

Logistic Regression

```
/****** impeach.sas *****/
options linesize=79 noovp formdlim='-';
title 'Clinton Impeachment Votes in the U. S. Senate (1999)';

proc format;
  value gfmt  0 = "Not Guilty" 1 = "Guilty";
  value pfmt  0 = "Democrat"  1 = "Republican";
  value ynfmt 0 = "No" 1 = "Yes";

data bill;
  infile 'impeach.dat';
  input name $ state $ perjury obstr nvguilt party ritewing cpercent
        nextelec firsterm;
  label perjury = "Vote on Perjury"
        obstr   = "Vote on Obstruction of Justice"
        nvguilt = "Number of votes for guilt (0-2)"
        ritewing = "Degree of ideological conservatism (0-100)"
        cpercent = "% vote Clinton received in 1996, home state"
        nextelec = "Year must run for re-election"
        firsterm = "First-term senator?";
  format perjury obstr gfmt.;
  format party pfmt.;
  format firsterm ynfmt.;
  /* Dummy variables for next election: Only use 2 at a time */
  if nextelec=2000 then next0=1; else next0=0;
  if nextelec=2002 then next2=1; else next2=0;
  if nextelec=2004 then next4=1; else next4=0;

proc univariate plot;
  var ritewing cpercent;
proc freq;
  tables state perjury obstr nvguilt party nextelec firsterm;
  tables (next0 next2 next4) * nextelec / norow nocol nopercnt;
  tables (party nextelec firsterm) * perjury / nocol nopercnt chisq;

proc logistic order=internal descending; /* Always use descending for 0-1 DV */
  title2 'Logistic regression on perjury vote';
  model perjury = ritewing cpercent next0 next2 firsterm;
  nextelec: test next0=next2=0;
  others: test cpercent=next0=next2=firsterm=0; /* Ctrl for ritewing */
  allvars: test ritewing=cpercent=next0=next2=firsterm=0;
  /* Just for comparison with Testing Global Null Hypothesis: BETA=0 */

/* I specified order=internal, because by default proc logistic uses
order=formatted -- that is, alphabetical order. This, with the descending
option, would have induced a model for the log odds of Not guilty, because
it's alphabetically the "highest" category -- not what we want. */

proc logistic order=internal descending; /* Always use descending for 0-1 DV */
  title2 'SAS will make your dummy variables';
  class nextelec / param=ref;
  /* Indicator dummy vars: Last category is reference */;
  model perjury = ritewing cpercent nextelec firsterm;
```

```

proc logistic;
  title2 'Compare LR test of 0.3661 p = 0.8327 from proc freq';
  class nextelec;
  model perjury=nextelec;

```

impeach.lst

Clinton Impeachment Votes in the U. S, Senate (1999) 1
 16:57 Saturday, February 24, 2007

The UNIVARIATE Procedure
 Variable: ritewing (Degree of ideological conservatism (0-100))

Moments

N	100	Sum Weights	100
Mean	47.31	Sum Observations	4731
Std Deviation	37.4779922	Variance	1404.5999
Skewness	0.05310893	Kurtosis	-1.6850447
Uncorrected SS	362879	Corrected SS	139055.39
Coeff Variation	79.2179078	Std Error Mean	3.74779922

Basic Statistical Measures

Location		Variability	
Mean	47.31000	Std Deviation	37.47799
Median	52.00000	Variance	1405
Mode	4.00000	Range	100.00000
		Interquartile Range	76.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 12.62341	Pr > t <.0001
Sign	M 46	Pr >= M <.0001
Signed Rank	S 2139	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	100
99%	100
95%	100
90%	96
75% Q3	84
50% Median	52
25% Q1	8
10%	4

5% 0
 1% 0
 0% Min 0

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
0	94	100	45
0	78	100	49

 Clinton Impeachment Votes in the U. S, Senate (1999) 2
 16:57 Saturday, February 24, 2007

The UNIVARIATE Procedure

Variable: ritewing (Degree of ideological conservatism (0-100))

