Exercise 1a

The MEANS Procedure

Analysis Variable : MPG_City MPG (City)						
Cylinders	Origin	Type	N Obs	N	Mean	Std Dev
4	Asia	Sedan	49	49	26.5102041	4.1085401
		Sports	8	8	21.5000000	2.9277002
	USA	Sedan	29	29	24.6551724	2.6087197
6	Asia	Sedan	41	41	18.9756098	1.3320624
		Sports	6	6	19.6666667	1.5055453
	USA	Sedan	45	45	19.2222222	1.6499158
		Sports	2	2	18.5000000	2.1213203

Mean of 4-cyl sedans > 6-cyl sedans.Among 4-cyl, mean of sedans > sports (26.5 > 21.5).For 4-cyl sedans, Asia > USA (26.5 > 24.7).For 6-cyl sedans, Asia and USA are almost the same.Cell sizes range from 49 (4-cyl Asia sedans) to 2 (6-cyl USA sports).
Exercise 1b

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Cylinders	2	4 6
Origin	2	Asia USA
Type	2	Sedan Sports

Number of Observations Read	180
Number of Observations Used	180

The GLM Procedure

Dependent Variable: MPG_City MPG (City)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1856.443851	618.814617	78.33	<.0001
Error	176	1390.356149	7.899751		
Corrected Total	179	3246.800000			

R-Square	Coeff Var	Root MSE	MPG_City Mean
0.571776	12.69872	2.810650	22.13333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Cylinders	1	1676.720719	1676.720719	212.25	<.0001
Origin	1	19.858239	19.858239	2.51	0.1147
Type	1	70.467407	70.467407	8.92	0.0032

We will keep cylinder and type as they are statistically significant. We will remove origin variable as it has a high p-value (0.1147) and is statistically insignificant.Overall the model is significant as p-value is less than 0.05.

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Cylinders	2	4 6
Type	2	Sedan Sports

Number of Observations Read	180
Number of Observations Used	180

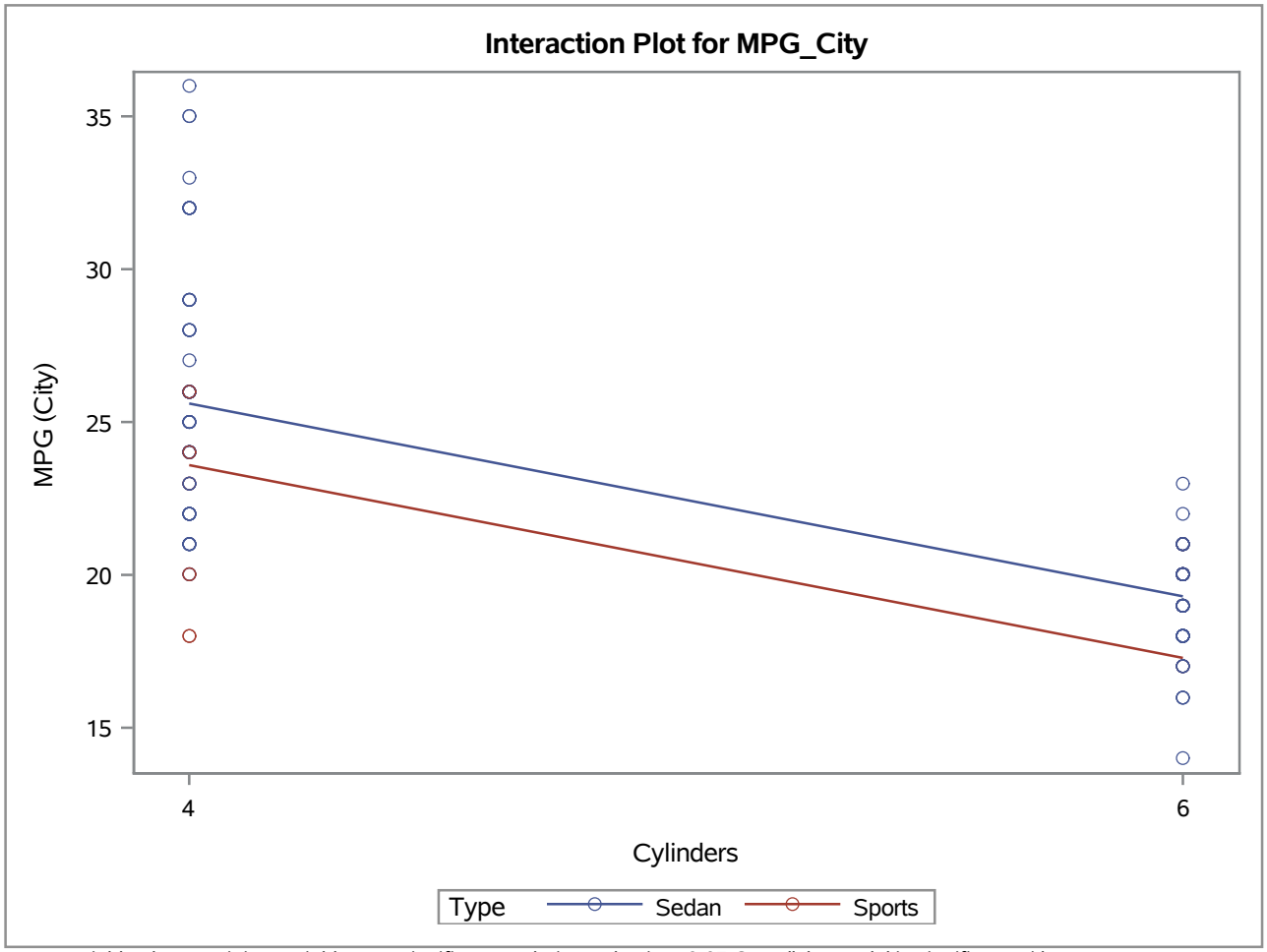
The GLM Procedure

Dependent Variable: MPG_City MPG (City)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1836.585612	918.292806	115.26	<.0001
Error	177	1410.214388	7.967313		
Corrected Total	179	3246.800000			

R-Square	Coeff Var	Root MSE	MPG_City Mean
0.565660	12.75290	2.822643	22.13333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Cylinders	1	1786.082869	1786.082869	224.18	<.0001
Type	1	59.183930	59.183930	7.43	0.0071



After removing type variable, the remaining variables are significant as their p-value is < 0.05.Overall the model is significant with p-value < 0.05 & F-value = 115.26.The percentage of variation explained by the model (R^2) = 56.5%.
Exercise 1c

The GLM Procedure

Dependent Variable: MPG_City MPG (City)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1913.379681	637.793227	84.18	<.0001
Error	176	1333.420319	7.576252		
Corrected Total	179	3246.800000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Cylinders	1	284.7924390	284.7924390	37.59	<.0001
Type	1	59.7701685	59.7701685	7.89	0.0055
Cylinders*Type	1	76.7940685	76.7940685	10.14	0.0017

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Cylinders	MPG_City LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
4	23.6602564	<.0001
6	19.2398256	

Least Squares Means for Effect Cylinders				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	4.420431	2.997538	5.843324

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Type	MPG_City LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Sedan	22.4625820	0.0055
Sports	20.4375000	

Least Squares Means for Effect Type				
i	**j**	**Difference Between Means**	**Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)**	
1	**2**	2.025082	0.602189	3.447975

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Cylinders	Type	MPG_City LSMEAN	LSMEAN Number
4	**Sedan**	25.8205128	1
4	**Sports**	21.5000000	2
6	**Sedan**	19.1046512	3
6	**Sports**	19.3750000	4

Least Squares Means for Effect Cylinders*Type				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	**2**	4.320513	1.670113	6.970913
1	**3**	6.715862	5.599564	7.832159
1	**4**	6.445513	3.795113	9.095913
2	**3**	2.395349	-0.243559	5.034257
2	**4**	2.125000	-1.444641	5.694641
3	**4**	-0.270349	-2.909257	2.368559

Final model used: mpg_city = Cylinders Type Cylinders*Type (Origin dropped).All the terms are highly significant as their p value is << 0.05.Overall model fit: $F = 84.18$, $p < .0001$. Therefore, model is highly significant.Variation explained: $R^2 = 0.0589$ so the model explains around 59% of the variability in city MPG.4-cylinder cars have much higher MPG (about 4.42) than 6-cylinder cars.Sedans get about 2 MPG more than sports cars, especially for 4-cylinder engines.The interaction means the effect of Type depends on Cylinders.Sedan vs. sports difference is large for 4-cyl but disappears for 6-cyl.Origin doesn't affect MPG once the other factors are included.
Exercise 2a

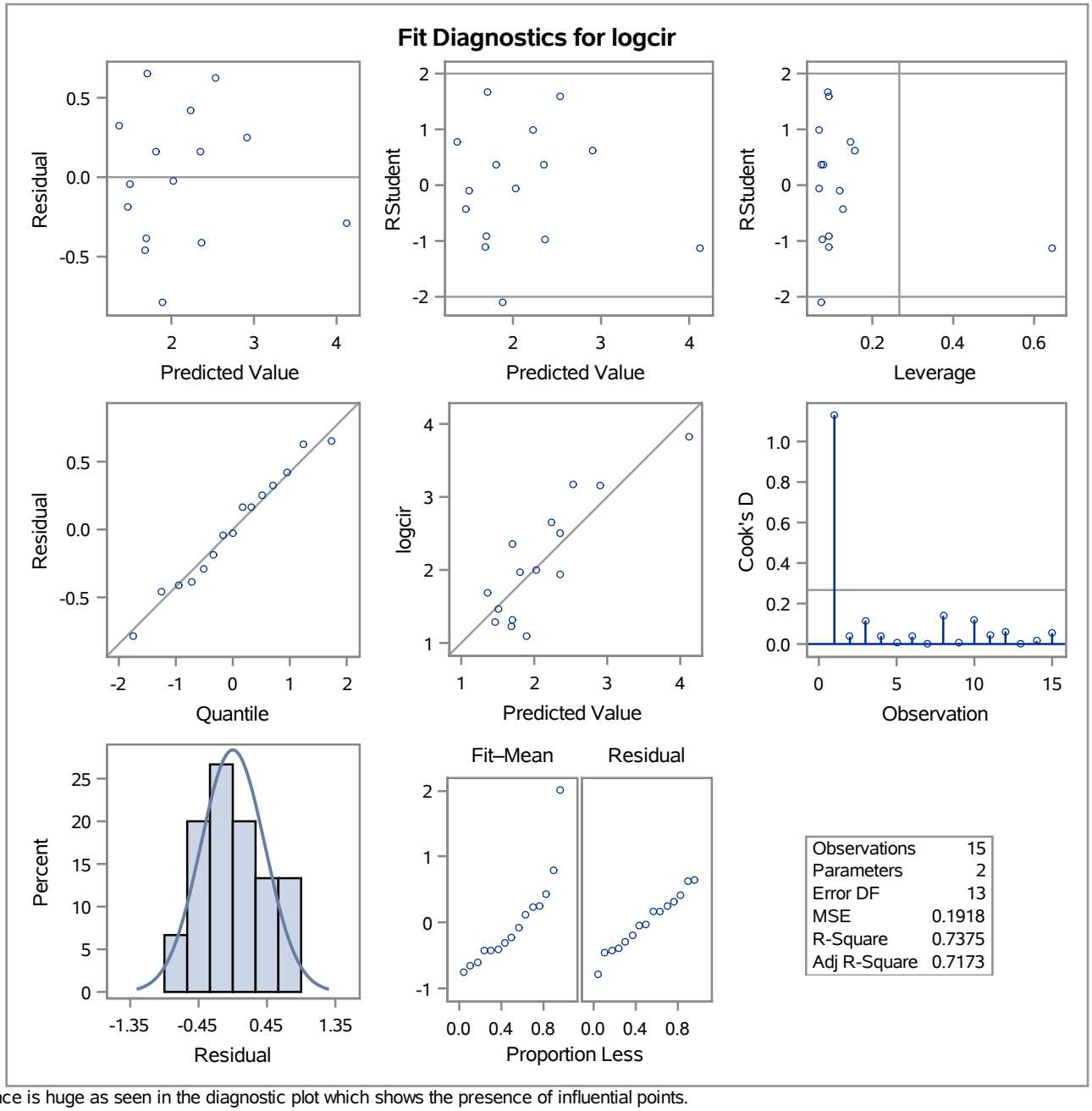
The REG Procedure
Model: MODEL1
Dependent Variable: logcir

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	7.00437	7.00437	36.53	<.0001
Error	13	2.49293	0.19176		
Corrected Total	14	9.49729			

Root MSE	0.43791	**R-Square**	0.7375
Dependent Mean	2.11062	**Adj R-Sq**	0.7173
Coeff Var	20.74780		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.96807	0.22028	4.39	0.0007
alcohol	1	0.12771	0.02113	6.04	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: logcir



Cook's distance is huge as seen in the diagnostic plot which shows the presence of influential points.

