

```

1      **EXAMPLE 1*****;
2      *** The 1994 salaries of all American league baseball players ***;
3      *** as reported in USA Today on April 5, 1994. ***;
4      *** Salaries include pro-rated signing bonuses. ***;
5      *****;
6      OPTIONS PS=256 LS=132 NOCENTER NODATE PAGENO=1;
7      dm 'log;clear;output;clear';
8      filename input 'C:\Geaghan\EXST\EXST7015New\Fall2002\SAS\07b-AnovaRBD_baseball.DAT';
9
10     DATA salaries; length team player $ 20; INFILE input MISSOEVER;
11     INPUT Team $ 1-20 Player $ 24-43 Position $ 47-48 read $ 50-51 Salary 60-68;
12     TITLE1 'EXST7015: Salaries of all American league baseball players (1994)';
13     TITLE2 'Salaries in thousands of dollars';
14     DH = 'No '; if read eq 'dh' then DH = 'Yes'; drop read;
15     salary = salary / 1000;
16     lsalary = log(salary);
17     CARDS;
NOTE: The infile INPUT is:
      File Name=C:\Geaghan\EXST\EXST7015New\Fall2002\SAS\07b-AnovaRBD_baseball.DAT,
      RECFM=V,LRECL=256
NOTE: 374 records were read from the infile INPUT.
      The minimum record length was 68.
      The maximum record length was 68.
NOTE: The data set WORK.SALARIES has 374 observations and 6 variables.
NOTE: DATA statement used:
      real time          0.11 seconds
      cpu time           0.09 seconds
17     !          RUN;
18     ;
19     PROC PRINT data=salaries; TITLE3 'RAW DATA LISTING'; RUN;

```

NOTE: There were 374 observations read from the data set WORK.SALARIES.
NOTE: The PROCEDURE PRINT printed pages 1-2.
NOTE: PROCEDURE PRINT used:
real time 0.06 seconds
cpu time 0.06 seconds

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
RAW DATA LISTING

Obs	team	player	Position	Salary	DH	lsalary
1	Baltimore Orioles	Rafael Palmeiro	1b	5406.60	No	8.59538
2	Baltimore Orioles	Cal Ripken	ss	5400.00	No	8.59415
3	Baltimore Orioles	Mike Devereaux	of	3375.00	No	8.12415
4	Baltimore Orioles	Sid Fernandez	p	3333.33	No	8.11173
5	Baltimore Orioles	Brady Anderson	of	3083.33	No	8.03377
6	Baltimore Orioles	Ben McDonald	p	2675.00	No	7.89170
. . .						
373	Toronto Blue Jays	Paul Spoljaric	p	109.00	No	4.69135
374	Toronto Blue Jays	Paul Menhart	p	109.00	No	4.69135

```

21     PROC MIXED DATA=salaries cl covtest; CLASSES team position dh;
22     TITLE3 'Analysis of Variance - PROC MIXED';
23     MODEL salary = position DH / htype=3 DDFM=Satterthwaite
outp=ResidData;
24     random team;
25     lsmeans position / pdiff adjust=tukey;
26     ods output diffs=ppp;
27     ods output lsmeans=mmm;
28     *ods listing exclude diffs;
29     *ods listing exclude lsmeans;
30     run;

```

WARNING: Length of CLASS variable team truncated to 16.
NOTE: Convergence criteria met.
NOTE: The data set WORK.MMM has 8 observations and 7 variables.
NOTE: The data set WORK.PPP has 28 observations and 10 variables.
NOTE: The data set WORK.RESIDDATA has 374 observations and 13 variables.

NOTE: The PROCEDURE MIXED printed page 3.