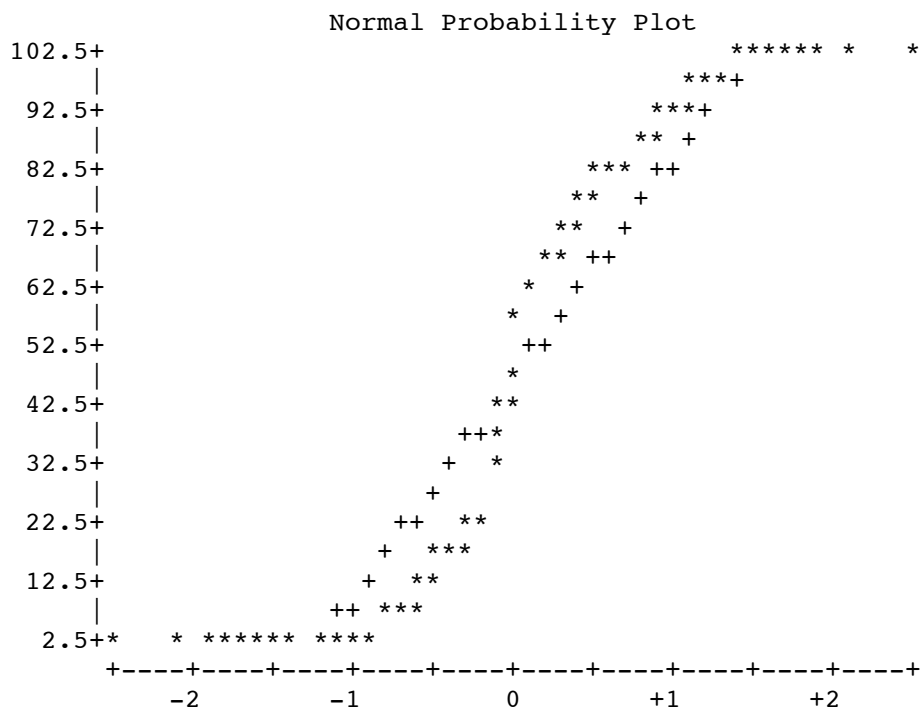
Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
0	61	100	66
0	59	100	71
0	44	100	85

Stem	Leaf	#	Boxplot
10	000000000	9	
9	6666	4	
9	22222	5	
8	88888	5	
8	00000444	8	
7	666	3	
7	2222	4	
6	88888	5	
6	00044	5	
5	68	2	
5			
4	8	1	
4	004	3	
3	6	1	
3	2	1	
2			
2	00014	5	
1	6666669	7	
1	2223	4	
0	8888888	7	
0	000000004444444444444444	21	

-----+-----+-----+-----+-----+
 Multiply Stem.Leaf by 10**+1

The UNIVARIATE Procedure
 Variable: ritewing (Degree of ideological conservatism (0-100))



The UNIVARIATE Procedure
 Variable: cpercent (% vote Clinton received in 1996, home state)

Moments

N	100	Sum Weights	100
Mean	47.26	Sum Observations	4726
Std Deviation	6.7264291	Variance	45.2448485
Skewness	-0.2317077	Kurtosis	-0.1940447
Uncorrected SS	227830	Corrected SS	4479.24
Coeff Variation	14.2328166	Std Error Mean	0.67264291

Basic Statistical Measures

Location		Variability	
Mean	47.26000	Std Deviation	6.72643
Median	48.00000	Variance	45.24485
Mode	44.00000	Range	29.00000
		Interquartile Range	8.00000

