

Homework assignment 8.

```

1 *****;
2 *** Speed of Kentucky Derby winner and track conditions ***;
3 *** for the years 1896 to 2000 ***;
4 *****;
5
6 dm'log;clear;output;clear';
7 options nodate nocenter nonumber ps=512 ls=99 nolabel;
8 ODS HTML style=minimal rs=none
9 ! BODY='C:\Geaghan\Current\EXST3201\Fall2005\SAS\KentuckyDerby01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\KentuckyDerby01.html
10 Title1 'Chapter 9 : Speed and track conditions for Kentucky Derby';
11 filename input1 'C:\Geaghan\Current\EXST3201\Datasets\ASCII\ex0920.csv';
12
13 data Derby; length Winner $ 16; infile input1 misover DSD dlm="," firstobs=2;
14 input year winner $ condition $ speed;
15 label Winner = 'Name of the Winner'
16 Speed = 'Speed (mph)'
17 Year = 'Year (1896-2000)'
18 Condition = 'Track condition (slow, good, fast)';
19 datalines;
NOTE: The infile INPUT1 is:
File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\ex0920.csv,
RECFM=V,LRECL=256
NOTE: 105 records were read from the infile INPUT1.
The minimum record length was 23.
The maximum record length was 48.
NOTE: The data set WORK.DERBY has 105 observations and 4 variables.
NOTE: DATA statement used (Total process time):
real time 0.01 seconds
cpu time 0.02 seconds
20 run;
21 PROC PRINT DATA=Derby; TITLE2 'Raw data Listing'; RUN;
NOTE: There were 105 observations read from the data set WORK.DERBY.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
real time 0.27 seconds
cpu time 0.02 seconds

```

Chapter 9 : Speed and track conditions for Kentucky Derby
Raw data Listing

Obs	Winner	year	condition	speed					
1	Ben Brush	1896	good	51.6634	35	Gallant Fox	1930	good	51.7241
2	Typhoon II	1897	slow	49.8113	36	Twenty Grand	1931	fast	54.1872
3	Plaudit	1898	good	51.1628	37	Burgoo King	1932	fast	52.7157
4	Manuel	1899	fast	50.0000	38	Brokers Tip	1933	good	52.0505
5	Lieut. Gibson	1900	fast	52.2772	39	Cavalcade	1934	fast	53.2258
6	His Eminence	1901	fast	51.6634	40	Omaha	1935	good	52.8000
7	Alan-a-Dale	1902	fast	51.2621	41	Bold Venture	1936	fast	53.4846
8	Judge Himes	1903	fast	51.1628	42	War Admiral	1937	fast	53.5714
9	Elwood	1904	fast	51.3619	43	Lawrin	1938	fast	52.8846
10	Agile	1905	slow	50.4780	44	Johnstown	1939	fast	53.4846
11	Sir Huon	1906	fast	51.2422	45	Gallahadion	1940	fast	52.8000
12	Pink Star	1907	slow	49.7738	46	Whirlaway	1941	fast	54.3657
13	Stone Street	1908	slow	48.8166	47	Shut Out	1942	fast	53.0547
14	Wintergreen	1909	slow	51.4821	48	Count Fleet	1943	fast	53.2258
15	Donau	1910	fast	52.2152	49	Pensive	1944	good	53.1401
16	Meridian	1911	fast	52.8000	50	Hoop Jr.	1945	slow	51.9685
17	Worth	1912	slow	51.0046	51	Assault	1946	slow	52.1327
18	Donerail	1913	fast	52.8846	52	Jet Pilot	1947	slow	52.0505
19	Old Rosebud	1914	fast	53.4846	53	Citation	1948	slow	52.6316
20	Regret	1915	fast	52.6316	54	Ponder	1949	fast	53.1401
21	George Smith	1916	fast	53.2258	55	Middleground	1950	fast	54.2763
22	*Omar Khayyam	1917	fast	52.9695	56	Count Turf	1951	fast	53.8336
23	Exterminator	1918	slow	50.4587	57	Hill Gail	1952	fast	54.2763
24	Sir Barton	1919	slow	50.8475	58	Dark Star	1953	fast	54.0984
25	Paul Jones	1920	slow	51.1628	59	Determine	1954	fast	53.6585
26	Behave Yourself	1921	fast	53.1401	60	Swaps	1955	fast	54.1872
27	Morvich	1922	fast	52.9695	61	Needles	1956	fast	53.4846
28	Zev	1923	fast	52.6316	62	Iron Liege	1957	fast	54.0098
29	Black Gold	1924	fast	52.7157	63	Tim Tam	1958	slow	52.8000
30	Flying Ebony	1925	slow	51.7241	64	*Tomy Lee	1959	fast	54.0098
31	Bubbling Over	1926	fast	53.3118	65	Venetian Way	1960	good	53.9216
32	Whiskery	1927	slow	52.3810	66	Carry Back	1961	good	53.2258
33	Reigh Count	1928	slow	50.6135	67	Decidedly	1962	fast	54.8173
34	Clyde Van Dusen	1929	slow	50.4587	68	Chateaugay	1963	fast	54.5455
					69	Northern Dancer	1964	fast	55.0000