Obs	country	alcohol	logcir	cd
1	France	24.7	3.83081	1.13082
2	Italy	15.2	3.16125	0.03682
3	W.Germany	12.3	3.16548	0.11551
4	Austria	10.9	1.94591	0.03954
5	Belgium	10.8	2.50960	0.00599
6	USA	9.9	2.65324	0.03663
7	Canada	8.3	2.00148	0.00014
8	E&W	7.2	1.09861	0.13956
9	Sweden	6.6	1.97408	0.00651
10	Japan	5.8	2.36085	0.12005
11	Netherlands	5.7	1.30833	0.04328
12	Ireland	5.6	1.22378	0.06202
13	Norway	4.2	1.45862	0.00084
14	Finland	3.9	1.28093	0.01475
15	Israel	3.1	1.68640	0.05440

Obs	country	alcohol	logcir	cd
1	France	24.7	3.83081	1.13082

France has a really high cook's distance. Therefore, it is a highly influential point.

The REG Procedure
Model: MODEL1
Dependent Variable: logcir

Number of Observations Read	14
Number of Observations Used	14

Analysis of Variance

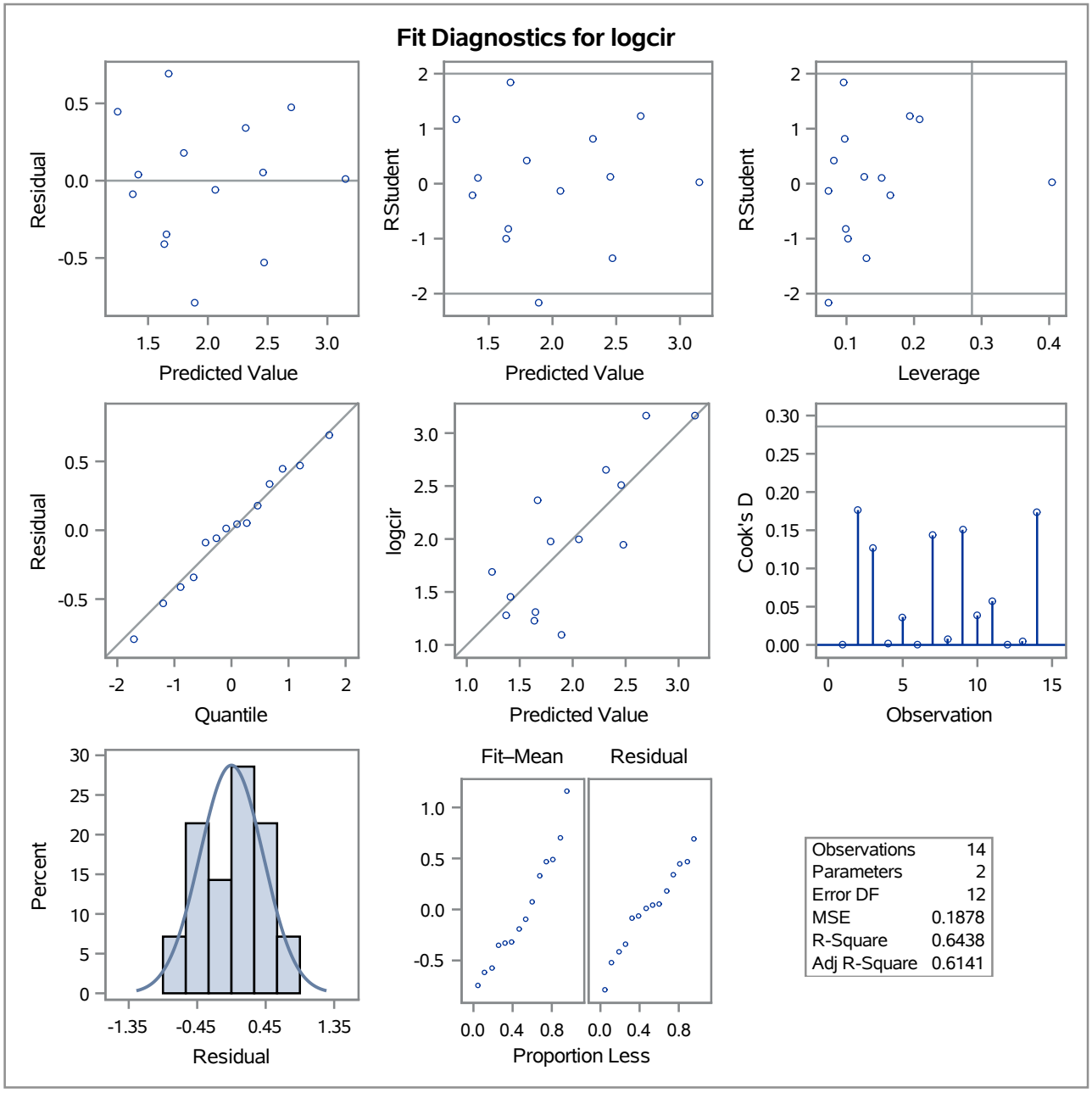
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4.07318	4.07318	21.69	0.0006
Error	12	2.25370	0.18781		
Corrected Total	13	6.32689			

Root MSE	0.43337	**R-Square**	0.6438
Dependent Mean	1.98775	**Adj R-Sq**	0.6141
Coeff Var	21.80194		

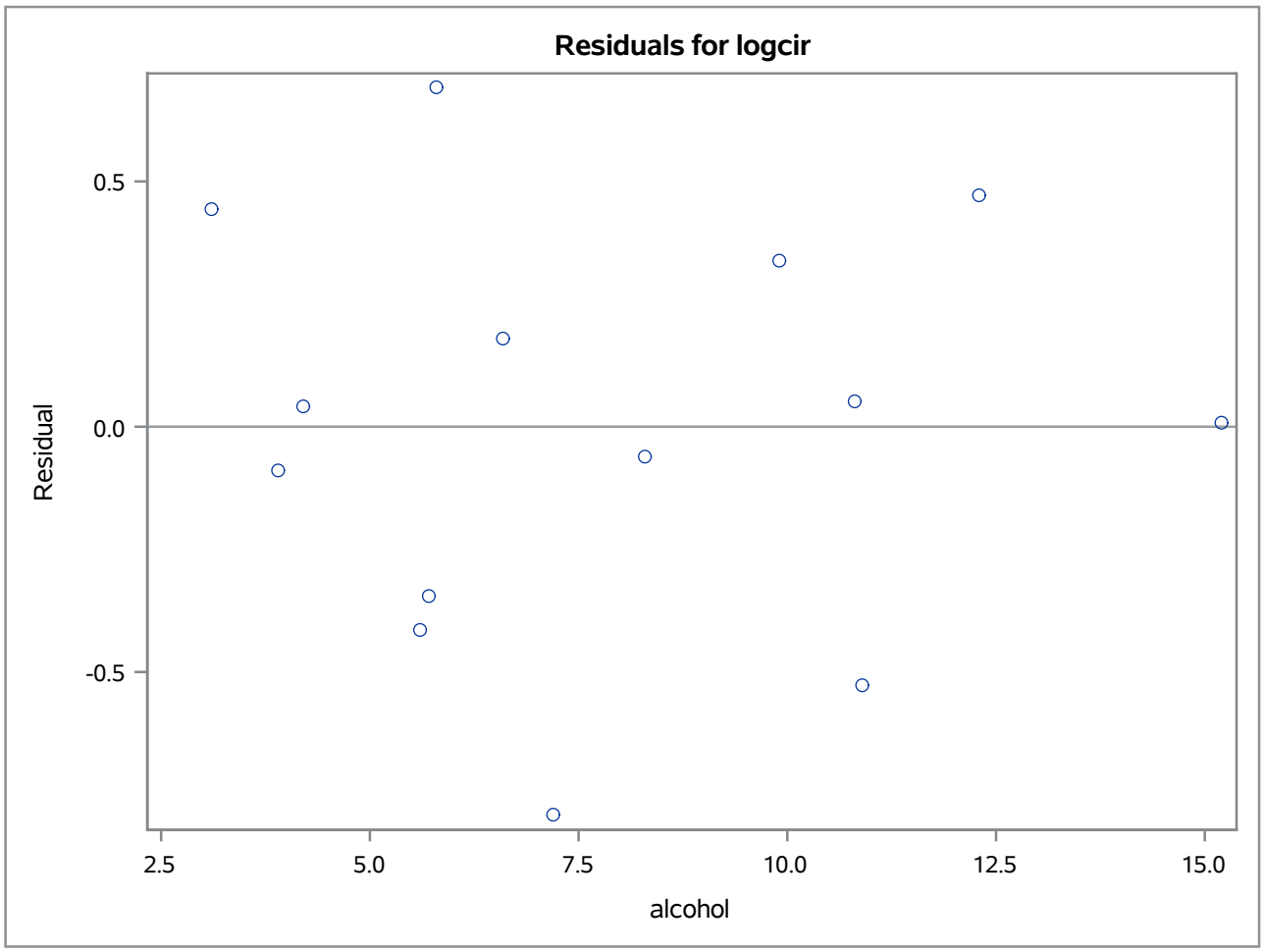
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.75356	0.28922	2.61	0.0230
alcohol	1	0.15780	0.03388	4.66	0.0006

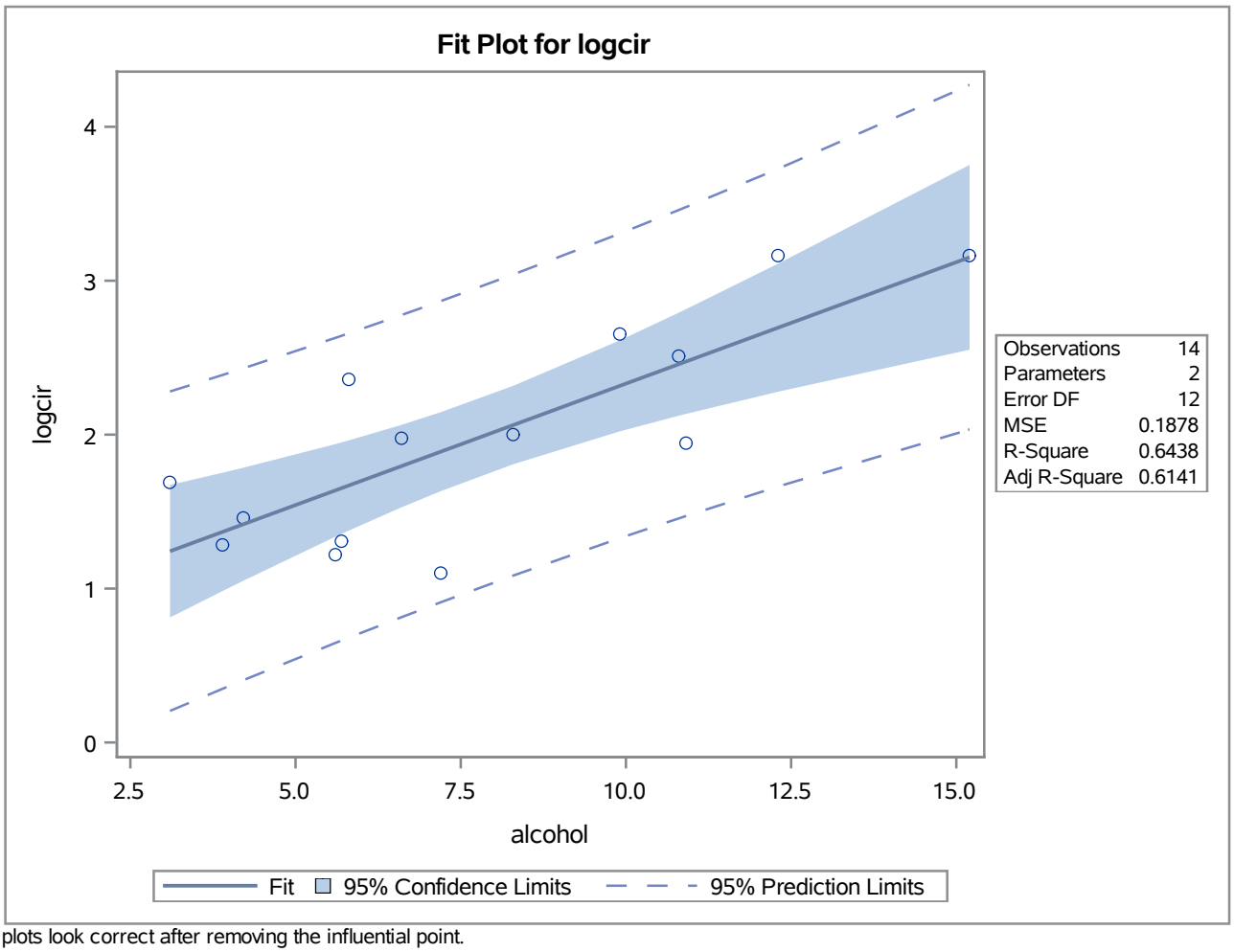
The REG Procedure
Model: MODEL1
Dependent Variable: logcir



