BIOS 625 HW #5 Solutions Sheet

[redacted]

Problem 1. Agresti 5.19.

$R = 1 : logit(\hat{\pi}) = 6.7 + .1A + 1.4S$. $R = 0 : logit(\hat{\pi}) = 7.0 + .1A + 1.2S$
The YS conditional odds ratio is $exp(1.4) = 4.1$ for blacks and $exp(1.2) = 3.3$ for whites. Note that .2, the coeff. of the cross-product term, is the difference between the log odds ratios 1.4 and 1.2. The coeff. of S of 1.2 is the log odds ratio between Y and S when R = 0 (whites), in which case the RS interaction does not enter the equation. The P-value of $P < .01$ for smoking represents the result of the test that the log odds ratio between Y and S for whites is 0.

Problem 2. Agresti 5.20

Part a. The estimated log odds ratio between race and driving after consuming a substantial amount of alcohol was $-.72$ in Grade 12 (i.e., for each gender, the estimated odds for blacks of driving after consuming a substantial amount of alcohol were $e^{-.72} = .49$ times the estimated odds for whites. The corresponding estimated log odds ratio was $-.72 + .74 = .02$ for Grade 9, $-.72 + .38 = .34$ for Grade 10, and $-.72 + .01 = -.71$ for Grade 11. i.e. there is essentially no association in Grade 9, but the association changes to an odds ratio of about .5 in Grades 11 and 12.

Problem 3. Agresti 5.24

Are people with more social ties less likely to get colds? Use logistic models to analyze the 2x2x2x2 contengency table on pp. 1943 of article by S. Cohen et al., J.Am.Med.Assoc. **277** (24).

See next several pages of SAS output:

HW5 Problem 3
Residuals vs predicted eta_i with LOESS Overlay

The LOESS Procedure
Selected Smoothing Parameter: 1
Dependent Variable: res

Stepwise Model Selection

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	109
2	Nonevent	167

Stepwise Selection Procedure

Class Level Information		
Class	**Value**	**Design Variables**
titer	**f<=2f**	1
	f>=4f	0
virus	**fHanks**	1
	fRV39f	0
social	**f1-5f**	1
	f>=6f	0

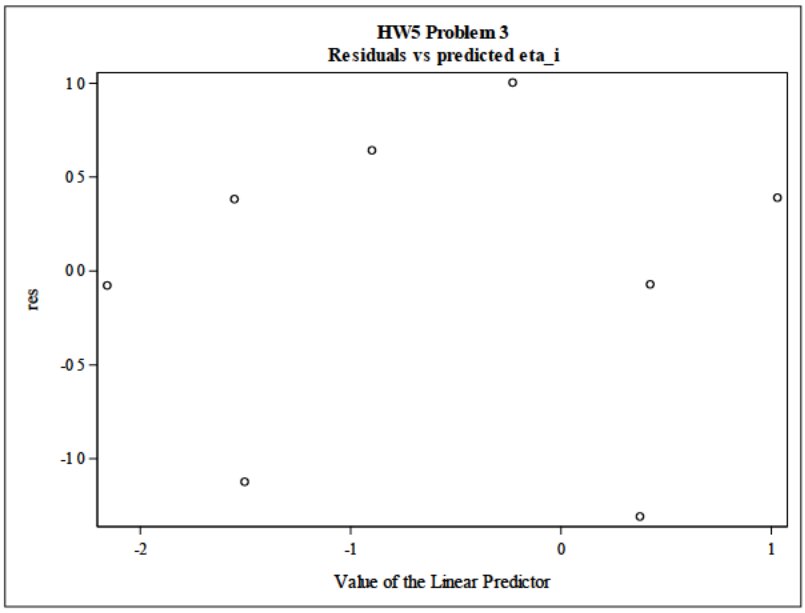
Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.5530	0.2880	29.0752	<.0001
titer	**f<=2f**	1	1.9280	0.2923	43.5231	<.0001
virus	**fHanks**	1	-0.6051	0.2805	4.6533	0.0310
social	**f1-5f**	1	0.6538	0.2782	5.5228	0.0188

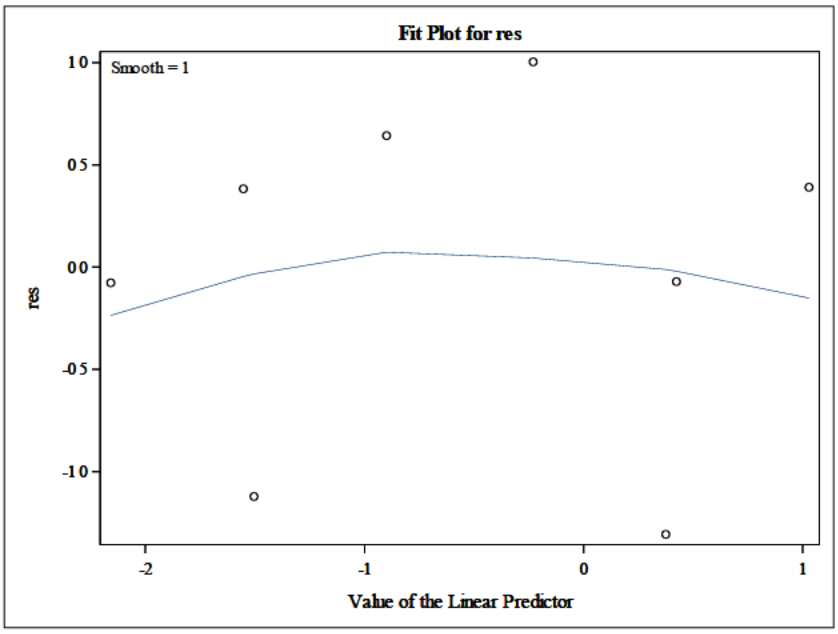
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
titer f<=2f vs f>=4f	6.876	3.878	12.193
virus fHanks vs fRV39f	0.546	0.315	0.946
social f1-5f vs f>=6f	1.923	1.115	3.317

HW5 Problem 3
Residuals vs predicted eta_i with LOESS Overlay

The LOESS Procedure
Selected Smoothing Parameter: 1
Dependent Variable: res

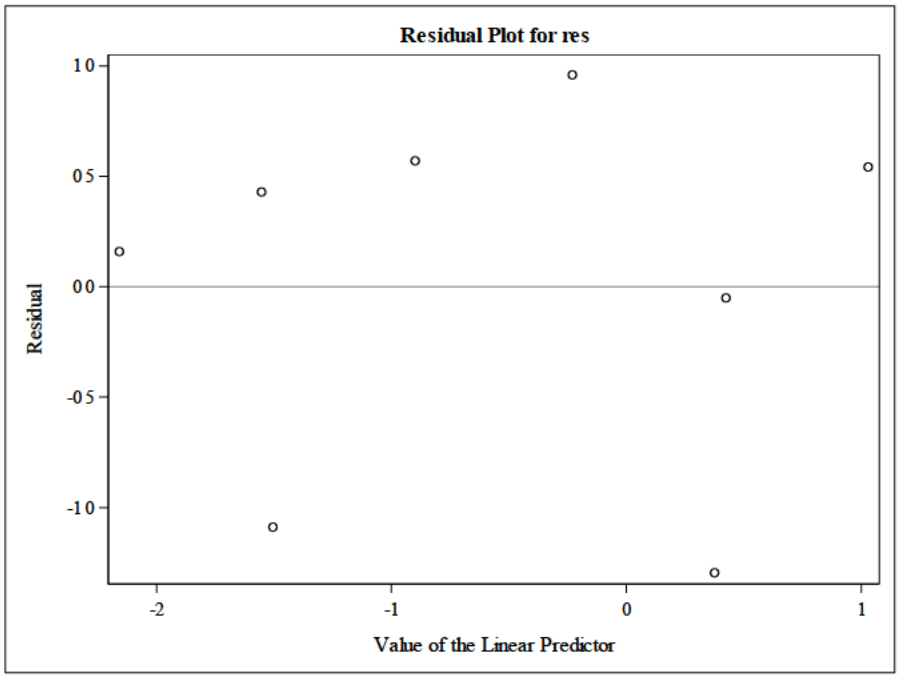
Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
2.1753	6	0.9029