70	Lucky Debonair	1965	fast	54.4554	88	Sunny's Halo	1983	fast	53.9216
71	Kauai King	1966	fast	54.0984	89	Swale	1984	fast	53.9216
72	Proud Clarion	1967	fast	54.7264	90	Spend a Buck	1985	fast	54.9085
73	Forward Pass**	1968	fast	54.0098	91	Ferdinand	1986	fast	53.7459
74	Majestic Prince	1969	fast	54.1872	92	Alysheba	1987	fast	53.4846
75	Dust Commander	1970	good	53.4846	93	Winning Colors	1988	fast	54.0098
76	Canonero II	1971	fast	53.5714	94	Sunday Silence	1989	slow	52.8000
77	Riva Ridge	1972	fast	54.1872	95	Unbridled	1990	good	54.0984
78	Secretariat	1973	fast	55.2764	96	Strike the Gold	1991	fast	53.6585
79	Cannonade	1974	fast	53.2258	97	Lil E. Tee	1992	fast	53.6585
80	Foolish Pleasure	1975	fast	54.0984	98	Sea Hero	1993	fast	53.9216
81	Bold Forbes	1976	fast	54.2763	99	Go for Gin	1994	slow	53.3981
82	Seattle Slew	1977	fast	54.0098	100	Thunder Gulch	1995	fast	54.4554
83	Affirmed	1978	fast	54.4554	101	Grindstone	1996	fast	54.5455
84	Spectacular Bid	1979	fast	53.9216	102	Silver Charm	1997	fast	53.9216
85	Genuine Risk	1980	fast	54.0984	103	Real Quiet	1998	fast	54.0098
86	Pleasant Colony	1981	fast	54.0984	104	Charismatic	1999	fast	53.6585
87	Gato Del Sol	1982	fast	53.9216	105	Fusaichi Pegasus	2000	fast	54.4900

```

22
23     options ps=65 ls=132;
24     proc plot data=Derby; TITLE2 'Plot of the raw data';
25         plot Speed * Condition = Winner;
26         plot Speed * Year = Winner;
27     RUN;
28     !     OPTIONS PS=256;

```

NOTE: There were 105 observations read from the data set WORK.DERBY.

NOTE: The PROCEDURE PLOT printed pages 2-3.

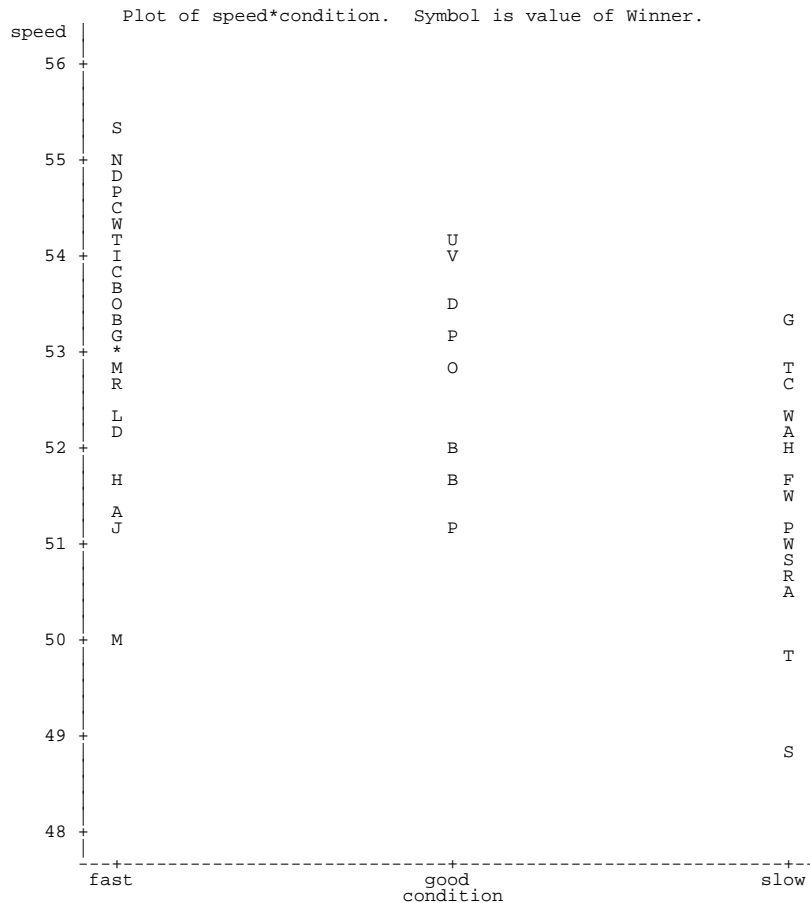
NOTE: PROCEDURE PLOT used (Total process time):

```

real time      0.33 seconds
cpu time       0.03 seconds

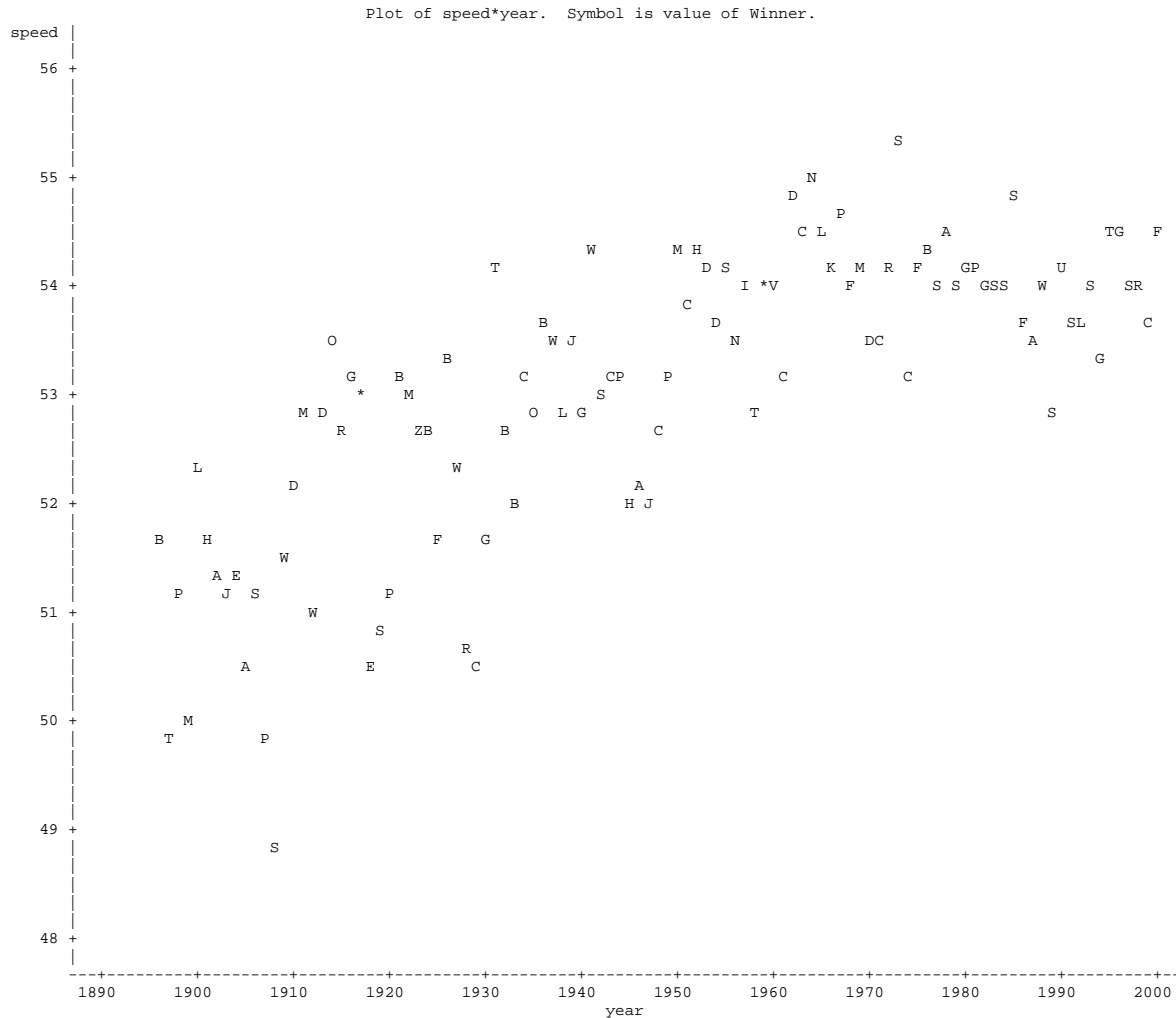
```