NOTE: PROCEDURE MIXED used:

```

real time          0.27 seconds
cpu time           0.27 seconds
31      %include 'C:\Geaghan\EXST\EXST7015New\Fall2002\SAS\pdmix800.sas';
647     %pdmix800(ppp,mmm,alpha=.05,sort=yes);

```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Analysis of Variance - PROC MIXED

The Mixed Procedure

Model Information	
Data Set	WORK.SALARIES
Dependent Variable	Salary
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information		
Class	Levels	Values
team	14	Baltimore Oriole Boston Red Sox California Angel Chicago White So Cleveland Indian Detroit Tigers Kansas City Roya Milwaukee Brewer Minnesota Twins New York Yankees Oakland Athletic Seattle Mariners Texas Rangers Toronto Blue Jay
Position	8	1b 2b 3b c if of p ss
DH	2	No Yes

Dimensions	
Covariance Parameters	2
Columns in X	11
Columns in Z	14
Subjects	1
Max Obs Per Subject	374
Observations Used	374
Observations Not Used	0
Total Observations	374

Iteration History				Criterion
Iteration	Evaluations	-2 Res Log Like		
0	1	6346.82951099		
1	2	6346.81578713		0.00000000

Convergence criteria met.

Covariance Parameter Estimates							
Cov Parm	Estimate	Standard Error	Z	Pr > Z	Alpha	Lower	Upper
team	3466.41	30458	0.11	0.4547	0.05	513.45	3.81E125
Residual	1924296	145057	13.27	<.0001	0.05	1668871	2243534

Fit Statistics	
-2 Res Log Likelihood	6346.8
AIC (smaller is better)	6350.8
AICC (smaller is better)	6350.8
BIC (smaller is better)	6352.1

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Position	7	359	2.75	0.0085
DH	1	364	4.45	0.0356

Least Squares Means

Effect	Position	Estimate	Standard Error	DF	t Value	Pr > t
Position	1b	2853.96	423.51	363	6.74	<.0001
Position	2b	2065.92	475.36	361	4.35	<.0001
Position	3b	1921.43	516.99	364	3.72	0.0002
Position	c	1615.79	437.88	364	3.69	0.0003
Position	if	1231.97	500.13	363	2.46	0.0142
Position	of	2296.24	383.25	361	5.99	<.0001
Position	p	1958.94	375.99	356	5.21	<.0001
Position	ss	1751.07	437.93	363	4.00	<.0001

Differences of Least Squares Means

Effect	Position	_Position	Estimate	Standard Error	DF	t Value	Pr > t	Adjustment	Adj P
Position	1b	2b	788.04	474.63	362	1.66	0.0977	Tukey-Kramer	0.7128
Position	1b	3b	932.53	488.80	357	1.91	0.0572	Tukey-Kramer	0.5464
Position	1b	c	1238.18	404.34	355	3.06	0.0024	Tukey-Kramer	0.0481
Position	1b	if	1621.99	471.06	363	3.44	0.0006	Tukey-Kramer	0.0147
Position	1b	of	557.73	351.72	356	1.59	0.1137	Tukey-Kramer	0.7587
Position	1b	p	895.02	336.27	356	2.66	0.0081	Tukey-Kramer	0.1380
Position	1b	ss	1102.89	404.37	356	2.73	0.0067	Tukey-Kramer	0.1178
Position	2b	3b	144.49	517.94	362	0.28	0.7804	Tukey-Kramer	1.0000
Position	2b	c	450.13	439.15	363	1.03	0.3060	Tukey-Kramer	0.9705
Position	2b	if	833.94	501.29	365	1.66	0.0971	Tukey-Kramer	0.7108
Position	2b	of	-230.32	391.74	363	-0.59	0.5569	Tukey-Kramer	0.9990
Position	2b	p	106.98	377.39	364	0.28	0.7770	Tukey-Kramer	1.0000
Position	2b	ss	314.85	439.12	362	0.72	0.4738	Tukey-Kramer	0.9965
Position	3b	c	305.64	446.79	358	0.68	0.4944	Tukey-Kramer	0.9974
Position	3b	if	689.46	508.05	365	1.36	0.1756	Tukey-Kramer	0.8758
Position	3b	of	-374.81	401.41	362	-0.93	0.3511	Tukey-Kramer	0.9826
Position	3b	p	-37.5081	386.28	360	-0.10	0.9227	Tukey-Kramer	1.0000
Position	3b	ss	170.36	446.80	359	0.38	0.7032	Tukey-Kramer	0.9999
Position	c	if	383.81	427.29	364	0.90	0.3696	Tukey-Kramer	0.9861
Position	c	of	-680.45	292.64	355	-2.33	0.0206	Tukey-Kramer	0.2828
Position	c	p	-343.15	271.57	354	-1.26	0.2072	Tukey-Kramer	0.9116
Position	c	ss	-135.29	352.39	355	-0.38	0.7013	Tukey-Kramer	0.9999
Position	if	of	-1064.26	379.47	364	-2.80	0.0053	Tukey-Kramer	0.0971
Position	if	p	-726.96	363.50	365	-2.00	0.0463	Tukey-Kramer	0.4834
Position	if	ss	-519.10	427.38	365	-1.21	0.2253	Tukey-Kramer	0.9273
Position	of	p	337.30	187.67	357	1.80	0.0731	Tukey-Kramer	0.6223
Position	of	ss	545.16	292.68	358	1.86	0.0633	Tukey-Kramer	0.5775
Position	p	ss	207.87	271.59	357	0.77	0.4446	Tukey-Kramer	0.9947