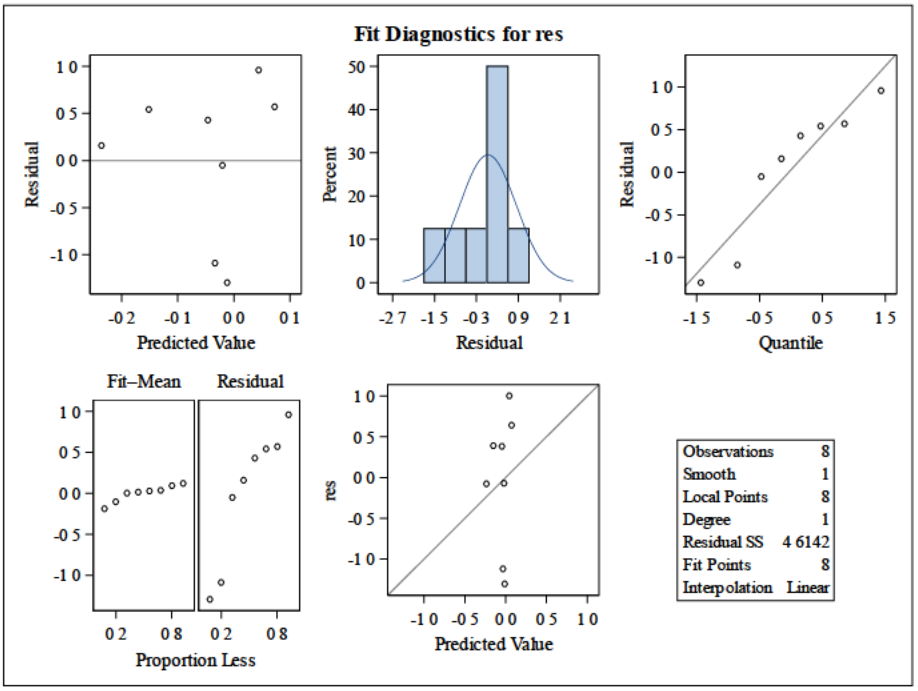


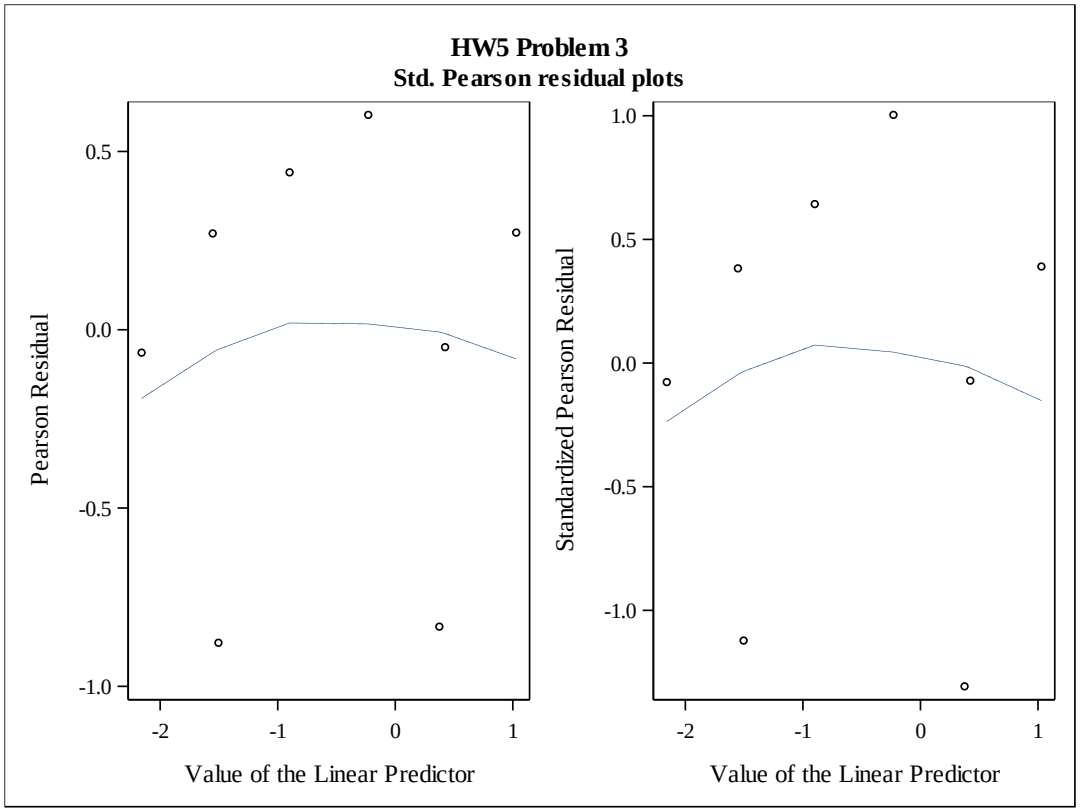


HW5 Problem 3
Residuals vs predicted eta_i with LOESS Overlay

The LOESS Procedure
Selected Smoothing Parameter: 1
Dependent Variable: res






HW5 Problem 3
Std. Pearson residual plots
Pearson Residual
0.5
0.0
-0.5
-1.0
-2
-1
0
1
Value of the Linear Predictor
Standardized Pearson Residual
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
Value of the Linear Predictor

Problem 4. Agresti 5.25

The derivative equals $\frac{\beta e^{\alpha+\beta x}}{[1+e^{\alpha+\beta x}]^2} = \beta\pi(x)(1-\pi(x))$

Problem 5. Agresti 5.26

The odds ratio e^β is approximately equal to the relative risk when the probability is near 0 and the complement is near 1, since

$$e^\beta = [\pi(x+1)/(1-(\pi(x+1)]/[\pi(x)/(1-\pi(x+1))] \approx \pi(x+1)/\pi(x)$$

Problem 6. Finney and Pregibon vasoconstriction

a. The AICs for the binary regressions using logit, probit and complimentary log-log links are 35.227, 35.287 and 32.622, respectively. Therefore, the complimentary log-log model has the smallest AIC.

b. In the fitted complimentary log-log model, we have $\hat{\beta}_0 = -2.9736$, $\hat{\beta}_1 = 3.9702$, and $\hat{\beta}_2 == 4.3361$. Therefore:

$$P(V=1) = 1 - exp[-exp\{-2.9736 + 3.9702 log(volume) + 4.3361 log(rate)\}]$$

$$P(V=1) = 1 - exp[-0.051 * volume^3.9702 * rate^4.3361]$$

Thus, increasing volume or rate increases the probability of vaso constriction.

c. The Hosmer-Lemeshow test statistic is 10.0705 (with $d.f. = 8$). This corresponds to a p-value of 0.2601. Therefore, there is no evidence of lack-of-fit.

d. An approach to detect ill-fit observations is to consider observations where the absolute value of the Pearson Residual is greater than 3. Using this approach, there are 2 observations that are considered ill-fit.

e. There are two influential observations according to the Dfbetas and Cis, and they are the same observations that have large Pearson residuals.

f. When we remove observations 4 and 18, the AIC value for the model drops from 32.622 to 13.32. None of the standardized Pearsons residuals are greater than 2.5, suggesting that no observations are especially ill-fitting. However, the coefficients change significantly. In particular, before we had $\hat{\beta}_0 = -2.9736$, $\hat{\beta}_1 = 3.9702$, and $\hat{\beta}_2 == 4.3361$. Now we have $\hat{\beta}_0 = -16.58384$, $\hat{\beta}_1 = 21.02387$, and $\hat{\beta}_2 = 25.30425$.

See next several pages of SAS output:

HW5, problem 6
Vaso Data Analysis [Cloglog]

Probability modeled is cons=1.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	56.040	32.622
SC	57.703	37.613
-2 Log L	54.040	26.622

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.9736	1.0934	7.3960	0.0065
lvol	1	3.9701	1.3825	8.2469	0.0041
lrate	1	4.3361	1.5589	7.7369	0.0054

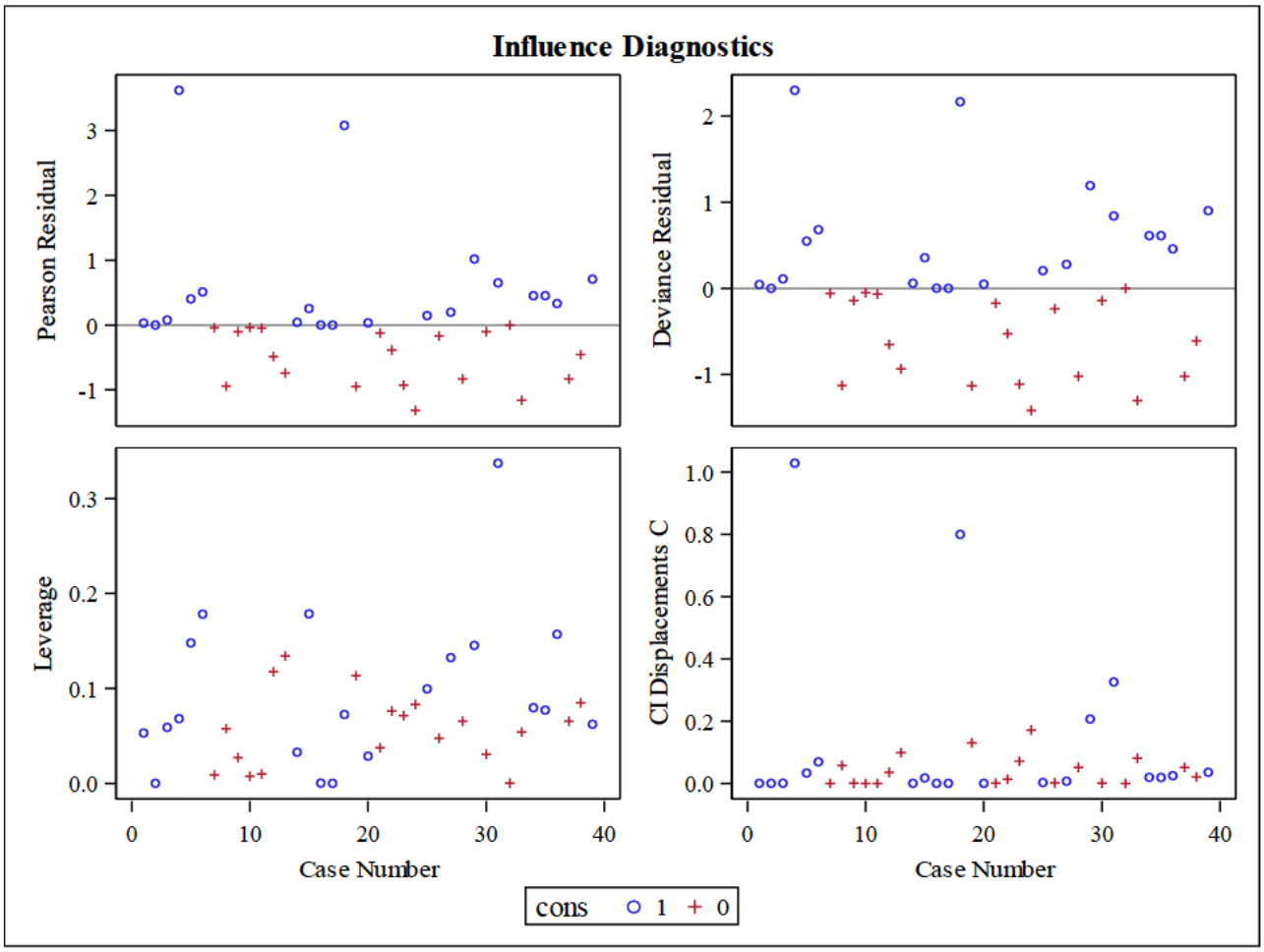
Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
10.0705	8	0.2601

