ELSE IF TEST=1 THEN DO;
IF SVARBL<XXX(I) THEN NNP(I)=NNP(I)+1;
IF SVARBL<XXX(I) THEN PPP(I)=PPP(I)+1; END;
END;

```

```

IF LASTREC THEN DO; /* CREATE FOOTNOTES */

```

```

DO I=1 TO 8H;
FOT(I)=ALP(I) || '-' || INPUT(XXX(I),8.);
END;
AFTNA=FOT1 || FOT2 || FOT3 || FOT4 || FOT5 ||
AFTNB=FOT9 || FOT10 || FOT11 || FOT12 || FOT13 ||
AFTNC=FOT17 || FOT18 || FOT19 || FOT20 || FOT21 ||
AFTND=FOT22 || FOT23 || FOT24;

```

```

CALL SYMPUT('FTNA',PUT(AFTNA,856.));
CALL SYMPUT('FTNB',PUT(AFTNB,856.));
CALL SYMPUT('FTNC',PUT(AFTNC,856.)); END;
DROP I K FOT1-FOT24;

```

```

IF LASTREC THEN OUTPUT; /* CREATE DATA SET FOR PLOTS */

```

```

DATA AAAA; /* CREATE DATA SET FOR PLOTS */
SET INAA;
ARRAY PPN(8H) PPN1-PPNH8;
ARRAY PPP(8H) PPP1-PPPH8;
ARRAY NNN(8H) NNN1-NNNH8;
ARRAY NNP(8H) NNP1-NNPNH;
ARRAY XXX(8H) XXX1-XXXH8;
ARRAY ALP(8H) ALP1-ALPH8;
DROP TEST I PPP1-PPPH8 PPN1-PPNH8 NNN1-NNNH8 NNP1-NNPNH
XXX1-XXXH8 ALP1-ALPH8 AFTNA AFTNB AFTNC;
FORMAT SENSIT FNEGFR SPECIF FPOSFR POSPRDV NEGPRDV F6.2;

```

```

DO I=1 TO 8H;
PPNO=PPN(I);
PPPO=PPP(I);
NNNO=NNN(I);
NNPO=NNP(I);
SVARBL=XXX(I);
ALPO=ALP(I);
SENSIT=PPPO/NUNP; /* SENSITIVITY */
FNEGFR=1-SENSIT; /* FALSE NEGATIVE FRACTION */
SPECIF=NNNO/NNPN; /* SPECIFICITY */
FPOSFR=1-SPECIF; /* FALSE POSITIVE FRACTION */
POSPRDV=PPPO/(PPPO+PPNO); /* POSITIVE PREDICTIVE VALUE */
NEGPRDV=NNNO/(NNNO+NNPO); /* NEGATIVE PREDICTIVE VALUE */
OUTPUT;

```

```

PROC PRINT DATA=AAAA;
VAR SVARBL NUNH NUNP PPNO PPPO NNNO NNPO ALPO
SENSIT FNEGFR SPECIF FPOSFR POSPRDV NEGPRDV;
TITLE1 'LISTING FOR SVARBL';
TITLE2 '-';
FOOTNOTE1 '-';
FOOTNOTE2 '-';
FOOTNOTE3 '-';

```

```

PROC PLOT;
PLOT SENSIT=SVARBL '*' SPECIF=SVARBL '*' /OVERLAY
VAXIS=.1 TO 1.0 BY .1;
TITLE1 'POSITIVE-SENSITIVITY NEGATIVE-SPECIFICITY';
TITLE2 'PLOT FOR SVARBL';

```

```

PROC PLOT;
PLOT POSPRDV=SVARBL '*' NEGPRDV=SVARBL '*' /OVERLAY
VAXIS=.2 TO 1.0 BY .1;
TITLE1 'POSITIVE AND NEGATIVE PREDICTIVE VALUES';
TITLE2 'PLOT FOR SVARBL';

```