The REG Procedure
Model: MODEL1
Dependent Variable: logcir



The REG Procedure
Model: MODEL1
Dependent Variable: logcir



The diagnost plots look correct after removing the influential point.
Exercise 2b
Overall F = 21.7, p = 0.0006 so model is highly significant.R^2 = 0.644 (Adj R^2 = 0.614) so the model explains 64% of the variability in log(cirrhosis deaths).Residual vs fitted: no funnel shape; variance looks roughly constant.QQ-plot / histogram: residuals are close to normal.Leverage & Cook's D: several moderate bars but none exceed cutoffs. No remaining unduly influential points.Slope for alcohol = 0.1578, p = 0.0006 (strongly positive).exp(0.1578) = 1.17.Each additional liter of alcohol per person per year is associated with 17% higher cirrhosis death rates.The raw model (chosen in class) has slightly higher R2(69% vs 64%), but shows worse diagnostics (non-constant variance, more influence).The log model gives a cleaner fit that better meets regression assumptions & has positive predictions.
Exercise 3a

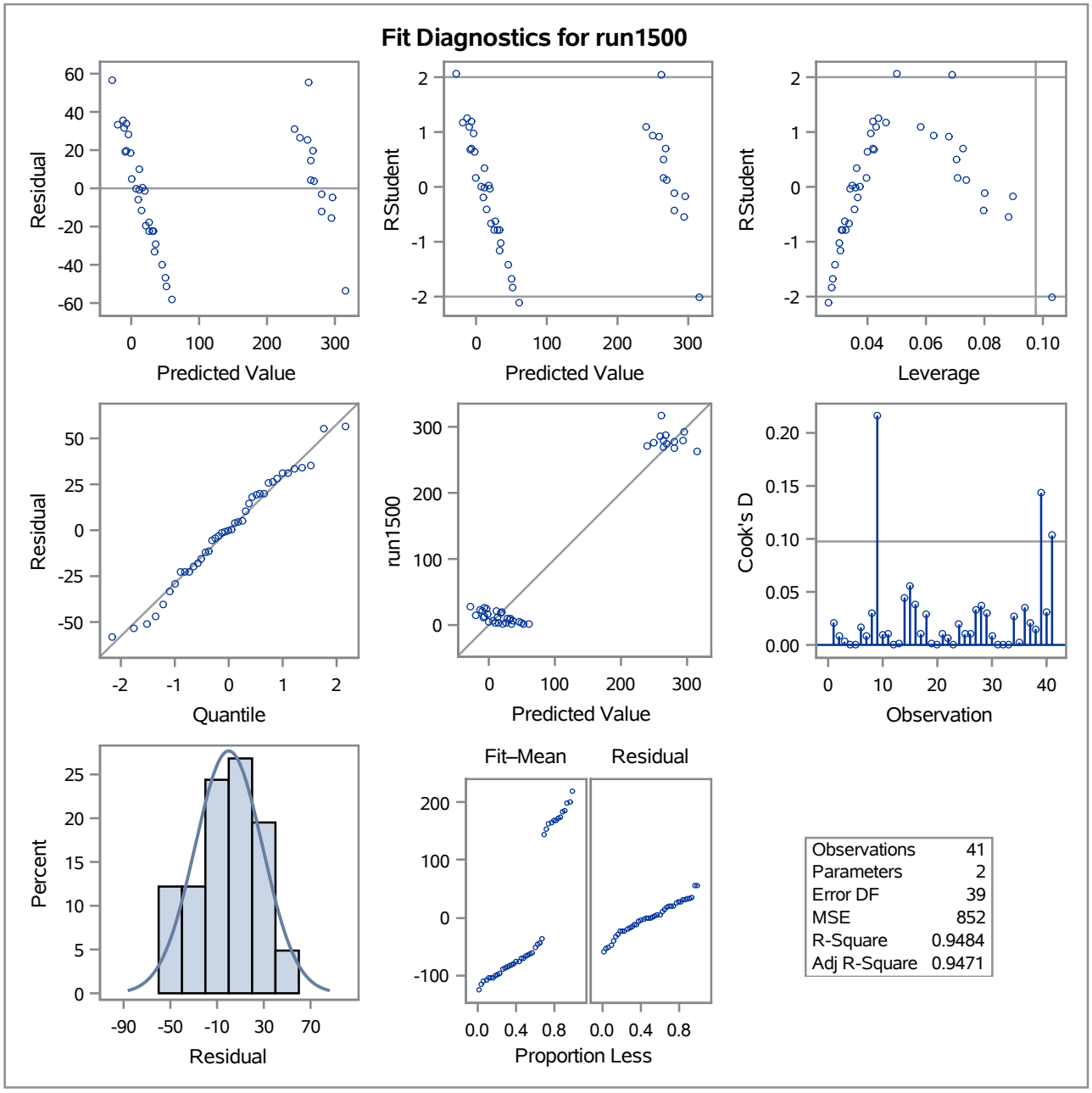
The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	610408	610408	716.44	<.0001
Error	39	33228	851.99956		
Corrected Total	40	643636			

Root MSE	29.18903	**R-Square**	0.9484
Dependent Mean	96.48268	**Adj R-Sq**	0.9471
Coeff Var	30.25313		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-491.85754	22.44825	-21.91	<.0001
run100	1	69.37974	2.59204	26.77	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: run1500



Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
1	7.58	14.83	2.07	49.81	14.69	43.75	5.02	63.19	291.70	1.00	0.02099
2	7.40	14.26	1.86	49.37	14.05	50.72	4.92	60.15	301.50	2.00	0.00807
3	7.30	14.77	2.04	48.37	14.09	48.95	4.92	50.31	300.20	3.00	0.00301
4	7.23	14.25	1.92	48.93	14.99	40.87	5.32	62.77	280.10	4.00	0.00077
5	7.09	15.19	2.10	50.42	15.31	46.26	4.72	63.44	276.40	5.00	0.00062
6	7.60	14.31	1.98	48.68	14.23	41.10	4.92	51.77	278.10	6.00	0.01649
7	11.13	7.30	13.48	2.01	48.62	14.17	45.67	4.42	55.37	268.00	0.00842
8	10.83	7.31	13.76	2.13	49.91	14.38	44.41	4.42	56.37	285.10	0.03002
9	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.10	0.21636
10	7.56	14.41	1.86	51.10	15.06	44.99	4.82	57.19	285.10	10.00	0.00996
11	6.97	14.09	1.95	49.48	14.48	42.10	4.72	55.40	282.00	11.00	0.01008
12	7.27	12.68	1.98	49.20	15.29	37.92	4.62	57.44	266.60	12.00	0.00001
13	11.36	6.80	13.46	1.86	51.16	15.67	40.49	5.02	54.68	291.70	0.00134
14	7.84	16.36	2.12	48.36	14.05	48.72	5.00	70.52	280.01	1.00	0.04473
15	7.96	15.23	2.06	49.19	14.13	50.11	4.90	69.71	282.00	2.00	0.05603
16	7.81	15.93	2.09	46.81	13.97	51.65	4.60	55.54	278.11	3.00	0.03832
17	7.47	15.73	2.15	48.97	14.56	48.34	4.40	58.46	265.42	4.00	0.01021
18	7.74	14.48	1.97	47.97	14.01	43.73	4.90	55.39	278.05	5.00	0.02878
19	10.91	7.14	15.31	2.12	49.40	14.95	45.62	4.70	63.45	269.54	0.00096
20	7.19	14.65	2.03	48.73	14.25	44.72	4.80	57.76	264.35	7.00	0.00000
21	7.53	14.26	1.88	48.81	14.80	42.05	5.40	61.33	276.33	8.00	0.01004
22	7.48	14.80	2.12	49.13	14.17	44.75	4.40	55.27	276.31	9.00	0.00663
23	10.98	7.49	14.01	1.94	49.76	14.25	42.43	5.10	56.32	273.56	0.00066
24	10.95	7.31	15.10	2.06	50.79	14.21	44.60	5.00	53.45	287.63	0.01935
25	10.90	7.30	14.77	1.88	50.30	14.34	44.41	5.00	60.89	278.82	0.00999
26	6.99	14.91	1.94	49.41	14.37	44.83	4.60	64.55	267.09	13.00	0.01060
27	6.81	15.24	1.91	49.27	14.01	49.02	4.20	61.52	272.74	14.00	0.03335
28	10.55	7.34	14.44	1.94	49.72	14.39	39.88	4.80	54.51	271.02	0.03680
29	10.68	7.50	14.97	1.94	49.12	15.01	40.35	4.60	59.26	275.71	0.02954
30	7.07	13.88	1.94	49.11	14.77	42.47	4.70	60.88	263.31	17.00	0.00859
31	7.34	13.55	1.97	49.65	14.78	45.13	4.50	60.79	272.63	18.00	0.00001
32	7.38	13.07	1.88	48.51	14.01	40.11	5.00	51.53	274.21	19.00	0.00003
33	11.14	6.61	15.69	2.03	51.04	14.88	41.90	4.80	65.82	277.94	0.00053
34	6.94	15.15	1.94	49.56	15.12	45.62	5.30	50.62	290.36	21.00	0.02716
35	7.26	14.57	1.85	48.61	14.41	40.95	4.40	60.71	269.70	22.00	0.00236
36	6.91	13.62	2.03	51.67	14.26	39.83	4.80	59.34	290.01	23.00	0.03534
37	7.03	13.22	1.85	49.34	15.38	40.22	4.50	58.36	263.08	24.00	0.02067
38	11.33	7.26	13.30	1.97	50.54	14.98	43.34	4.50	52.92	278.67	0.01509

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
39	10.86	7.07	14.81	1.94	51.16	14.96	46.07	4.70	53.05	317.00	0.14349
40	6.99	13.53	1.85	50.95	15.09	43.01	4.50	60.00	281.70	27.00	0.03077
41	6.68	14.92	1.94	53.20	15.39	48.66	4.40	58.62	296.12	28.00	0.10329

The residuals are approximately normal & there seem to be no influential points.The model is highly significant ($p < 0.0001$) with an F-value = 716.44, showing a very strong linear relationship between 100m and 1500m dash times.The $R^2 = 0.9484$ means that about 95% of the variation in the 1500m time can be explained by the 100m time.The positive slope (69.38) suggests that athletes who take longer to complete the 100m dash also tend to take longer for the 1500m run.
Exercise 3b

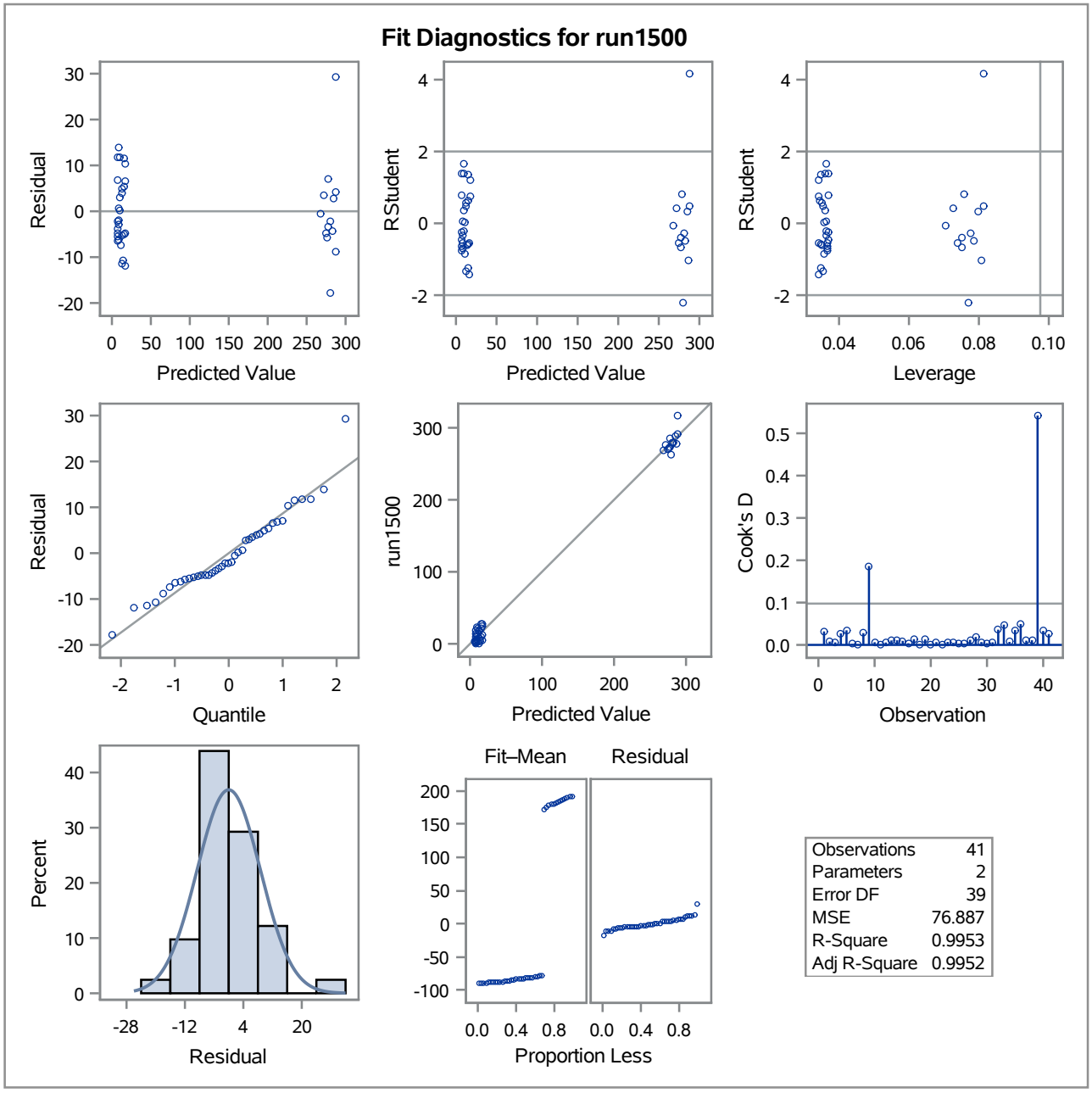
The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	640637	640637	8332.20	<.0001
Error	39	2998.59149	76.88696		
Corrected Total	40	643636			