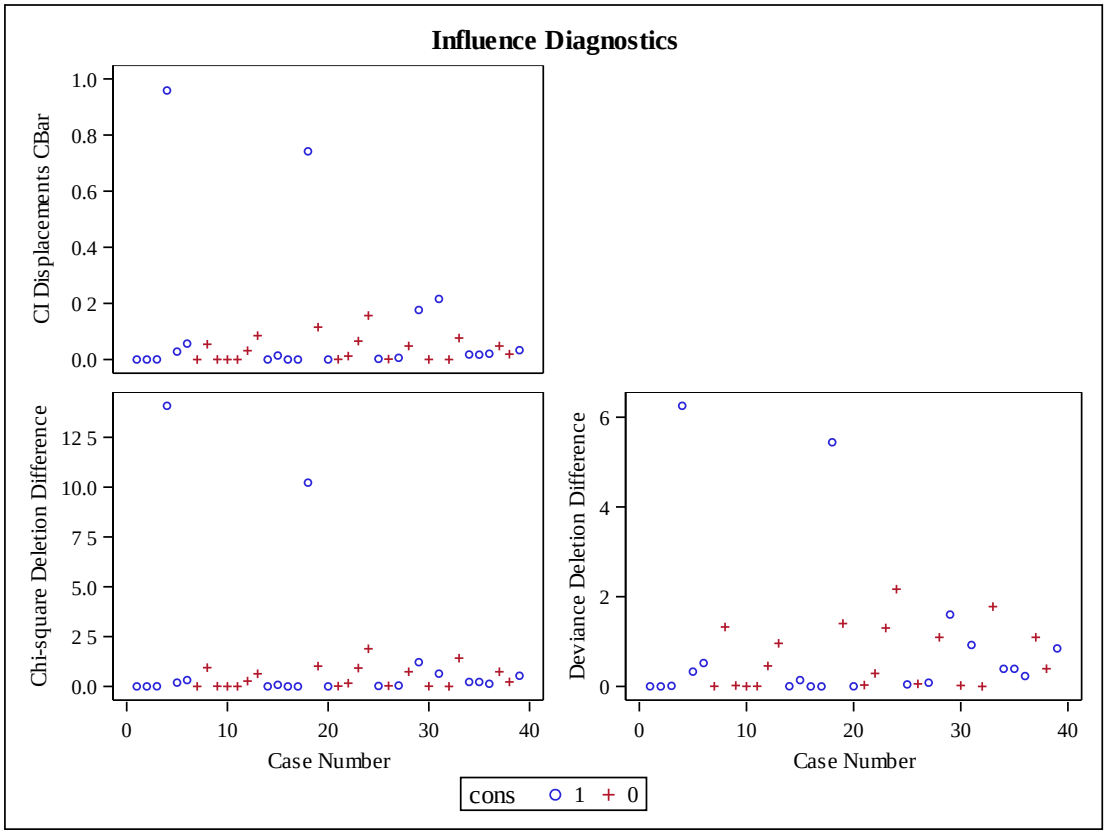
HW5, problem 6
Vaso Data Analysis [Cloglog]

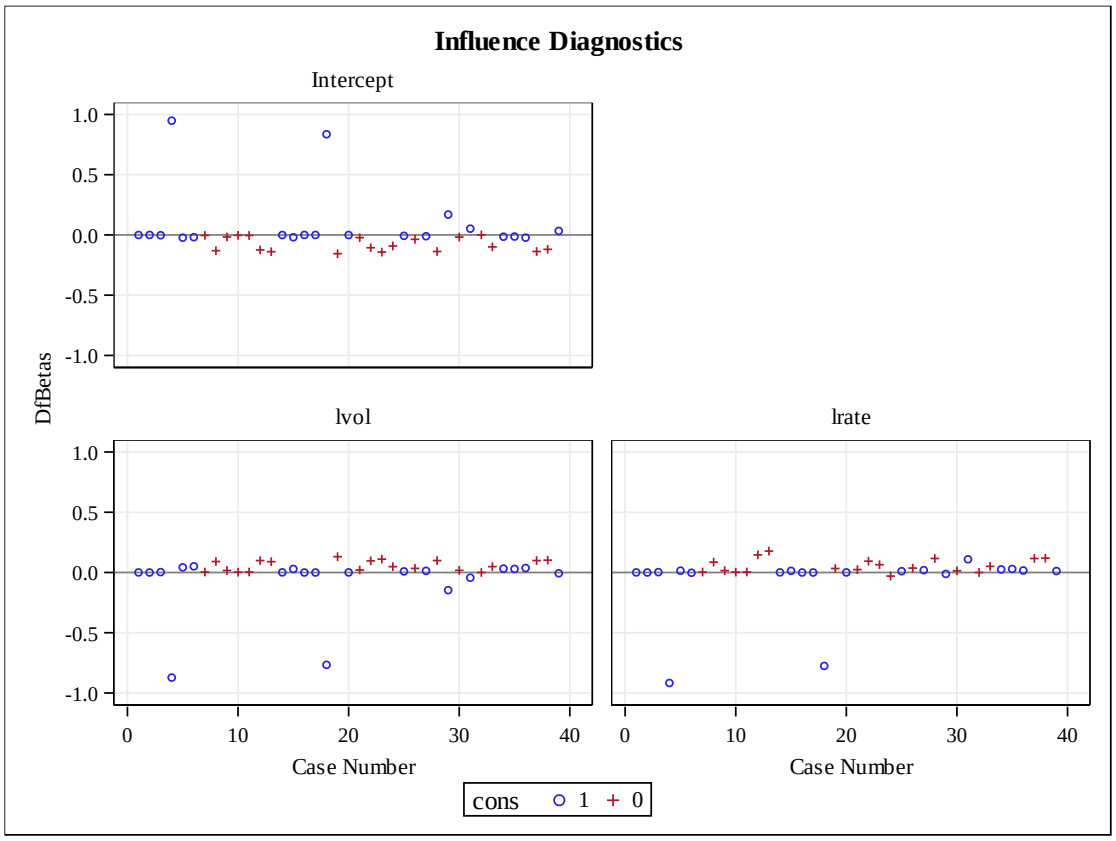
The LOGISTIC Procedure

Regression Diagnostics

Case Number	Covariates		Pearson Residual	Deviance Residual	lvol DfBeta	lrate DfBeta	Confidence Interval Displacement C	Confidence Interval Displacement CBar
	lvol	lrate						
4	0.4055	-0.2877	3.6225	2.3012	-0.8722	-0.9173	1.0287	0.9587
18	0.3471	-0.1625	3.0796	2.1679	-0.7663	-0.7754	0.7999	0.7419




Influence Diagnostics
CI Displacements CBar
Chi-square Deletion Difference
Deviance Deletion Difference
Case Number
Case Number
cons 1 0


Influence Diagnostics
Intercept
lvol
lrate
DfBetas
Case Number
Case Number
cons 1 0

HW5, problem 6
Vaso Data Analysis [Cloglog] Without Observations 4 and 18

Number of Observations Read	37
Number of Observations Used	37

Response Profile		
Ordered Value	**cons**	**Total Frequency**
1	1	18
2	0	19

Probability modeled is cons=1.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	53.266	13.325
SC	54.877	18.158
-2 Log L	51.266	7.325

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	43.9410	2	<.0001
Score	18.7712	2	<.0001
Wald	3.1492	2	0.2071

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-16.5838	10.3652	2.5598	0.1096
lvol	1	21.0237	13.0935	2.5782	0.1083
lrate	1	25.3041	16.7511	2.2819	0.1309

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
0.6550	5	0.9853

Problem 7. Agresti 6.4

For table 10.1, treating marijuana; Set the value as 1 if the predictor use of alcohol is YES and 0 otherwise; 1 if the predictor use of cigarette is YES and 0 otherwise; female as 1 and male as 0; white as 1 and 0 other race. Using a backwards elimination, the final model is composed of the predictors alcohol, cigarette and gender. All interaction terms were non-significant. Therefore, the fitted model is:

$$logit(\hat{\pi}) = -5.1883 + 3.0201 * alcohol + 2.8591 * cigarette - 0.3279 * gender$$

The Pearson GOF statistic yields a p-value of 0.8781, which indicates theres no evidence of gross lack of fit in this model. The odds of marijuana use among alcohol users is $exp(3.0201) = 20.494$ times of the odds among non-alcohol users when keeping the remaining parameters constant; the odds of marijuana use among smokers is $exp(2.8591) = 17.446$ times of the odds among non-smokers when keeping the remaining parameters constant. And the odds of marijuana use among males is $1/exp(0.3279) = 1.388$ times the odds among females when keeping the remaining parameters constant.

