STAT 512 | SPRING'17

PROJECT

RENT FOR LAND PLANTED TO ALFALFA

PROFESSOR- DR WALID SHARABATI

SONAKSHI MALHOTRA |HARSHA DESHMUKH | GRADEN YOUNG | ANDREW MCGOWEN

Model

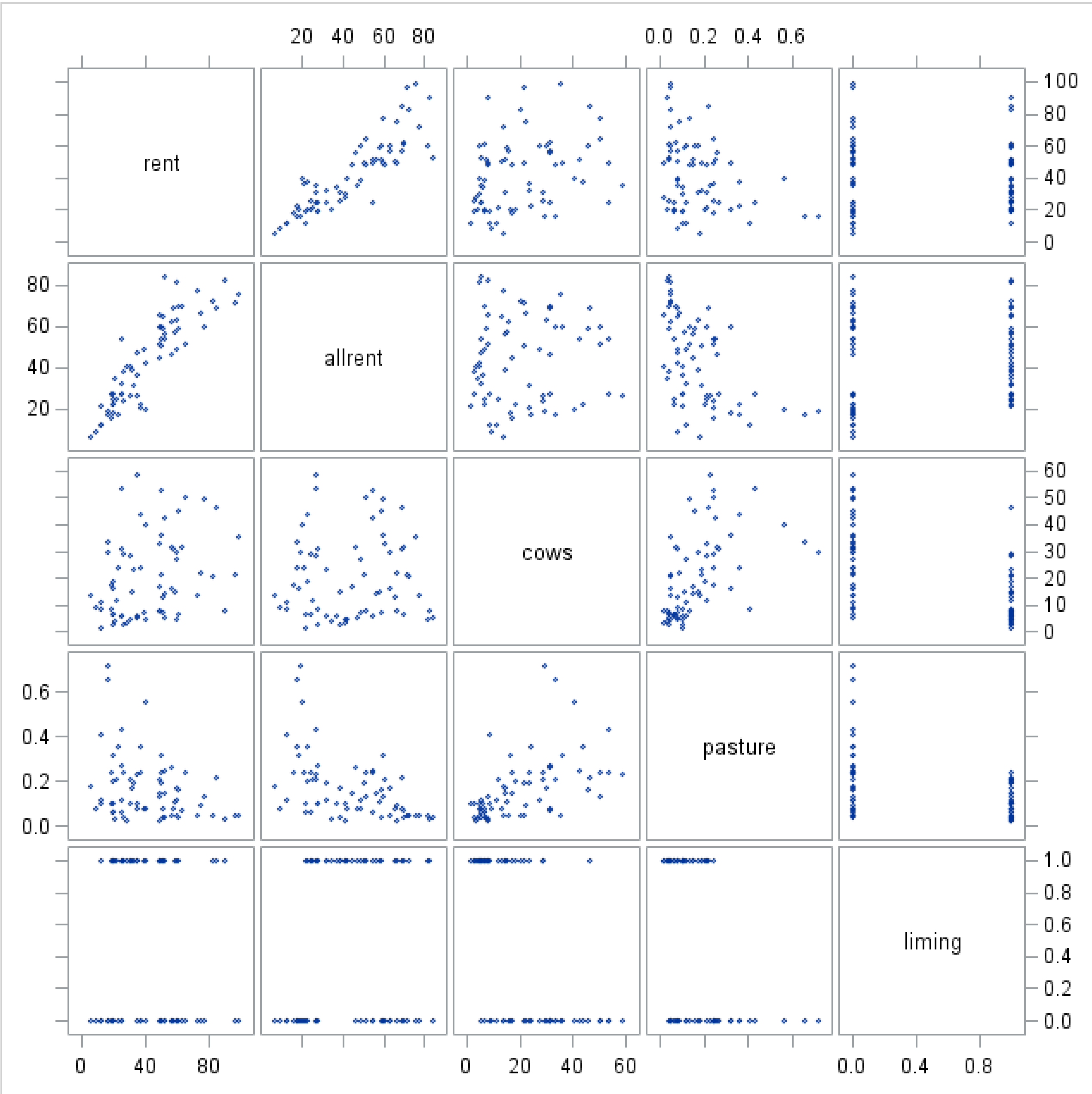
Response variable: Rent

Predictor variable: allrent cows pasture liming

Obs	rent	allrent	cows	pasture	liming
1	18.38	15.5	17.25	0.24	0
2	20	22.29	18.51	0.2	1
3	11.5	12.36	11.13	0.12	0
4	25	31.84	5.54	0.12	1
5	52.5	83.9	5.44	0.04	0
6	82.5	72.25	20.37	0.05	1
7	25	27.14	31.2	0.27	0
8	30.67	40.41	4.29	0.1	1
9	12	12.42	8.69	0.41	0
10	61.25	69.42	6.63	0.04	1
11	60	48.46	27.4	0.12	0
12	57.5	69	31.23	0.08	0
13	31	26.09	28.5	0.21	1
14	60	62.83	29.98	0.17	0
15	72.5	77.06	13.59	0.05	0
16	60.33	58.83	45.46	0.16	0
17	49.75	59.48	35.9	0.32	0
18	8.5	9	8.89	0.08	0
19	36.5	20.64	23.81	0.24	0
20	60	81.4	4.54	0.05	1
21	16.25	18.92	29.62	0.72	0
22	50	50.32	21.36	0.19	1
23	11.5	21.33	1.53	0.1	1
24	35	46.85	5.42	0.08	1
25	75	65.94	22.1	0.09	0
26	31.56	38.68	14.55	0.17	1
27	48.5	51.19	7.59	0.13	1
28	77.5	59.42	49.86	0.13	0
29	21.67	24.64	11.46	0.21	1
30	19.75	26.94	2.48	0.1	1
31	56	46.2	31.62	0.26	0
32	25	26.86	53.73	0.43	0
33	40	20	40.18	0.56	0
34	56.67	62.52	15.89	0.05	0
35	51.79	56	14.25	0.15	1
36	96.67	71.41	21.37	0.05	0
37	50.83	65	13.24	0.08	1