Root MSE	8.76852	**R-Square**	0.9953
Dependent Mean	96.48268	**Adj R-Sq**	0.9952
Coeff Var	9.08818		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-98.54861	2.53779	-38.83	<.0001
run400	1	7.54808	0.08269	91.28	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: run1500



Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
1	7.58	14.83	2.07	49.81	14.69	43.75	5.02	63.19	291.70	1.00	0.03182
2	7.40	14.26	1.86	49.37	14.05	50.72	4.92	60.15	301.50	2.00	0.00780
3	7.30	14.77	2.04	48.37	14.09	48.95	4.92	50.31	300.20	3.00	0.00593
4	7.23	14.25	1.92	48.93	14.99	40.87	5.32	62.77	280.10	4.00	0.02733
5	7.09	15.19	2.10	50.42	15.31	46.26	4.72	63.44	276.40	5.00	0.03446
6	7.60	14.31	1.98	48.68	14.23	41.10	4.92	51.77	278.10	6.00	0.00208
7	11.13	7.30	13.48	2.01	48.62	14.17	45.67	4.42	55.37	268.00	0.00010
8	10.83	7.31	13.76	2.13	49.91	14.38	44.41	4.42	56.37	285.10	0.02772
9	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.10	0.18623
10	7.56	14.41	1.86	51.10	15.06	44.99	4.82	57.19	285.10	10.00	0.00637
11	6.97	14.09	1.95	49.48	14.48	42.10	4.72	55.40	282.00	11.00	0.00002
12	7.27	12.68	1.98	49.20	15.29	37.92	4.62	57.44	266.60	12.00	0.00565
13	11.36	6.80	13.46	1.86	51.16	15.67	40.49	5.02	54.68	291.70	0.01049
14	7.84	16.36	2.12	48.36	14.05	48.72	5.00	70.52	280.01	1.00	0.01089
15	7.96	15.23	2.06	49.19	14.13	50.11	4.90	69.71	282.00	2.00	0.00956
16	7.81	15.93	2.09	46.81	13.97	51.65	4.60	55.54	278.11	3.00	0.00393
17	7.47	15.73	2.15	48.97	14.56	48.34	4.40	58.46	265.42	4.00	0.01349
18	7.74	14.48	1.97	47.97	14.01	43.73	4.90	55.39	278.05	5.00	0.00125
19	10.91	7.14	15.31	2.12	49.40	14.95	45.62	4.70	63.45	269.54	0.01281
20	7.19	14.65	2.03	48.73	14.25	44.72	4.80	57.76	264.35	7.00	0.00103
21	7.53	14.26	1.88	48.81	14.80	42.05	5.40	61.33	276.33	8.00	0.00656
22	7.48	14.80	2.12	49.13	14.17	44.75	4.40	55.27	276.31	9.00	0.00009
23	10.98	7.49	14.01	1.94	49.76	14.25	42.43	5.10	56.32	273.56	0.00695
24	10.95	7.31	15.10	2.06	50.79	14.21	44.60	5.00	53.45	287.63	0.00484
25	10.90	7.30	14.77	1.88	50.30	14.34	44.41	5.00	60.89	278.82	0.00314
26	6.99	14.91	1.94	49.41	14.37	44.83	4.60	64.55	267.09	13.00	0.00240
27	6.81	15.24	1.91	49.27	14.01	49.02	4.20	61.52	272.74	14.00	0.01194
28	10.55	7.34	14.44	1.94	49.72	14.39	39.88	4.80	54.51	271.02	0.01869
29	10.68	7.50	14.97	1.94	49.12	15.01	40.35	4.60	59.26	275.71	0.00671
30	7.07	13.88	1.94	49.11	14.77	42.47	4.70	60.88	263.31	17.00	0.00407
31	7.34	13.55	1.97	49.65	14.78	45.13	4.50	60.79	272.63	18.00	0.00613
32	7.38	13.07	1.88	48.51	14.01	40.11	5.00	51.53	274.21	19.00	0.03595
33	11.14	6.61	15.69	2.03	51.04	14.88	41.90	4.80	65.82	277.94	0.04783
34	6.94	15.15	1.94	49.56	15.12	45.62	5.30	50.62	290.36	21.00	0.00710
35	7.26	14.57	1.85	48.61	14.41	40.95	4.40	60.71	269.70	22.00	0.03497
36	6.91	13.62	2.03	51.67	14.26	39.83	4.80	59.34	290.01	23.00	0.04922
37	7.03	13.22	1.85	49.34	15.38	40.22	4.50	58.36	263.08	24.00	0.00992
38	11.33	7.26	13.30	1.97	50.54	14.98	43.34	4.50	52.92	278.67	0.01094

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
39	10.86	7.07	14.81	1.94	51.16	14.96	46.07	4.70	53.05	317.00	0.54191
40	6.99	13.53	1.85	50.95	15.09	43.01	4.50	60.00	281.70	27.00	0.03282
41	6.68	14.92	1.94	53.20	15.39	48.66	4.40	58.62	296.12	28.00	0.02563

The residuals are approximately normal & there seem to be no influential points.Model 1 (run100 ~ run1500)R^2 = 0.9484 so 94.8% of variation explained.F = 716.44, p < 0.0001 (strong relationship)Slope = 69.38 so athletes with higher 100m times also have higher 1500m timesModel 2 (run400 ~ run1500)R^2 = 0.9953 so 99.5% of variation explained.F = 8332.20, p < 0.0001 (even stronger relationship)Slope = 7.55 so athletes with higher 400m times tend to have higher 1500m times.The run400 model explains more variation (99.5% vs. 94.8%) and has a better overall fit.Diagnostics also show fewer residual issues for the 400m model.Therefore, run400 is the better predictor of 1500m performance
Exercise 4a

The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Number of Observations Read	41
Number of Observations Used	41

Analysis of Variance

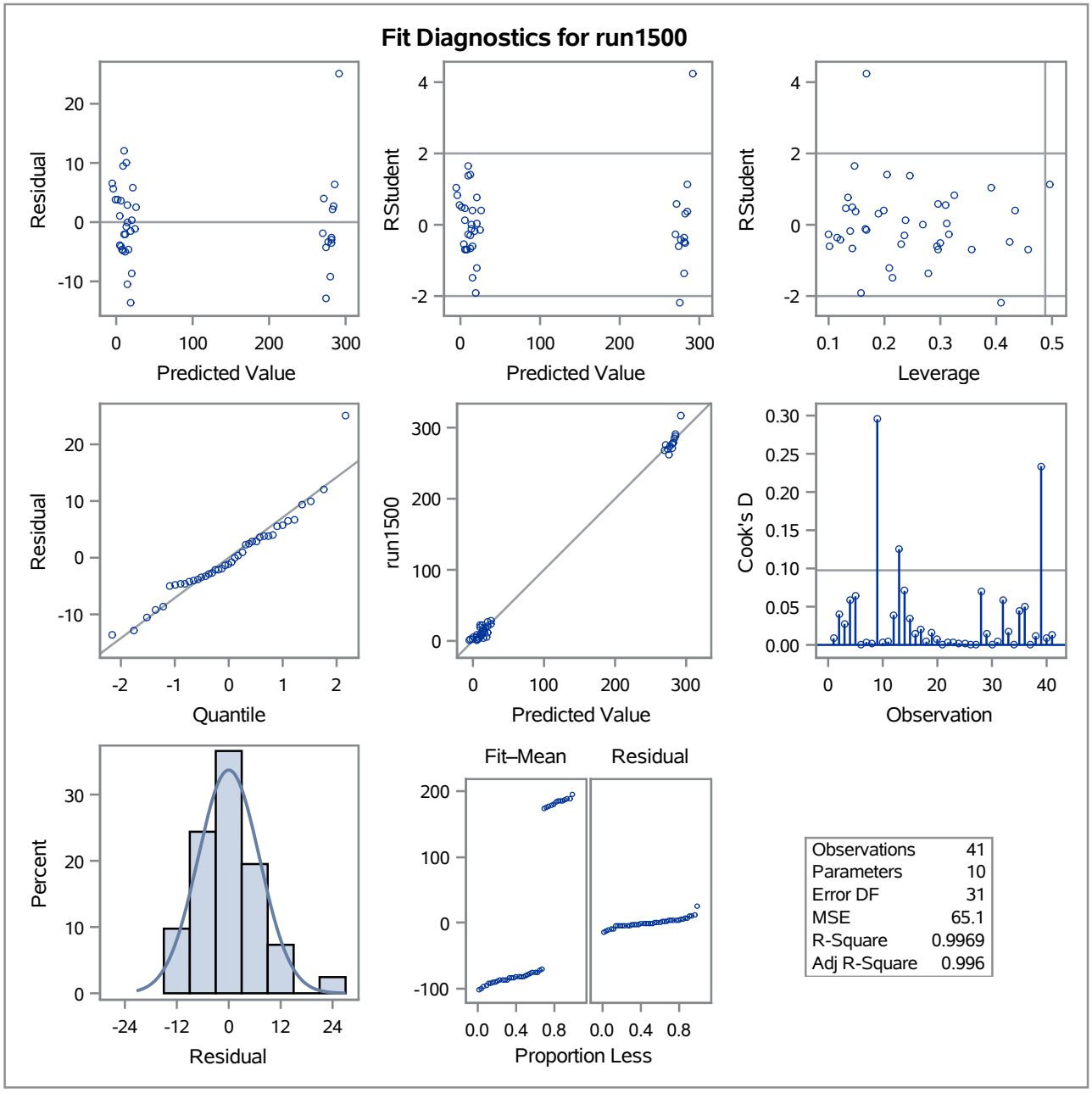
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	641618	71291	1095.09	<.0001
Error	31	2018.11071	65.10035		
Corrected Total	40	643636			

Root MSE	8.06848	**R-Square**	0.9969
Dependent Mean	96.48268	**Adj R-Sq**	0.9960
Coeff Var	8.36262		

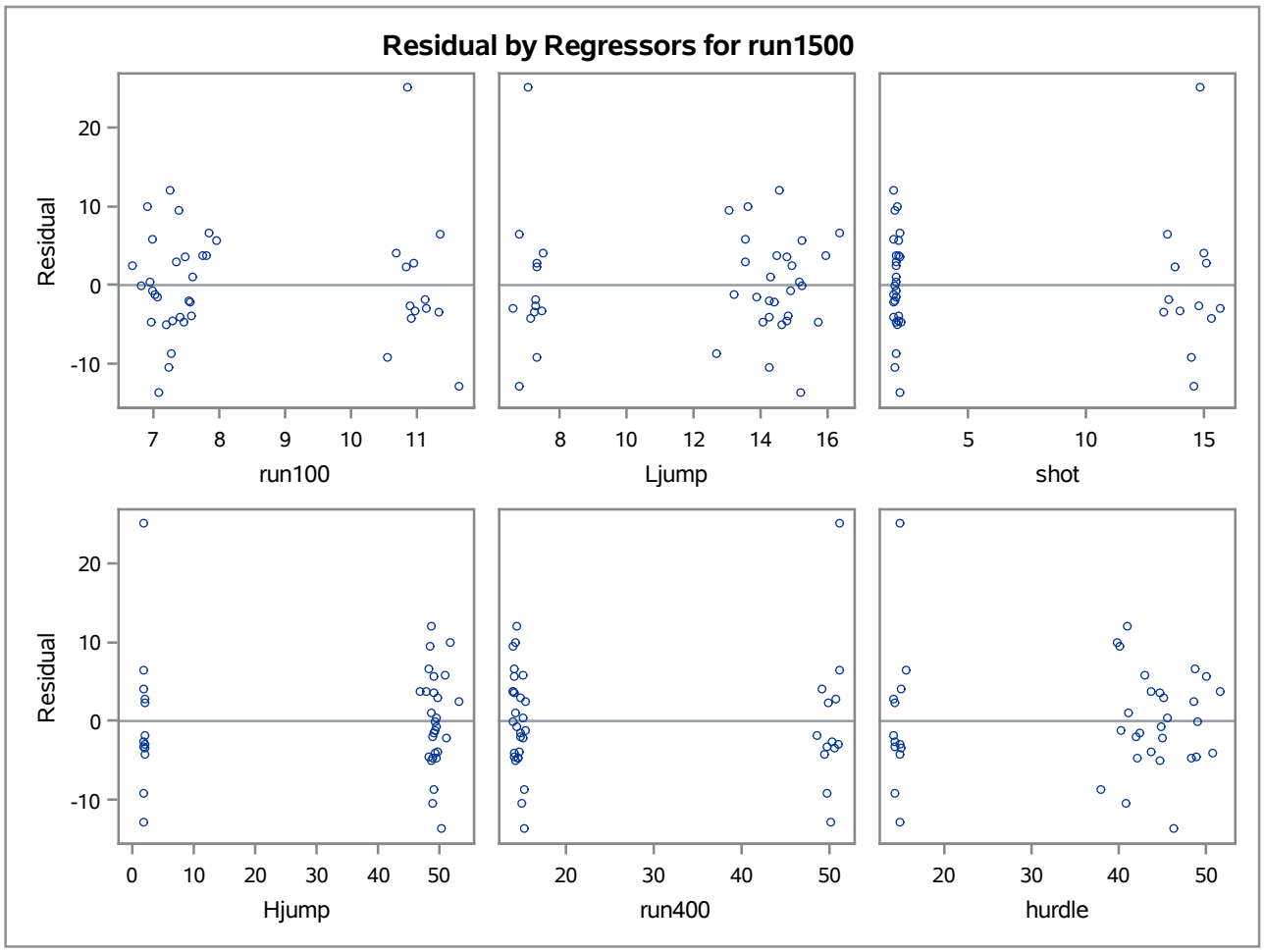
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	57.24623	100.84589	0.57	0.5744	0
run100	1	-14.19205	5.26869	-2.69	0.0113	54.07259
Ljump	1	-2.18717	2.83363	-0.77	0.4460	61.31661
shot	1	-2.81471	3.15344	-0.89	0.3790	211.72929
Hjump	1	0.05632	1.52467	0.04	0.9708	713.61440
run400	1	8.10473	1.57536	5.14	<.0001	428.66555
hurdle	1	0.39981	0.67888	0.59	0.5602	59.09633
discus	1	0.71353	0.93930	0.76	0.4532	182.48772
polevlt	1	-0.19900	0.37067	-0.54	0.5952	56.56097
javelin	1	-0.12662	0.18571	-0.68	0.5004	233.34061