See next several pages of SAS output:

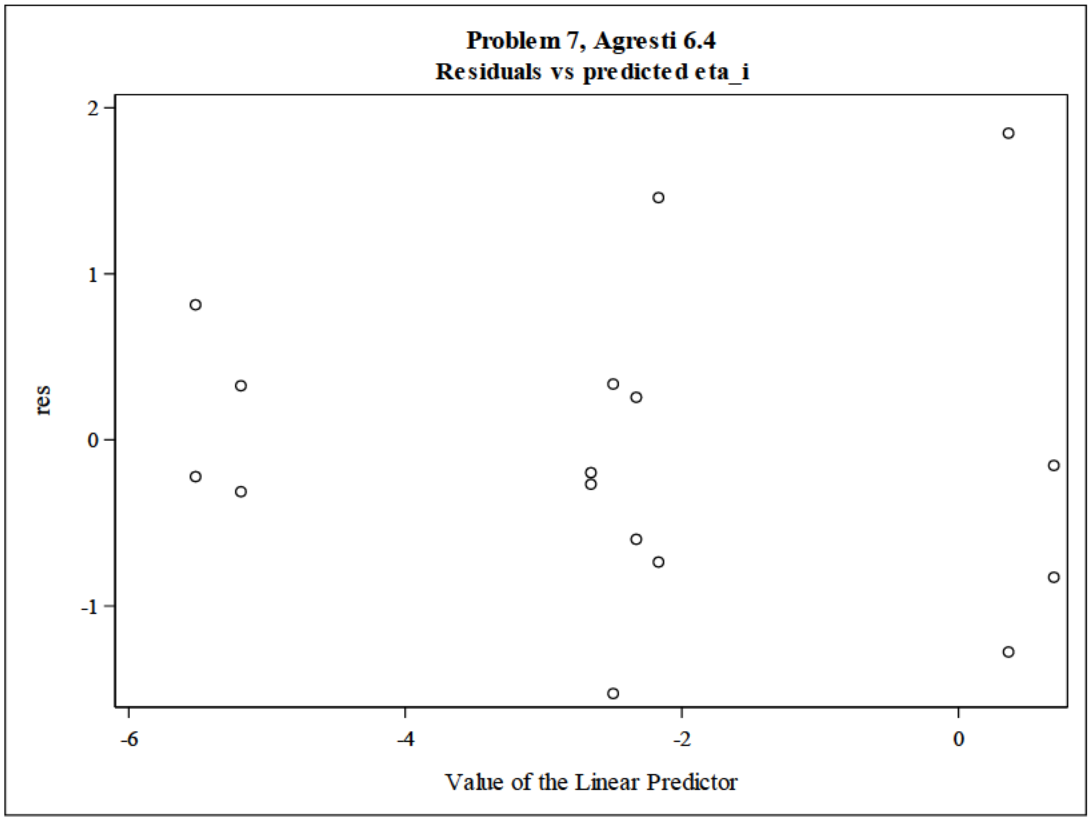
Problem 7, Agresti 6.4
Stepwise Regression on Marijuana Data

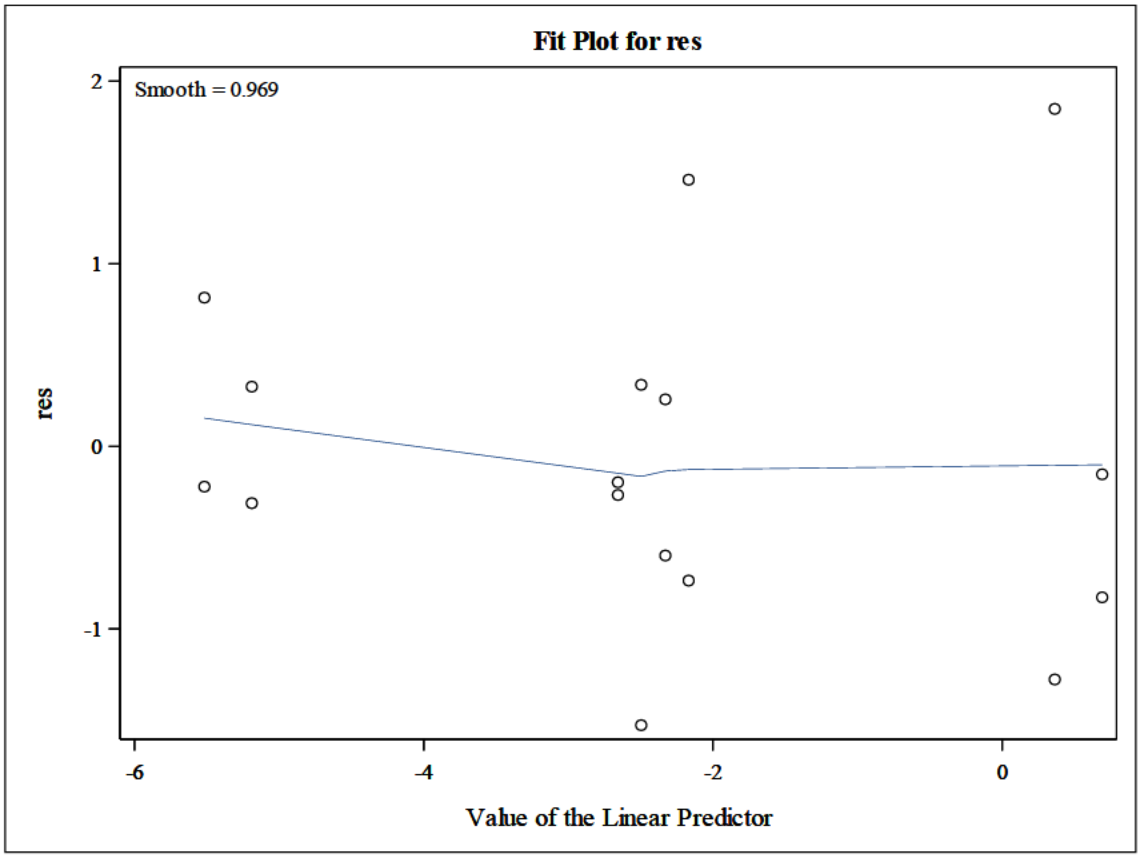
Final Model Selected

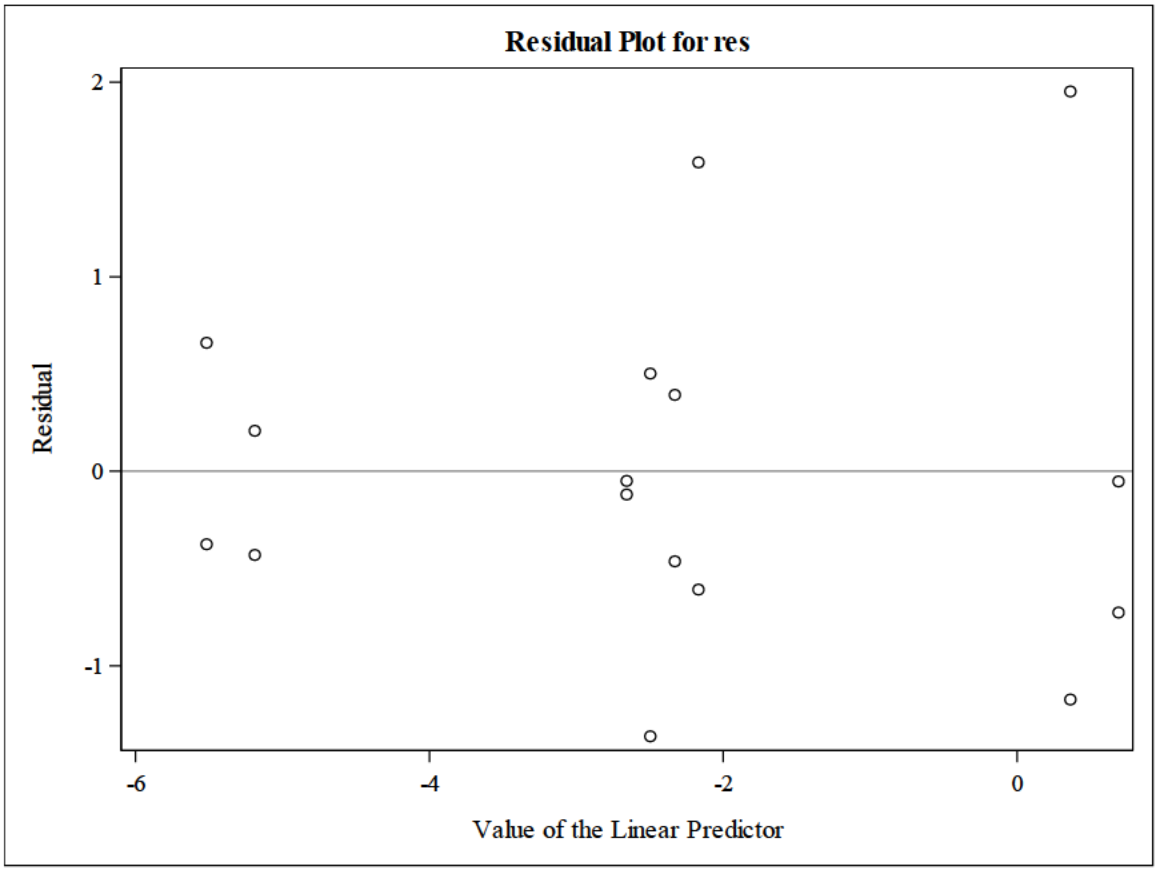
Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-5.1883	0.4769	118.3642	<.0001
a	**1**	1	3.0201	0.4653	42.1249	<.0001
c	**1**	1	2.8591	0.1642	303.0914	<.0001
g	**1**	1	-0.3279	0.1026	10.2200	0.0014