```

PROC PLOT;
PLOT SENSIT=FPOSFR=ALPO /
VAXIS=0 TO 1.0 BY .1 HAXIS=0 TO 1.0 BY .1;
TITLE1 'SENSITIVITY VERSUS FALSE POSITIVE FRACTION';
TITLE2 'ROC CURVE FOR SVARBL';
FOOTNOTE1 'AFTNA';
FOOTNOTE2 'AFTNB';
FOOTNOTE3 'AFTNC';

```

```

PROC LOGISTIC DATA=ADATA;
MODEL TEST=SVARBL / CTABLE;
OUTPUT OUT=PRED P=PHAT LOWER=LCL UPPER=UCL;
TITLE1 'LOGISTIC MODEL FOR SVARBL';
FOOTNOTE1 '-';
FOOTNOTE2 '-';
FOOTNOTE3 '-';

```

```

DATA PRED;
SET PRED; /* ROUND AND TRIM ENDPOINTS FOR PRINTS AND PLOTS */
PHAT=ROUND(PHAT,.01);
LCL=ROUND(LCL,.01);
UCL=ROUND(UCL,.01);
IF PHAT>.97 THEN DELETE;
IF PHAT<.03 THEN DELETE;
DROP LEVEL;

```

```

PROC SORT DATA=PRED; BY PHAT;

```

```

DATA SETONE; /* ELIMINATE DUPLICATE POINTS */
SET PRED; BY PHAT;
IF FIRST.PHAT;

```

```

PROC PRINT DATA=SETONE;
PLOT DATA=SETONE;
PLOT PHAT=SVARBL='*' UCL=SVARBL='*' LCL=SVARBL='*' /OVERLAY;
TITLE1 'LOGISTIC MODEL FOR SVARBL';
TITLE2 'PROBABILITY OF TRUE POSITIVE';

```