Chapter 9 : Speed and track conditions for Kentucky Derby
Plot of the raw data



NOTE: 60 obs hidden.

Chapter 9 : Speed and track conditions for Kentucky Derby
 Plot of the raw data



```

29      PROC GLM DATA=Derby; class condition;
30          TITLE2 'Analysis of Covariance with GLM - linear';
31          MODEL Speed = Year | Condition;
32          output out=next1 r=resid p=YHat;
33      RUN;
33      !      OPTIONS PS=45;
34
35          options ps=52 ls=132;
NOTE: The data set WORK.NEXT1 has 105 observations and 6 variables.
NOTE: The PROCEDURE GLM printed pages 4-5.
NOTE: PROCEDURE GLM used (Total process time):
      real time          0.19 seconds
      cpu time           0.08 seconds
    
```

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - linear

The GLM Procedure

Class Level Information

Class	Levels	Values
condition	3	fast good slow

Number of Observations Read 105
 Number of Observations Used 105

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - linear

The GLM Procedure

Dependent Variable: speed

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	143.7858750	28.7571750	70.04	<.0001
Error	99	40.6500458	0.4106065		
Corrected Total	104	184.4359208			

R-Square Coeff Var Root MSE speed Mean
 0.779598 1.208063 0.640786 53.04243

Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	1	103.5684162	103.5684162	252.23	<.0001
condition	2	38.1346829	19.0673414	46.44	<.0001
year*condition	2	2.0827760	1.0413880	2.54	0.0843

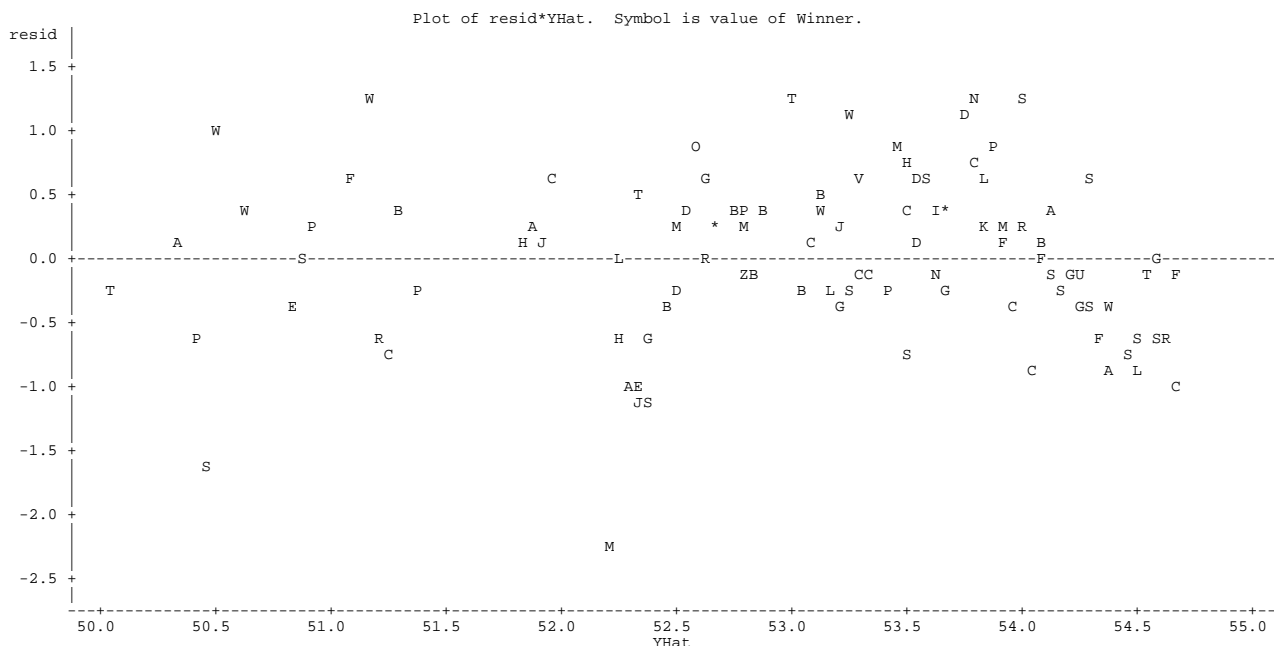
Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	1	40.61645387	40.61645387	98.92	<.0001
condition	2	2.34155232	1.17077616	2.85	0.0625
year*condition	2	2.08277597	1.04138799	2.54	0.0843

```

36      proc plot data=next1;  TITLE3 'Residual analysis';
37          plot resid * YHat = Winner / vref=0;
38          plot resid * Year = Winner / vref=0;
39          plot resid * Condition = Winner / vref=0;
40      RUN;
40      !      OPTIONS PS=256;
    
```

