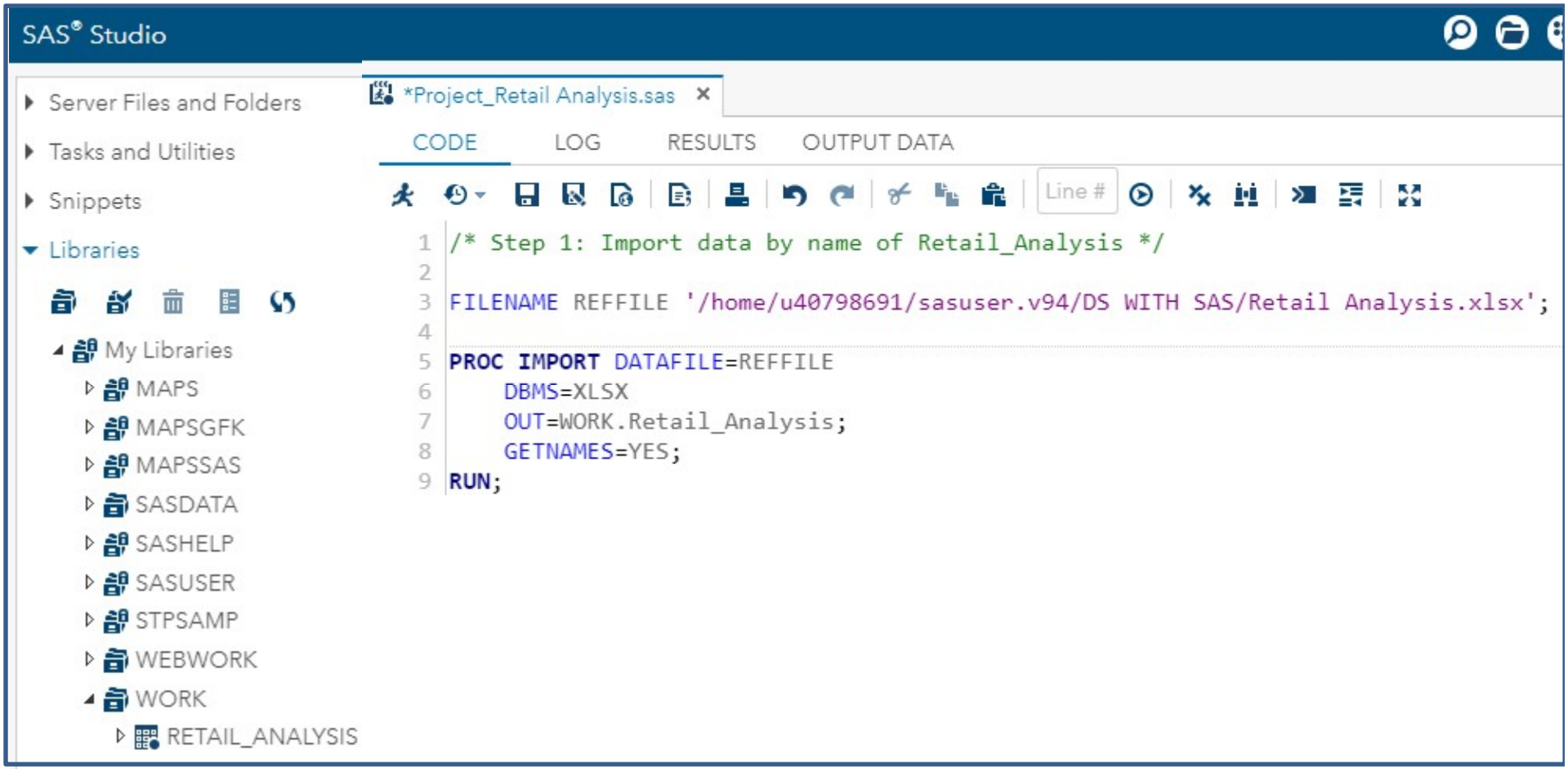
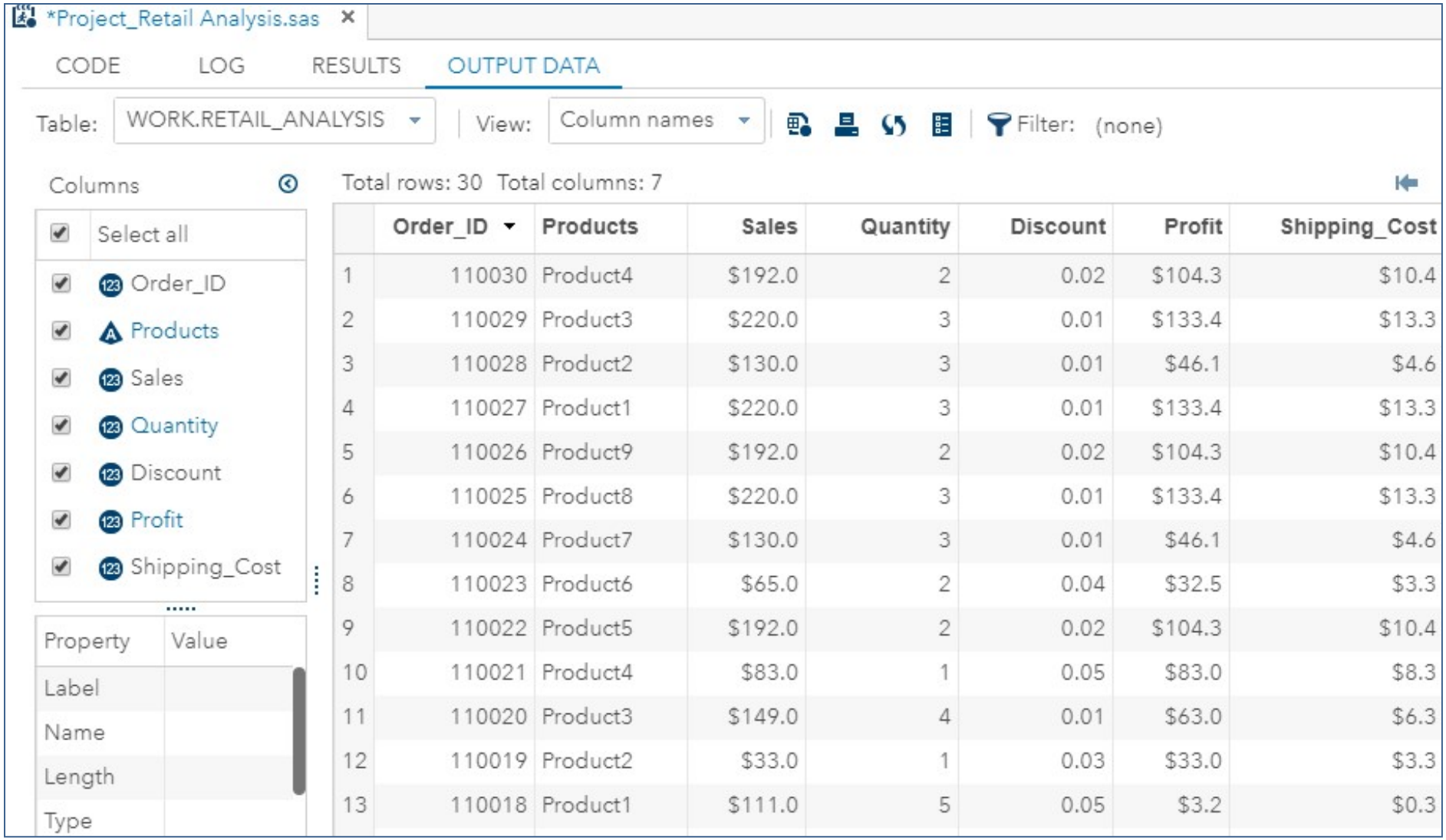
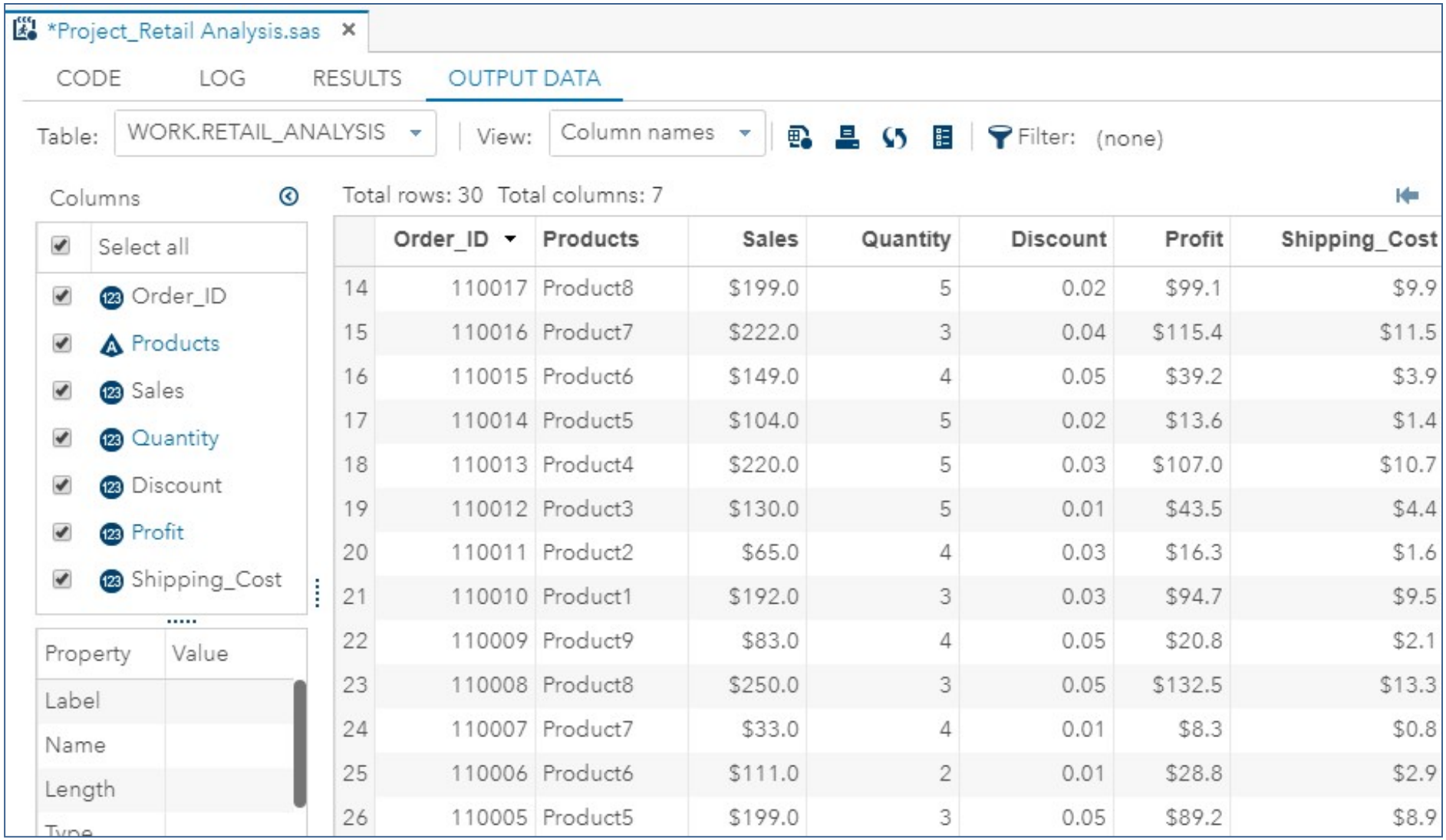
Project : Retail Analysis