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Analysis of Variance - PROC MIXED

Effect=Position	Method=Tukey-Kramer(P<.05)	Comparison Group=1	Standard	Letter	MinSig	MaxSig	AvgSig
Obs	Position	Estimate	Error	Group	Diff	Diff	Diff
1	1b	2853.96	423.51	A	572.129	1579.01	1203.47
2	of	2296.24	383.25	AB	572.129	1579.01	1203.47
3	2b	2065.92	475.36	AB	572.129	1579.01	1203.47
4	p	1958.94	375.99	AB	572.129	1579.01	1203.47
5	3b	1921.43	516.99	AB	572.129	1579.01	1203.47
6	ss	1751.07	437.93	AB	572.129	1579.01	1203.47
7	c	1615.79	437.88	B	572.129	1579.01	1203.47
8	if	1231.97	500.13	B	572.129	1579.01	1203.47

```

650      proc sort data=salaries; by position; run;
NOTE: There were 374 observations read from the data set WORK.SALARIES.
NOTE: The data set WORK.SALARIES has 374 observations and 6 variables.
NOTE: PROCEDURE SORT used:
      real time          0.04 seconds
      cpu time           0.04 seconds
651      proc means data=salaries noprint; by position; var salary;
652          output out = next n=m mean=mean var=var std=std stderr=stderr;
653      run;
NOTE: There were 374 observations read from the data set WORK.SALARIES.
NOTE: The data set WORK.NEXT has 8 observations and 8 variables.
NOTE: PROCEDURE MEANS used:
      real time          0.03 seconds
      cpu time           0.03 seconds
654      proc sort data=next; by descending mean; run;
NOTE: There were 8 observations read from the data set WORK.NEXT.
NOTE: The data set WORK.NEXT has 8 observations and 8 variables.
NOTE: PROCEDURE SORT used:
      real time          0.04 seconds
      cpu time           0.04 seconds
655      proc print data=next;
656          TITLE3 'Listing of means by position';
657      run;
NOTE: There were 8 observations read from the data set WORK.NEXT.
NOTE: The PROCEDURE PRINT printed page 5.
NOTE: PROCEDURE PRINT used:
      The SAS System
      real time          0.01 seconds
      cpu time           0.01 seconds

```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Listing of means by position

Obs	Position	_TYPE_	_FREQ_	m	mean	var	std	stderr
1	1b	0	20	20	2246.55	3361252.63	1833.37	409.954
2	of	0	82	82	1555.90	2630761.85	1621.96	179.116
3	2b	0	15	15	1411.89	1998451.89	1413.67	365.007
4	p	0	165	165	1199.79	1890770.20	1375.05	107.048
5	3b	0	14	14	1165.40	1542338.81	1241.91	331.914
6	ss	0	31	31	992.22	1434743.33	1197.81	215.133
7	c	0	31	31	857.87	797114.46	892.81	160.354
8	if	0	16	16	468.84	671899.22	819.69	204.924

```
659      OPTIONS PS=52 ls=111;
660      proc plot data=ResidData; plot resid*pred;
661          TITLE3 'Analysis of Variance - PROC MIXED';
662          TITLE4 'Residual plot';
663      run;
663      !      OPTIONS PS=256 ls=80;
664
```

NOTE: There were 374 observations read from the data set WORK.RESIDDATA.