```

XMACRO FALSEPOS;

```

LISTING FOR CONTVA

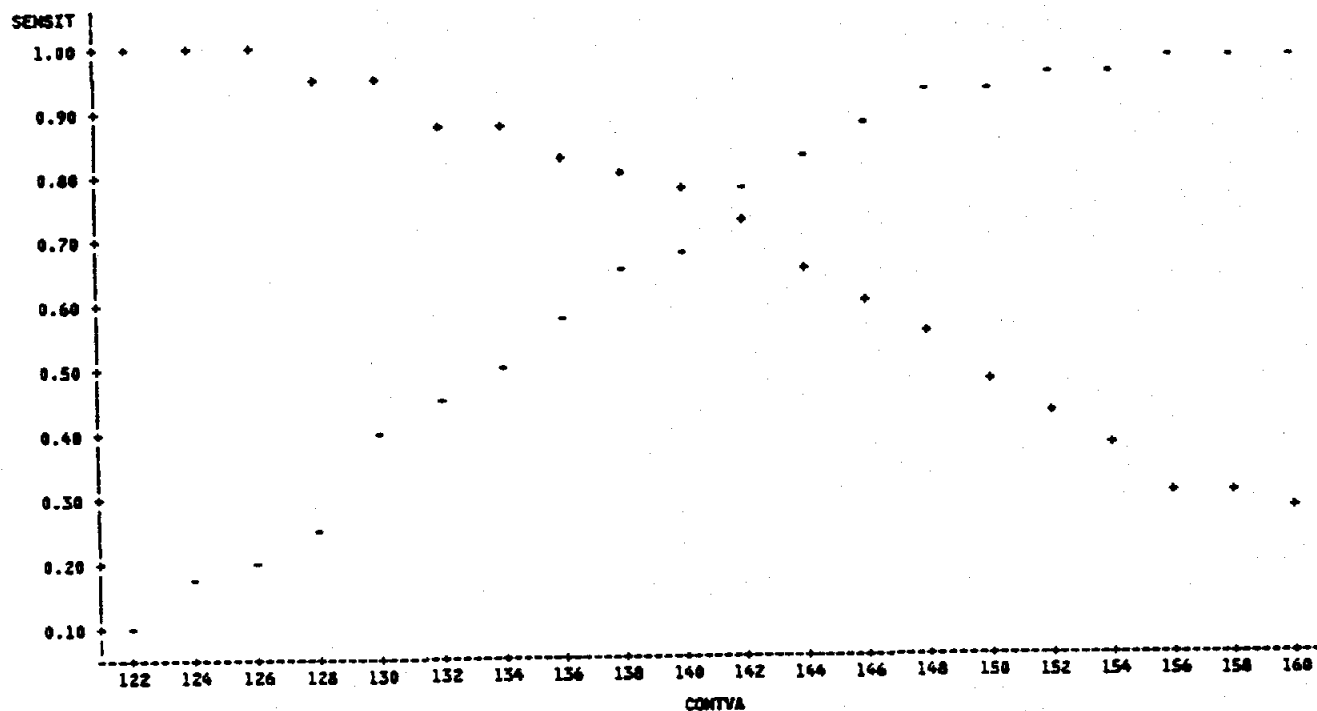
1

OBS	CONTVA	NUNH	NUNP	PPNO	PPPO	NNNO	NNPO	ALPO	SENSIT	FNEGFR	SPECIF	FPOSFR	POSPRDV	NEGPRDV
1	122	65	88	58	87	7	1	A	0.99	0.01	0.11	0.89	0.68	0.88
2	124	65	88	53	87	12	1	B	0.99	0.01	0.18	0.82	0.62	0.92
3	126	65	88	52	87	13	1	C	0.99	0.01	0.20	0.80	0.63	0.93
4	128	65	88	49	84	16	4	D	0.95	0.05	0.25	0.75	0.63	0.88
5	130	65	88	39	83	26	5	E	0.94	0.06	0.40	0.60	0.68	0.84
6	132	65	88	34	78	29	10	F	0.89	0.11	0.45	0.55	0.68	0.74
7	134	65	88	33	77	32	11	G	0.88	0.13	0.49	0.51	0.78	0.74
8	136	65	88	28	72	37	16	H	0.82	0.18	0.57	0.43	0.72	0.70
9	138	65	88	22	70	43	18	I	0.80	0.20	0.66	0.34	0.76	0.70
10	140	65	88	21	68	44	20	J	0.77	0.23	0.68	0.32	0.76	0.69
11	142	65	88	14	63	51	25	K	0.72	0.28	0.78	0.22	0.82	0.67
12	144	65	88	11	58	54	30	L	0.66	0.34	0.83	0.17	0.84	0.64