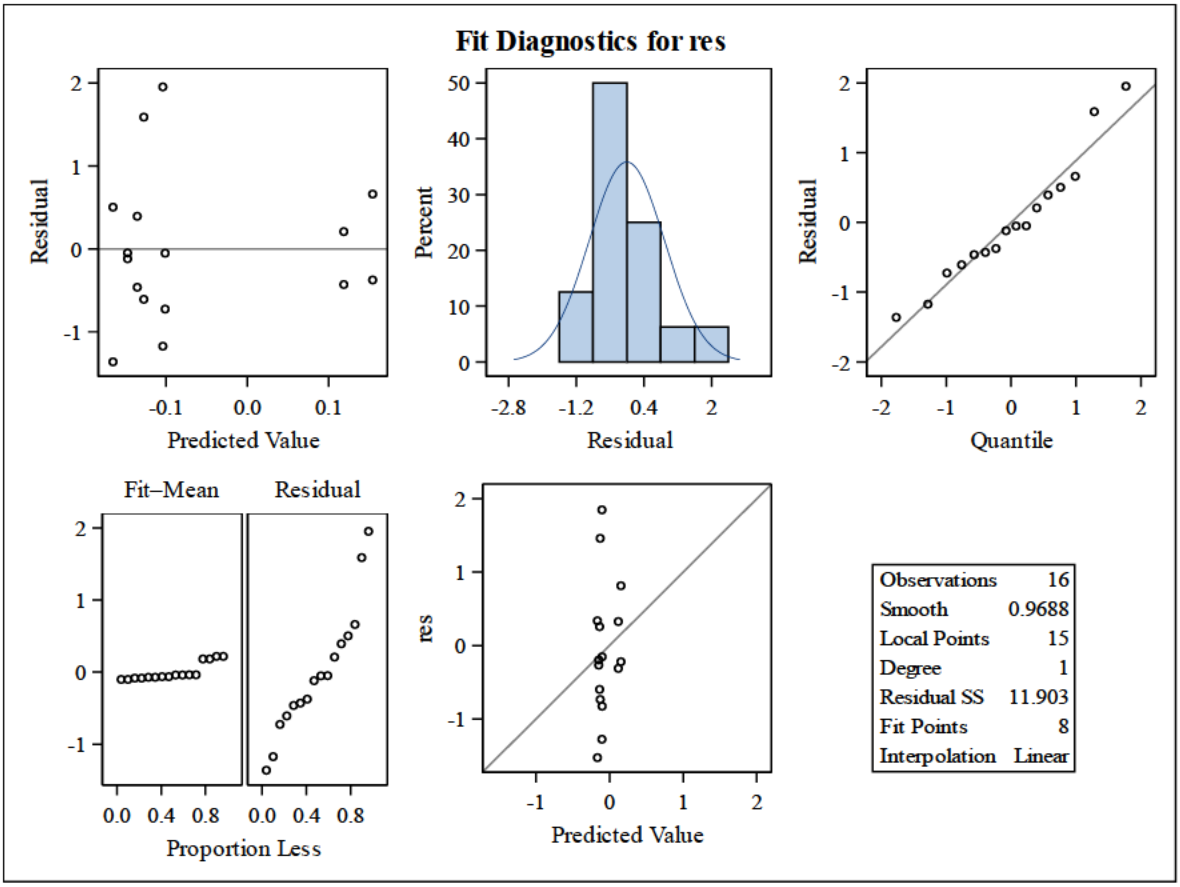
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
a 1 vs 2	20.494	8.233	51.016
c 1 vs 2	17.446	12.645	24.071
g 1 vs 2	0.720	0.589	0.881

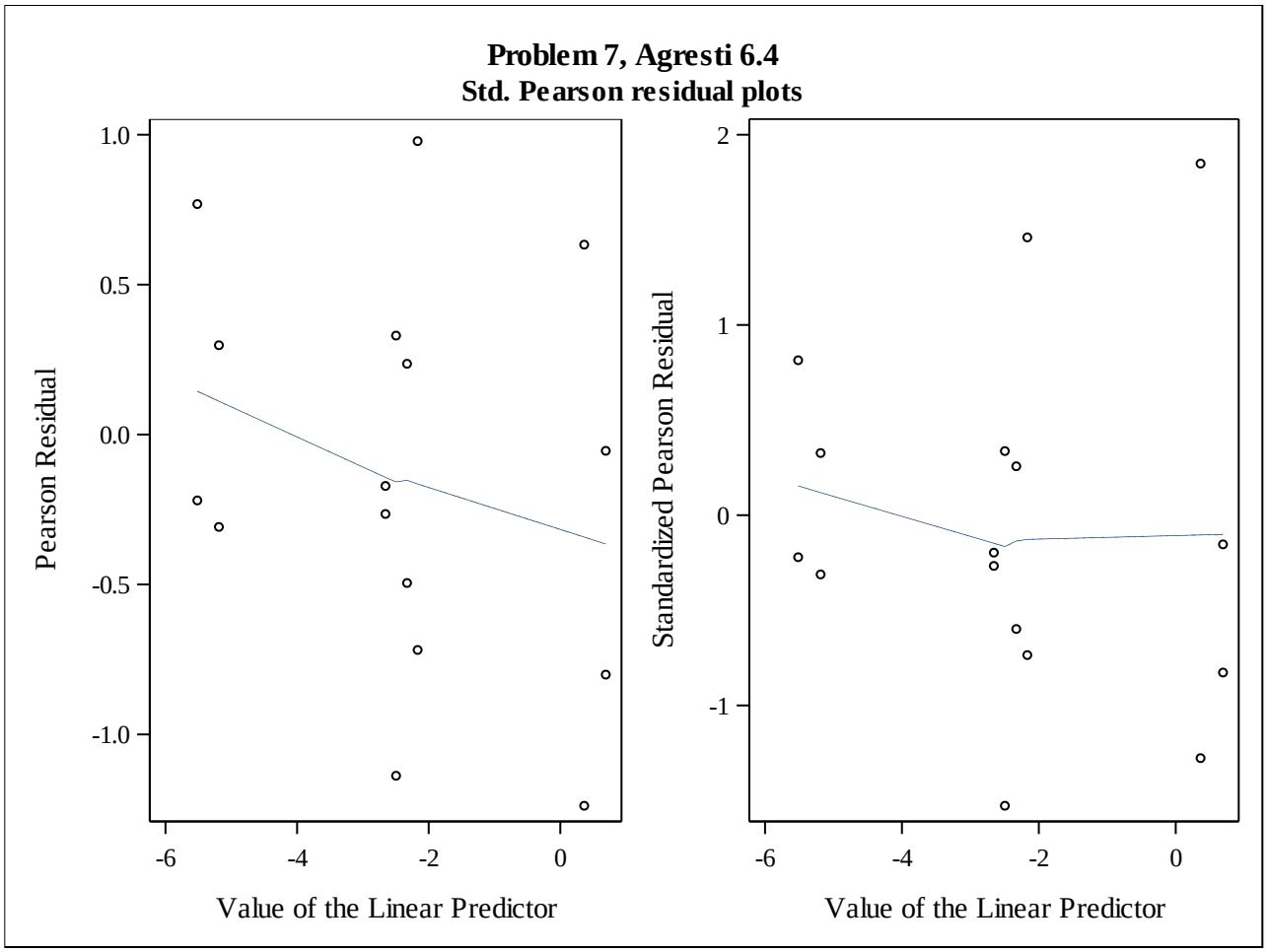
Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
1.8966	3	0.5941


Problem 7, Agresti 6.4
Residuals vs predicted eta_i
res
2
1
0
-1
-6
-4
-2
0
Value of the Linear Predictor


Fit Plot for res
Smooth = 0.969
res
2
1
0
-1
-6
-4
-2
0
Value of the Linear Predictor


Residual Plot for res
Residual
2
1
0
-1
-6
-4
-2
0
Value of the Linear Predictor


Fit Diagnostics for res
Residual
Predicted Value
Percent
Residual
Residual
Quantile
Fit–Mean
Residual
Proportion Less
res
Predicted Value
Observations 16
Smooth 0.9688
Local Points 15
Degree 1
Residual SS 11.903
Fit Points 8
Interpolation Linear


Problem 7, Agresti 6.4
Std. Pearson residual plots
Pearson Residual
1.0
0.5
0.0
-0.5
-1.0
-6
-4
-2
0
Value of the Linear Predictor
Standardized Pearson Residual
2
1
0
-1
-6
-4
-2
0
Value of the Linear Predictor

Problem 8. Dixon and Massey

First, we fit the model with all main effects and 2-way interactions. Choosing the predictors that have p-value greater than 0.05, we get a model with the predictors age and weight. The fitted model is given by:

$$logit(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 * Age + \hat{\beta}_2 * weight$$

$$logit(\hat{\pi}) = -7.5128 + 0.0636 * Age + 0.0160 * weight$$

A backward elimination can also be done to find the best model. The Hosmer-Lemeshow test has a p-value equal to 0.7941, which indicates that there is no gross lack of fit in this model. The odds of having an incident increases by a multiplicative factor of exp(0.0636) = 1.066 (95%C.I is (1.025,1.108))for every unit increase in age while holding weight constant; and the odds the odds of having an incident increases by a multiplicative factor of 1.016 95%C.I is (1.000,1.032) for every unit increase in weight while holding age constant.

See next several pages of SAS output:

Problem 8, heart.sas
Stepwise Regression on Heart Data

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-7.5128	1.7093	19.3176	<.0001
AG	1	0.0636	0.0197	10.4389	0.0012
W	1	0.0160	0.00795	4.0464	0.0443

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
AG	1.066	1.025	1.108
W	1.016	1.000	1.032

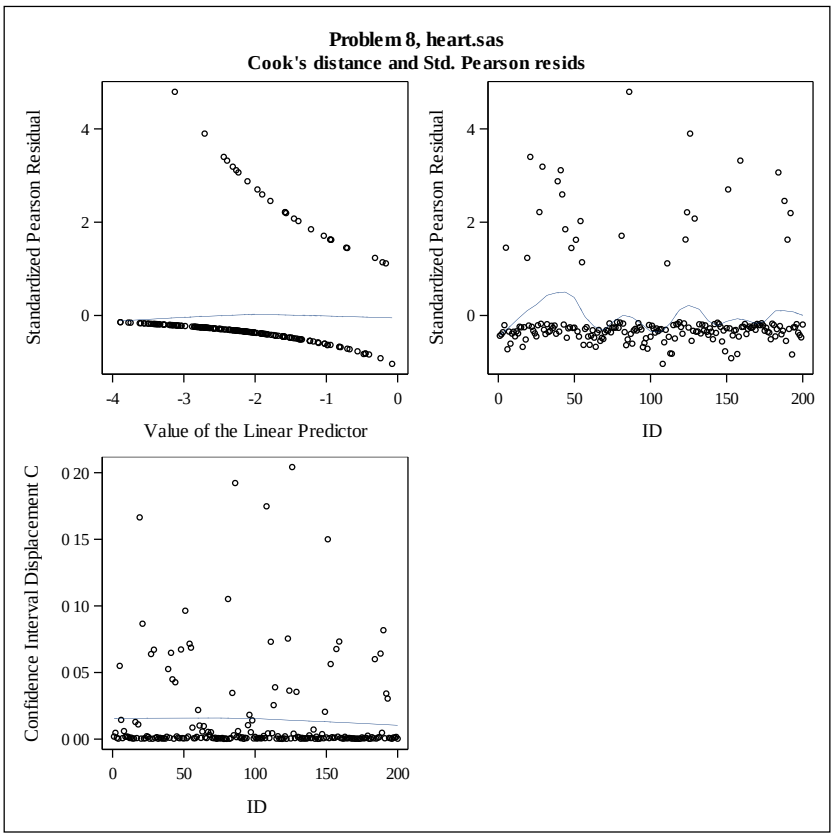
Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
4.6510	8	0.7941