NOTE: The PROCEDURE PLOT printed page 6.

NOTE: PROCEDURE PLOT used:

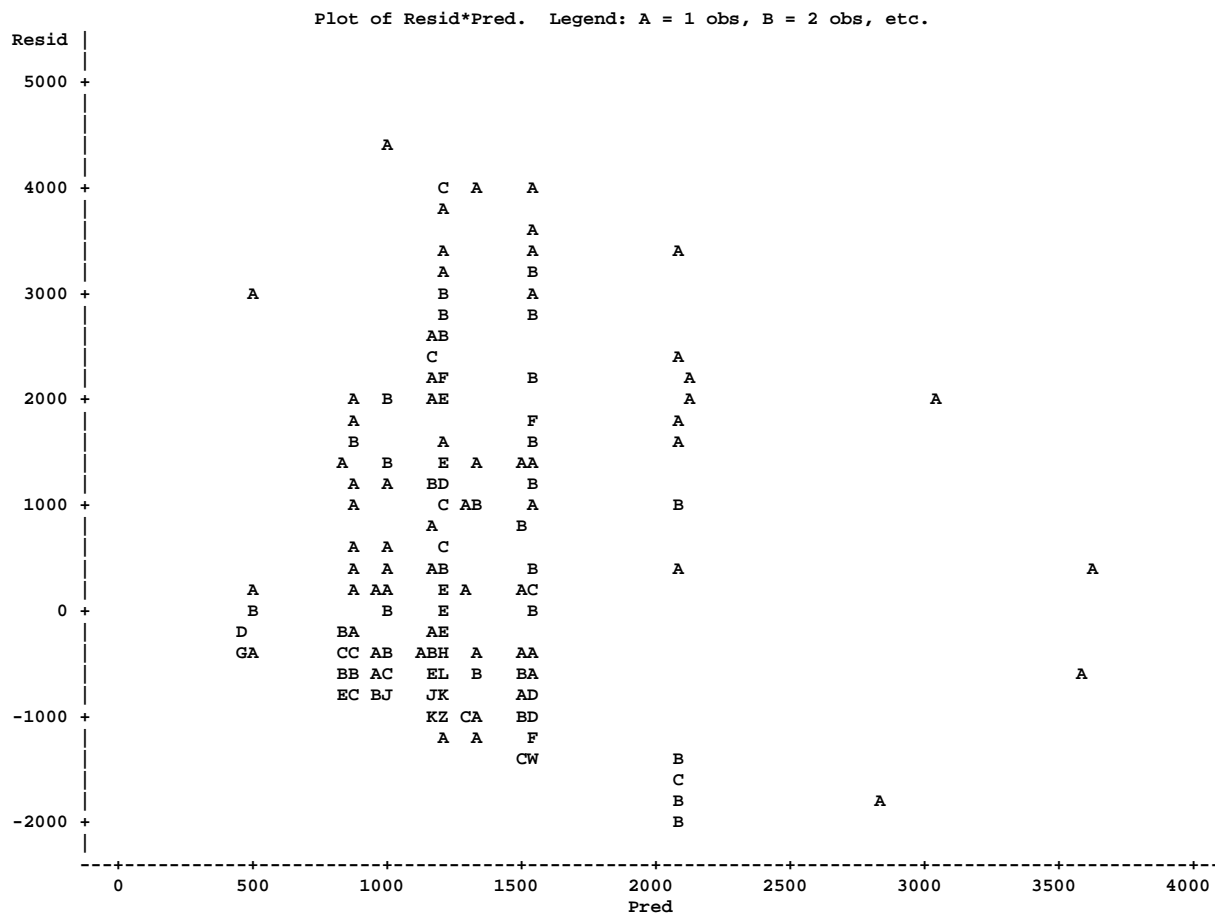
```
real time      0.02 seconds
cpu time       0.02 seconds
```

EXST7015: Salaries of all American league baseball players (1994)

Salaries in thousands of dollars

Analysis of Variance - PROC MIXED

Residual plot



```
665      PROC UNIVARIATE DATA=ResidData PLOT NORMAL; VAR resid;
666          TITLE4 'Residual analysis with PROC UNIVARIATE';
667      RUN;
```

NOTE: The PROCEDURE UNIVARIATE printed page 7.