1. Import the required dataset


SAS® Studio
Server Files and Folders
Tasks and Utilities
Snippets
Libraries
My Libraries
MAPS
MAPSGFK
MAPSSAS
SASDATA
SASHELP
SASUSER
STPSAMP
WEBWORK
WORK
RETAIL_ANALYSIS
*Project_Retail Analysis.sas
CODE
LOG
RESULTS
OUTPUT DATA
Line #
1 /* Step 1: Import data by name of Retail_Analysis */
2
3 FILENAME REFFILE '/home/u40798691/sasuser.v94/DS WITH SAS/Retail Analysis.xlsx';
4
5 PROC IMPORT DATAFILE=REFFILE
6 DBMS=XLSX
7 OUT=WORK.Retail_Analysis;
8 GETNAMES=YES;
9 RUN;



	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost
1	110030	Product4	$192.0	2	0.02	$104.3	$10.4
2	110029	Product3	$220.0	3	0.01	$133.4	$13.3
3	110028	Product2	$130.0	3	0.01	$46.1	$4.6
4	110027	Product1	$220.0	3	0.01	$133.4	$13.3
5	110026	Product9	$192.0	2	0.02	$104.3	$10.4
6	110025	Product8	$220.0	3	0.01	$133.4	$13.3
7	110024	Product7	$130.0	3	0.01	$46.1	$4.6
8	110023	Product6	$65.0	2	0.04	$32.5	$3.3
9	110022	Product5	$192.0	2	0.02	$104.3	$10.4
10	110021	Product4	$83.0	1	0.05	$83.0	$8.3
11	110020	Product3	$149.0	4	0.01	$63.0	$6.3
12	110019	Product2	$33.0	1	0.03	$33.0	$3.3
13	110018	Product1	$111.0	5	0.05	$3.2	$0.3



	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost
14	110017	Product8	$199.0	5	0.02	$99.1	$9.9
15	110016	Product7	$222.0	3	0.04	$115.4	$11.5
16	110015	Product6	$149.0	4	0.05	$39.2	$3.9
17	110014	Product5	$104.0	5	0.02	$13.6	$1.4
18	110013	Product4	$220.0	5	0.03	$107.0	$10.7
19	110012	Product3	$130.0	5	0.01	$43.5	$4.4
20	110011	Product2	$65.0	4	0.03	$16.3	$1.6
21	110010	Product1	$192.0	3	0.03	$94.7	$9.5
22	110009	Product9	$83.0	4	0.05	$20.8	$2.1
23	110008	Product8	$250.0	3	0.05	$132.5	$13.3
24	110007	Product7	$33.0	4	0.01	$8.3	$0.8
25	110006	Product6	$111.0	2	0.01	$28.8	$2.9
26	110005	Product5	$199.0	3	0.05	$89.2	$8.9

```
14 PROC PRINT DATA=Retail_Analysis;
15 RUN;
```

Obs	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6
2	110002	Product2	$104.0	1	0.03	$20.9	$2.1
3	110003	Product3	$149.0	4	0.01	$63.0	$6.3
4	110004	Product4	$222.0	4	0.03	$115.4	$11.5
5	110005	Product5	$199.0	3	0.05	$89.2	$8.9
6	110006	Product6	$111.0	2	0.01	$28.8	$2.9
7	110007	Product7	$33.0	4	0.01	$8.3	$0.8
8	110008	Product8	$250.0	3	0.05	$132.5	$13.3
9	110009	Product9	$83.0	4	0.05	$20.8	$2.1
10	110010	Product1	$192.0	3	0.03	$94.7	$9.5
11	110011	Product2	$65.0	4	0.03	$16.3	$1.6
12	110012	Product3	$130.0	5	0.01	$43.5	$4.4
13	110013	Product4	$220.0	5	0.03	$107.0	$10.7
14	110014	Product5	$104.0	5	0.02	$13.6	$1.4
15	110015	Product6	$149.0	4	0.05	$39.2	$3.9
16	110016	Product7	$222.0	3	0.04	$115.4	$11.5
17	110017	Product8	$199.0	5	0.02	$99.1	$9.9
18	110018	Product1	$111.0	5	0.05	$3.2	$0.3
19	110019	Product2	$33.0	1	0.03	$33.0	$3.3
20	110020	Product3	$149.0	4	0.01	$63.0	$6.3
21	110021	Product4	$83.0	1	0.05	$83.0	$8.3
22	110022	Product5	$192.0	2	0.02	$104.3	$10.4
23	110023	Product6	$65.0	2	0.04	$32.5	$3.3
24	110024	Product7	$130.0	3	0.01	$46.1	$4.6
25	110025	Product8	$220.0	3	0.01	$133.4	$13.3
26	110026	Product9	$192.0	2	0.02	$104.3	$10.4
27	110027	Product1	$220.0	3	0.01	$133.4	$13.3
28	110028	Product2	$130.0	3	0.01	$46.1	$4.6
29	110029	Product3	$220.0	3	0.01	$133.4	$13.3
30	110030	Product4	$192.0	2	0.02	$104.3	$10.4

2. Perform descriptive statistics for the dataset

CONTENTS

```
11 PROC CONTENTS DATA=WORK.Retail_Analysis;
12 RUN;
```

The CONTENTS Procedure

Data Set Name	WORK.RETAIL_ANALYSIS	Observations	30
Member Type	DATA	Variables	7
Engine	V9	Indexes	0
Created	08/04/2019 16:48:09	Observation Length	56
Last Modified	08/04/2019 16:48:09	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	2334
Obs in First Data Page	30
Number of Data Set Repairs	0
Filename	/saswork/SAS_work5C9D0001C5C9_odaws02-prod-sg/SAS_work69570001C5C9_odaws02-prod-sg/retail_analysis.sas7bdat
Release Created	9.0401M5
Host Created	Linux
Inode Number	1610614235
Access Permission	rw-r--r--
Owner Name	u40798691
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
5	Discount	Num	8	BEST.		Discount
1	Order_ID	Num	8	BEST.		Order_ID
2	Products	Char	8	$8.	$8.	Products
6	Profit	Num	8	NLMNY15.1		Profit
4	Quantity	Num	8	BEST.		Quantity
3	Sales	Num	8	NLMNY15.1		Sales
7	Shipping_Cost	Num	8	NLMNY15.1		Shipping_Cost

CORRELATION

```
17  PROC CORR DATA = Retail_Analysis;
18  RUN;
```

The CORR Procedure

6 Variables:	Order_ID Sales Quantity Discount Profit Shipping_Cost

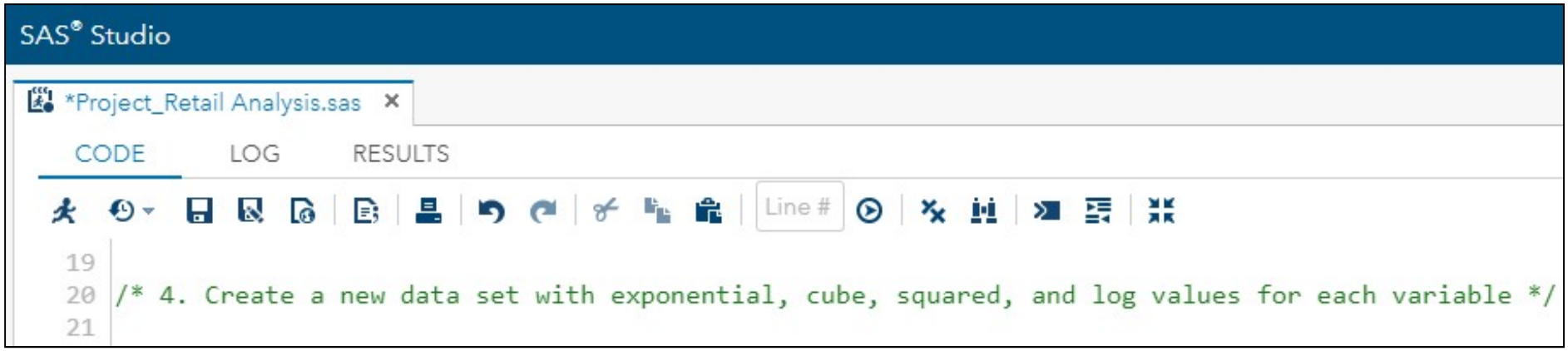
Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Order_ID	30	110016	8.80341	3300465	110001	110030	Order_ID
Sales	30	152.96667	63.17599	4589	33.00000	250.00000	Sales
Quantity	30	3.16667	1.23409	95.00000	1.00000	5.00000	Quantity
Discount	30	0.02567	0.01547	0.77000	0.01000	0.05000	Discount
Profit	30	72.10633	44.60090	2163	3.25000	135.60000	Profit
Shipping_Cost	30	7.21063	4.46009	216.31900	0.32500	13.56000	Shipping_Cost

Pearson Correlation Coefficients, N = 30
Prob > |r| under H0: Rho=0

	Order_ID	Sales	Quantity	Discount	Profit	Shipping_Cost
Order_ID Order_ID	1.00000	0.08386 0.6595	-0.16663 0.3788	-0.20388 0.2799	0.20600 0.2748	0.20600 0.2748
Sales Sales	0.08386 0.6595	1.00000	0.08897 0.6401	-0.09791 0.6067	0.89522 <.0001	0.89522 <.0001
Quantity Quantity	-0.16663 0.3788	0.08897 0.6401	1.00000	-0.03312 0.8621	-0.17876 0.3446	-0.17876 0.3446
Discount Discount	-0.20388 0.2799	-0.09791 0.6067	-0.03312 0.8621	1.00000	-0.13694 0.4705	-0.13694 0.4705
Profit Profit	0.20600 0.2748	0.89522 <.0001	-0.17876 0.3446	-0.13694 0.4705	1.00000	1.00000 <.0001
Shipping_Cost Shipping_Cost	0.20600 0.2748	0.89522 <.0001	-0.17876 0.3446	-0.13694 0.4705	1.00000 <.0001	1.00000