Problem 8, heart.sas
Backward Regression on Heart Data

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-7.5128	1.7093	19.3176	<.0001
AG	1	0.0636	0.0197	10.4389	0.0012
W	1	0.0160	0.00795	4.0464	0.0443

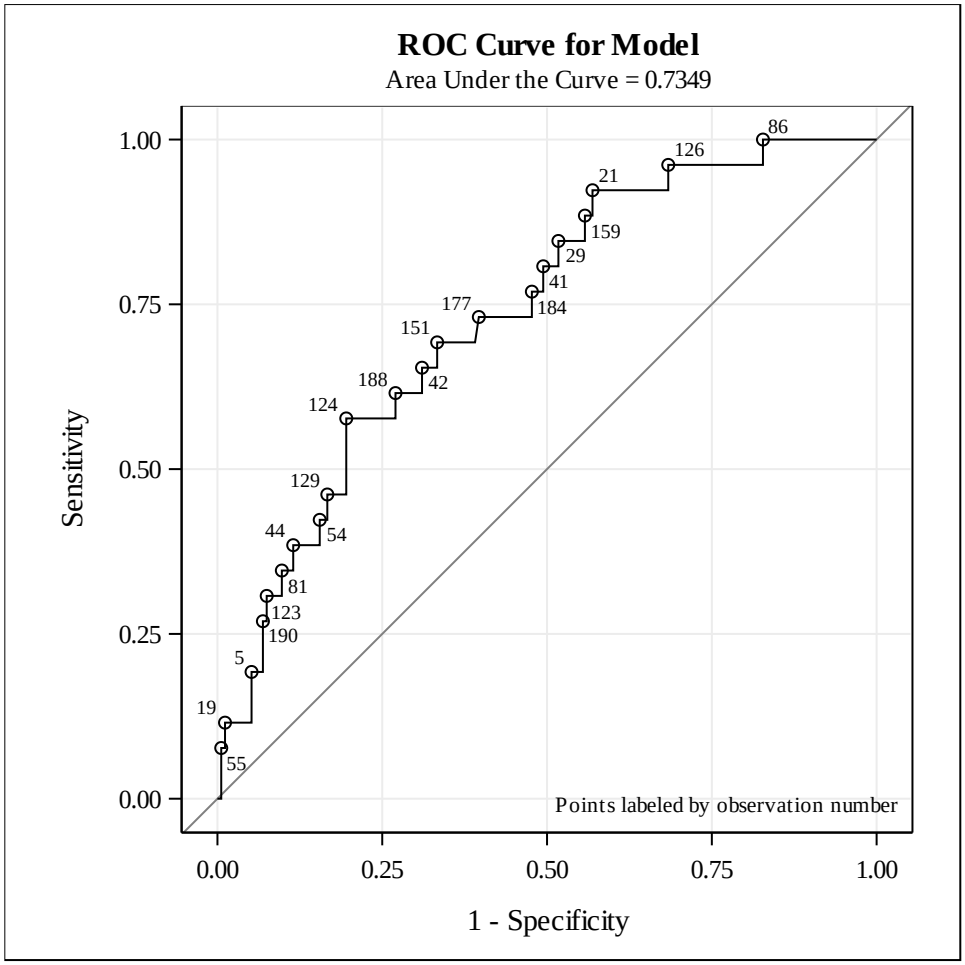
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
AG	1.066	1.025	1.108
W	1.016	1.000	1.032

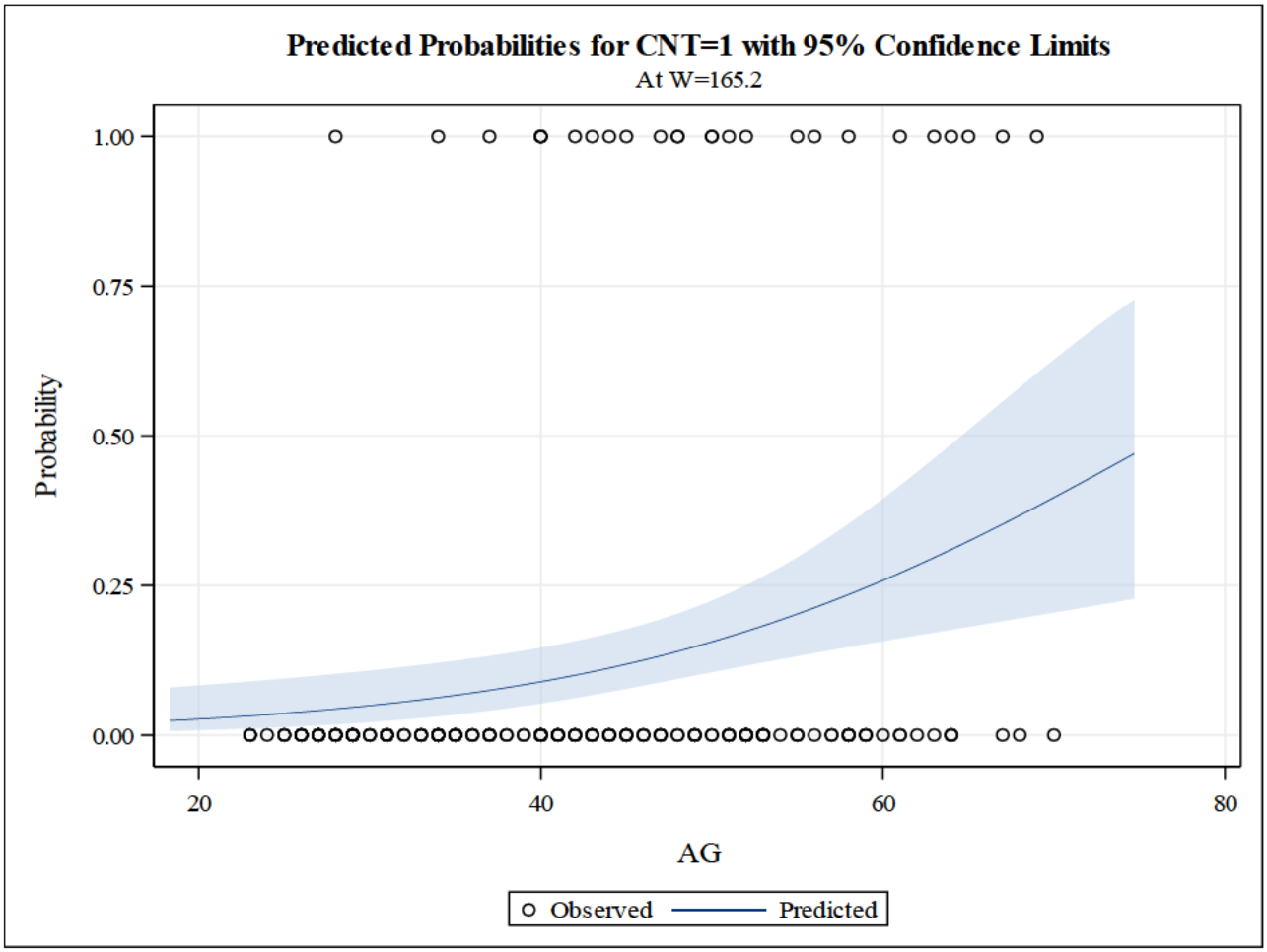
Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
4.6510	8	0.7941


Problem 8, heart.sas
Cook's distance and Std. Pearson resids
Standardized Pearson Residual
Value of the Linear Predictor
Standardized Pearson Residual
ID
Confidence Interval Displacement C
ID

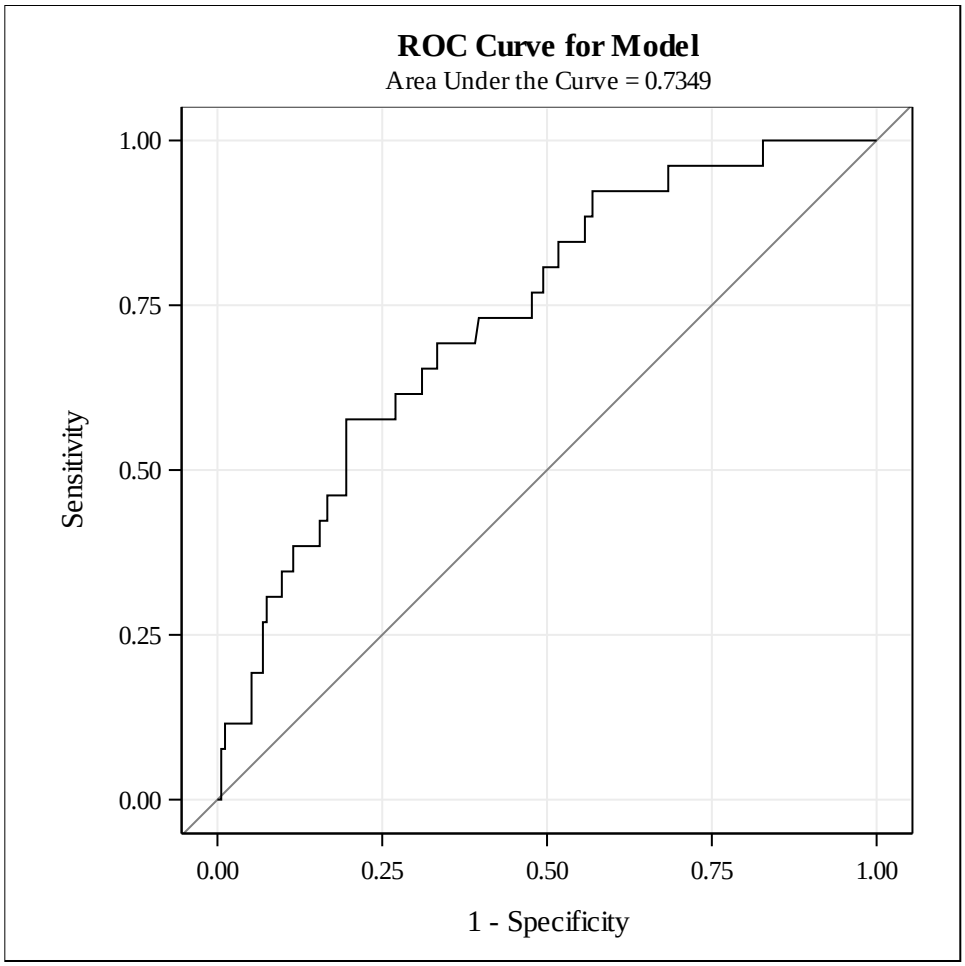
Problem 8, heart.sas
Cook's distance and Std. Pearson resids