Tests for Location: Mu0=0

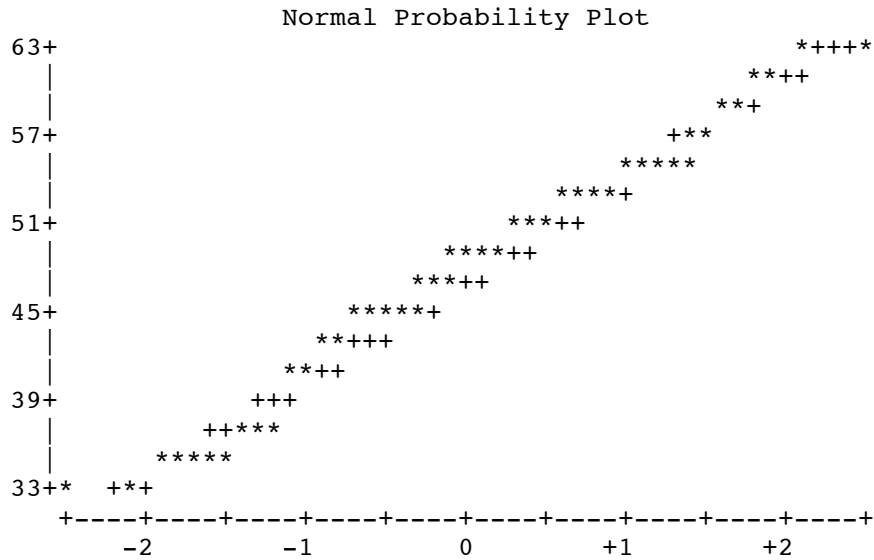
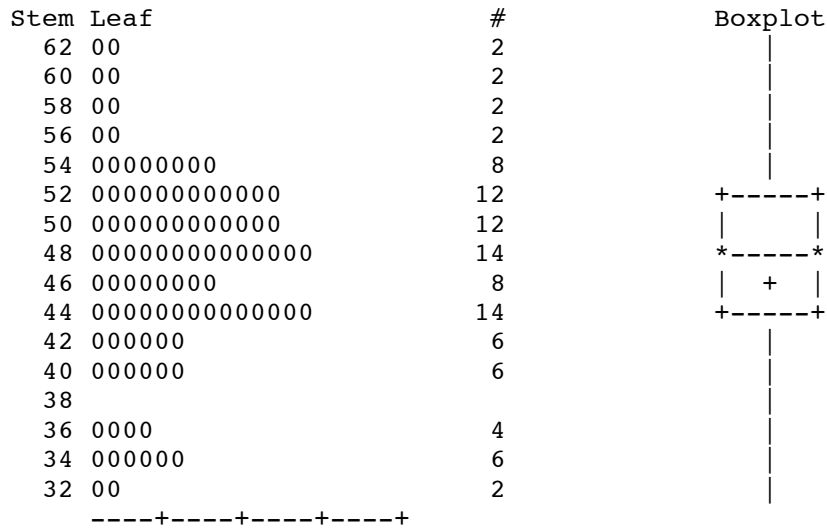
Test	-Statistic-	-----p Value-----
Student's t	t 70.26016	Pr > t <.0001
Sign	M 50	Pr >= M <.0001
Signed Rank	S 2525	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	62.0
99%	62.0
95%	59.0
90%	54.0
75% Q3	52.0
50% Median	48.0
25% Q1	44.0
10%	36.5
5%	34.0
1%	33.0
0% Min	33.0

Extreme Observations

----Lowest----		----Highest----	
Value	Obs	Value	Obs
33	88	59	64
33	87	60	77
34	24	60	78
34	23	62	41
34	4	62	42



Clinton Impeachment Votes in the U. S, Senate (1999) 6
 16:57 Saturday, February 24, 2007

The FREQ Procedure

state	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AK	2	2.00	2	2.00
AL	2	2.00	4	4.00
AR	2	2.00	6	6.00
AZ	2	2.00	8	8.00
CA	2	2.00	10	10.00

CO	2	2.00	12	12.00
CT	2	2.00	14	14.00
DE	2	2.00	16	16.00
FL	2	2.00	18	18.00
GA	2	2.00	20	20.00
HI	2	2.00	22	22.00
IA	2	2.00	24	24.00
ID	2	2.00	26	26.00
IL	2	2.00	28	28.00
IN	2	2.00	30	30.00
KS	2	2.00	32	32.00
KY	2	2.00	34	34.00
LA	2	2.00	36	36.00
MA	2	2.00	38	38.00
MD	2	2.00	40	40.00
ME	2	2.00	42	42.00
MI	2	2.00	44	44.00
MN	2	2.00	46	46.00
MO	2	2.00	48	48.00
MS	2	2.00	50	50.00
MT	2	2.00	52	52.00
NC	2	2.00	54	54.00
ND	2	2.00	56	56.00
NE	2	2.00	58	58.00
NH	2	2.00	60	60.00
NJ	2	2.00	62	62.00
NM	2	2.00	64	64.00
NV	2	2.00	66	66.00
NY	2	2.00	68	68.00
OH	2	2.00	70	70.00
OK	2	2.00	72	72.00
OR	2	2.00	74	74.00
PA	2	2.00	76	76.00
RI	2	2.00	78	78.00
SC	2	2.00	80	80.00
SD	2	2.00	82	82.00
TN	2	2.00	84	84.00
TX	2	2.00	86	86.00
UT	2	2.00	88	88.00
VA	2	2.00	90	90.00
VT	2	2.00	92	92.00
WA	2	2.00	94	94.00
WI	2	2.00	96	96.00
WV	2	2.00	98	98.00
WY	2	2.00	100	100.00

Clinton Impeachment Votes in the U. S, Senate (1999) 7
16:57 Saturday, February 24, 2007

The FREQ Procedure

Vote on Perjury

perjury	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Not Guilty	55	55.00	55	55.00
Guilty	45	45.00	100	100.00

Vote on Obstruction of Justice

obstr	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Not Guilty	50	50.00	50	50.00
Guilty	50	50.00	100	100.00

Number of votes for guilt (0-2)

nvguilt	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	50	50.00	50	50.00
1	5	5.00	55	55.00
2	45	45.00	100	100.00

party	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Democrat	45	45.00	45	45.00
Republican	55	55.00	100	100.00