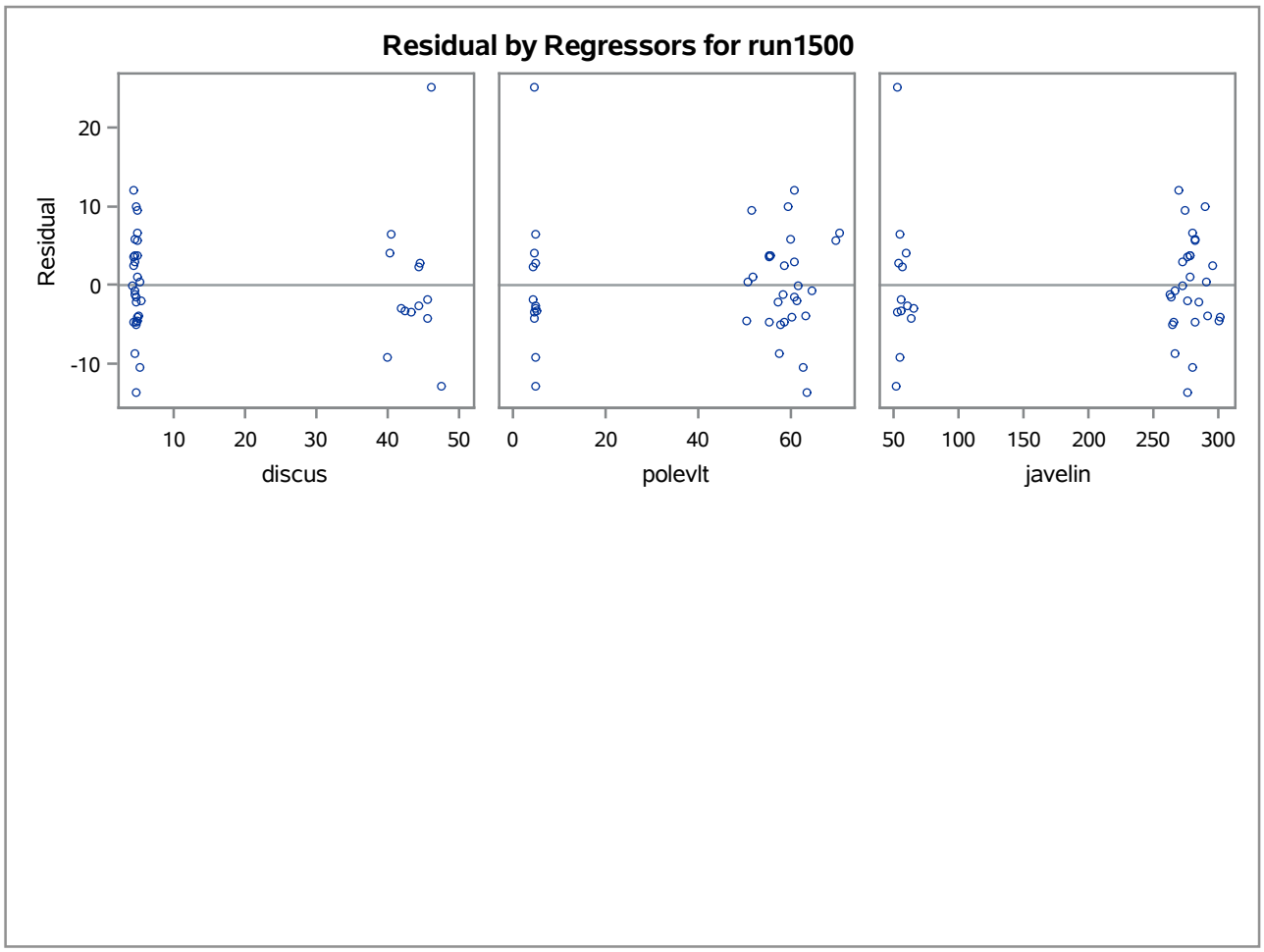
The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
1	7.58	14.83	2.07	49.81	14.69	43.75	5.02	63.19	291.70	1.00	0.00871
2	7.40	14.26	1.86	49.37	14.05	50.72	4.92	60.15	301.50	2.00	0.04018
3	7.30	14.77	2.04	48.37	14.09	48.95	4.92	50.31	300.20	3.00	0.02756
4	7.23	14.25	1.92	48.93	14.99	40.87	5.32	62.77	280.10	4.00	0.05907
5	7.09	15.19	2.10	50.42	15.31	46.26	4.72	63.44	276.40	5.00	0.06358
6	7.60	14.31	1.98	48.68	14.23	41.10	4.92	51.77	278.10	6.00	0.00061
7	11.13	7.30	13.48	2.01	48.62	14.17	45.67	4.42	55.37	268.00	0.00376
8	10.83	7.31	13.76	2.13	49.91	14.38	44.41	4.42	56.37	285.10	0.00218
9	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.10	0.29529
10	7.56	14.41	1.86	51.10	15.06	44.99	4.82	57.19	285.10	10.00	0.00269
11	6.97	14.09	1.95	49.48	14.48	42.10	4.72	55.40	282.00	11.00	0.00420
12	7.27	12.68	1.98	49.20	15.29	37.92	4.62	57.44	266.60	12.00	0.03908
13	11.36	6.80	13.46	1.86	51.16	15.67	40.49	5.02	54.68	291.70	0.12530
14	7.84	16.36	2.12	48.36	14.05	48.72	5.00	70.52	280.01	1.00	0.07069
15	7.96	15.23	2.06	49.19	14.13	50.11	4.90	69.71	282.00	2.00	0.03435
16	7.81	15.93	2.09	46.81	13.97	51.65	4.60	55.54	278.11	3.00	0.01434
17	7.47	15.73	2.15	48.97	14.56	48.34	4.40	58.46	265.42	4.00	0.02071
18	7.74	14.48	1.97	47.97	14.01	43.73	4.90	55.39	278.05	5.00	0.00431
19	10.91	7.14	15.31	2.12	49.40	14.95	45.62	4.70	63.45	269.54	0.01602
20	7.19	14.65	2.03	48.73	14.25	44.72	4.80	57.76	264.35	7.00	0.00748
21	7.53	14.26	1.88	48.81	14.80	42.05	5.40	61.33	276.33	8.00	0.00080
22	7.48	14.80	2.12	49.13	14.17	44.75	4.40	55.27	276.31	9.00	0.00343
23	10.98	7.49	14.01	1.94	49.76	14.25	42.43	5.10	56.32	273.56	0.00256
24	10.95	7.31	15.10	2.06	50.79	14.21	44.60	5.00	53.45	287.63	0.00242
25	10.90	7.30	14.77	1.88	50.30	14.34	44.41	5.00	60.89	278.82	0.00164
26	6.99	14.91	1.94	49.41	14.37	44.83	4.60	64.55	267.09	13.00	0.00021
27	6.81	15.24	1.91	49.27	14.01	49.02	4.20	61.52	272.74	14.00	0.00000
28	10.55	7.34	14.44	1.94	49.72	14.39	39.88	4.80	54.51	271.02	0.06987
29	10.68	7.50	14.97	1.94	49.12	15.01	40.35	4.60	59.26	275.71	0.01454
30	7.07	13.88	1.94	49.11	14.77	42.47	4.70	60.88	263.31	17.00	0.00060
31	7.34	13.55	1.97	49.65	14.78	45.13	4.50	60.79	272.63	18.00	0.00393
32	7.38	13.07	1.88	48.51	14.01	40.11	5.00	51.53	274.21	19.00	0.05903
33	11.14	6.61	15.69	2.03	51.04	14.88	41.90	4.80	65.82	277.94	0.01680
34	6.94	15.15	1.94	49.56	15.12	45.62	5.30	50.62	290.36	21.00	0.00011
35	7.26	14.57	1.85	48.61	14.41	40.95	4.40	60.71	269.70	22.00	0.04502
36	6.91	13.62	2.03	51.67	14.26	39.83	4.80	59.34	290.01	23.00	0.05000
37	7.03	13.22	1.85	49.34	15.38	40.22	4.50	58.36	263.08	24.00	0.00049
38	11.33	7.26	13.30	1.97	50.54	14.98	43.34	4.50	52.92	278.67	0.01150

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
39	10.86	7.07	14.81	1.94	51.16	14.96	46.07	4.70	53.05	317.00	0.23387
40	6.99	13.53	1.85	50.95	15.09	43.01	4.50	60.00	281.70	27.00	0.00927
41	6.68	14.92	1.94	53.20	15.39	48.66	4.40	58.62	296.12	28.00	0.01293

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
9	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.1	0.29529
13	11.36	6.80	13.46	1.86	51.16	15.67	40.49	5.02	54.68	291.7	0.12530
39	10.86	7.07	14.81	1.94	51.16	14.96	46.07	4.70	53.05	317.0	0.23387

Obs 9,39 and 41 have high cook's distance. Therefore, they are influential points.

The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Number of Observations Read	38
Number of Observations Used	38

Analysis of Variance

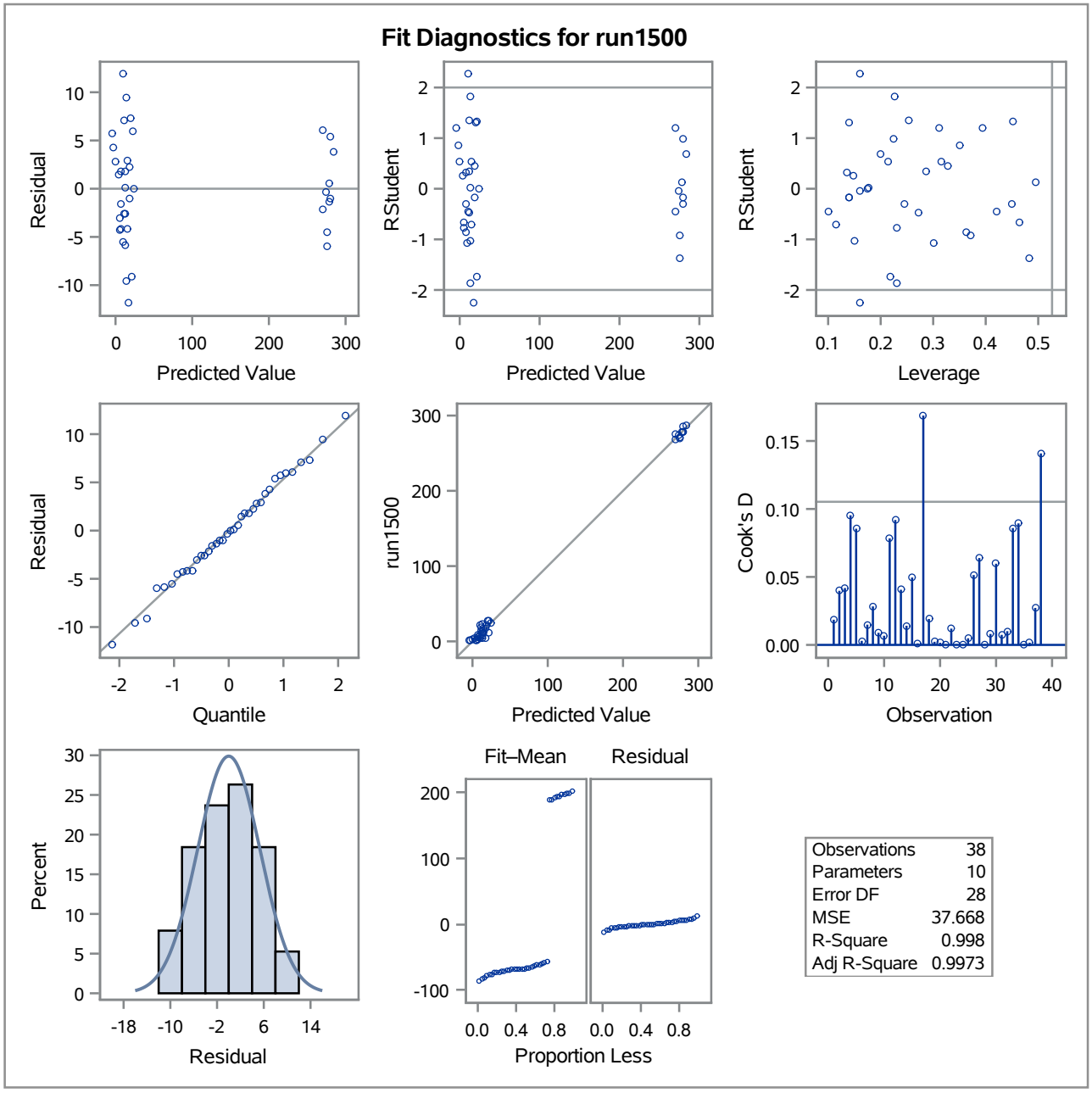
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	519521	57725	1532.44	<.0001
Error	28	1054.71187	37.66828		
Corrected Total	37	520575			