NOTE: PROCEDURE UNIVARIATE used:

```
real time      0.02 seconds
cpu time       0.02 seconds
```

```
667      !      QUIT;
```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Analysis of Variance - PROC MIXED
Residual analysis with PROC UNIVARIATE

The UNIVARIATE Procedure
Variable: Resid

Moments			
N	374	Sum Weights	374
Mean	0	Sum Observations	0
Std Deviation	1371.12189	Variance	1879975.25
Skewness	1.14864767	Kurtosis	0.44981134
Uncorrected SS	701230767	Corrected SS	701230767
Coeff Variation	.	Std Error Mean	70.8990253

Basic Statistical Measures			
Location		Variability	
Mean	0.00	Std Deviation	1371
Median	-580.87	Variance	1879975
Mode	-1082.71	Range	6394
		Interquartile Range	1919

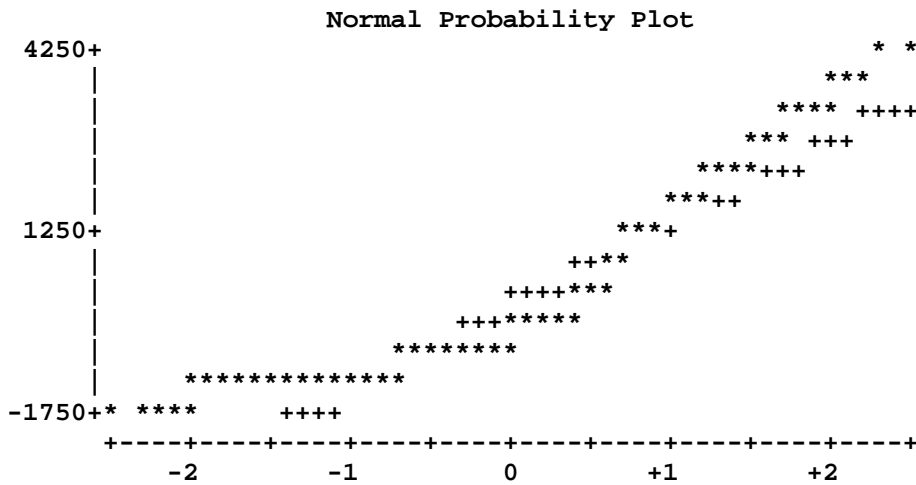
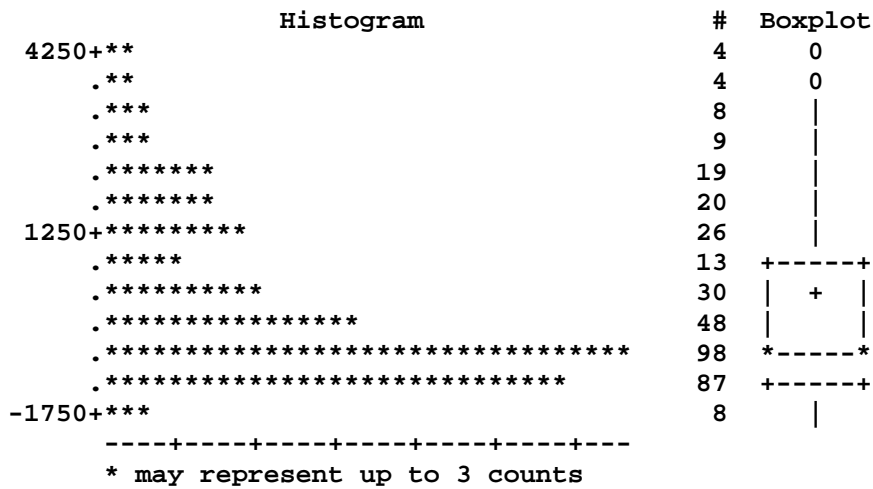
NOTE: The mode displayed is the smallest of 2 modes with a count of 3.