38	34.33	36.28	5.85	0.1	1
39	48.75	59.88	32.99	0.21	0
40	25.8	23.62	28.89	0.24	1
41	20	24.2	6.29	0.06	1
42	16	17.09	33.34	0.66	0

43	48.67	44.56	16.7	0.15	1
44	20.78	34.46	4.2	0.03	1
45	32.5	31.55	23.47	0.19	1
46	19	26.94	8.28	0.1	1
47	51.5	58.71	7.4	0.04	1
48	49.17	65.74	7.71	0.02	1
49	85	69.05	46.18	0.22	1
50	58.75	57.54	14.98	0.11	1
51	19.33	21.73	6.58	0.06	0
52	5	6.17	13.68	0.18	0
53	65	51	50.5	0.24	0
54	20	18.25	16.12	0.32	0
55	62.5	69.88	31.48	0.07	0
56	35	26.68	58.6	0.23	0
57	99.17	75.73	35.43	0.05	0
58	40.25	41.77	4.53	0.08	1
59	39.17	48.5	6.82	0.08	1
60	37.5	21.89	43.7	0.36	0
61	26.25	38.33	2.83	0.04	1
62	52.14	53.95	42.54	0.25	0
63	22.5	17.17	24.16	0.36	0
64	90	82	7.89	0.03	1
65	28	40.6	3.27	0.02	1
66	50	53.89	53.16	0.24	0
67	24.5	54.17	5.57	0.06	1

Analysis of Data before regression:



Var	rent	allrent	cows	pasture	liming
rent	1	0.87577	0.30857	-0.32338	-0.08895
allrent	0.87577	1	0.04872	-0.49982	0.08896
cows	0.30857	0.04872	1	0.5226	-0.58344
pasture	-0.32338	-0.49982	0.5226	1	-0.42678
liming	-0.08895	0.08896	-0.58344	-0.42678	1

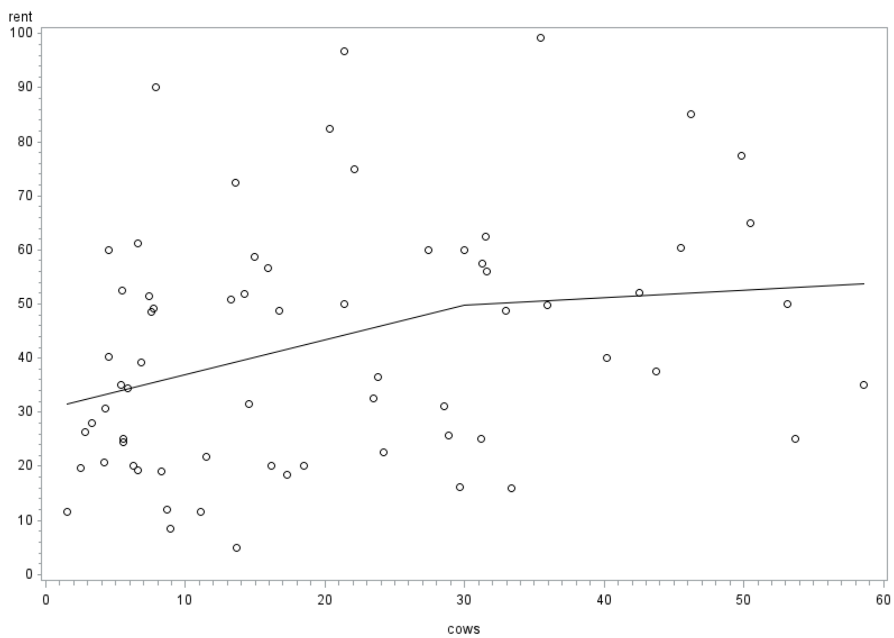
The predictor variables do not seem to have major collinearity issues. Pasture, however, does not seem to be linearly related to the response variable.