Obs	ID	AG	S	D	Ch	H	W	eta	phat	lcl	ucl	p	h2	q	c
21	21	37	110	70	312	71	170	-2.43996	0.080176	0.044684	0.139733	3.387130	0.007438	3.39979	0.08662
29	29	40	110	74	336	68	166	-2.31323	0.090033	0.053777	0.146943	3.179150	0.006555	3.18962	0.06713
41	41	40	130	90	520	68	169	-2.26522	0.094045	0.056668	0.152103	3.103750	0.006636	3.11410	0.06478
86	86	34	110	80	214	67	139	-3.12677	0.042017	0.017700	0.096464	4.774950	0.008292	4.79487	0.19223
126	126	28	120	86	386	70	189	-2.70816	0.062494	0.025570	0.144813	3.873190	0.013255	3.89912	0.20423
159	159	40	110	70	244	70	161	-2.39324	0.083690	0.048877	0.139663	3.308910	0.006602	3.31989	0.07325
184	184	43	138	94	320	65	159	-2.23450	0.096695	0.059371	0.153653	3.056430	0.006345	3.06618	0.06003





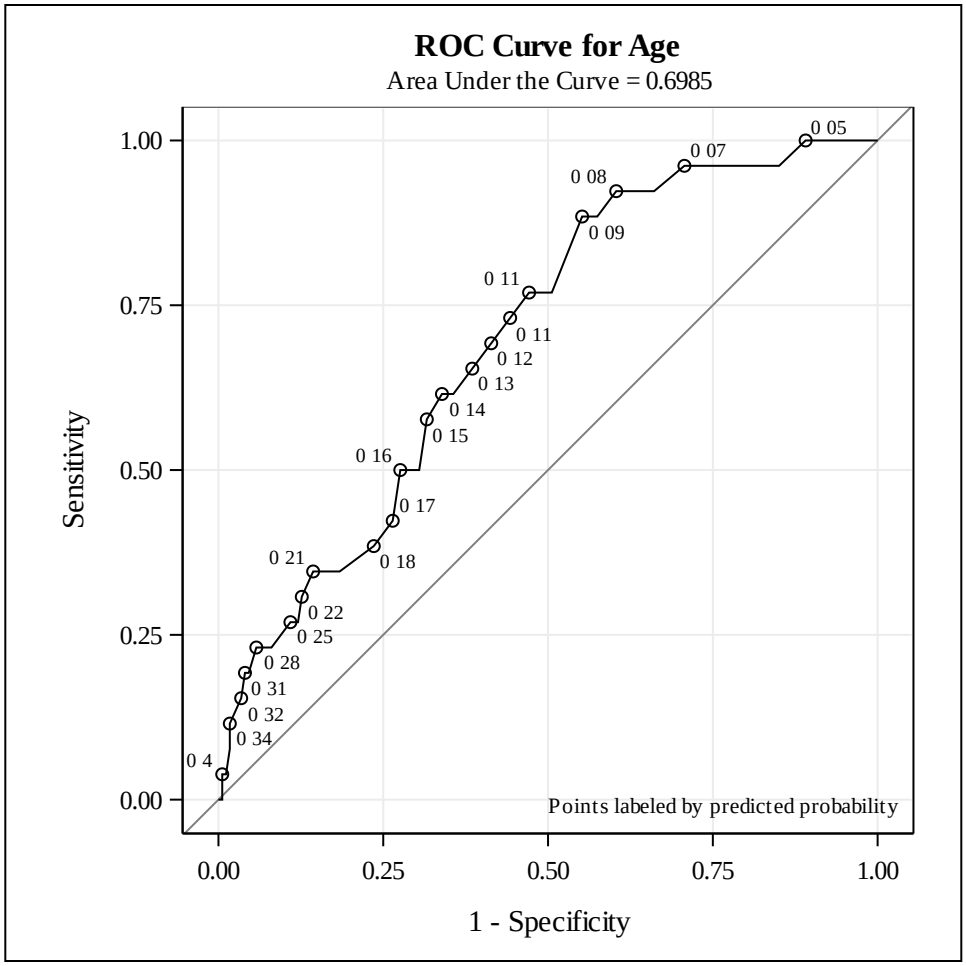
R-Square	0.0759	Max-rescaled R-Square	0.1410



ROC Model: Age

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-4.7925	0.9595	24.9481	<.0001
AG	1	0.0633	0.0192	10.8270	0.0010

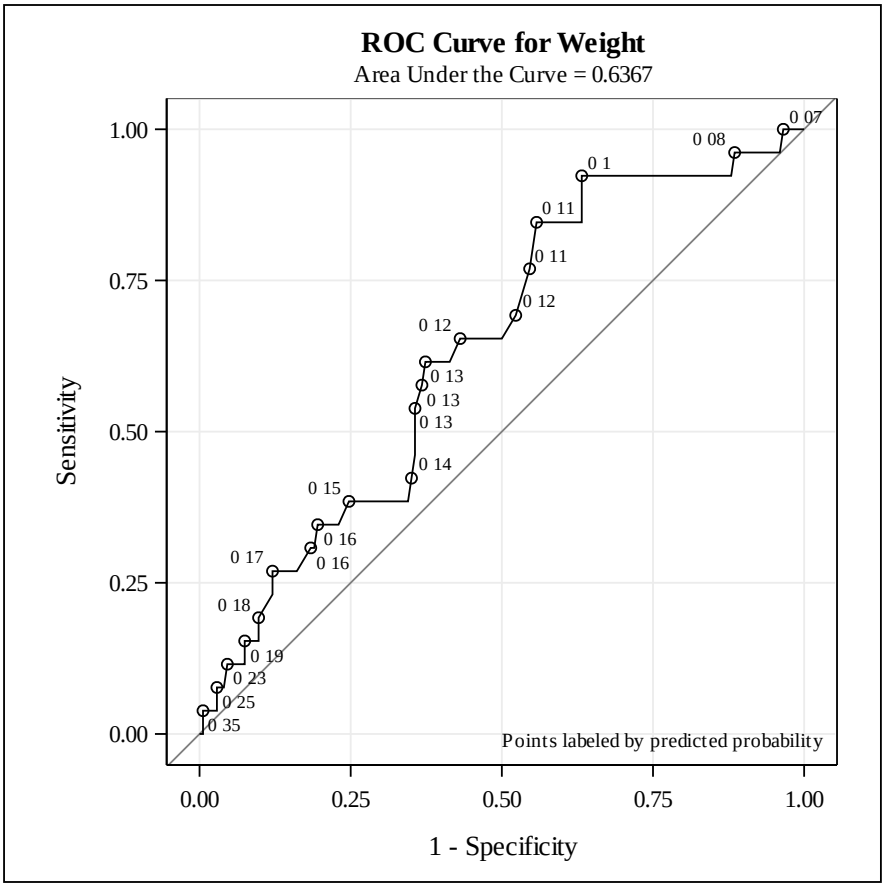
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
AG	1.065	1.026	1.106

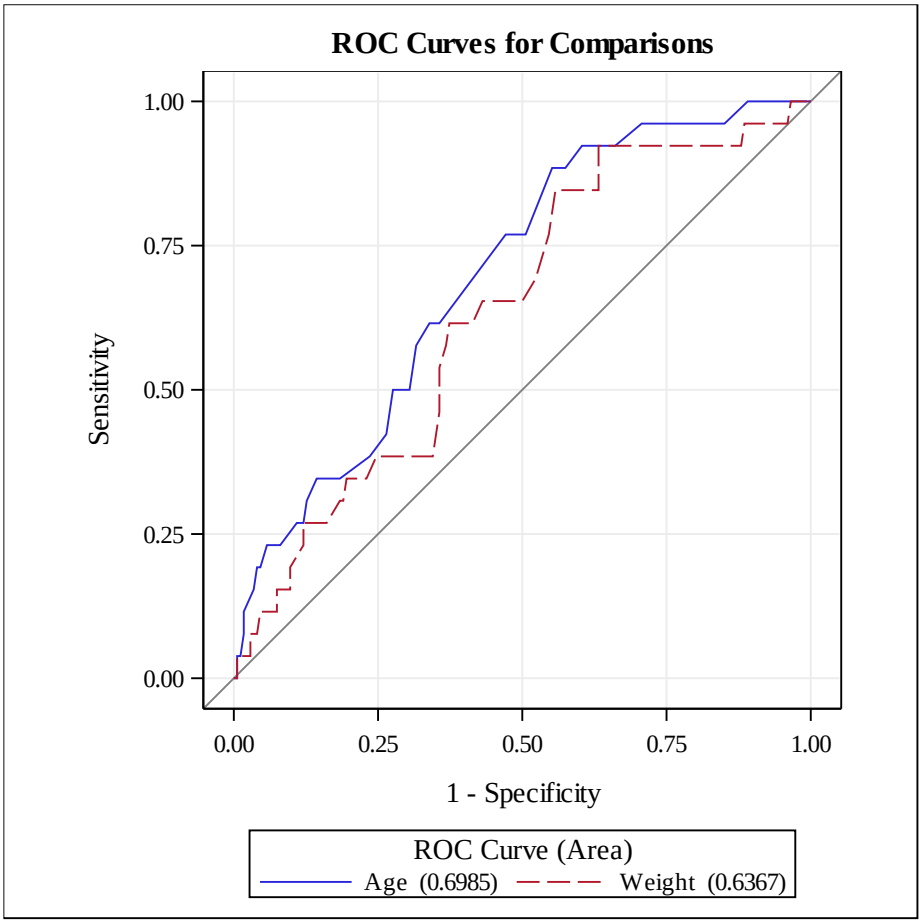


ROC Model: Weight

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-4.7244	1.3692	11.9064	0.0006
W	1	0.0167	0.00783	4.5536	0.0328