Tests for Location: Mu0=0			
Test	-Statistic-	-----p Value-----	
Student's t	t 0	Pr > t	1.0000
Sign	M -54	Pr >= M	<.0001
Signed Rank	S -3878.5	Pr >= S	0.0637

Tests for Normality			
Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.864381	Pr < W	<0.0001
Kolmogorov-Smirnov	D 0.187785	Pr > D	<0.0100
Cramer-von Mises	W-Sq 3.64595	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 19.53307	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	4400.707
99%	4017.597
95%	2878.646
90%	2118.449
75% Q3	904.436
50% Median	-580.868
25% Q1	-1014.244
10%	-1277.183
5%	-1415.098
1%	-1848.584
0% Min	-1992.824

Extreme Observations			
-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-1992.82	183	3953.94	347
-1940.49	76	4017.60	348
-1849.27	306	4029.02	240
-1848.58	202	4093.31	80
-1831.02	92	4400.71	2

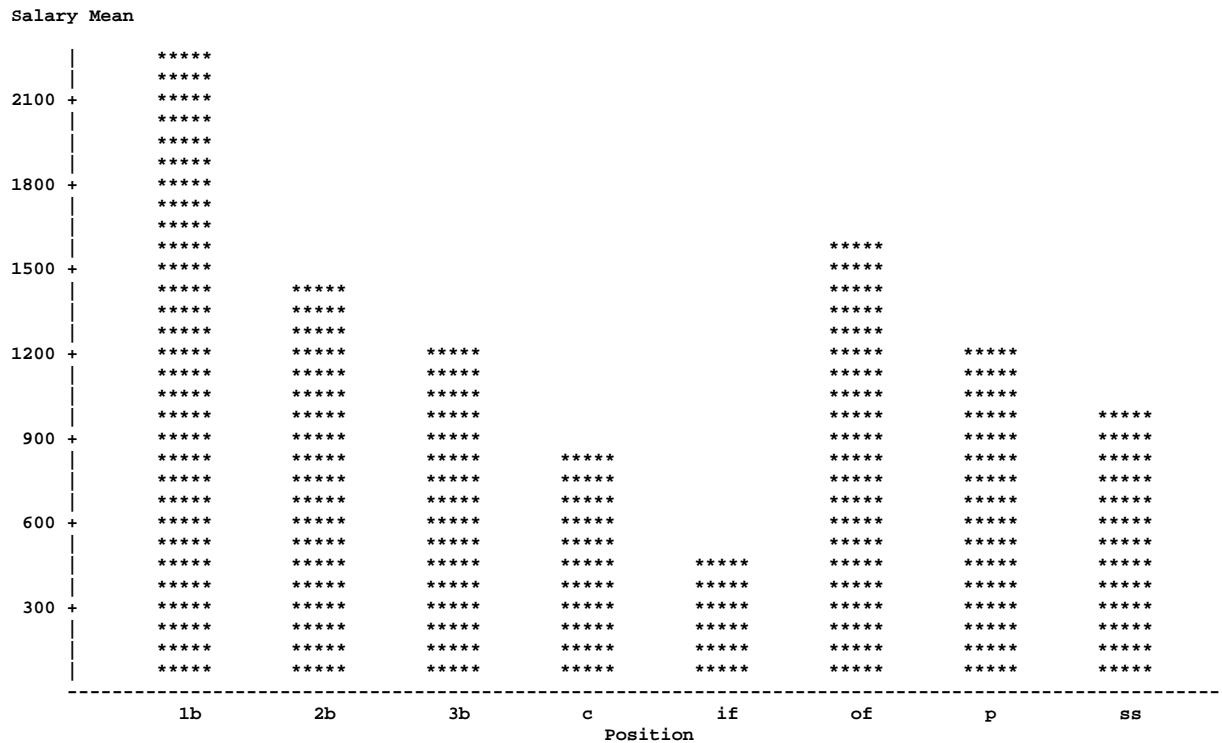


```

669      OPTIONS PS=52 ls=111;
670      PROC CHART DATA=salaries;
671          TITLE3 'Histogram of baseball player salaries';
672          VBAR position / SUMVAR=salary TYPE=MEAN; RUN;
NOTE: The PROCEDURE CHART printed page 8.
NOTE: PROCEDURE CHART used:
      real time          0.01 seconds
      cpu time           0.01 seconds
673      run; quit; OPTIONS PS=256 ls=80;