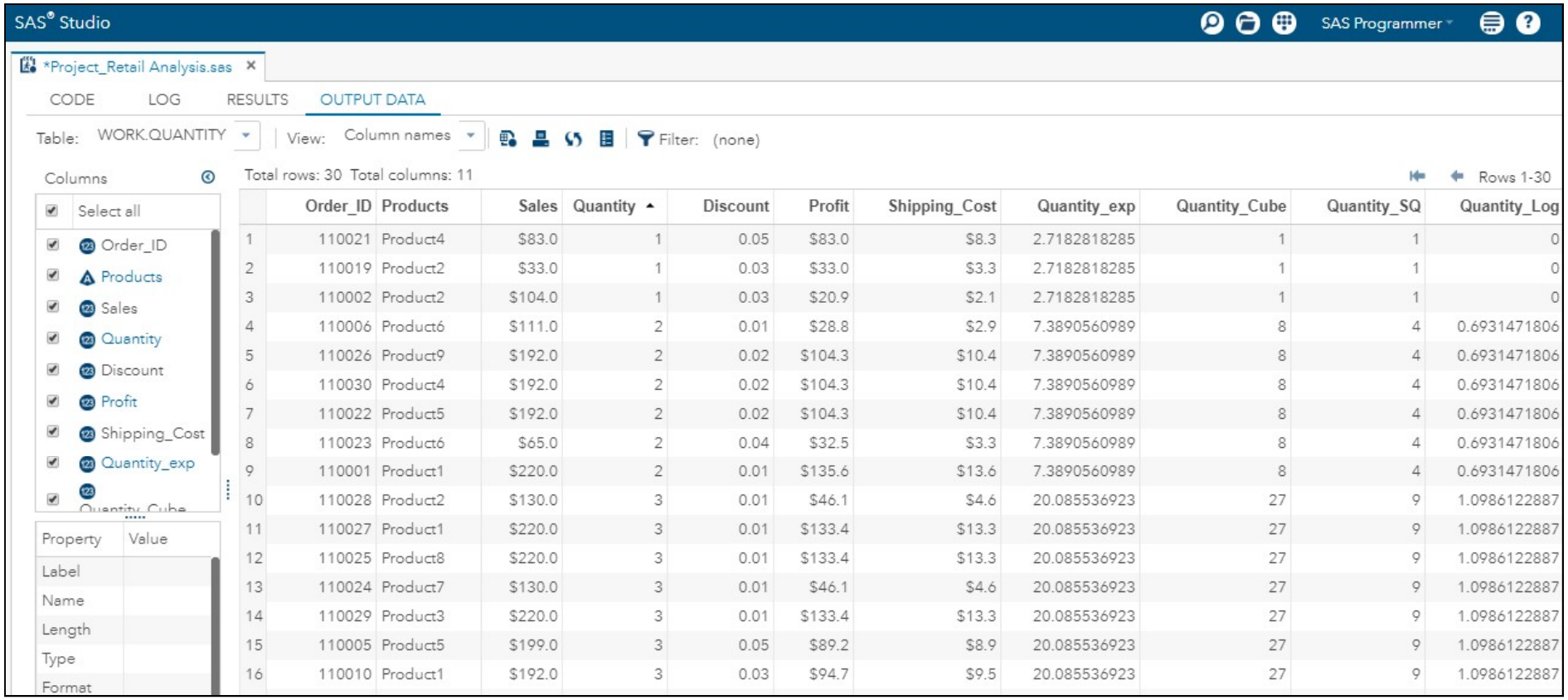
3. Create a new data set with exponential, cube, squared, and log values for each variable




```
23 DATA QUANTITY;
24 SET Retail_Analysis;
25 Quantity_exp = exp(Quantity);
26 Quantity_Cube = Quantity*Quantity*Quantity;
27 Quantity_SQ = Quantity*Quantity;
28 Quantity_Log = log(Quantity);
29 RUN;
30
31 DATA DISCOUNT;
32 SET Retail_Analysis;
33 Discount_exp = exp(Discount);
34 Discount_Cube = Discount*Discount*Discount;
35 Discount_SQ = Discount*Discount;
36 Discount_Log = log(Discount);
37 RUN;
```

```
39 DATA PROFIT;
40 SET Retail_Analysis;
41 Profit_exp = exp(Profit);
42 Profit_Cube = Profit*Profit*Profit;
43 Profit_SQ = Profit*Profit;
44 Profit_Log = log(Profit);
45 RUN;
46
47 DATA SHIPPING_COST;
48 SET Retail_Analysis;
49 Shipping_Cost_exp = exp(Shipping_Cost);
50 Shipping_Cost_Cube = Shipping_Cost*Shipping_Cost*Shipping_Cost;
51 Shipping_Cost_SQ = Shipping_Cost*Shipping_Cost;
52 Shipping_Cost_Log = log(Shipping_Cost);
53 RUN;
```

Quantity



	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Quantity_exp	Quantity_Cube	Quantity_SQ	Quantity_Log
1	110021	Product4	$83.0	1	0.05	$83.0	$8.3	2.7182818285	1	1	0
2	110019	Product2	$33.0	1	0.03	$33.0	$3.3	2.7182818285	1	1	0
3	110002	Product2	$104.0	1	0.03	$20.9	$2.1	2.7182818285	1	1	0
4	110006	Product6	$111.0	2	0.01	$28.8	$2.9	7.3890560989	8	4	0.6931471806
5	110026	Product9	$192.0	2	0.02	$104.3	$10.4	7.3890560989	8	4	0.6931471806
6	110030	Product4	$192.0	2	0.02	$104.3	$10.4	7.3890560989	8	4	0.6931471806
7	110022	Product5	$192.0	2	0.02	$104.3	$10.4	7.3890560989	8	4	0.6931471806
8	110023	Product6	$65.0	2	0.04	$32.5	$3.3	7.3890560989	8	4	0.6931471806
9	110001	Product1	$220.0	2	0.01	$135.6	$13.6	7.3890560989	8	4	0.6931471806
10	110028	Product2	$130.0	3	0.01	$46.1	$4.6	20.085536923	27	9	1.0986122887
11	110027	Product1	$220.0	3	0.01	$133.4	$13.3	20.085536923	27	9	1.0986122887
12	110025	Product8	$220.0	3	0.01	$133.4	$13.3	20.085536923	27	9	1.0986122887
13	110024	Product7	$130.0	3	0.01	$46.1	$4.6	20.085536923	27	9	1.0986122887
14	110029	Product3	$220.0	3	0.01	$133.4	$13.3	20.085536923	27	9	1.0986122887
15	110005	Product5	$199.0	3	0.05	$89.2	$8.9	20.085536923	27	9	1.0986122887
16	110010	Product1	$192.0	3	0.03	$94.7	$9.5	20.085536923	27	9	1.0986122887

17	110008	Product8	$250.0	3	0.05	$132.5	$13.3	20.085536923	27	9	1.0986122887
18	110016	Product7	$222.0	3	0.04	$115.4	$11.5	20.085536923	27	9	1.0986122887
19	110015	Product6	$149.0	4	0.05	$39.2	$3.9	54.598150033	64	16	1.3862943611
20	110007	Product7	$33.0	4	0.01	$8.3	$0.8	54.598150033	64	16	1.3862943611
21	110003	Product3	$149.0	4	0.01	$63.0	$6.3	54.598150033	64	16	1.3862943611
22	110011	Product2	$65.0	4	0.03	$16.3	$1.6	54.598150033	64	16	1.3862943611
23	110020	Product3	$149.0	4	0.01	$63.0	$6.3	54.598150033	64	16	1.3862943611
24	110009	Product9	$83.0	4	0.05	$20.8	$2.1	54.598150033	64	16	1.3862943611
25	110004	Product4	$222.0	4	0.03	$115.4	$11.5	54.598150033	64	16	1.3862943611
26	110014	Product5	$104.0	5	0.02	$13.6	$1.4	148.4131591	125	25	1.6094379124
27	110013	Product4	$220.0	5	0.03	$107.0	$10.7	148.4131591	125	25	1.6094379124
28	110012	Product3	$130.0	5	0.01	$43.5	$4.4	148.4131591	125	25	1.6094379124
29	110018	Product1	$111.0	5	0.05	$3.2	$0.3	148.4131591	125	25	1.6094379124
30	110017	Product8	$199.0	5	0.02	$99.1	$9.9	148.4131591	125	25	1.6094379124

Discount

Total rows: 30 Total columns: 11 Rows 1-30

	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Discount_exp	Discount_Cube	Discount_SQ	Discount_Log
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6	1.0100501671	1E-6	0.0001	-4.605170186
2	110002	Product2	$104.0	1	0.03	$20.9	$2.1	1.030454534	0.000027	0.0009	-3.506557897
3	110003	Product3	$149.0	4	0.01	$63.0	$6.3	1.0100501671	1E-6	0.0001	-4.605170186
4	110004	Product4	$222.0	4	0.03	$115.4	$11.5	1.030454534	0.000027	0.0009	-3.506557897
5	110005	Product5	$199.0	3	0.05	$89.2	$8.9	1.0512710964	0.000125	0.0025	-2.995732274
6	110006	Product6	$111.0	2	0.01	$28.8	$2.9	1.0100501671	1E-6	0.0001	-4.605170186
7	110007	Product7	$33.0	4	0.01	$8.3	$0.8	1.0100501671	1E-6	0.0001	-4.605170186
8	110008	Product8	$250.0	3	0.05	$132.5	$13.3	1.0512710964	0.000125	0.0025	-2.995732274
9	110009	Product9	$83.0	4	0.05	$20.8	$2.1	1.0512710964	0.000125	0.0025	-2.995732274
10	110010	Product1	$192.0	3	0.03	$94.7	$9.5	1.030454534	0.000027	0.0009	-3.506557897
11	110011	Product2	$65.0	4	0.03	$16.3	$1.6	1.030454534	0.000027	0.0009	-3.506557897
12	110012	Product3	$130.0	5	0.01	$43.5	$4.4	1.0100501671	1E-6	0.0001	-4.605170186
13	110013	Product4	$220.0	5	0.03	$107.0	$10.7	1.030454534	0.000027	0.0009	-3.506557897
14	110014	Product5	$104.0	5	0.02	$13.6	$1.4	1.02020134	8E-6	0.0004	-3.912023005
15	110015	Product6	$149.0	4	0.05	$39.2	$3.9	1.0512710964	0.000125	0.0025	-2.995732274
16	110016	Product7	$222.0	3	0.04	$115.4	$11.5	1.0408107742	0.000064	0.0016	-3.218875825

Profit

Total rows: 30 Total columns: 11

	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Profit_exp	Profit_Cube	Profit_SQ	Profit_Log
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6	7.768403E58	2493326.016	18387.36	4.9097093755
2	110002	Product2	$104.0	1	0.03	$20.9	$2.1	1169684627.2	9103.145472	435.9744	3.038791763
3	110003	Product3	$149.0	4	0.01	$63.0	$6.3	2.3873942E27	250523.58246	3974.0416	4.1437694455
4	110004	Product4	$222.0	4	0.03	$115.4	$11.5	1.2595385E50	1535202.7587	13307.9296	4.7480576735
5	110005	Product5	$199.0	3	0.05	$89.2	$8.9	5.2161859E38	708539.46088	7947.7225	4.4903203443
6	110006	Product6	$111.0	2	0.01	$28.8	$2.9	3.1549697E12	23838.140152	828.2884	3.3596807015
7	110007	Product7	$33.0	4	0.01	$8.3	$0.8	3827.6258214	561.515625	68.0625	2.1102132003
8	110008	Product8	$250.0	3	0.05	$132.5	$13.3	3.4996036E57	2326203.125	17556.25	4.8865826454
9	110009	Product9	$83.0	4	0.05	$20.8	$2.1	1027094726.7	8934.171875	430.5625	3.0325462467
10	110010	Product1	$192.0	3	0.03	$94.7	$9.5	1.3689051E41	849816.32205	8971.8784	4.5509251711
11	110011	Product2	$65.0	4	0.03	$16.3	$1.6	11409991.764	4291.015625	264.0625	2.7880929088
12	110012	Product3	$130.0	5	0.01	$43.5	$4.4	7.7948895E18	82312.875	1892.25	3.7727609381
13	110013	Product4	$220.0	5	0.03	$107.0	$10.7	2.9478784E46	1225043	11449	4.6728288345
14	110014	Product5	$104.0	5	0.02	$13.6	$1.4	806129.75912	2515.456	184.96	2.6100697927
15	110015	Product6	$149.0	4	0.05	$39.2	$3.9	1.0576542E17	60236.288	1536.64	3.6686767468
16	110016	Product7	$222.0	3	0.04	$115.4	$11.5	1.2595385E50	1535202.7587	13307.9296	4.7480576735