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
W	1.017	1.001	1.033


ROC Curve for Weight
Area Under the Curve = 0.6367
Sensitivity
1 - Specificity
0.00
0.25
0.50
0.75
1.00
0 07
0 08
0 1
0 11
0 11
0 12
0 12
0 13
0 13
0 13
0 14
0 15
0 16
0 16
0 17
0 18
0 19
0 23
0 25
0 35
Points labeled by predicted probability


ROC Curves for Comparisons
Sensitivity
1 - Specificity
0.00
0.25
0.50
0.75
1.00
ROC Curve (Area)
Age (0.6985)
Weight (0.6367)

ROC Association Statistics							
	Mann-Whitney						
ROC Model	Area	Standard Error	95% Wald Confidence Limits		Somers' D	Gamma	Tau-a
Age	0.6985	0.0504	0.5998	0.7972	0.3970	0.4065	0.0903
Weight	0.6367	0.0551	0.5287	0.7447	0.2734	0.2758	0.0622

ROC Contrast Coefficients	
ROC Model	Row1
Age	1
Weight	-1

ROC Contrast Test Results			
Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Weight	1	0.6822	0.4088

ROC Contrast Estimation and Testing Results by Row						
Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
Age - Weight	0.0618	0.0748	-0.0848	0.2084	0.6822	0.4088

```
/* HW5, Problem 3 */
/*ods rtf file="SAS Output_HW5 Prob 3_files.rtf";*/

title1 'HW5 Problem 3';

data colds;
input colds total titer$ virus$ social$;
datalines;
25 33 f<=2f fRV39f f1-5f
20 38 f<=2f fRV39f f>=6f
18 30 f<=2f fHanksf f1-5f
21 43 f<=2f fHanksf f>=6f
11 34 f>=4f fRV39f f1-5f
8 42 f>=4f fRV39f f>=6f
3 26 f>=4f fHanksf f1-5f
3 30 f>=4f fHanksf f>=6f
;
run;
proc print;
run;

proc logistic data=colds outest=betas covout;
title2 'Stepwise Regression on Cold Data';
class titer virus social/param=ref;
model colds/total = titer virus social titer*virus virus*social titer*social
                   / selection=stepwise
                     slentry=0.05
                     slstay=0.05
                     details
                     lackfit;
output out=pred p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h xbeta=eta predprob=(individual c
run;



/* Creating the standardized residuals from the "pred" dataset */
data pred2; set pred; res = p/sqrt(1-h); run;

/* Usual plot of stand. Residuals vs predicted eta_i */

proc sgplot data=pred2;
          title2 "Residuals vs predicted eta_i";
      scatter y=res x=eta;
run;


/* Now, doing the LOESS overlay on the r_i vs eta_i fit. You can learn about smoothing parameter selec
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_loess_sect004.
Lot to learn!   */

proc loess data=pred2;
title2 "Residuals vs predicted eta_i with LOESS Overlay";
model res = eta;
run;
```

```
/* Also, you can try this one below */

proc sgscatter data=pred;
title2 "Std. Pearson residual plots";
plot p*eta q*eta / loess;
run;
/*ods rtf close;*/

/* These are some extras you can ignore */


proc print data=betas;
      title2 'Parameter Estimates and Covariance Matrix';
run;
proc print data=pred;
      title2 'Predicted Probabilities and 95% Confidence Limits';
run;




/* HW5, problem 6 */
/*ods rtf bodytitle file='SAS Output HW5 Problem 6.rtf';*/
title 'HW5, problem 6';
data vaso;
 input cons volume rate;
 datalines;
/* DATALINES not shown here. */
.
.
.
;
run;

data vaso1; set vaso;
lvol = log(volume);
lrate = log(rate);
run;


/*proc print data=vaso1;*/
/*run;*/

proc logistic data=vaso1 descending;
title2 'Vaso Data Analysis [Logit]';
model cons = lvol lrate/link = logit aggregate lackfit;
output out=pred p=phat lower=lcl upper=ucl reschi=p h=h xbeta=eta;
run;


proc logistic data=vaso1 descending;
title2 'Vaso Data Analysis [Probit]';
model cons = lvol lrate/link = probit aggregate lackfit;
output out=pred p=phat lower=lcl upper=ucl reschi=p h=h xbeta=eta;
run;
```

```
proc logistic data=vaso1 descending;
title2 'Vaso Data Analysis [Cloglog]';
model cons = lvol lrate/link = cloglog aggregate lackfit influence iplots;
output out=pred p=phat lower=lcl upper=ucl reschi=p h=h xbeta=eta;
run;


/* Note that the 2 abberrant observations are 4 ans 18. Both have high Pearson, Dbbeta, and Cis. */

data vaso2; set vaso1;
 if _n_=4 then delete; if _n_=18 then delete;

proc logistic descending data=vaso2;
 model cons = lvol lrate / link=cloglog lackfit;
run;

ods rtf close;

/* Aside */

proc logistic data=vaso1 descending;
model cons = lvol lrate/link = cloglog aggregate lackfit influence;
output out=pred reschi = u stdreschi=r xbeta=eta p=p;
run;

data pred1; set pred; keep r u;
run;

proc print data=pred1;
run;

title;
```

```
/* Problem 7, Agresti 6.4     */

/*a = alcohol; c = cigarette use; m = marijuana use; r = race [1 = white, 2 = black]; g = gender [1 = 

title "Problem 7, Agresti 6.4";
/*ods rtf bodytitle file='SAS Output HW5 Problem 7.rtf';*/
data mari;
input a c m r g count total;
datalines;
1 1 1 1 1 405  673
1 1 1 2 1  23  46
1 2 1 1 1  13  231
1 2 1 2 1   2  21
2 1 1 1 1   1  18
2 1 1 2 1   0   1
2 2 1 1 1   1  118
2 2 1 2 1   0   12
1 1 1 1 2 453  681
1 1 1 2 2  30  49
1 2 1 1 2  28  229
1 2 1 2 2   1  19
2 1 1 1 2   1  18
2 1 1 2 2   1  9
2 2 1 1 2   1  134
2 2 1 2 2   0  17
;
run;

proc logistic data= mari outest=betas covout;
title2 'Stepwise Regression on Marijuana Data';
class a c r g/param=ref;
model count/total = a|c|r|g@2 / selection=stepwise slentry=0.05 slstay=0.05 details lackfit;
output out=pred p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h xbeta=eta predprob=(individual c
run;

/* Creating the standardized residuals from the "pred" dataset */
data pred2; set pred; res = p/sqrt(1-h);

/* Usual plot of stand. Residuals vs predicted eta_i */
proc sgplot data=pred2;
          title2 "Residuals vs predicted eta_i";
      scatter y=res x=eta;
run;

proc loess data=pred2;
title2 'residuals vs. eta';
model res = eta;
run;

/* Also, you can try this one below */
proc sgscatter data=pred;
title2 "Std. Pearson residual plots";
plot p*eta q*eta / loess;
run;
/*ods rtf close;*/
```

```
/* Problem 8, heart.sas  */
title "Problem 8, heart.sas";
/*ods rtf bodytitle file='SAS Output HW5 Problem 8.rtf';*/

data heart;
 input ID AG S D Ch H W CNT garbage;
 datalines;
/* DATALINES not shown here (very long) */
.
.
.
;
run;

/* (a) Setpwise elimination, you can also try Backward Elimination */
proc logistic data=heart descending;
title2 'Stepwise Regression on Heart Data';
model CNT = AG|S|D|Ch|H|W@2
                   / selection=stepwise
                     slentry=0.05
                     slstay=0.05
                     details
                     lackfit
                                        scale = none;
output out=predict p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h2 xbeta=eta c = c predprob=(in
run;


proc logistic data=heart descending;
title2 'Backward Regression on Heart Data';
model CNT = AG|S|D|Ch|H|W@2
                   / selection=B fast ctable
                     slentry=0.05
                     slstay=0.05
                     details
                     lackfit;
output out=predict p=phat lower=lcl upper=ucl stdreschi = q reschi=p h=h2 c = c xbeta=eta predprob=(in
run;



/* Plots, and also comment from the Table on Outliers. Interpretation in details, wrt odds ratios */

proc sgscatter data=predict;
title2 "Cook's distance and Std. Pearson resids";
plot q*eta q*id c*id/loess ;
proc print data=predict(where=(c>0.2 or q>3 or q<-3));
*var y width color c r;
run;
```

```
/* Now, predictive accuracy */
proc logistic data = heart descending plots(only)=(roc(id=obs) effect);
model CNT = AG W/scale = none details lackfit rsquare;
run;

proc logistic data = heart descending plots;
model CNT = AG W/scale = none details lackfit rsquare outroc=ROCcurve ctable;
run;

/* Below find individual ROC curves for each covariate */

proc logistic data=heart plots=roc(id=prob);
      model CNT(event='1') = AG W / nofit;
      roc 'Age' AG;
      roc 'Weight' W;
      roccontrast reference('Weight') / estimate e;
run;
/*ods rtf close;*/


/* The GAM below is some extras */
*proc gam plots(clm);
proc gam data=heart;
 model CNT(EVENT='1') = spline(AG) spline(Ch) spline(W) / dist=binomial link=logit;
run;
*ods graphics off;
*ods html close;
```