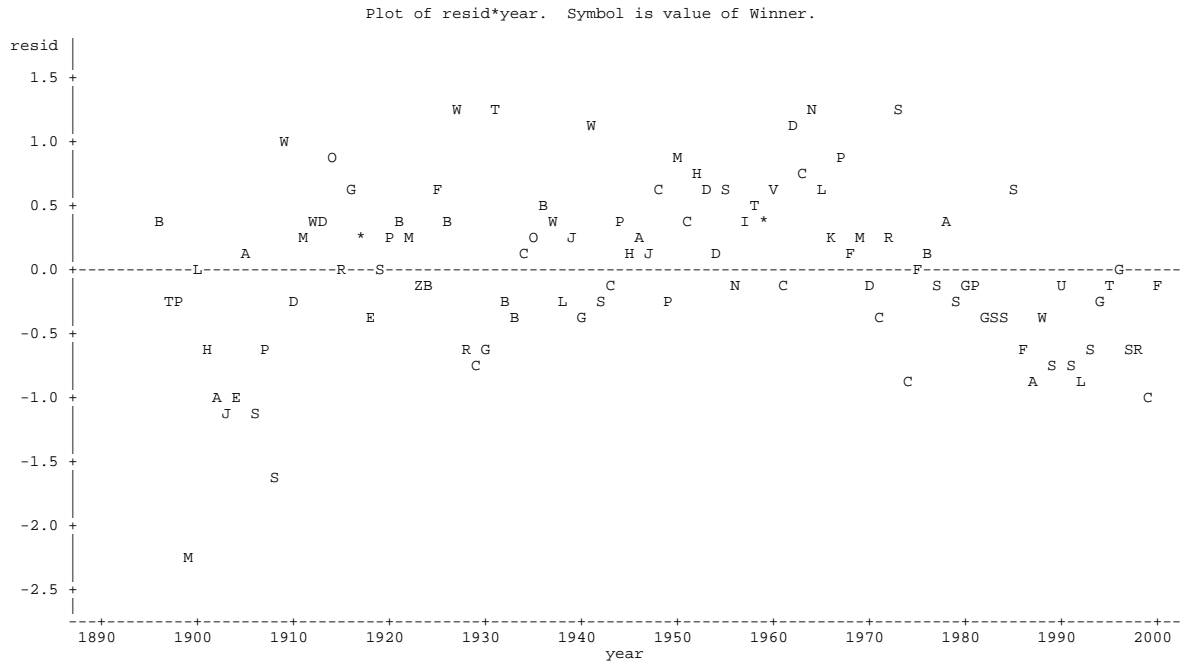
NOTE: There were 105 observations read from the data set WORK.NEXT1.
 NOTE: The PROCEDURE PLOT printed pages 6-8.
 NOTE: PROCEDURE PLOT used (Total process time):
 real time 0.06 seconds
 cpu time 0.01 seconds

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - linear
 Residual analysis

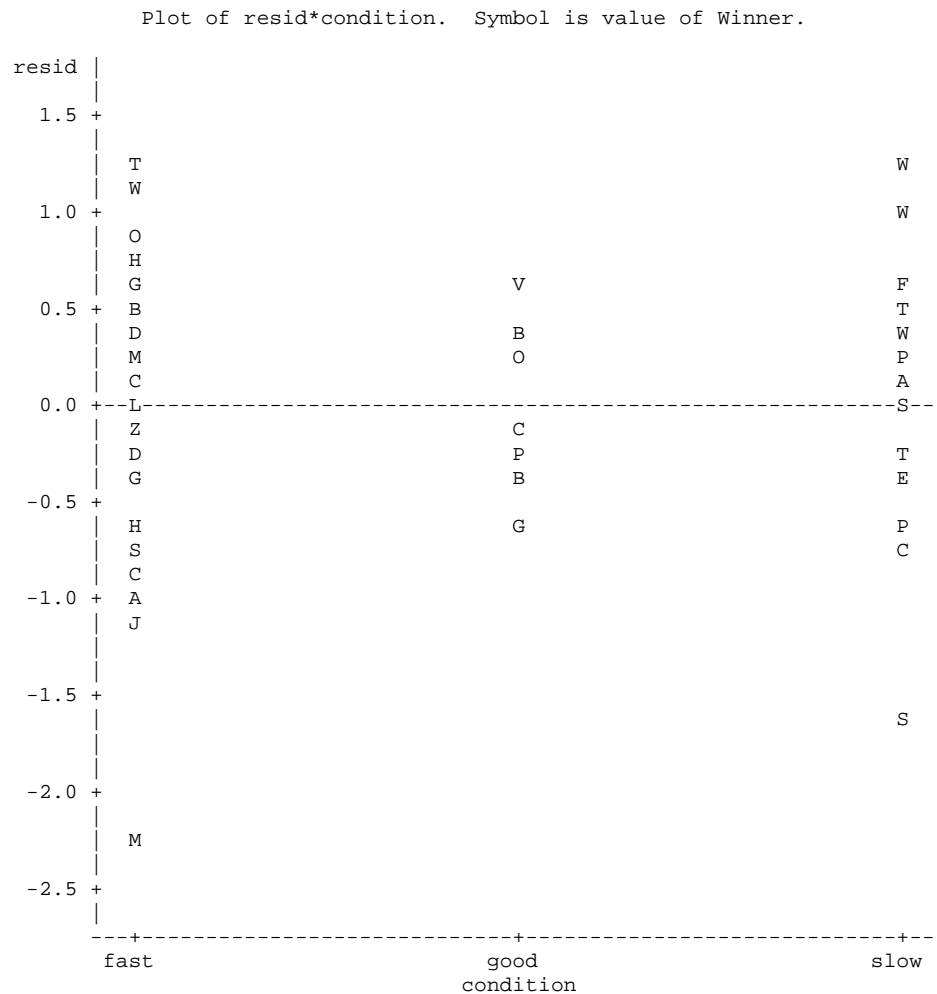


NOTE: 4 obs hidden.