Shipping_Cost

Total rows: 30 Total columns: 11 Rows 1-30

	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Shipping_Cost_exp	Shipping_Cost_Cube	Shipping_Cost_SQ	Shipping_Cost_Log
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6	774520.95915	2493.326016	183.8736	2.6071242825
2	110002	Product2	$104.0	1	0.03	$20.9	$2.1	8.068761493	9.103145472	4.359744	0.73620667
3	110003	Product3	$149.0	4	0.01	$63.0	$6.3	546.75456016	250.52358246	39.740416	1.8411843526
4	110004	Product4	$222.0	4	0.03	$115.4	$11.5	102334.28116	1535.2027587	133.079296	2.4454725805
5	110005	Product5	$199.0	3	0.05	$89.2	$8.9	7442.782129	708.53946087	79.477225	2.1877352513
6	110006	Product6	$111.0	2	0.01	$28.8	$2.9	17.778680238	23.838140152	8.282884	1.0570956085
7	110007	Product7	$33.0	4	0.01	$8.3	$0.8	2.2818807653	0.561515625	0.680625	-0.192371893
8	110008	Product8	$250.0	3	0.05	$132.5	$13.3	568070.04002	2326.203125	175.5625	2.5839975524
9	110009	Product9	$83.0	4	0.05	$20.8	$2.1	7.964546459	8.934171875	4.305625	0.7299611537
10	110010	Product1	$192.0	3	0.03	$94.7	$9.5	12990.842953	849.81632205	89.718784	2.2483400781
11	110011	Product2	$65.0	4	0.03	$16.3	$1.6	5.0784190372	4.291015625	2.640625	0.4855078158
12	110012	Product3	$130.0	5	0.01	$43.5	$4.4	77.478462925	82.312875	18.9225	1.4701758451
13	110013	Product4	$220.0	5	0.03	$107.0	$10.7	44355.85513	1225.043	114.49	2.3702437415
14	110014	Product5	$104.0	5	0.02	$13.6	$1.4	3.8961933018	2.515456	1.8496	0.3074846997
15	110015	Product6	$149.0	4	0.05	$39.2	$3.9	50.400444778	60.236288	15.3664	1.3660916538
16	110016	Product7	$222.0	3	0.04	$115.4	$11.5	102334.28116	1535.2027587	133.079296	2.4454725805
17	110017	Product8	$199.0	5	0.02	$99.1	$9.9	20130.673991	973.242271	98.2081	2.2935443483

4. Perform regression test

```
/* Data Preparation */

/* Step 2: Create a new variable Total_Sales = sales*quantity*/

DATA Retail_Analysis;
set Retail_Analysis;
Total_Sales = sales*quantity;
run;
```

ULTS OUTPUT DATA

SIS View: Column names Filter: (none)

Total rows: 30 Total columns: 8

	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Total_Sales
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6	440
2	110002	Product2	$104.0	1	0.03	$20.9	$2.1	104
3	110003	Product3	$149.0	4	0.01	$63.0	$6.3	596
4	110004	Product4	$222.0	4	0.03	$115.4	$11.5	888
5	110005	Product5	$199.0	3	0.05	$89.2	$8.9	597
6	110006	Product6	$111.0	2	0.01	$28.8	$2.9	222
7	110007	Product7	$33.0	4	0.01	$8.3	$0.8	132
8	110008	Product8	$250.0	3	0.05	$132.5	$13.3	750
9	110009	Product9	$83.0	4	0.05	$20.8	$2.1	332
10	110010	Product1	$192.0	3	0.03	$94.7	$9.5	576
11	110011	Product2	$65.0	4	0.03	$16.3	$1.6	260
12	110012	Product3	$130.0	5	0.01	$43.5	$4.4	650
13	110013	Product4	$220.0	5	0.03	$107.0	$10.7	1100
14	110014	Product5	$104.0	5	0.02	$13.6	$1.4	520
15	110015	Product6	$149.0	4	0.05	$39.2	$3.9	596
16	110016	Product7	$222.0	3	0.04	$115.4	$11.5	666
17	110017	Product8	$199.0	5	0.02	$99.1	$9.9	995
18	110018	Product1	$111.0	5	0.05	$3.2	$0.3	555
19	110019	Product2	$33.0	1	0.03	$33.0	$3.3	33
20	110020	Product3	$149.0	4	0.01	$63.0	$6.3	596
21	110021	Product4	$83.0	1	0.05	$83.0	$8.3	83
22	110022	Product5	$192.0	2	0.02	$104.3	$10.4	384
23	110023	Product6	$65.0	2	0.04	$32.5	$3.3	130

```
64 /* Step 3: Now predict total_sales */
65
66 PROC REG DATA=Retail_Analysis;
67 MODEL Total_Sales= Quantity Discount Profit Shipping_Cost;
68 run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1823605	607868	72.63	<.0001
Error	26	217595	8369.04811		
Corrected Total	29	2041201			

Root MSE	91.48250	R-Square	0.8934
Dependent Mean	491.10000	Adj R-Sq	0.8811
Coeff Var	18.62808		

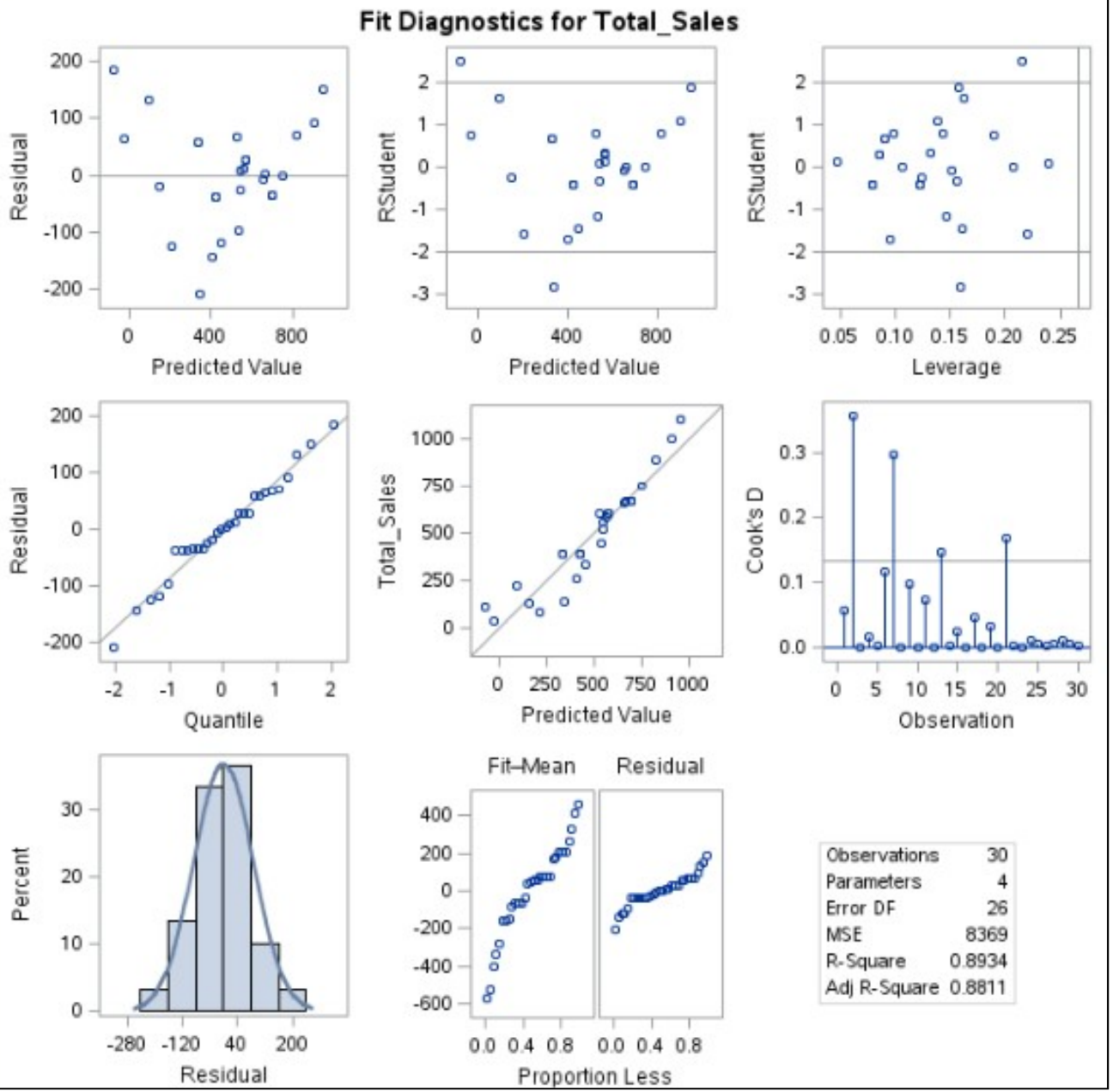
Note: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

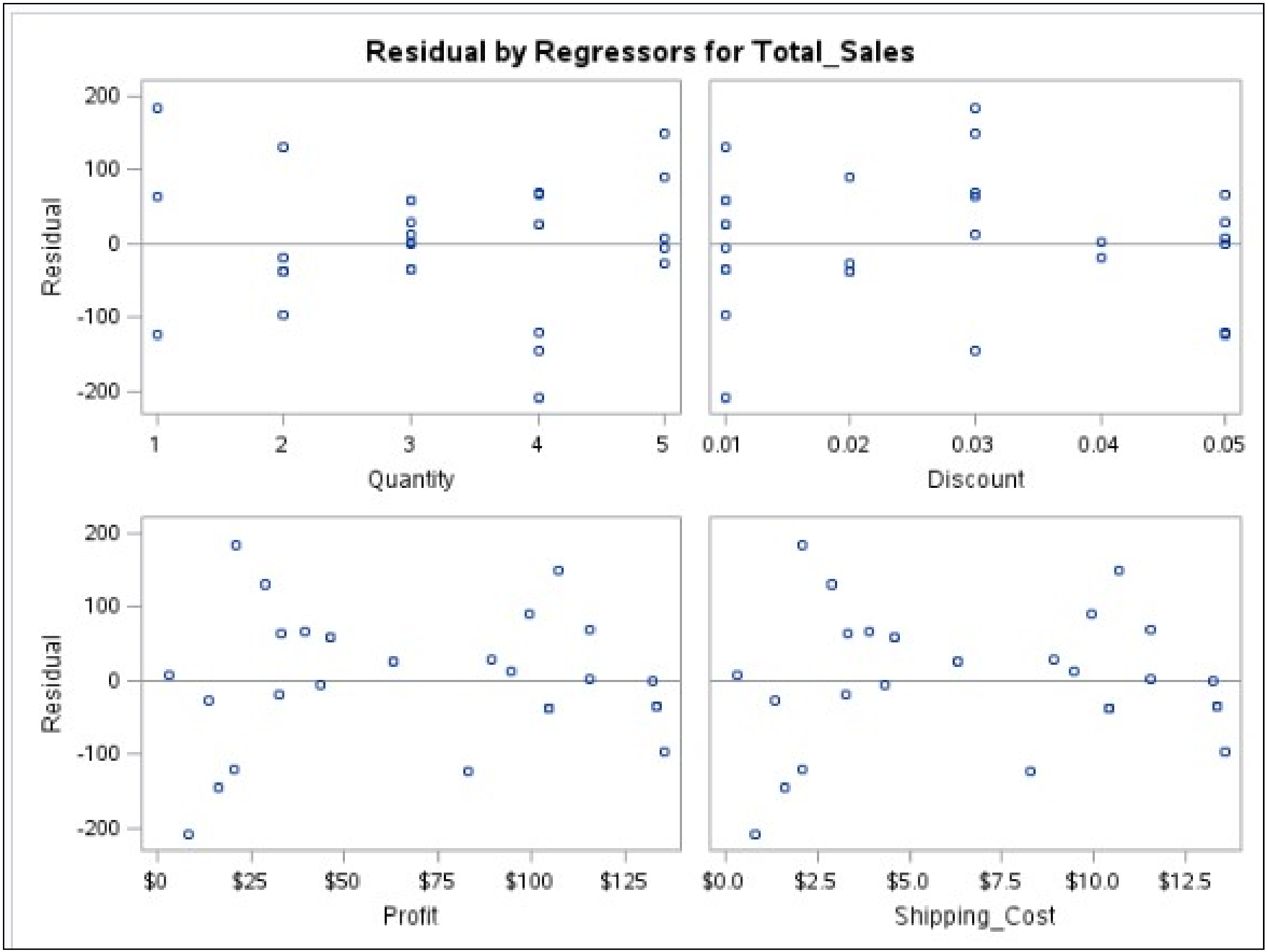
Note: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

Shipping_Cost =	0.1 * Profit

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-378.71173	68.58927	-5.52	<.0001
Quantity	Quantity	1	167.95517	14.01532	11.98	<.0001
Discount	Discount	1	1439.79360	1110.79329	1.30	0.2063
Profit	Profit	B	4.17438	0.39127	10.67	<.0001
Shipping_Cost	Shipping_Cost	0	0	.	.	.

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales




Residual by Regressors for Total_Sales
Residual
200
100
0
-100
-200
1
2
3
4
5
Quantity
0.01
0.02
0.03
0.04
0.05
Discount
$0
$25
$50
$75
$100
$125
Profit
$0.0
$2.5
$5.0
$7.5
$10.0
$12.5
Shipping_Cost