13	146	65	88	8	53	57	35	M	0.60	0.40	0.88	0.12	0.87	0.62
14	148	65	88	5	49	60	39	N	0.56	0.44	0.92	0.08	0.91	0.61
15	150	65	88	3	42	60	46	O	0.48	0.52	0.92	0.08	0.89	0.57
16	152	65	88	3	38	62	50	P	0.43	0.57	0.95	0.05	0.93	0.55
17	154	65	88	3	32	62	56	Q	0.36	0.64	0.95	0.05	0.91	0.53
18	156	65	88	1	27	64	61	R	0.31	0.69	0.98	0.02	0.96	0.51
19	158	65	88	1	27	64	61	S	0.31	0.69	0.98	0.02	0.96	0.51
20	160	65	88	1	24	64	64	T	0.27	0.73	0.98	0.02	0.96	0.50

POSITIVE-SENSITIVITY NEGATIVE-SPECIFICITY
PLOT FOR CONTVA

2

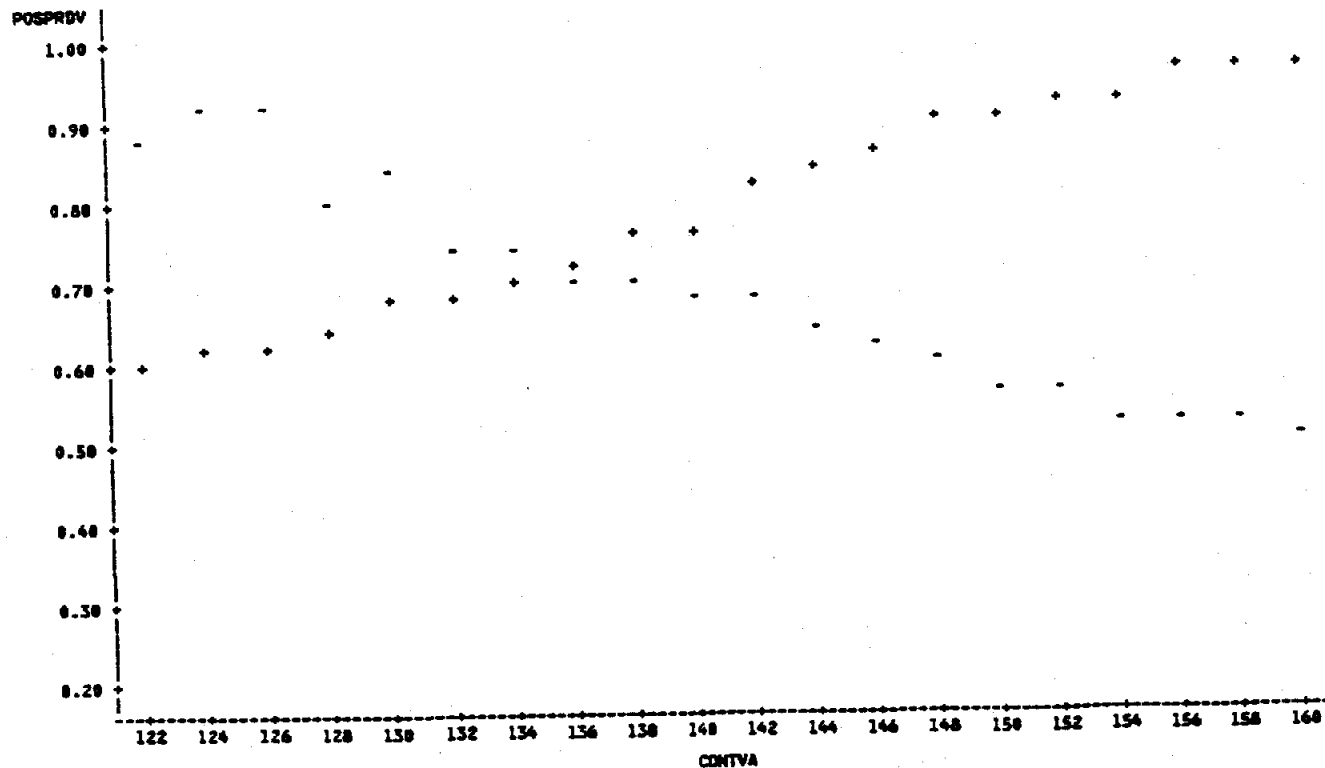
Plot of SENSIT=CONTVA. Symbol used is '+'.
Plot of SPECIF=CONTVA. Symbol used is '-'.



POSITIVE AND NEGATIVE PREDICTIVE VALUES
PLOT FOR CONTVA

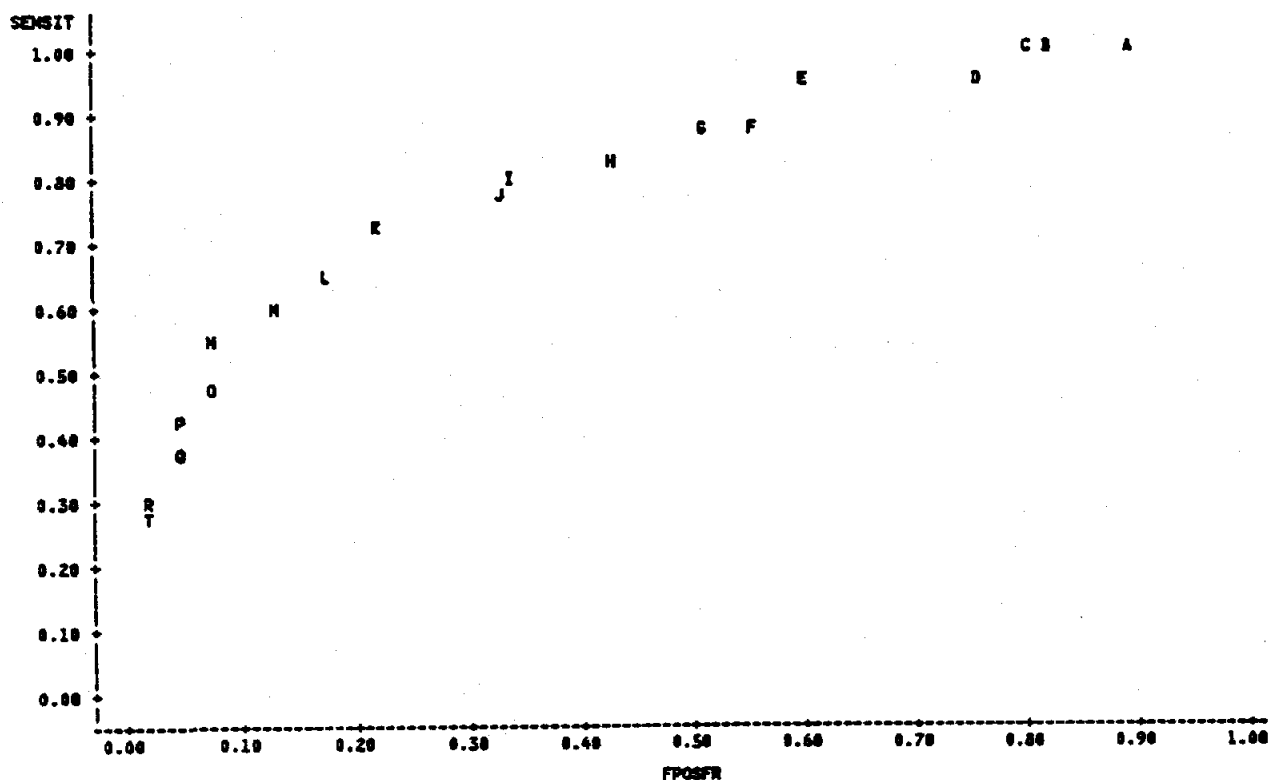
3

Plot of POSPRDV=CONTVA. Symbol used is '+'.
Plot of NEGPRDV=CONTVA. Symbol used is '-'.