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - linear
 Residual analysis



Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - linear
 Residual analysis



NOTE: 66 obs hidden.

```

41          PROC UNIVARIATE DATA=NEXT1 NORMAL PLOT; VAR resid; RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 9.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time          0.06 seconds
      cpu time           0.02 seconds
42

```

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - linear
 Residual analysis

The UNIVARIATE Procedure
 Variable: resid

Moments

N	105	Sum Weights	105
Mean	0	Sum Observations	0
Std Deviation	0.62519263	Variance	0.39086583
Skewness	-0.4247796	Kurtosis	0.80774296
Uncorrected SS	40.6500458	Corrected SS	40.6500458
Coeff Variation	.	Std Error Mean	0.06101255

Basic Statistical Measures

Location		Variability	
Mean	0.000000	Std Deviation	0.62519
Median	0.022672	Variance	0.39087
Mode	.	Range	3.47678
		Interquartile Range	0.74486

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 0	Pr > t	1.0000
Sign	M 0.5	Pr >= M	1.0000
Signed Rank	S 91.5	Pr >= S	0.7715

Tests for Normality

Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.982403	Pr < W	0.1784
Kolmogorov-Smirnov	D 0.05245	Pr > D	>0.1500
Cramer-von Mises	W-Sq 0.039921	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq 0.276695	Pr > A-Sq	>0.2500

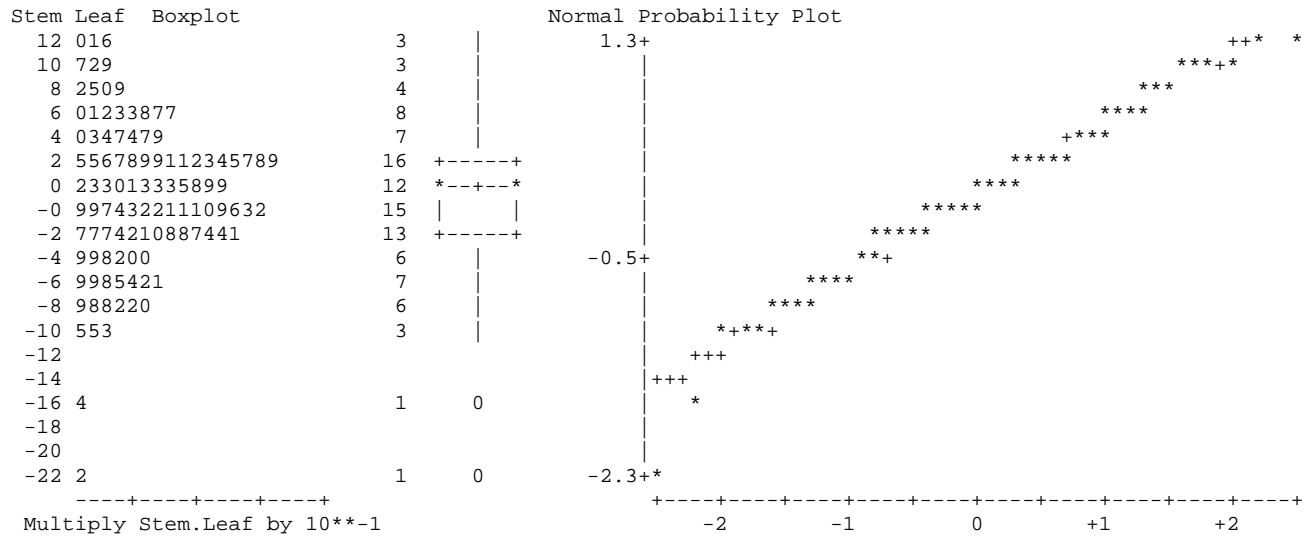
Quantiles (Definition 5)

Quantile	Estimate
100% Max	1.2569786
99%	1.2115248
95%	1.0653810
90%	0.7692421
75% Q3	0.3795203
50% Median	0.0226722
25% Q1	-0.3653388
10%	-0.7986058
5%	-0.9931567
1%	-1.6418310
0% Min	-2.2198062

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-2.21981	4	1.12454	46
-1.64183	13	1.18918	36

```
-1.15429      8      1.19947      69
-1.14780     11      1.21152      32
-1.03063      7      1.25698      78
```



```
43 PROC GLM DATA=Derby; class condition;
44 TITLE2 'Analysis of Covariance with GLM - quadratic';
45 TITLE3 '*** ==> Best model <=== ***';
46 MODEL Speed = Year | Condition year*year year*year*condition / solution;
47 output out=next2 r=resid p=YHat;
48 RUN;
48 ! OPTIONS PS=45;
49
50 options ps=52 ls=132;
NOTE: The data set WORK.NEXT2 has 105 observations and 6 variables.
NOTE: The PROCEDURE GLM printed pages 10-11.
NOTE: PROCEDURE GLM used (Total process time):
      real time          0.11 seconds
      cpu time           0.06 seconds
```

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - quadratic
 *** ==> Best model <=== ***

The GLM Procedure

Class Level Information

Class	Levels	Values
condition	3	fast good slow

Number of Observations Read 105
 Number of Observations Used 105

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - quadratic
 *** ==> Best model <=== ***

The GLM Procedure

Dependent Variable: speed

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	157.1190601	19.6398825	69.02	<.0001
Error	96	27.3168607	0.2845506		
Corrected Total	104	184.4359208			

R-Square	Coeff Var	Root MSE	speed Mean
0.851890	1.005672	0.533433	53.04243

Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	1	103.5684162	103.5684162	363.97	<.0001
condition	2	38.1346829	19.0673414	67.01	<.0001
year*condition	2	2.0827760	1.0413880	3.66	0.0294
year*year	1	11.4570011	11.4570011	40.26	<.0001
year*year*condition	2	1.8761840	0.9380920	3.30	0.0412