```
/* Step 4 */:

PROC REG DATA=Retail_Analysis;
MODEL Total_Sales= Quantity Discount Profit ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance

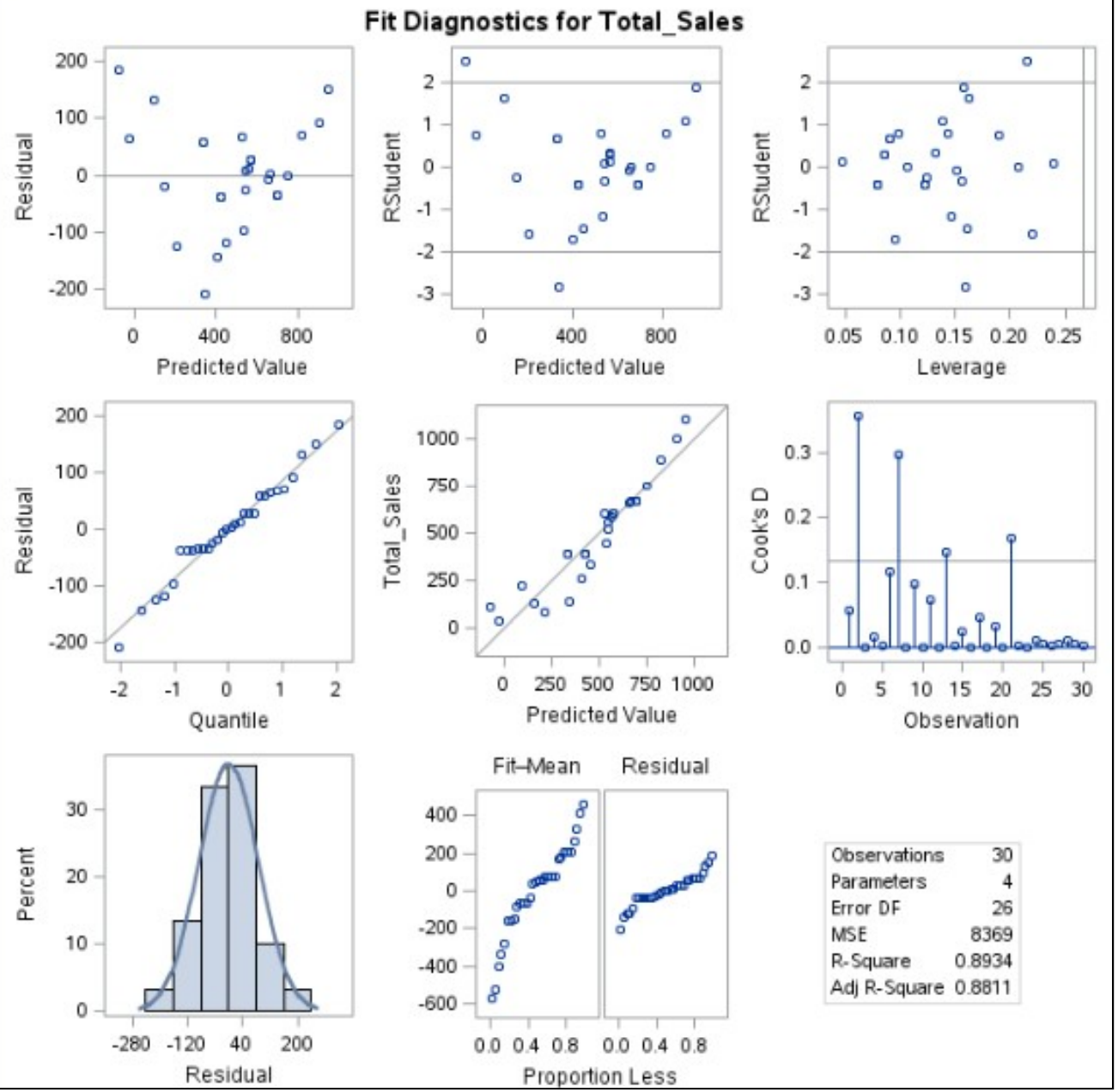
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1823605	607868	72.63	<.0001
Error	26	217595	8369.04811		
Corrected Total	29	2041201			

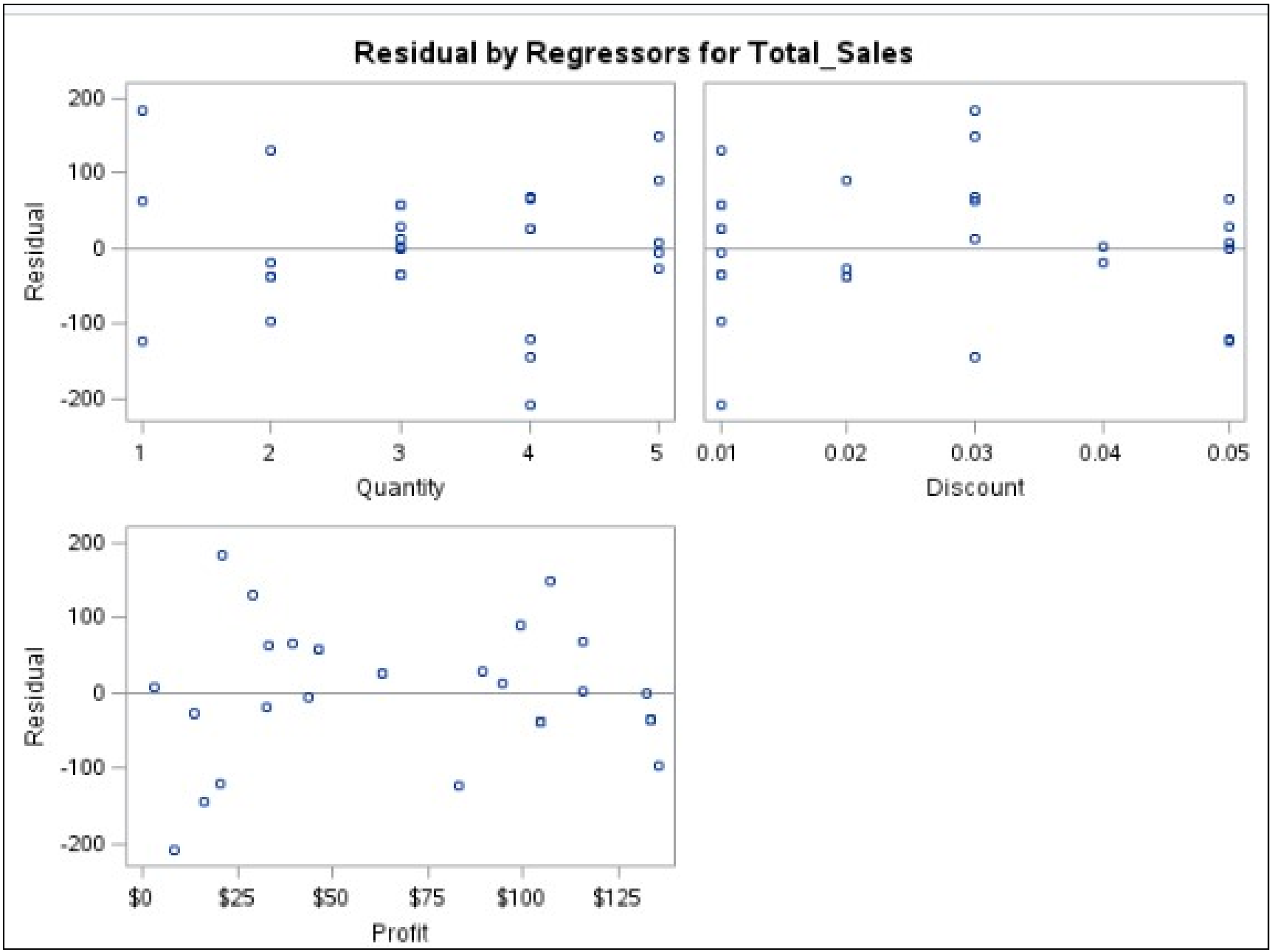
Root MSE	91.48250	R-Square	0.8934
Dependent Mean	491.10000	Adj R-Sq	0.8811
Coeff Var	18.62808		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-378.71173	68.58927	-5.52	<.0001
Quantity	Quantity	1	167.95517	14.01532	11.98	<.0001
Discount	Discount	1	1439.79360	1110.79329	1.30	0.2063
Profit	Profit	1	4.17438	0.39127	10.67	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales




Residual by Regressors for Total_Sales
Residual
200
100
0
-100
-200
1
2
3
4
5
Quantity
0.01
0.02
0.03
0.04
0.05
Discount
$0
$25
$50
$75
$100
$125
Profit

```
/* Step 5: Discount is insignificant in the results. Drop discount and rerun model */
/* Below model if final model with quantity and profit as significant variables
and Adj Rsquare as .88 i.e 88% accuracy*/

PROC REG DATA=Retail_Analysis;
MODEL Total_Sales= Quantity Profit ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance

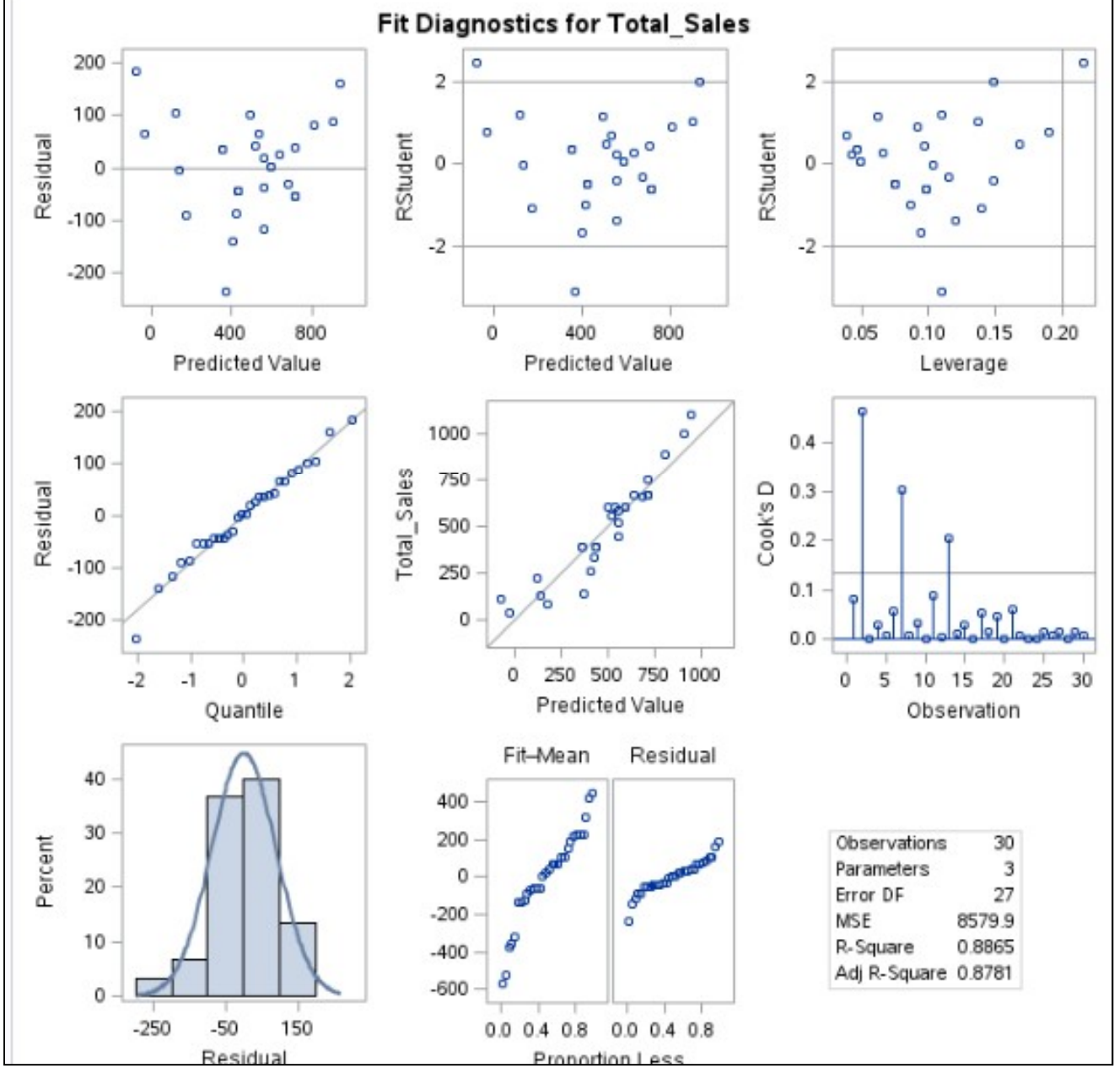
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1809545	904772	105.45	<.0001
Error	27	231656	8579.85375		
Corrected Total	29	2041201			

Root MSE	92.62750	R-Square	0.8865
Dependent Mean	491.10000	Adj R-Sq	0.8781
Coeff Var	18.86123		

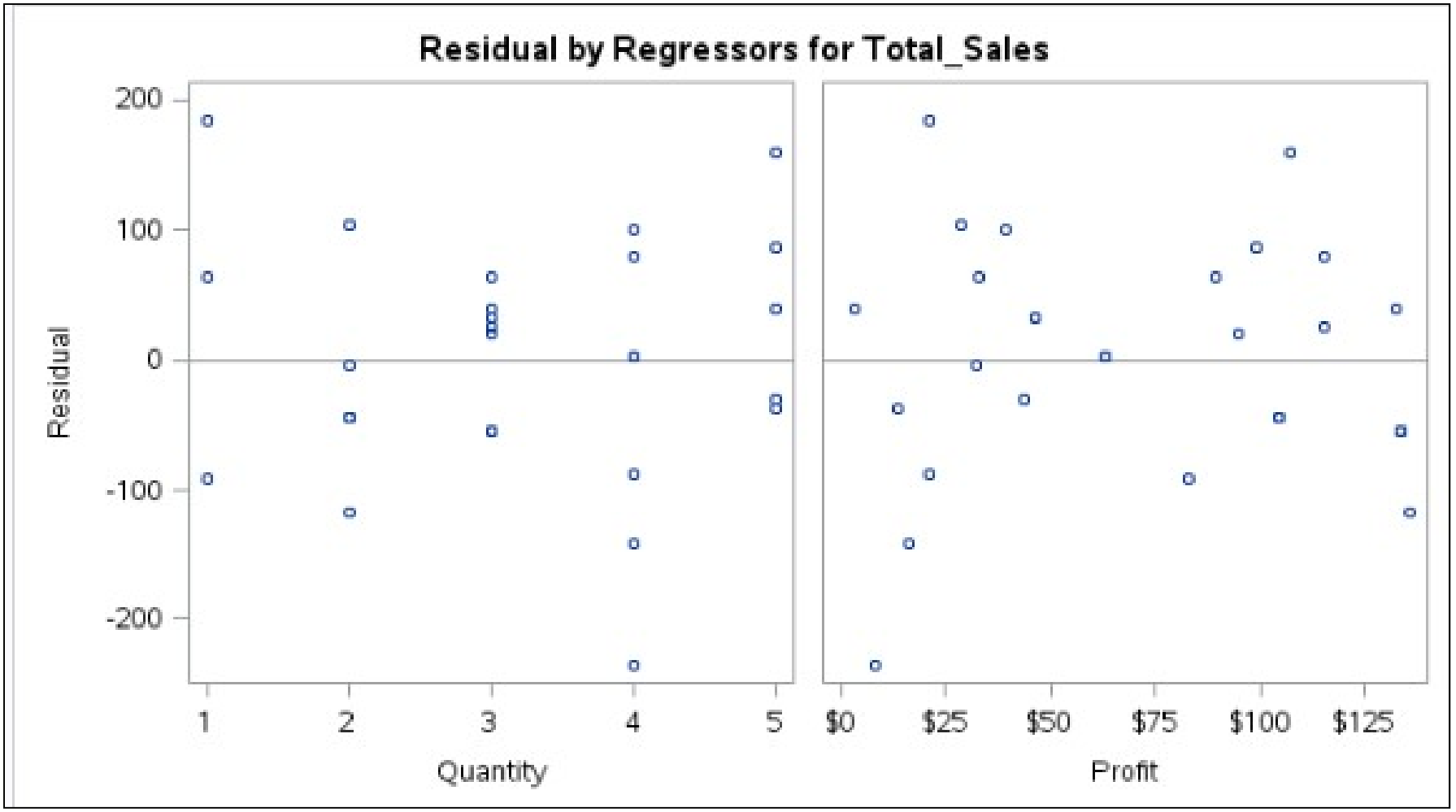
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-333.04421	59.58531	-5.59	<.0001
Quantity	Quantity	1	166.88150	14.16593	11.78	<.0001
Profit	Profit	1	4.10070	0.39197	10.46	<.0001

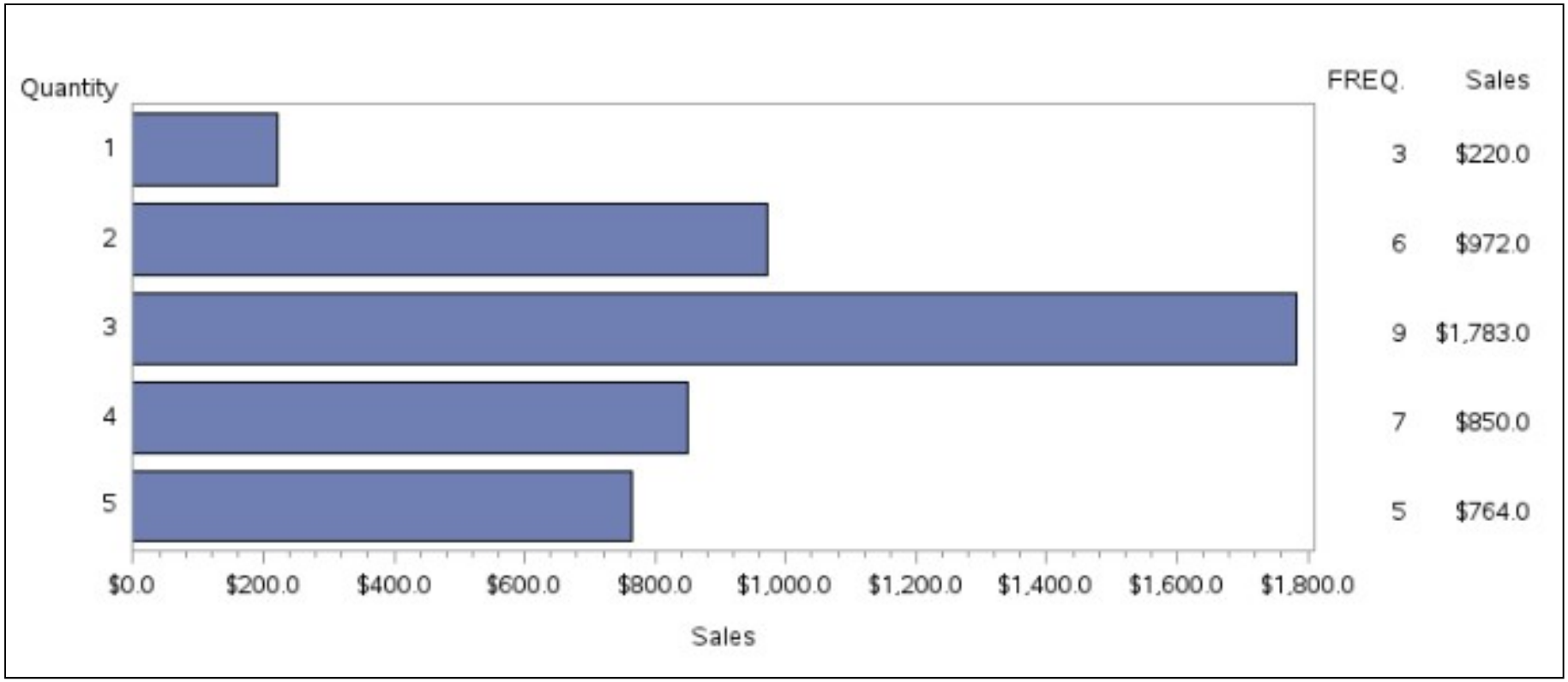
The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales



Observations	30
Parameters	3
Error DF	27
MSE	8579.9
R-Square	0.8865
Adj R-Square	0.8781

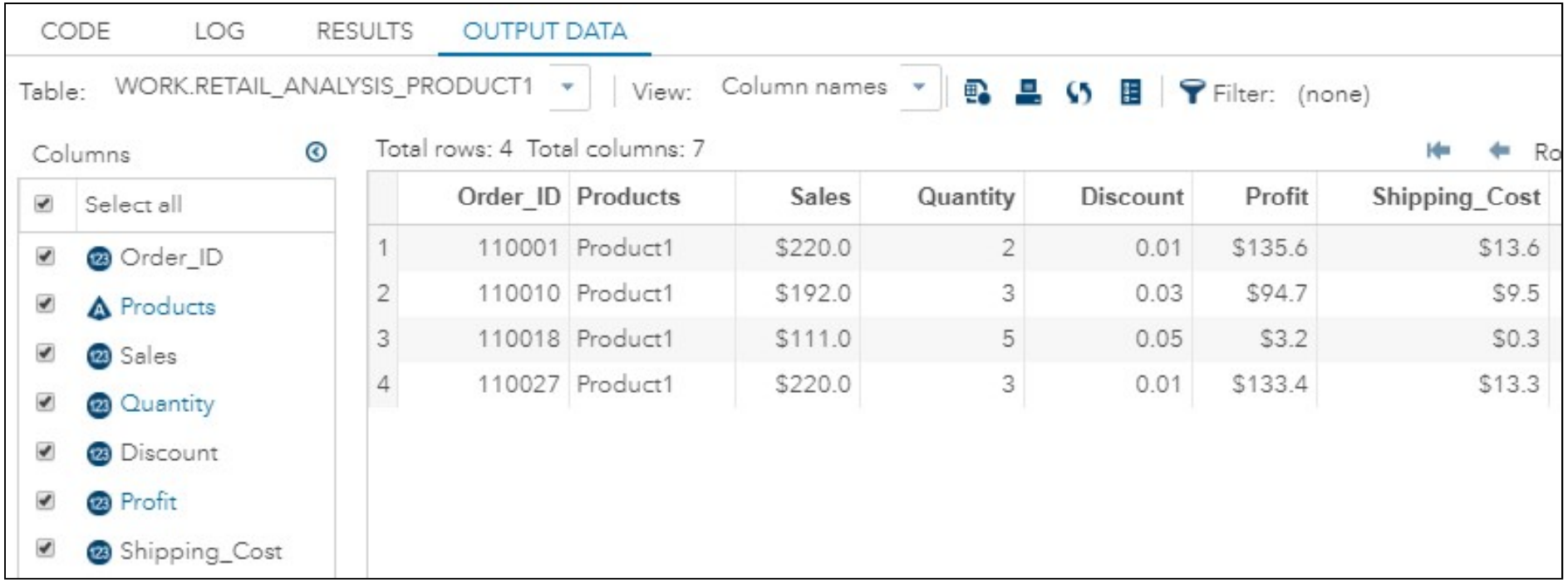

Residual by Regressors for Total_Sales
Residual
200
100
0
-100
-200
1
2
3
4
5
Quantity
$0
$25
$50
$75
$100
$125
Profit