```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Histogram of baseball player salaries



```

675 PROC MIXED DATA=salaries cl; CLASSES team position dh;
676 TITLE3 'Analysis of Variance - PROC MIXED with logarithmic
transformation';
677 MODEL lsalary = position DH / htype=3 DDFM=Satterthwaite outp=ResidData2;
678 random team;
679 run;
WARNING: Length of CLASS variable team truncated to 16.
NOTE: Convergence criteria met.
NOTE: The data set WORK.RESIDDATA2 has 374 observations and 13 variables.
NOTE: The PROCEDURE MIXED printed page 9.
NOTE: PROCEDURE MIXED used:
real time 0.17 seconds
cpu time 0.16 seconds
679 ! quit;

```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Analysis of Variance - PROC MIXED with logarithmic transformation

The Mixed Procedure

Model Information	
Data Set	WORK.SALARIES
Dependent Variable	lsalary
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
team	14	Baltimore Oriole Boston Red Sox California Angel Chicago White So Cleveland Indian Detroit Tigers Kansas City Roya Milwaukee Brewer Minnesota Twins New York Yankees Oakland Athletic Seattle Mariners Texas Rangers Toronto Blue Jay
Position	8	1b 2b 3b c if of p ss
DH	2	No Yes

Dimensions

Covariance Parameters	2
Columns in X	11
Columns in Z	14
Subjects	1
Max Obs Per Subject	374
Observations Used	374
Observations Not Used	0
Total Observations	374

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	1221.23422138	
1	2	1220.79934573	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
team	0.01687	0.05	0.002732	846.73
Residual	1.5200	0.05	1.3181	1.7723

Fit Statistics

-2 Res Log Likelihood	1220.8
AIC (smaller is better)	1224.8
AICC (smaller is better)	1224.8
BIC (smaller is better)	1226.1

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Position	7	358	2.18	0.0356
DH	1	362	3.06	0.0811

```

680      PROC UNIVARIATE DATA=ResidData2 PLOT NORMAL; VAR resid;
681      TITLE4 'Residual analysis with PROC UNIVARIATE';
682      RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 10.
NOTE: PROCEDURE UNIVARIATE used:
      real time          0.03 seconds
      cpu time           0.03 seconds
682      !      QUIT;

```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Analysis of Variance - PROC MIXED with logarithmic transformation
Residual analysis with PROC UNIVARIATE

The UNIVARIATE Procedure
Variable: Resid

Moments			
N	374	Sum Weights	374
Mean	0	Sum Observations	0
Std Deviation	1.21463657	Variance	1.47534199
Skewness	0.11338944	Kurtosis	-1.2644419
Uncorrected SS	550.302561	Corrected SS	550.302561
Coeff Variation	.	Std Error Mean	0.06280736

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	1.21464
Median	-0.09396	Variance	1.47534
Mode	-1.63399	Range	4.82406
		Interquartile Range	2.19735

NOTE: The mode displayed is the smallest of 2 modes with a count of 3.

Tests for Location: Mu0=0			
Test	-Statistic-	-----p Value-----	
Student's t	t 0	Pr > t	1.0000
Sign	M -11	Pr >= M	0.2775
Signed Rank	S -32.5	Pr >= S	0.9876

Tests for Normality			
Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.946761	Pr < W	<0.0001
Kolmogorov-Smirnov	D 0.084031	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.874675	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 6.050037	Pr > A-Sq	<0.0050