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	1	2.93641115	2.93641115	10.32	0.0018
condition	2	1.86621254	0.93310627	3.28	0.0419
year*condition	2	1.87218831	0.93609416	3.29	0.0415
year*year	1	2.75466449	2.75466449	9.68	0.0025
year*year*condition	2	1.87618404	0.93809202	3.30	0.0412

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-1176.386286 B	589.5339348	-2.00	0.0488
year	1.225935 B	0.6063550	2.02	0.0460
condition fast	-738.162017 B	658.2706525	-1.12	0.2649
condition good	1181.908988 B	934.0279139	1.27	0.2088
condition slow	0.000000 B	.	.	.
year*condition fast	0.769054 B	0.6767239	1.14	0.2586
year*condition good	-1.208707 B	0.9621921	-1.26	0.2121
year*condition slow	0.000000 B	.	.	.
year*year	-0.000306 B	0.0001559	-1.96	0.0529
year*year*condition fast	-0.000200 B	0.0001739	-1.15	0.2534
year*year*condition good	0.000309 B	0.0002478	1.25	0.2151
year*year*condition slow	0.000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

```

51      proc plot data=next2;  TITLE3 'Residual analysis';
52          plot resid * YHat = Winner / vref=0;
53          plot resid * Year = Winner / vref=0;
54          plot resid * Condition = Winner / vref=0;
55      RUN;
56      !      OPTIONS PS=256;

```

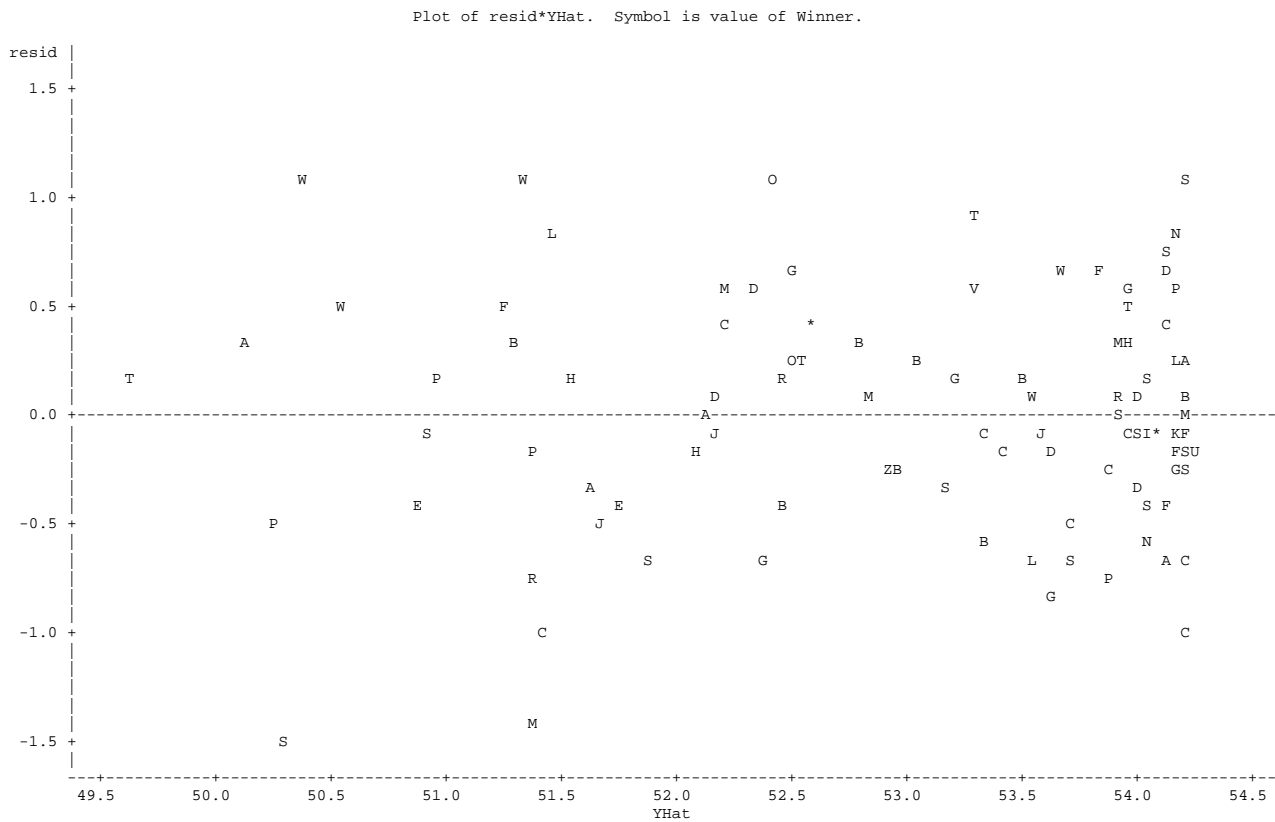
NOTE: There were 105 observations read from the data set WORK.NEXT2.

NOTE: The PROCEDURE PLOT printed pages 12-14.