```
89  PROC GCHART DATA = WORK.Retail_Analysis;
90  HBAR Quantity/type = sum sumvar = Sales;
91  RUN;
```



Product 1

```
14 DATA WORK.RETAIL_ANALYSIS_Product1;
15 SET Retail_Analysis;
16 WHERE Products = 'Product1';
17 RUN;
```



	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6
2	110010	Product1	$192.0	3	0.03	$94.7	$9.5
3	110018	Product1	$111.0	5	0.05	$3.2	$0.3
4	110027	Product1	$220.0	3	0.01	$133.4	$13.3

```
20 DATA RETAIL_ANALYSIS_Product1;
21 SET RETAIL_ANALYSIS_Product1;
22 Total_Sales_Product1 = sales*quantity;
23 run;
```

CODE LOG RESULTS OUTPUT DATA

Table: WORK.RETAIL_ANALYSIS_PRODUCT1 | View: Column names | Filter: (none)

Columns

- Select all
- Order_ID
- Products
- Sales
- Quantity
- Discount
- Profit
- Shipping_Cost
- Total_Sales_Product1

Total rows: 4 Total columns: 8

	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Total_Sales_Product1
1	110001	Product1	$220.0	2	0.01	$135.6	$13.6	440
2	110010	Product1	$192.0	3	0.03	$94.7	$9.5	576
3	110018	Product1	$111.0	5	0.05	$3.2	$0.3	555
4	110027	Product1	$220.0	3	0.01	$133.4	$13.3	660

```
26 PROC REG DATA=RETAIL_ANALYSIS_Product1;
27 MODEL Total_Sales_Product1 = Quantity Profit ;
28 run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales_Product1

Number of Observations Read	4
Number of Observations Used	4

Analysis of Variance

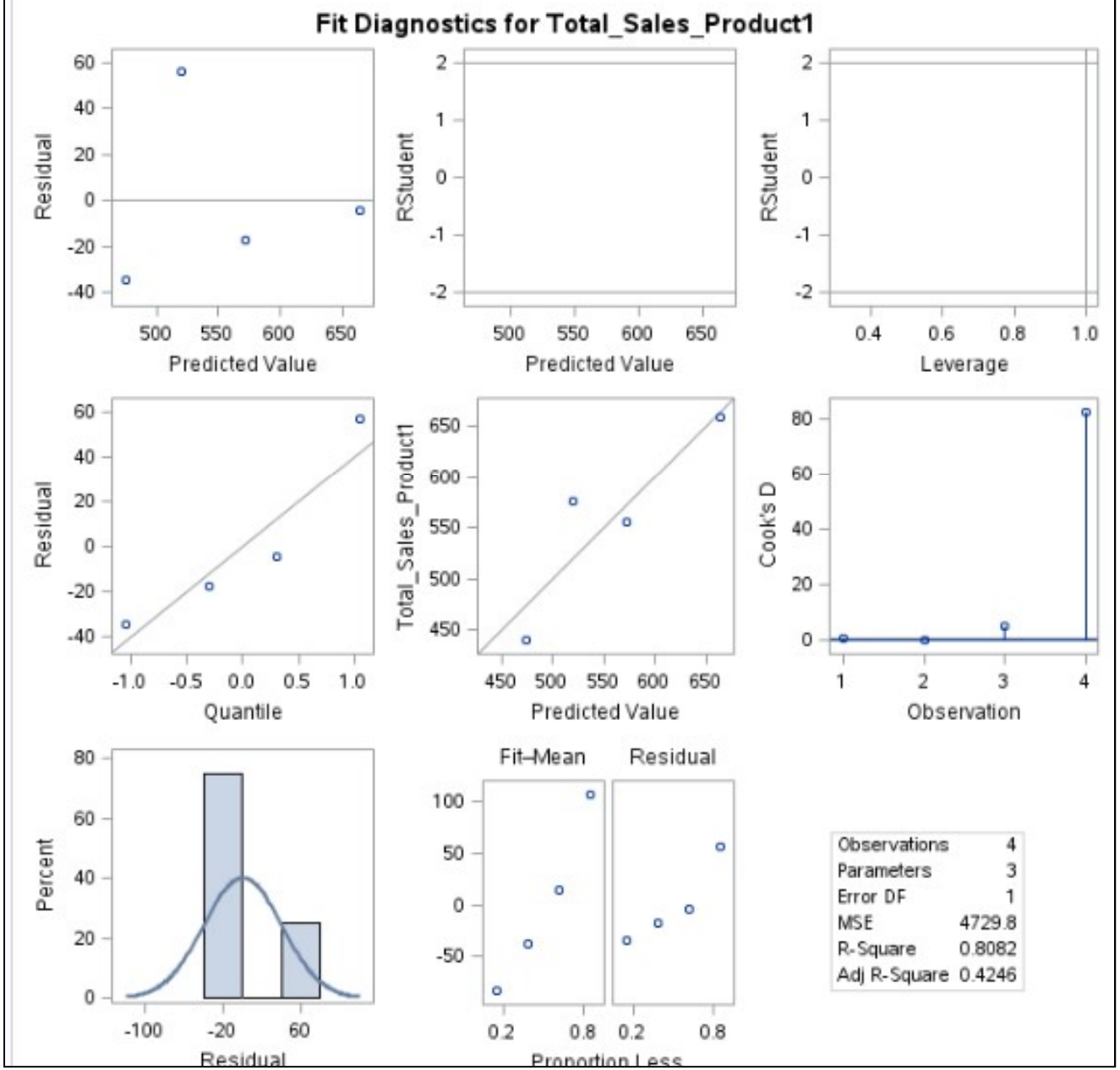
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	19931	9965.49488	2.11	0.4379
Error	1	4729.76024	4729.76024		
Corrected Total	3	24661			

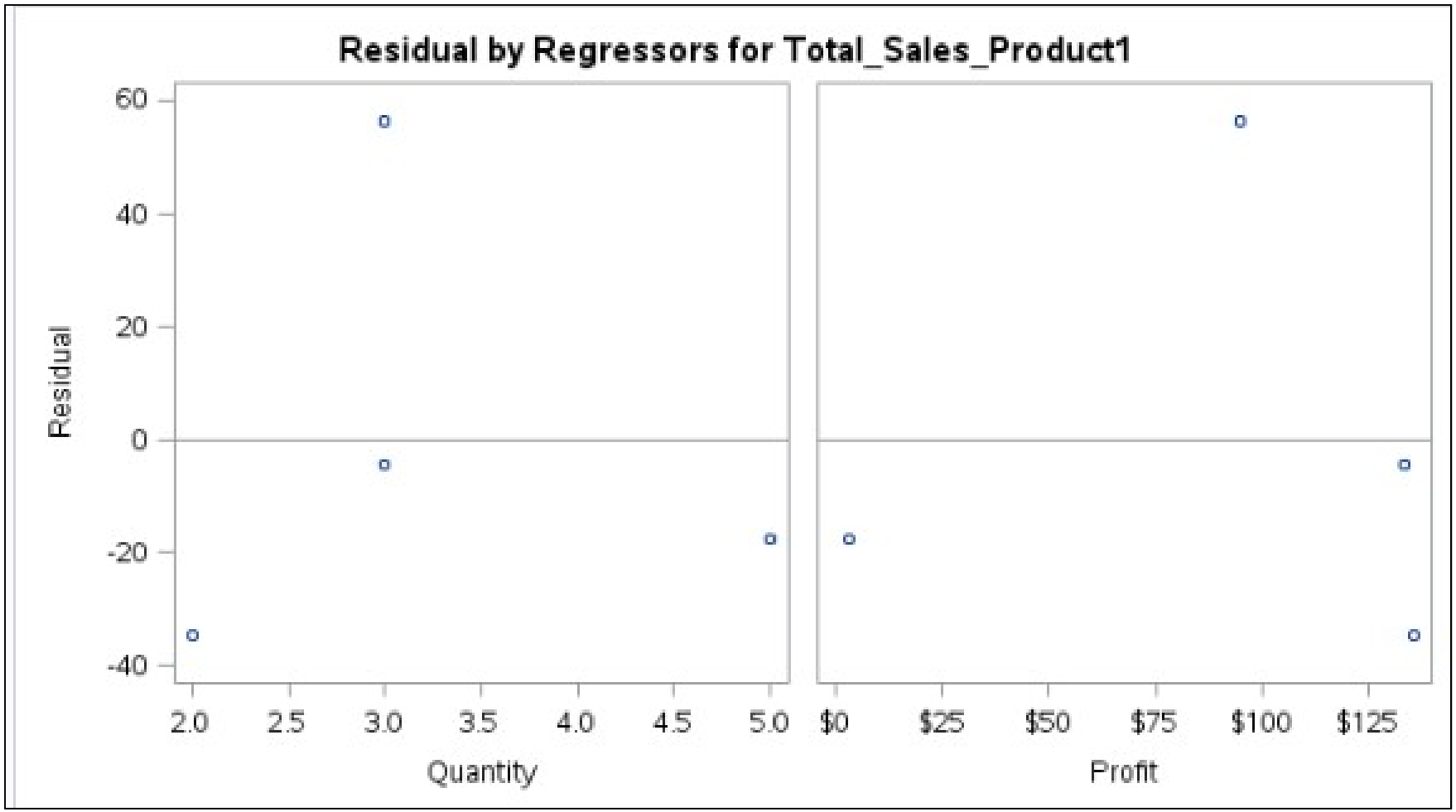
Root MSE	68.77325	R-Square	0.8082
Dependent Mean	557.75000	Adj R-Sq	0.4246
Coeff Var	12.33048		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-428.88298	488.15863	-0.88	0.5411
Quantity	Quantity	1	197.82022	96.44401	2.05	0.2888
Profit	Profit	1	3.74654	1.96020	1.91	0.3069

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales_Product1




Residual by Regressors for Total_Sales_Product1
Residual
60
40
20
0
-20
-40
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Quantity
$0
$25
$50
$75
$100
$125
Profit

Product 2

```
14 DATA WORK.RETAIL_ANALYSIS_Product2;
15 SET Retail_Analysis;
16 WHERE Products = 'Product2';
17 RUN;
```

CODE LOG RESULTS OUTPUT DATA

Table: WORK.RETAIL_ANALYSIS_PRODUCT2 | View: Column names | Filter: (none)