SENSITIVITY VERSUS FALSE POSITIVE FRACTION ROC CURVE FOR CONTVA

Plot of SENSIT=FPOSFR. Symbol is value of ALPO.



NOTE: 1 obs hidden.

A=122 B=124 C=126 D=128 E=130 F=132 G=134 H=136
I=138 J=140 K=142 L=144 M=146 N=148 O=150 P=152
Q=154 R=156 S=158 T=160

LOGISTIC MODEL FOR CONTVA

The LOGISTIC Procedure

Data Set: WORK.MWMM
Response Variable: TEST
Response Levels: 2
Number of Observations: 153
Link Function: Logit

Response Profile

Ordered Value	TEST	Count
1	1	68
2	2	65

Simple Statistics for Explanatory Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
CONTVA	144.183667	17.793937	109.000	221.000

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	210.632	161.302	.
SC	213.663	167.342	.
-2 LOG L	208.632	157.302	51.331 with 1 DF (p=0.0001)
Score	.	.	37.839 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate
INTERCPT	-14.1693	2.6283	29.2483	0.0001	
CONTVA	0.1024	0.0167	29.9353	0.0001	1.004487

LOGISTIC MODEL FOR CONTVA

The LOGISTIC Procedure

Classification Table

Observed		Predicted		Total
		EVENT	NO EVENT	
Observed	EVENT	69	19	88
	NO EVENT	21	44	65
Total		90	63	153

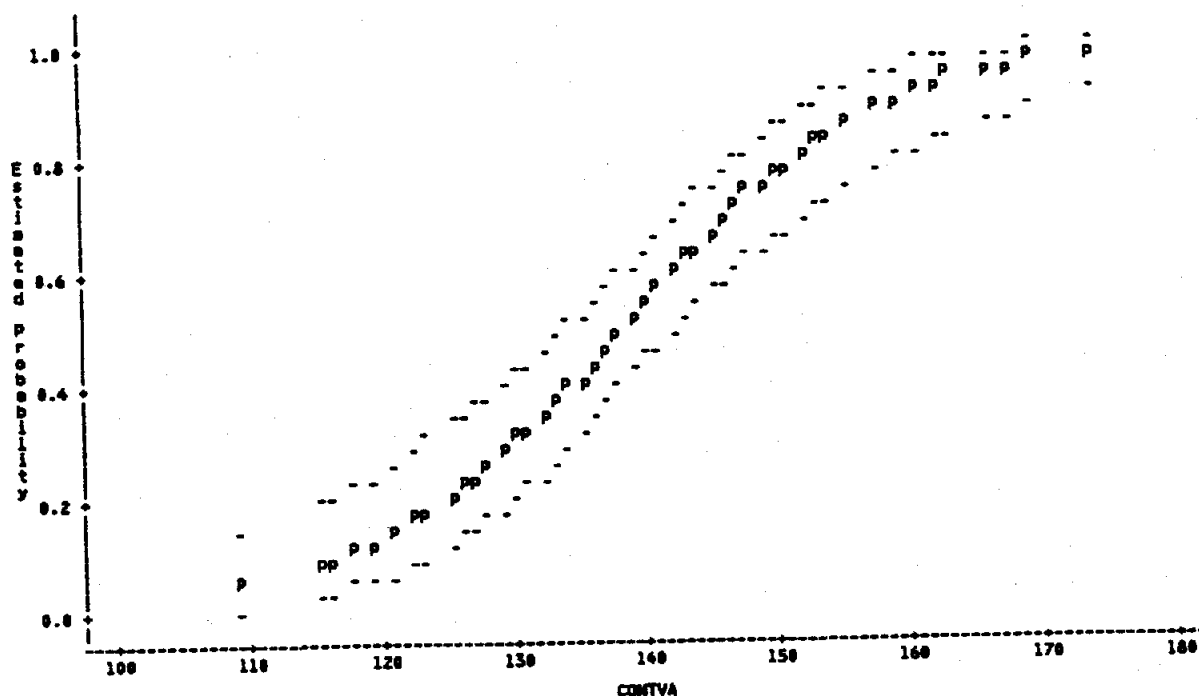
Sensitivity= 78.42 Specificity= 67.72 Correct= 73.92
False Positive Rate= 23.32 False Negative Rate= 30.22

NOTE: An EVENT is an outcome whose ordered response value is 1.

LOGISTIC MODEL FOR CONTVA

PROBABILITY OF TRUE POSITIVE

Plot of PMATH=CONTVA. Symbol used is 'p'.
Plot of UCLM=CONTVA. Symbol used is '-.-'.
Plot of LCLM=CONTVA. Symbol used is '-.-'.



* SAS and SAS/STAT are registered trademarks of SAS Institute, Inc., Cary, NC, USA.

Alfred J. Anderson, M.S.
Aurora Health Care
3031 West Montana
Milwaukee, WI 53215

(414) 647-3547
(414) 647-7354 FAX