Root MSE	6.13745	**R-Square**	0.9980
Dependent Mean	81.18395	**Adj R-Sq**	0.9973
Coeff Var	7.55993		

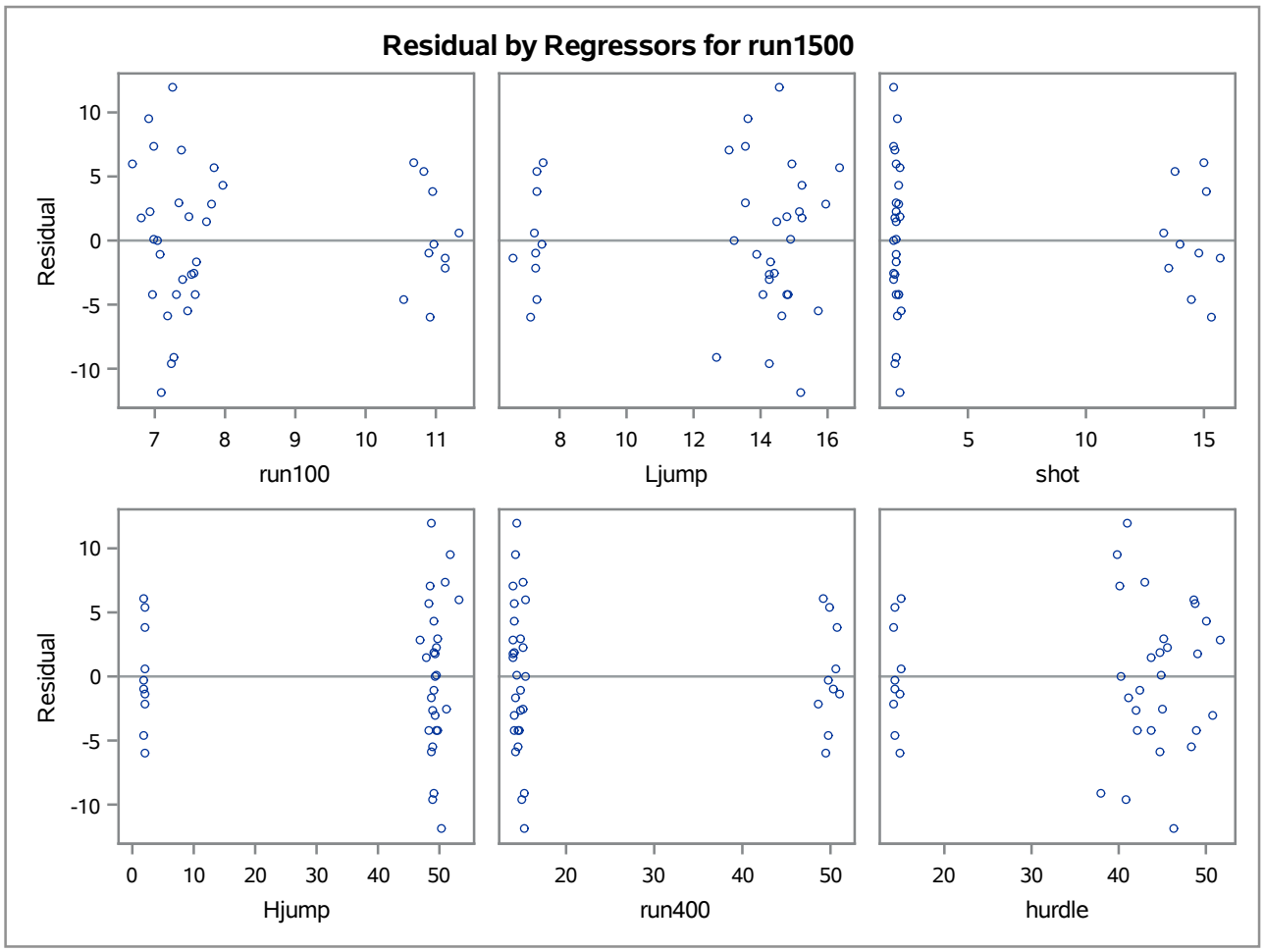
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	52.95603	79.89420	0.66	0.5129
run100	1	-10.68008	4.42349	-2.41	0.0225
Ljump	1	-2.27069	2.20229	-1.03	0.3113
shot	1	-0.55251	2.68293	-0.21	0.8383
Hjump	1	0.54813	1.27271	0.43	0.6700
run400	1	6.37092	1.43252	4.45	0.0001
hurdle	1	0.25824	0.53198	0.49	0.6311
discus	1	1.24166	0.93664	1.33	0.1957
polevlt	1	-0.28891	0.28416	-1.02	0.3180
javelin	1	-0.17867	0.14762	-1.21	0.2363

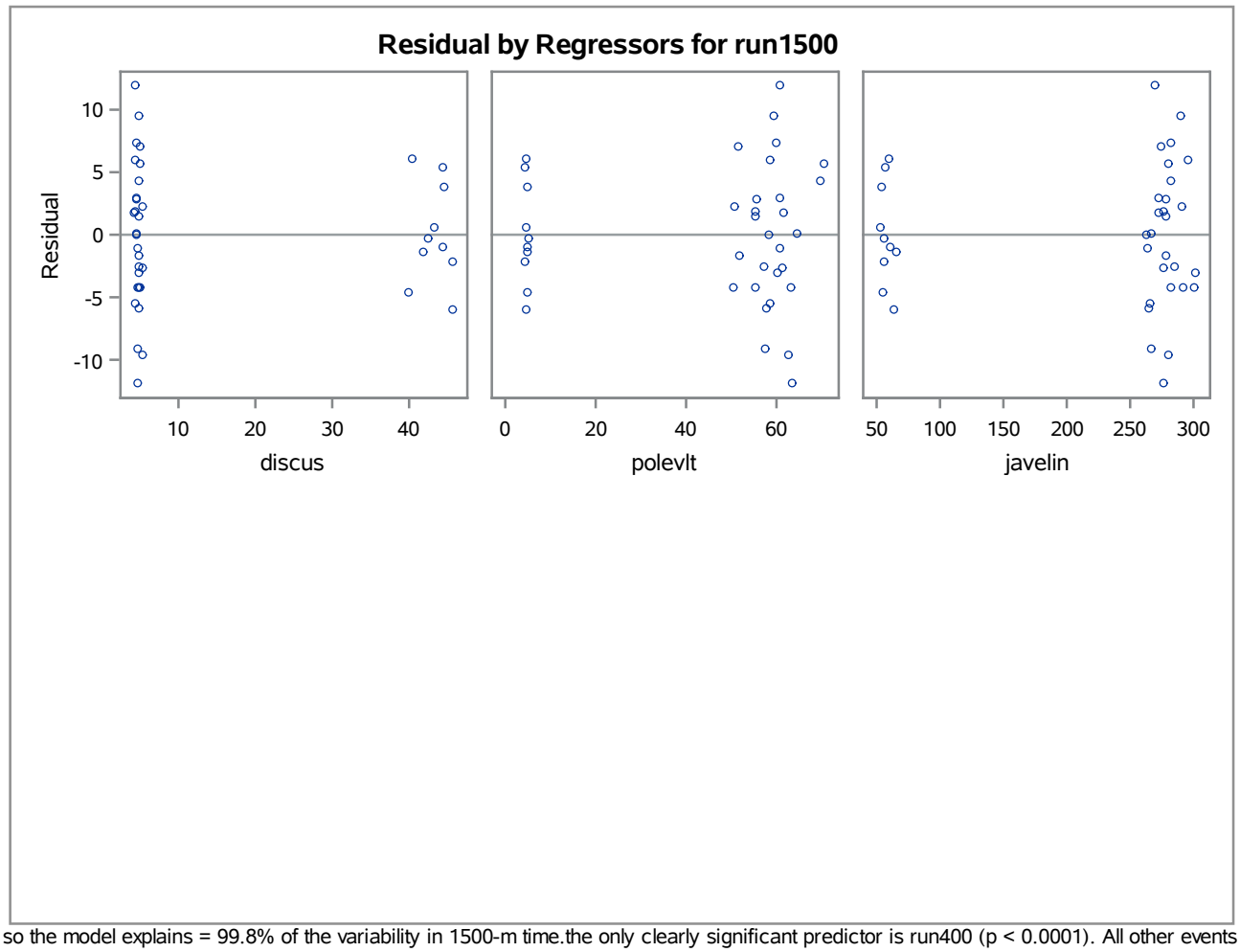
The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



R2 = 0.9981 so the model explains = 99.8% of the variability in 1500-m time.the only clearly significant predictor is run400 ($p < 0.0001$). All other events have p-values > 0.05.QQ/Residual plots: approximately normal.keep run400 (strong, stable, highly significant predictor).drop run100, Ljump, shot, Hjump, hurdle, discus, polevlt, javelin (all non-significant).
Exercise 4b

The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Number of Observations Read	41
Number of Observations Used	41

Forward Selection: Step 1

Variable run400 Entered: R-Square = 0.9953 and C(p) = 9.0611

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	640637	640637	8332.20	<.0001
Error	39	2998.59149	76.88696		
Corrected Total	40	643636			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-98.54861	2.53779	115942	1507.96	<.0001
run400	7.54808	0.08269	640637	8332.20	<.0001

Bounds on condition number: 1, 1

Forward Selection: Step 2

Variable run100 Entered: R-Square = 0.9965 and C(p) = -0.7004

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	641403	320702	5457.74	<.0001
Error	38	2232.91414	58.76090		
Corrected Total	40	643636			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-22.13355	21.28493	63.53983	1.08	0.3050
run100	-13.21161	3.65997	765.67735	13.03	0.0009
run400	8.92662	0.38867	30995	527.48	<.0001