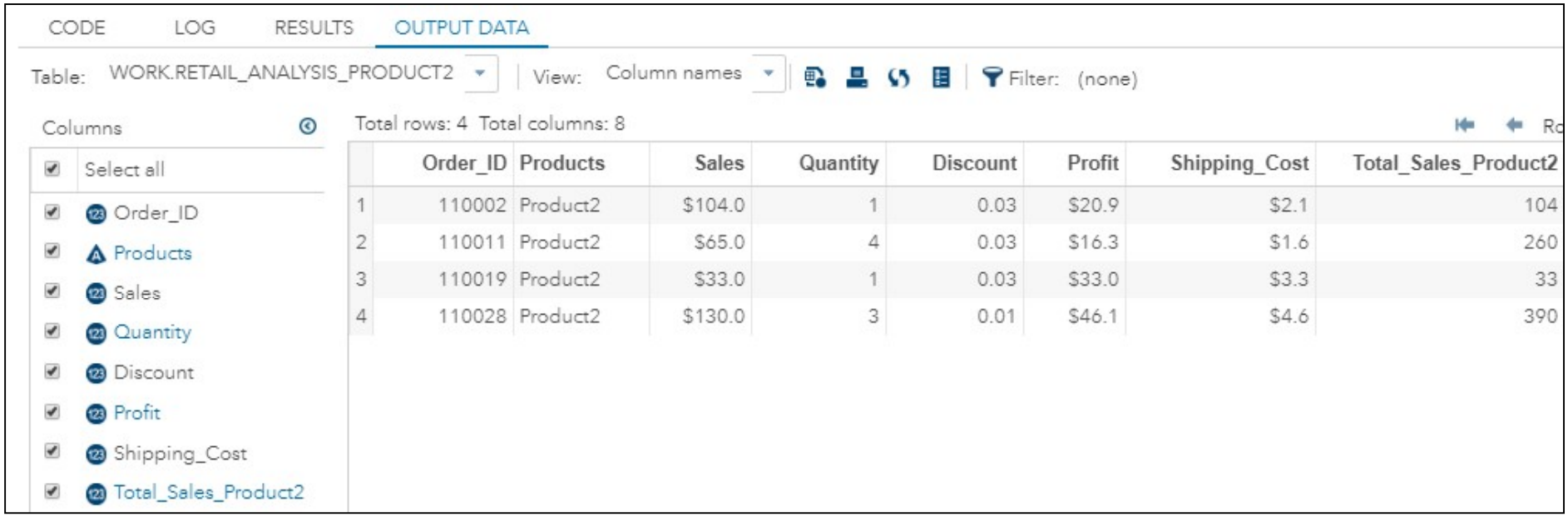
Columns

- Select all
- Order_ID
- Products
- Sales
- Quantity
- Discount
- Profit
- Shipping_Cost

Total rows: 4 Total columns: 7

	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost
1	110002	Product2	$104.0	1	0.03	$20.9	$2.1
2	110011	Product2	$65.0	4	0.03	$16.3	$1.6
3	110019	Product2	$33.0	1	0.03	$33.0	$3.3
4	110028	Product2	$130.0	3	0.01	$46.1	$4.6

```
20 DATA RETAIL_ANALYSIS_Product2;
21 SET RETAIL_ANALYSIS_Product2;
22 Total_Sales_Product2 = sales*quantity;
23 run;
```



	Order_ID	Products	Sales	Quantity	Discount	Profit	Shipping_Cost	Total_Sales_Product2
1	110002	Product2	$104.0	1	0.03	$20.9	$2.1	104
2	110011	Product2	$65.0	4	0.03	$16.3	$1.6	260
3	110019	Product2	$33.0	1	0.03	$33.0	$3.3	33
4	110028	Product2	$130.0	3	0.01	$46.1	$4.6	390

```
26 PROC REG DATA=RETAIL_ANALYSIS_Product2;
27 MODEL Total_Sales_Product2 = Quantity Profit ;
28 run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales_Product2

Number of Observations Read	4
Number of Observations Used	4

Analysis of Variance

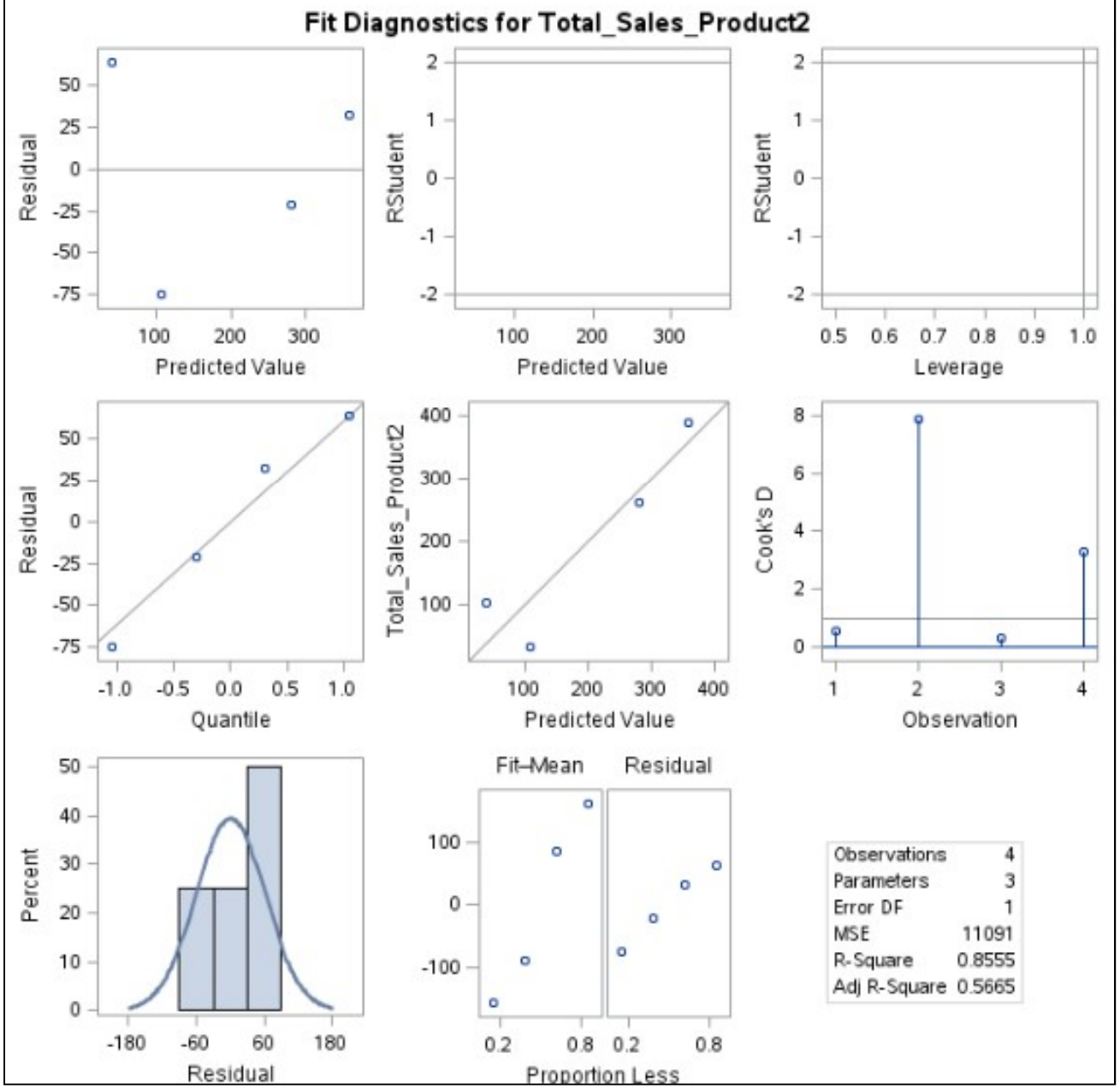
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	65672	32836	2.96	0.3801
Error	1	11091	11091		
Corrected Total	3	76763			

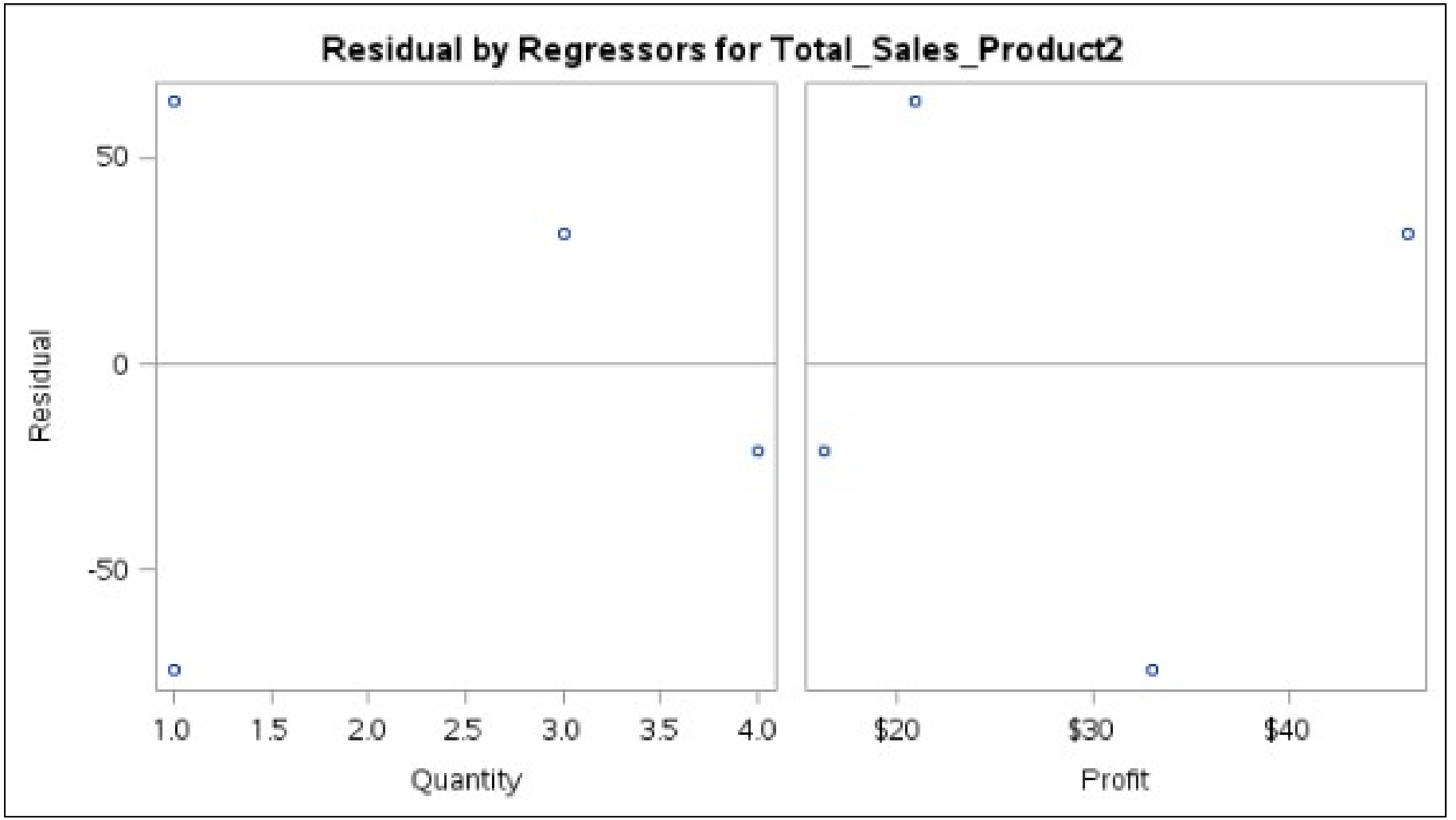
Root MSE	105.31403	R-Square	0.8555
Dependent Mean	196.75000	Adj R-Sq	0.5665
Coeff Var	53.52683		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-164.75652	174.39173	-0.94	0.5181
Quantity	Quantity	1	88.93949	40.64102	2.19	0.2729
Profit	Profit	1	5.55425	4.55693	1.22	0.4374

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales_Product2




Residual by Regressors for Total_Sales_Product2
Residual
50
0
-50
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Quantity
$20
$30
$40
Profit