Pasture and liming seem to have a negative impact on the response variable.

Part I

Question 1) **Choose a predictor of your choice and conduct piecewise SLR to model the relationship with response variable. Be smart about choosing the point at which the two pieces change slope, if none of the predictors has curved relationship with the response then the center should be your point where the two pieces meet. Determine whether the two pieces are the same.**

```
proc sort data=alfalfa;
by cows;
symbol1 v=C i=sm70;
proc gplot data= alfalfa;
plot rent*cows;
run;
data alfalfaone; set alfalfa;
if cows le 30
then cslope = 0;
if cows gt 30
then cslope = cows-30;
run;
proc print data=alfalfaone;
run;

proc reg data=alfalfaone;
model rent = cows cslope;
output out = alfalfaout p = renthat;
run;
symbol1 v=circle i=none c=black;
symbol2 v=none i =join c=black;
proc sort data=alfalfaout; by cows;
proc gplot data=alfalfaout;
plot (rent renthat)*cows/overlay;
run;
```



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3464.85967	1732.42984	3.67	0.0310
Error	64	30205	471.95794		
Corrected Total	66	33670			

Root MSE	21.72459	R-Square	0.1029
Dependent Mean	42.16612	Adj R-Sq	0.0749
Coeff Var	51.52144		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	30.55531	5.40907	5.65	<.0001
cows	1	0.64038	0.30564	2.10	0.0401
cslope	1	-0.50385	0.68027	-0.74	0.4616

<u>Checking if both lines are the same</u>

Both lines will be same is slope of cslope is zero.

```
proc reg data=alfalfaone;
model rent = cows cslope;
test cslope;
run;
```

Result: yes, both lines are the same since we cannot reject the null hypothesis that slope of cslope is zero.

The SAS System

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable rent				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	258.91381	0.55	0.4616
Denominator	64	471.95794		

Question 2)

Creation of sum using 'allrent' and 'cows' as predictor variables.

```
data alfalfa2;
set alfalfa;
sum=allrent +cows;
run;
```

Run the regression using all predictor variables except the variables used to create sum

Predict response a) using all explanatory variables b) and sum

Output		Case A: pasture liming		Case B: sum pasture liming	
Obs	Dependent Variable	Predicted Value	Residual	Predicted Value2	Residual2
1	18.38	43.44	-25.06	15.883	2.497
2	20	33.7592	-13.7592	25.0855	-5.0855
3	11.5	51.7274	-40.2274	13.3206	-1.8206
4	25	39.2841	-14.2841	25.3796	-0.3796
5	52.5	57.2523	-4.7523	63.9929	-11.4929
6	82.5	44.1184	38.3816	67.9879	14.5121
7	25	41.3681	-16.3681	33.4509	-8.4509
8	30.67	40.6653	-9.9953	31.3986	-0.7286
9	12	31.6995	-19.6995	1.5071	10.4929
10	61.25	44.809	16.441	56.2846	4.9654
11	60	51.7274	8.2726	51.4079	8.5921
12	57.5	54.4898	3.0102	70.5222	-13.0222
13	31	33.0685	-2.0685	34.7669	-3.7669
14	60	48.2743	11.7257	61.9967	-1.9967
15	72.5	56.5617	15.9383	64.5979	7.9021
16	60.33	48.9649	11.3651	70.6935	-10.3635
17	49.75	37.9151	11.8349	58.6507	-8.9007
18	8.5	54.4898	-45.9898	10.6386	-2.1386
19	36.5	43.44	-6.94	24.3921	12.1079
20	60	44.1184	15.8816	63.1297	-3.1297
21	16.25	10.2905	5.9595	10.6782	5.5718
22	50	34.4498	15.5502	47.8913	2.1087
23	11.5	40.6653	-29.1653	15.5149	-4.0149
24	35	42.0465	-7.0465	37.5994	-2.5994
25	75	53.7992	21.2008	61.3091	13.6909
26	31.56	35.831	-4.271	35.1685	-3.6085
27	48.5	38.5935	9.9065	40.5955	7.9045
28	77.5	51.0367	26.4633	75.3656	2.1344
29	21.67	33.0685	-11.3985	21.3196	0.3504
30	19.75	40.6653	-20.9153	20.2858	-0.5358
31	56	42.0587	13.9413	47.9658	8.0342
32	25	30.3183	-5.3183	44.0699	-19.0699
33	40	21.3403	18.6597	24.7065	15.2935
34	56.67	56.5617	0.1083	55.6961	0.9739