Year must run for re-election

nextelec	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2000	33	33.00	33	33.00
2002	33	33.00	66	66.00
2004	34	34.00	100	100.00

First-term senator?

firstterm	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	62	62.00	62	62.00
Yes	38	38.00	100	100.00

Table of next0 by nextelec

next0	nextelec(Year must run for re-election)			
Frequency	2000	2002	2004	Total
0	0	33	34	67
1	33	0	0	33
Total	33	33	34	100

Table of next2 by nextelec

next2	nextelec(Year must run for re-election)			
Frequency	2000	2002	2004	Total
0	33	0	34	67
1	0	33	0	33
Total	33	33	34	100

Table of next4 by nextelec

next4	nextelec(Year must run for re-election)			
Frequency	2000	2002	2004	Total
0	33	33	0	66
1	0	0	34	34
Total	33	33	34	100

Table of party by perjury

party	perjury(Vote on Perjury)		Total
	Not Guilty	Guilty	
Democrat	45 100.00	0 0.00	45
Republican	10 18.18	45 81.82	55
Total	55	45	100

Clinton Impeachment Votes in the U. S, Senate (1999) 9
 16:57 Saturday, February 24, 2007

The FREQ Procedure

Statistics for Table of party by perjury

Statistic	DF	Value	Prob
Chi-Square	1	66.9421	<.0001
Likelihood Ratio Chi-Square	1	85.4724	<.0001
Continuity Adj. Chi-Square	1	63.6772	<.0001
Mantel-Haenszel Chi-Square	1	66.2727	<.0001
Phi Coefficient		0.8182	
Contingency Coefficient		0.6332	
Cramer's V		0.8182	

Fisher's Exact Test

Cell (1,1) Frequency (F)	45
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	4.760E-19
Table Probability (P)	1.110E-16
Two-sided Pr <= P	4.821E-19

Sample Size = 100

Table of nextelec by perjury

```
nextelec(Year must run for re-election)
perjury(Vote on Perjury)
```

Frequency Row Pct	Not Guilty	Guilty	Total
2000	18 54.55	15 45.45	33
2002	17 51.52	16 48.48	33
2004	20 58.82	14 41.18	34
Total	55	45	100

Clinton Impeachment Votes in the U. S, Senate (1999) 10
 16:57 Saturday, February 24, 2007

The FREQ Procedure

Statistics for Table of nextelec by perjury

Statistic	DF	Value	Prob
Chi-Square	2	0.3655	0.8330
Likelihood Ratio Chi-Square	2	0.3661	0.8327
Mantel-Haenszel Chi-Square	1	0.1255	0.7231
Phi Coefficient		0.0605	
Contingency Coefficient		0.0603	
Cramer's V		0.0605	

Sample Size = 100

Table of firsterm by perjury

```

firsterm(First-term senator?)
perjury(Vote on Perjury)
Frequency|
Row Pct  |Not Guil|Guilty  | Total
          |ty      |        |
-----+-----+-----+
No       |      39|      23|      62
          | 62.90 | 37.10 |
-----+-----+-----+
Yes      |      16|      22|      38
          | 42.11 | 57.89 |
-----+-----+-----+
Total    |      55|      45|     100
    
```

Clinton Impeachment Votes in the U. S, Senate (1999) 11
 16:57 Saturday, February 24, 2007

The FREQ Procedure

Statistics for Table of firsterm by perjury

Statistic	DF	Value	Prob
Chi-Square	1	4.1176	0.0424
Likelihood Ratio Chi-Square	1	4.1258	0.0422
Continuity Adj. Chi-Square	1	3.3201	0.0684
Mantel-Haenszel Chi-Square	1	4.0764	0.0435
Phi Coefficient		0.2029	
Contingency Coefficient		0.1989	
Cramer's V		0.2029	

Fisher's Exact Test

Cell (1,1) Frequency (F)	39
Left-sided Pr <= F	0.9874
Right-sided Pr >= F	0.0342
Table Probability (P)	0.0216
Two-sided Pr <= P	0.0619

Sample Size = 100

Clinton Impeachment Votes in the U. S, Senate (1999) 12
Logistic regression on perjury vote
16:57 Saturday, February 24, 2007

The LOGISTIC Procedure

Model Information

Data Set	WORK.BILL	
Response Variable	perjury	Vote on Perjury
Number of Response Levels	2	
Number of Observations	100	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	perjury	Total Frequency
1	Guilty	45
2	Not Guilty	55

Probability modeled is perjury='Guilty'.