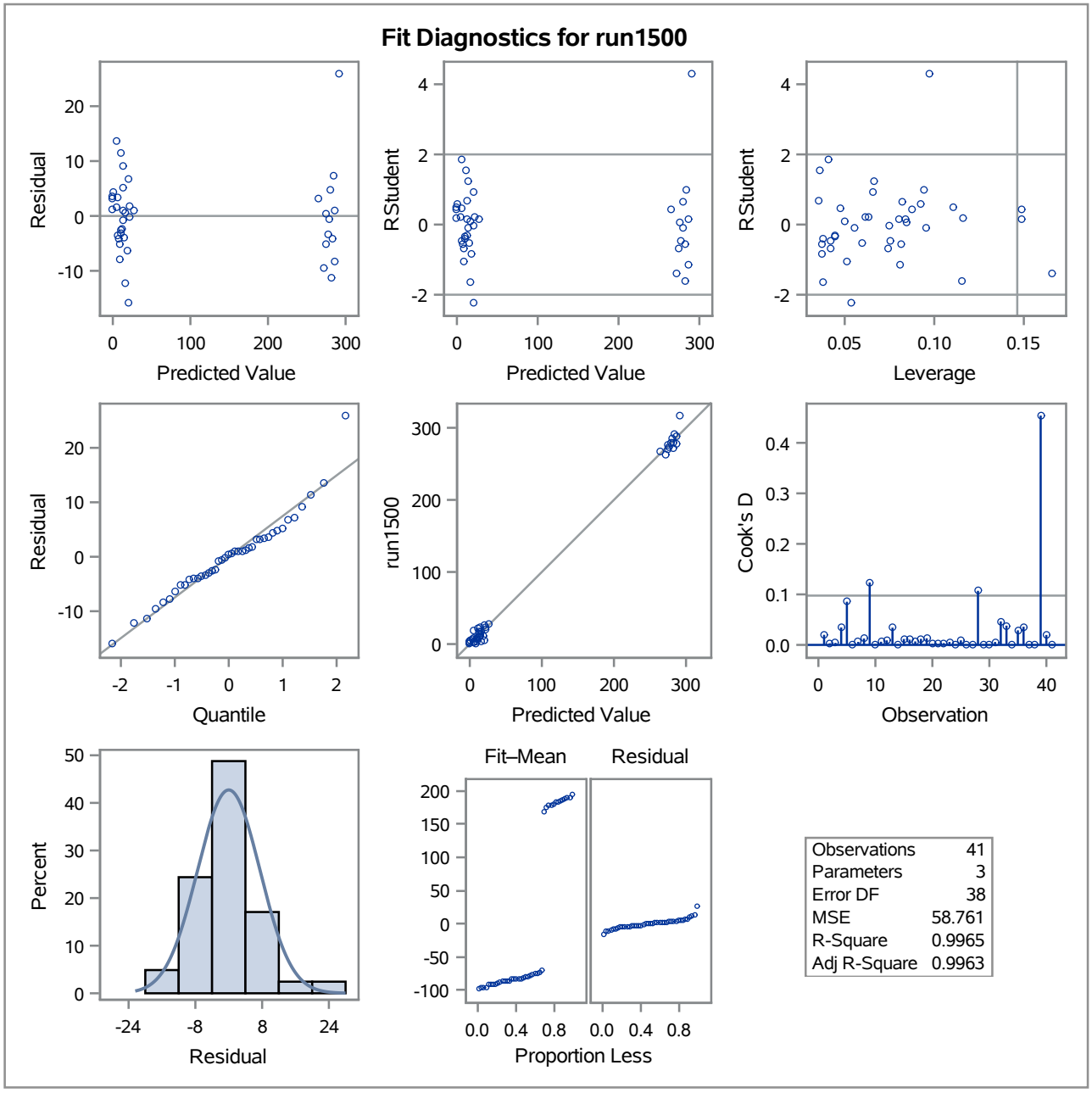
Bounds on condition number: 28.908, 115.63

No other variable met the 0.0500 significance level for entry into the model.

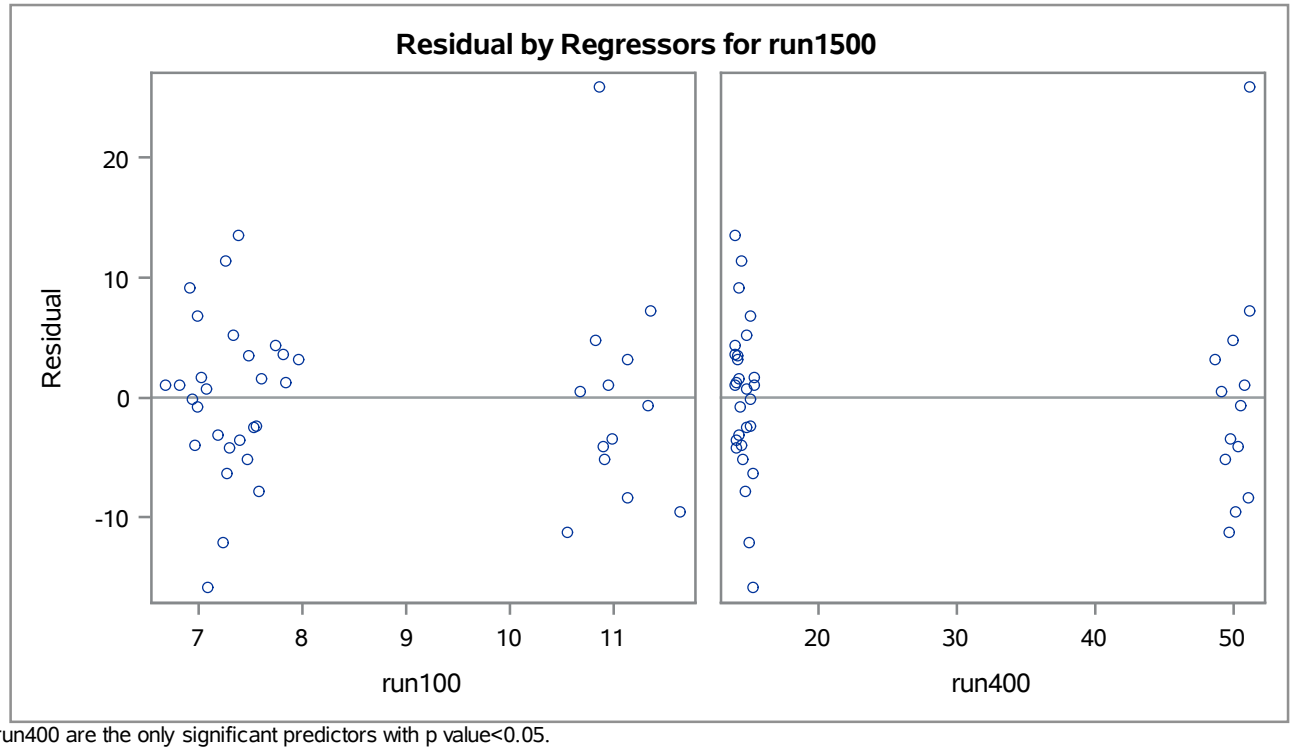
The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Summary of Forward Selection							
Step	**Variable Entered**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	run400	1	0.9953	0.9953	9.0611	8332.20	<.0001
2	run100	2	0.0012	0.9965	-0.7004	13.03	0.0009

The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



run100 and run400 are the only significant predictors with p value<0.05.

The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Number of Observations Read	41
Number of Observations Used	41

Analysis of Variance

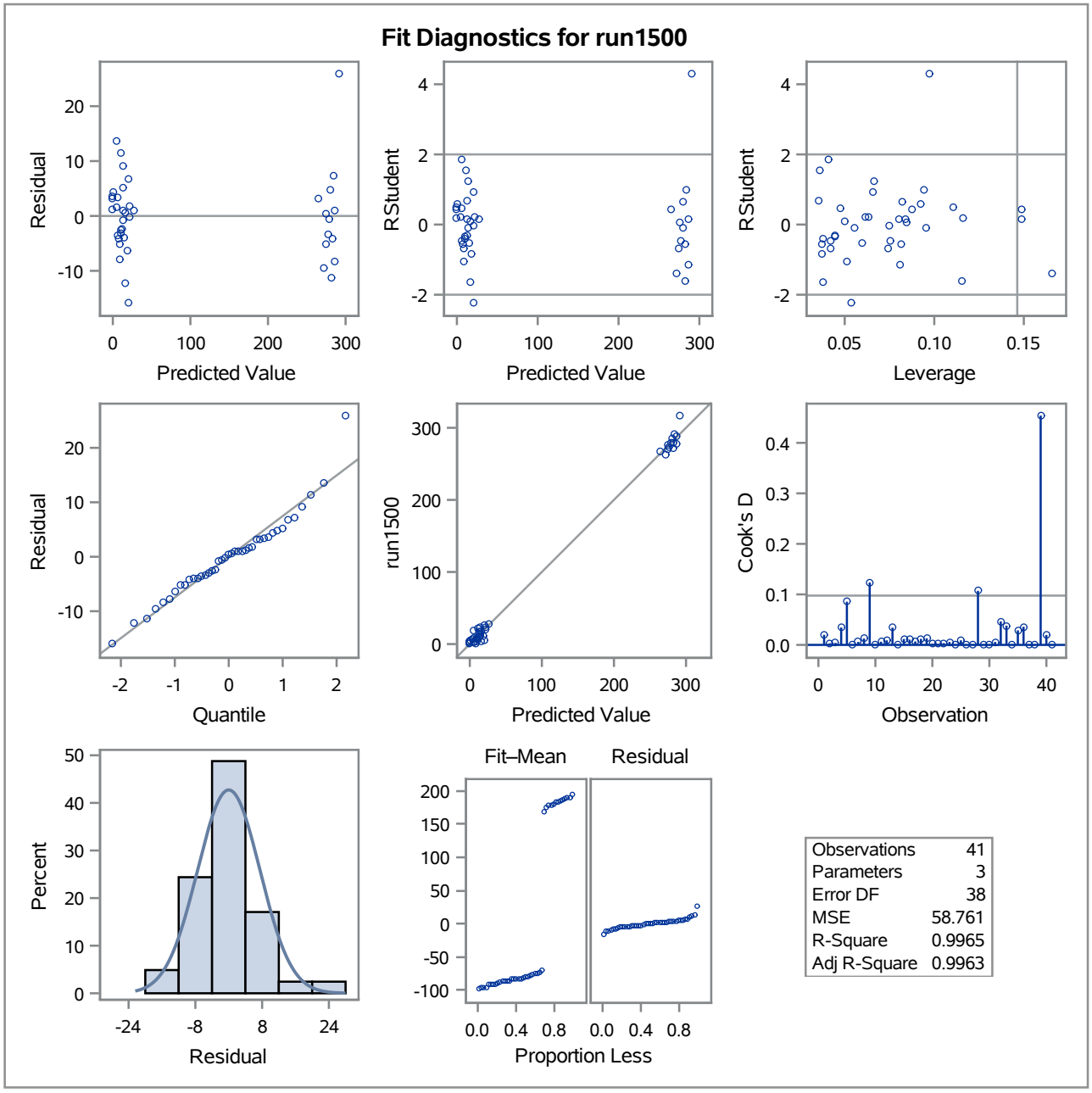
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	641403	320702	5457.74	<.0001
Error	38	2232.91414	58.76090		
Corrected Total	40	643636			

Root MSE	7.66557	**R-Square**	0.9965
Dependent Mean	96.48268	**Adj R-Sq**	0.9963
Coeff Var	7.94502		

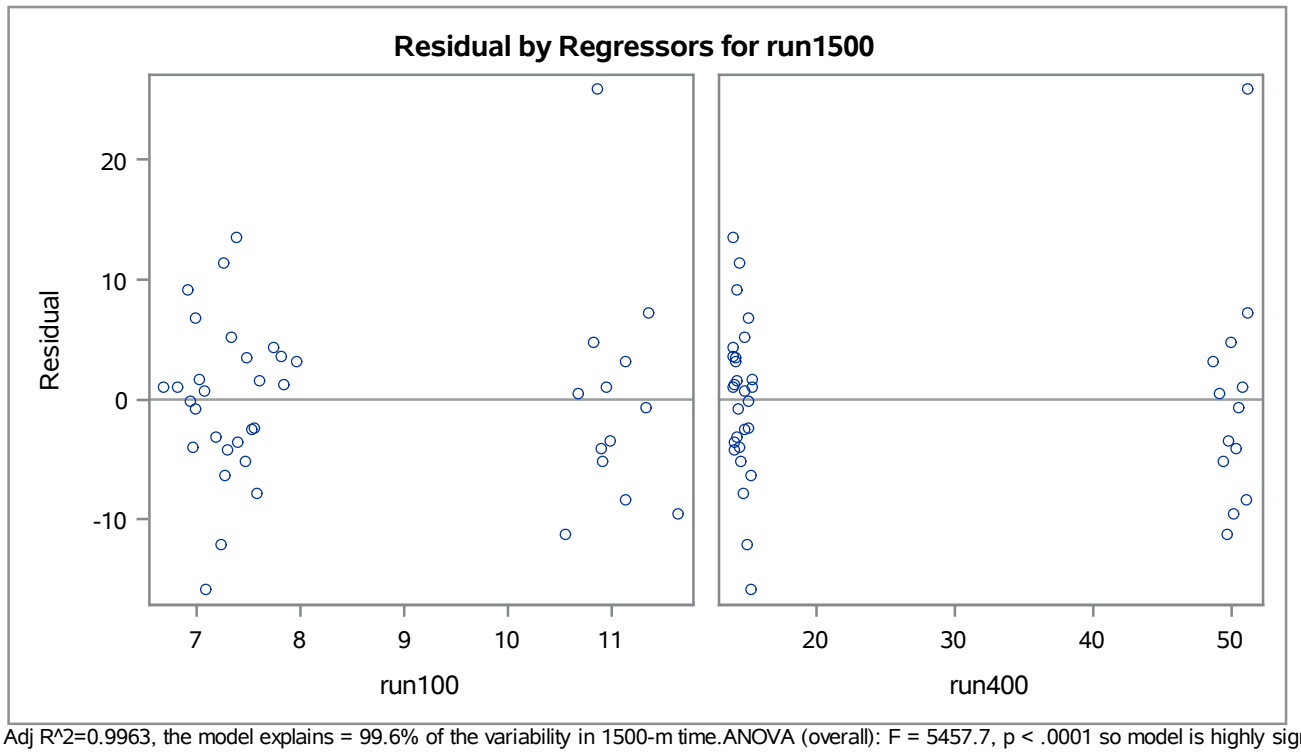
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-22.13355	21.28493	-1.04	0.3050
run100	1	-13.21161	3.65997	-3.61	0.0009
run400	1	8.92662	0.38867	22.97	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



R^2: 0.9965, Adj R^2=0.9963, the model explains = 99.6% of the variability in 1500-m time.ANOVA (overall): F = 5457.7, p < .0001 so model is highly significant.All the terms have p value < 0.05 so they are significant.The histogram plot looks roughly normal but there can be some influential points.