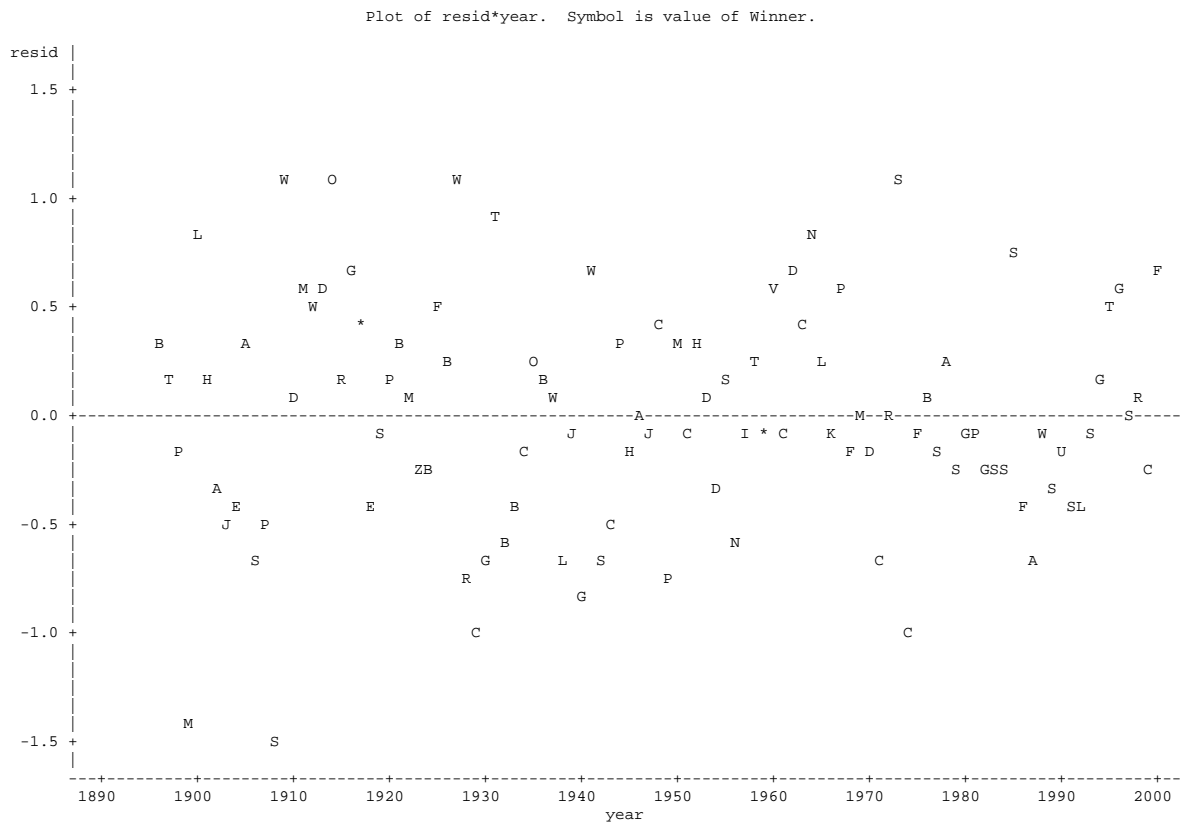
NOTE: PROCEDURE PLOT used (Total process time):

real time	0.06 seconds
cpu time	0.01 seconds

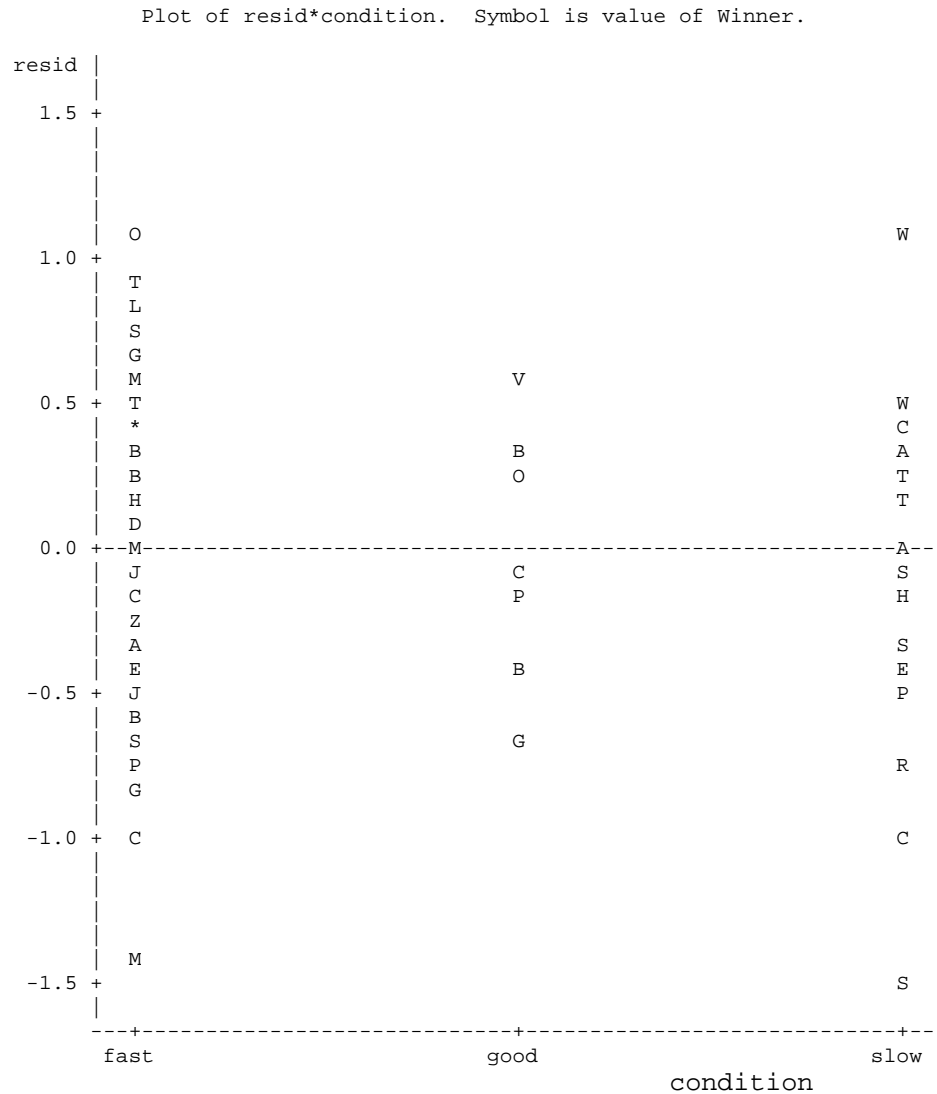
Chapter 9 : Speed and track conditions for Kentucky Derby
Analysis of Covariance with GLM - quadratic
Residual analysis



Chapter 9 : Speed and track conditions for Kentucky Derby
Analysis of Covariance with GLM - quadratic
Residual analysis



Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - quadratic
 Residual analysis



NOTE: 58 obs hidden.

```
56          PROC UNIVARIATE DATA=NEXT2 NORMAL PLOT; VAR resid; RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 15.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time          0.10 seconds
      cpu time           0.04 seconds
```