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	139.628	47.162
SC	142.233	62.793
-2 Log L	137.628	35.162

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	102.4657	5	<.0001
Score	75.7079	5	<.0001
Wald	21.0387	5	0.0008

Logistic regression on perjury vote

16:57 Saturday, February 24, 2007

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.1127	3.9359	0.6254	0.4290
ritewing	1	0.1073	0.0241	19.7476	<.0001
cpercent	1	-0.0671	0.0774	0.7533	0.3854
next0	1	0.6387	1.1484	0.3093	0.5781
next2	1	-0.8046	1.0946	0.5403	0.4623
firstterm	1	0.1713	0.9203	0.0346	0.8524

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
ritewing	1.113	1.062	1.167
cpercent	0.935	0.804	1.088
next0	1.894	0.199	17.985
next2	0.447	0.052	3.822
firstterm	1.187	0.195	7.207

Association of Predicted Probabilities and Observed Responses

Percent Concordant	97.4	Somers' D	0.949
Percent Discordant	2.5	Gamma	0.950
Percent Tied	0.1	Tau-a	0.474
Pairs	2475	c	0.974

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
nextelec	1.4733	2	0.4787
others	1.7315	4	0.7850
allvars	21.0387	5	0.0008

The LOGISTIC Procedure

Model Information

Data Set	WORK.BILL	
Response Variable	perjury	Vote on Perjury
Number of Response Levels	2	
Number of Observations	100	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	perjury	Total Frequency
1	Guilty	45
2	Not Guilty	55

Probability modeled is perjury='Guilty'.

Class Level Information

Class	Value	Design Variables	
		1	2
nextelec	2000	1	0
	2002	0	1
	2004	0	0

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	139.628	47.162
SC	142.233	62.793
-2 Log L	137.628	35.162

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	102.4657	5	<.0001
Score	75.7079	5	<.0001
Wald	21.0387	5	0.0008

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
ritewing	1	19.7476	<.0001
cpercent	1	0.7533	0.3854
nextelec	2	1.4733	0.4787
firstterm	1	0.0346	0.8524

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.1127	3.9359	0.6254	0.4290
ritewing	1	0.1073	0.0241	19.7476	<.0001
cpercent	1	-0.0671	0.0774	0.7533	0.3854
nextelec 2000	1	0.6387	1.1484	0.3093	0.5781
nextelec 2002	1	-0.8046	1.0946	0.5403	0.4623
firstterm	1	0.1713	0.9203	0.0346	0.8524

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
ritewing	1.113	1.062	1.167
cpercent	0.935	0.804	1.088
nextelec 2000 vs 2004	1.894	0.199	17.985
nextelec 2002 vs 2004	0.447	0.052	3.822
firstterm	1.187	0.195	7.207

Association of Predicted Probabilities and Observed Responses

Percent Concordant	97.4	Somers' D	0.949
Percent Discordant	2.5	Gamma	0.950
Percent Tied	0.1	Tau-a	0.474
Pairs	2475	c	0.974

Clinton Impeachment Votes in the U. S, Senate (1999) 16
Compare LR test of 0.3661 p = 0.8327 from proc freq
16:57 Saturday, February 24, 2007

The LOGISTIC Procedure

Model Information

Data Set	WORK.BILL	
Response Variable	perjury	Vote on Perjury
Number of Response Levels	2	
Number of Observations	100	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	perjury	Total Frequency
1	Guilty	45
2	Not Guilty	55

Probability modeled is perjury='Guilty'.

Class Level Information

Class	Value	Design Variables	
		1	2
nextelec	2000	1	0
	2002	0	1
	2004	-1	-1

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	139.628	143.262
SC	142.233	151.077
-2 Log L	137.628	137.262

Clinton Impeachment Votes in the U. S, Senate (1999)
 Compare LR test of 0.3661 p = 0.8327 from proc freq

17

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.3661	2	0.8327
Score	0.3655	2	0.8330
Wald	0.3648	2	0.8333

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
nextelec	2	0.3648	0.8333

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.1999	0.2014	0.9851	0.3209
nextelec 2000	1	0.0176	0.2851	0.0038	0.9509
nextelec 2002	1	0.1392	0.2846	0.2394	0.6246

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
nextelec 2000 vs 2004	1.190	0.452 3.132
nextelec 2002 vs 2004	1.345	0.512 3.531

Association of Predicted Probabilities and Observed Responses

Percent Concordant	36.7	Somers' D	0.066
Percent Discordant	30.1	Gamma	0.099
Percent Tied	33.2	Tau-a	0.033
Pairs	2475	c	0.533