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
1	7.58	14.83	2.07	49.81	14.69	43.75	5.02	63.19	291.70	1.00	0.01992
2	7.40	14.26	1.86	49.37	14.05	50.72	4.92	60.15	301.50	2.00	0.00321
3	7.30	14.77	2.04	48.37	14.09	48.95	4.92	50.31	300.20	3.00	0.00400
4	7.23	14.25	1.92	48.93	14.99	40.87	5.32	62.77	280.10	4.00	0.03445
5	7.09	15.19	2.10	50.42	15.31	46.26	4.72	63.44	276.40	5.00	0.08543
6	7.60	14.31	1.98	48.68	14.23	41.10	4.92	51.77	278.10	6.00	0.00091
7	11.13	7.30	13.48	2.01	48.62	14.17	45.67	4.42	55.37	268.00	0.00598
8	10.83	7.31	13.76	2.13	49.91	14.38	44.41	4.42	56.37	285.10	0.01263
9	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.10	0.12358
10	7.56	14.41	1.86	51.10	15.06	44.99	4.82	57.19	285.10	10.00	0.00161
11	6.97	14.09	1.95	49.48	14.48	42.10	4.72	55.40	282.00	11.00	0.00627
12	7.27	12.68	1.98	49.20	15.29	37.92	4.62	57.44	266.60	12.00	0.00900
13	11.36	6.80	13.46	1.86	51.16	15.67	40.49	5.02	54.68	291.70	0.03411
14	7.84	16.36	2.12	48.36	14.05	48.72	5.00	70.52	280.01	1.00	0.00141
15	7.96	15.23	2.06	49.19	14.13	50.11	4.90	69.71	282.00	2.00	0.01167
16	7.81	15.93	2.09	46.81	13.97	51.65	4.60	55.54	278.11	3.00	0.01034
17	7.47	15.73	2.15	48.97	14.56	48.34	4.40	58.46	265.42	4.00	0.00689
18	7.74	14.48	1.97	47.97	14.01	43.73	4.90	55.39	278.05	5.00	0.01195
19	10.91	7.14	15.31	2.12	49.40	14.95	45.62	4.70	63.45	269.54	0.01304
20	7.19	14.65	2.03	48.73	14.25	44.72	4.80	57.76	264.35	7.00	0.00220
21	7.53	14.26	1.88	48.81	14.80	42.05	5.40	61.33	276.33	8.00	0.00172
22	7.48	14.80	2.12	49.13	14.17	44.75	4.40	55.27	276.31	9.00	0.00355
23	10.98	7.49	14.01	1.94	49.76	14.25	42.43	5.10	56.32	273.56	0.00588
24	10.95	7.31	15.10	2.06	50.79	14.21	44.60	5.00	53.45	287.63	0.00062
25	10.90	7.30	14.77	1.88	50.30	14.34	44.41	5.00	60.89	278.82	0.00900
26	6.99	14.91	1.94	49.41	14.37	44.83	4.60	64.55	267.09	13.00	0.00022
27	6.81	15.24	1.91	49.27	14.01	49.02	4.20	61.52	272.74	14.00	0.00058
28	10.55	7.34	14.44	1.94	49.72	14.39	39.88	4.80	54.51	271.02	0.10706
29	10.68	7.50	14.97	1.94	49.12	15.01	40.35	4.60	59.26	275.71	0.00013
30	7.07	13.88	1.94	49.11	14.77	42.47	4.70	60.88	263.31	17.00	0.00015
31	7.34	13.55	1.97	49.65	14.78	45.13	4.50	60.79	272.63	18.00	0.00576
32	7.38	13.07	1.88	48.51	14.01	40.11	5.00	51.53	274.21	19.00	0.04652
33	11.14	6.61	15.69	2.03	51.04	14.88	41.90	4.80	65.82	277.94	0.03810
34	6.94	15.15	1.94	49.56	15.12	45.62	5.30	50.62	290.36	21.00	0.00001
35	7.26	14.57	1.85	48.61	14.41	40.95	4.40	60.71	269.70	22.00	0.02880
36	6.91	13.62	2.03	51.67	14.26	39.83	4.80	59.34	290.01	23.00	0.03589
37	7.03	13.22	1.85	49.34	15.38	40.22	4.50	58.36	263.08	24.00	0.00121
38	11.33	7.26	13.30	1.97	50.54	14.98	43.34	4.50	52.92	278.67	0.00029

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
39	10.86	7.07	14.81	1.94	51.16	14.96	46.07	4.70	53.05	317.00	0.45400
40	6.99	13.53	1.85	50.95	15.09	43.01	4.50	60.00	281.70	27.00	0.01960
41	6.68	14.92	1.94	53.20	15.39	48.66	4.40	58.62	296.12	28.00	0.00118

Obs	run100	Ljump	shot	Hjump	run400	hurdle	discus	polevlt	javelin	run1500	cd
9	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.10	0.12358
28	10.55	7.34	14.44	1.94	49.72	14.39	39.88	4.80	54.51	271.02	0.10706
39	10.86	7.07	14.81	1.94	51.16	14.96	46.07	4.70	53.05	317.00	0.45400

Obs 9,39 and 28 have high cook's distance. Therefore, they are influential points.

The REG Procedure
Model: MODEL1
Dependent Variable: run1500

Number of Observations Read	38
Number of Observations Used	38

Analysis of Variance

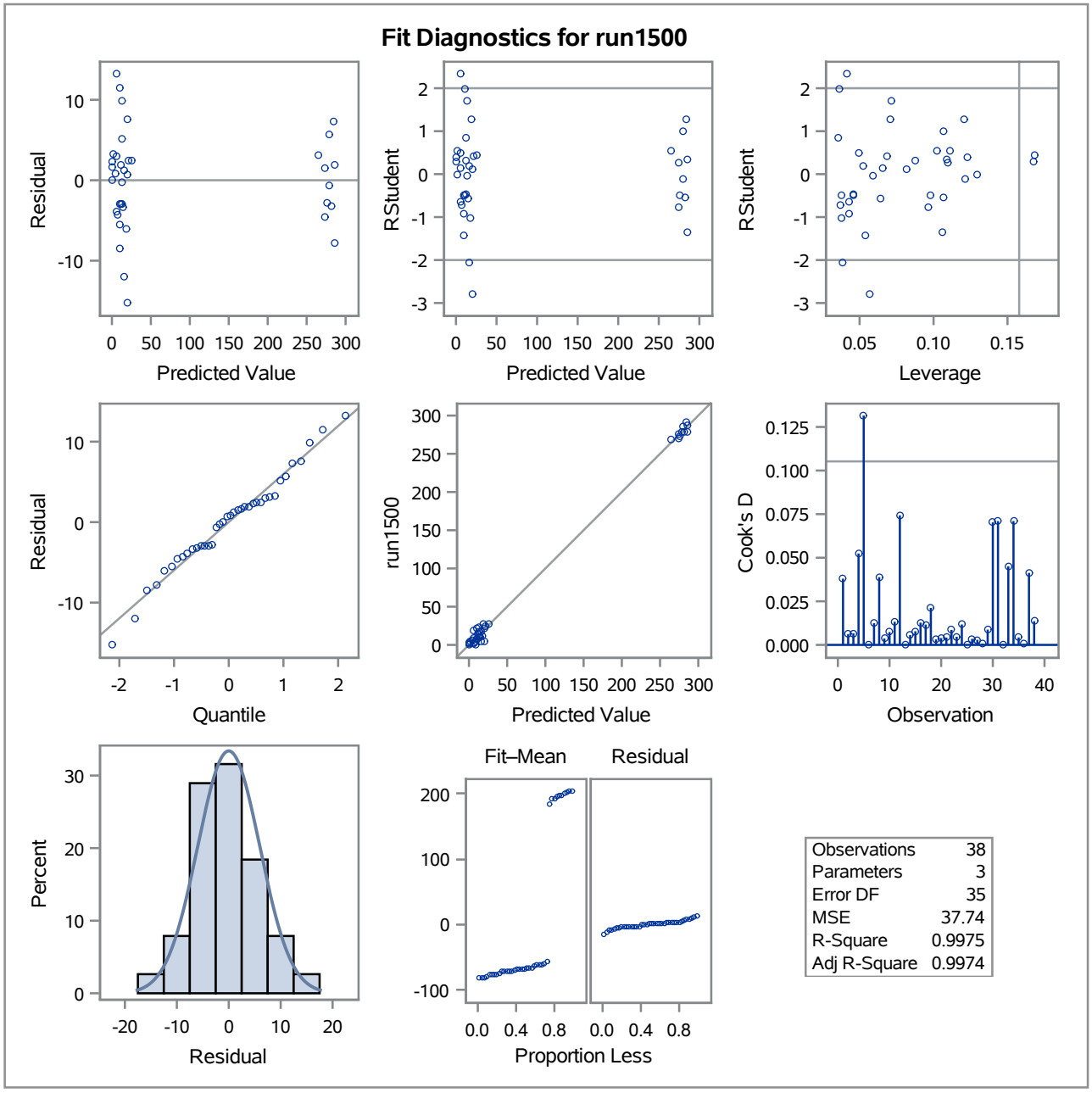
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	527522	263761	6988.88	<.0001
Error	35	1320.90360	37.74010		
Corrected Total	37	528843			

Root MSE	6.14330	**R-Square**	0.9975
Dependent Mean	81.72816	**Adj R-Sq**	0.9974
Coeff Var	7.51675		

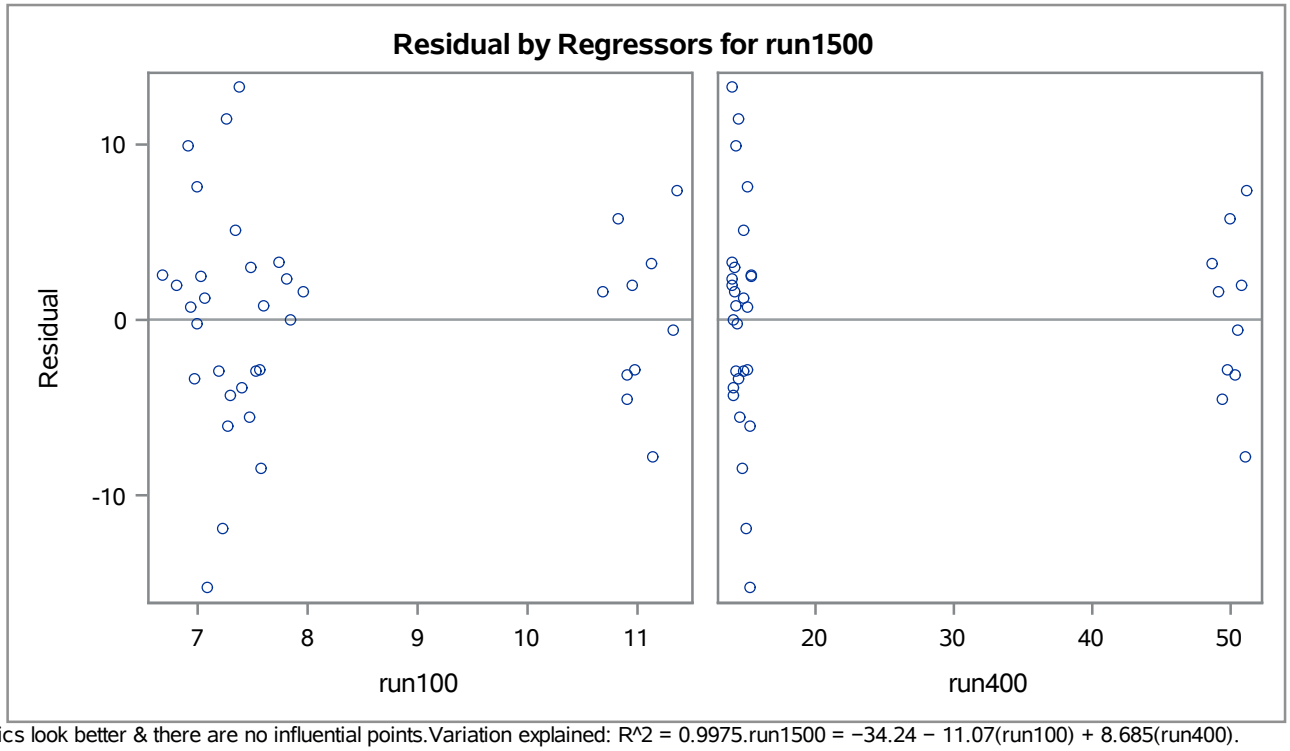
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-34.23835	18.43081	-1.86	0.0716
run100	1	-11.06790	3.17246	-3.49	0.0013
run400	1	8.68491	0.33765	25.72	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The REG Procedure
Model: MODEL1
Dependent Variable: run1500



The diagnostics look better & there are no influential points.Variation explained: R^2 = 0.9975.run1500 = −34.24 − 11.07(run100) + 8.685(run400).