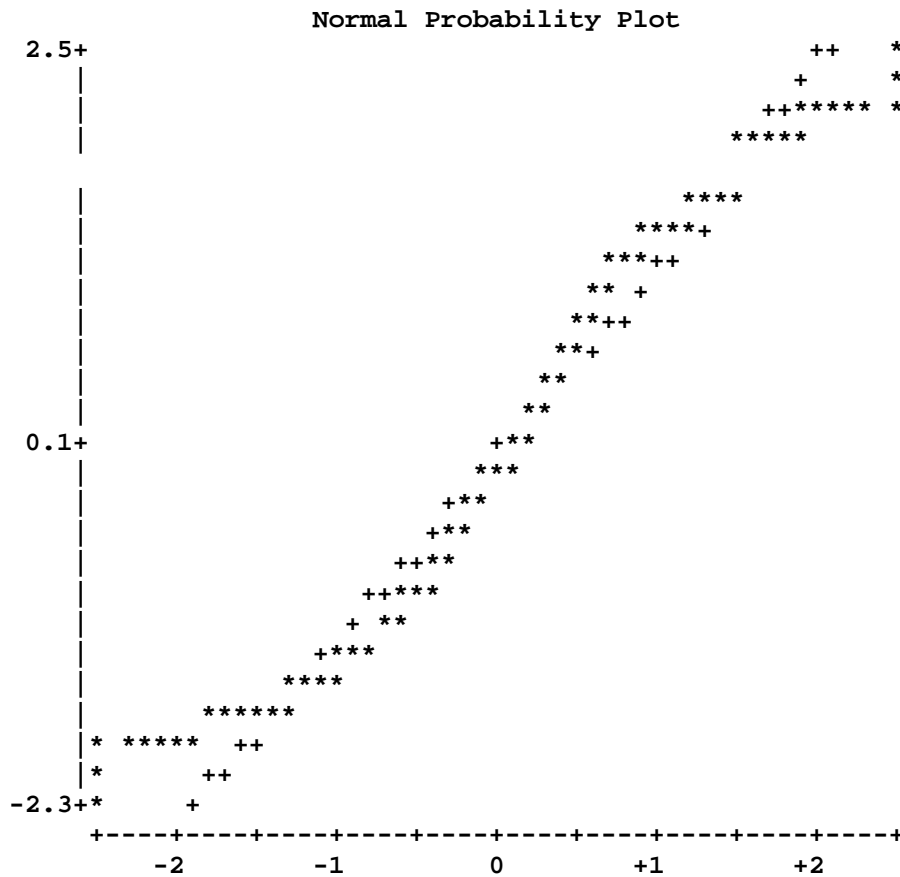
Quantiles (Definition 5)	
Quantile	Estimate
100% Max	2.4878558
99%	2.1264322
95%	1.9098152
90%	1.6593573
75% Q3	1.0782959
50% Median	-0.0939615
25% Q1	-1.1190570
10%	-1.5963910
5%	-1.6813634
1%	-1.9400166
0% Min	-2.3362008

Extreme Observations			
-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-2.33620	9	2.11858	247
-2.04187	3	2.12643	190
-1.94872	121	2.14542	213
-1.94002	109	2.30579	344
-1.94002	108	2.48786	85

Stem Leaf	#	Boxplot
24 9	1	
22 1	1	
20 23349235	8	
18 1157790123355889	16	
16 024556667700245668	18	
14 0001233566688899000134566667	28	
12 24455667891245889	17	
10 001223345589989	15	+-----+
8 23334668888044778	17	
6 23346802245666	14	
4 135688812478	12	
2 1111266133579	13	
0 0012223457706899	16	+
-0 97655544422200099876544222	26	*-----*
-2 988653332222109776640	21	
-4 75443100643	11	
-6 9888655543199210	16	
-8 7655422109876553111	19	
-10 98755432221975543210	20	+-----+
-12 9665554200098887776521	22	
-14 876655543188776544442221	24	
-16 855540885444333322222110000	27	
-18 5444987653	10	
-20 4	1	
-22 4	1	

-----+-----+-----+-----+-----

Multiply Stem.Leaf by 10**-1



```
684     OPTIONS PS=52 ls=111;
685     proc plot data=ResidData2; plot resid*pred;
686         TITLE4 'Residual plot';
687     run;
```

EXST7015: Salaries of all American league baseball players (1994)
Salaries in thousands of dollars
Analysis of Variance - PROC MIXED with logarithmic transformation
Residual plot