Chapter 9 : Speed and track conditions for Kentucky Derby
 Analysis of Covariance with GLM - quadratic
 Residual analysis

The UNIVARIATE Procedure
 Variable: resid

Moments			
N	105	Sum Weights	105
Mean	0	Sum Observations	0
Std Deviation	0.51250573	Variance	0.26266212
Skewness	-0.1157465	Kurtosis	0.21748374
Uncorrected SS	27.3168607	Corrected SS	27.3168607
Coeff Variation	.	Std Error Mean	0.05001544

Basic Statistical Measures

Location		Variability	
Mean	0.00000	Std Deviation	0.51251
Median	-0.07481	Variance	0.26266
Mode	.	Range	2.60589
		Interquartile Range	0.68680

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 0	Pr > t	1.0000
Sign	M -5.5	Pr >= M	0.3291
Signed Rank	S -5.5	Pr >= S	0.9861

Tests for Normality

Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.990125	Pr < W	0.6419
Kolmogorov-Smirnov	D 0.062791	Pr > D	>0.1500
Cramer-von Mises	W-Sq 0.037907	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq 0.236589	Pr > A-Sq	>0.2500

Quantiles (Definition 5)

Quantile	Estimate
100% Max	1.1207308
99%	1.0825225
95%	0.8444068
90%	0.6833044
75% Q3	0.3437084
50% Median	-0.0748126
25% Q1	-0.3430915
10%	-0.6332637
5%	-0.7640386
1%	-1.3832697
0% Min	-1.4851635

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-1.485163	13	0.905388	36
-1.383270	4	1.051401	32
-0.977254	79	1.073519	78
-0.966192	34	1.082523	19
-0.829274	45	1.120731	14

Stem Leaf Boxplot

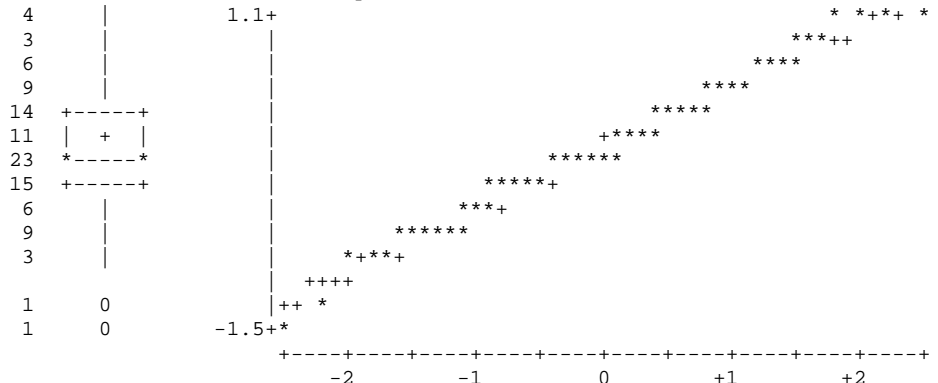
```

10 5782
 8 241
 6 248017
 4 027894589
 2 016679991446679
 0 56811236779
-0 98853321110998888751100
-2 998855477544320
-4 610710
-6 657443331
-8 873
-10
-12 8
-14 9

```

Multiply Stem.Leaf by 10**-1

Normal Probability Plot



```

58      data derby; set derby;
59          if condition eq 'fast' then do; Cond1 = 1; Cond2 = 0; YC1 = year*cond1; YC2 = year*cond2;
60              Year2 = year*year; YC12 = year*year*cond1; YC22 = year*year*cond2; end;
61          if condition eq 'good' then do; Cond1 = 0; Cond2 = 1; YC1 = year*cond1; YC2 = year*cond2;
62              Year2 = year*year; YC12 = year*year*cond1; YC22 = year*year*cond2; end;
63          if condition eq 'slow' then do; Cond1 = 0; Cond2 = 0; YC1 = year*cond1; YC2 = year*cond2;
64              Year2 = year*year; YC12 = year*year*cond1; YC22 = year*year*cond2; end;
65      run;

```

NOTE: There were 105 observations read from the data set WORK.DERBY.

NOTE: The data set WORK.DERBY has 105 observations and 11 variables.

NOTE: DATA statement used (Total process time):

```

real time      0.02 seconds
cpu time       0.03 seconds

```

```

66
67      PROC REG DATA=Derby lineprinter;
68          TITLE2 'Analysis of Covariance with REG';
69          MODEL Speed = Year Cond1 Cond2 YC1 YC2 year2 YC12 YC22;
70          output out=next3 r=resid p=YHat;
71      RUN;
72
73      ODS HTML close;
74      quit;

```

NOTE: The data set WORK.NEXT3 has 105 observations and 13 variables.

NOTE: The PROCEDURE REG printed page 16.

NOTE: PROCEDURE REG used (Total process time):

```

real time      0.18 seconds
cpu time       0.05 seconds

```

Chapter 9 : Speed and track conditions for Kentucky Derby Analysis of Covariance with REG

The REG Procedure

Model: MODEL1

Dependent Variable: speed

```

Number of Observations Read      105
Number of Observations Used      105

```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	157.11906	19.63988	69.02	<.0001
Error	96	27.31686	0.28455		
Corrected Total	104	184.43592			

```

Root MSE      0.53343      R-Square      0.8519
Dependent Mean 53.04243      Adj R-Sq      0.8395
Coeff Var      1.00567

```

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-1176.38627	589.53395	-2.00	0.0488
year	1	1.22594	0.60636	2.02	0.0460
Cond1	1	-738.16203	658.27068	-1.12	0.2649
Cond2	1	1181.90897	934.02793	1.27	0.2088
YC1	1	0.76905	0.67672	1.14	0.2586
YC2	1	-1.20871	0.96219	-1.26	0.2121
Year2	1	-0.00030556	0.00015589	-1.96	0.0529
YC12	1	-0.00019983	0.00017389	-1.15	0.2534
YC22	1	0.00030921	0.00024776	1.25	0.2151