35	51.79	37.2122	14.5778	48.242	3.548
36	96.67	56.5617	40.1083	66.147	30.523
37	50.83	42.0465	8.7835	56.4866	-5.6566
38	34.33	40.6653	-6.3353	29.5295	4.8005
39	48.75	45.5118	3.2382	60.6497	-11.8997
40	25.8	30.9967	-5.1967	32.2111	-6.4111
41	20	43.4278	-23.4278	22.4547	-2.4547
42	16	14.4341	1.5659	14.1388	1.8612
43	48.67	37.2122	11.4578	41.7038	6.9662
44	20.78	45.4996	-24.7196	29.4396	-8.6596
45	32.5	34.4498	-1.9498	35.7749	-3.2749
46	19	40.6653	-21.6653	24.504	-5.504
47	51.5	44.809	6.691	49.0555	2.4445
48	49.17	46.1902	2.9798	55.0891	-5.9191
49	85	32.3779	52.6221	78.521	6.479
50	58.75	39.9747	18.7753	51.2836	7.4664
51	19.33	55.871	-36.541	18.9121	0.4179
52	5	47.5837	-42.5837	8.5873	-3.5873
53	65	43.44	21.56	65.883	-0.883
54	20	37.9151	-17.9151	14.2798	5.7202
55	62.5	55.1804	7.3196	71.6917	-9.1917
56	35	44.1306	-9.1306	54.4343	-19.4343
57	99.17	56.5617	42.6083	79.5143	19.6557
58	40.25	42.0465	-1.7965	33.2576	6.9924
59	39.17	42.0465	-2.8765	39.8176	-0.6476
60	37.5	35.1526	2.3474	35.5945	1.9055
61	26.25	44.809	-18.559	30.9101	-4.6601
62	52.14	42.7494	9.3906	61.8917	-9.7517
63	22.5	35.1526	-12.6526	17.9509	4.5491
64	90	45.4996	44.5004	66.6978	23.3022
65	28	46.1902	-18.1902	33.5763	-5.5763
66	50	43.44	6.56	69.9194	-19.9194
67	24.5	43.4278	-18.9278	43.7275	-19.2275

Calculate extra sum of squares for the comparison of these two analyses.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5641.64264	2820.82132	6.44	0.0028
Error	64	28029	437.94571		
Corrected Total	66	33670			

Root MSE	20.92715	R-Square	0.1676
Dependent Mean	42.16612	Adj R-Sq	0.1415
Coeff Var	49.63025		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	60.01473	5.78250	10.38	<.0001
pasture	1	-69.06150	19.71274	-3.50	0.0008
liming	1	-12.44327	5.65473	-2.20	0.0314

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	27503	9167.81454	93.66	<.0001
Error	63	6166.72438	97.88451		
Corrected Total	66	33670			

Root MSE	9.89366	R-Square	0.8168
Dependent Mean	42.16612	Adj R-Sq	0.8081
Coeff Var	23.46353		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.40906	4.83540	0.08	0.9329
sum	1	0.72727	0.04866	14.94	<.0001
pasture	1	-34.76756	9.59788	-3.62	0.0006
liming	1	1.95718	2.84172	0.69	0.4935

SSM(sum|pasture liming) = SSM(sum pasture liming) – SSM(pasture liming)

= 27503 – 5641.64264

= 21861.3574 = **21862**

F statistic: $\frac{\text{SSM(sum|pasture liming)}}{dof}$ / MSE(full model) = $(\frac{21861.3574}{1})$ / 97.88451 = 223.338 = **223.34**

Degree of freedom of this F statistic = **1,63**

(b) **Use test statement to obtain the same statistics. Give the test statistic, degree of freedom, p-value and conclusion.**

```
proc reg data=alfalfa2;
model rent = sum pasture liming/ p;
id sum pasture liming;
remove : test sum;
run;
```

Test remove Results for Dependent Variable rent				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	21862	223.34	<.0001
Denominator	63	97.88451		

Conclusion: Since p value is less than the α value we conclude that the statistic is significant and we reject the null hypothesis: H_0 = Sum = 0

<u>(c) Compare the test statistic and p-value from the test statement with the individual t-test for the coefficient of SUM variable in full model. Explain the relationship.</u>

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.40906	4.83540	0.08	0.9329
sum	1	0.72727	0.04866	14.94	<.0001
pasture	1	-34.76756	9.59788	-3.62	0.0006
liming	1	1.95718	2.84172	0.69	0.4935

The p-value from the General Linear test we conducted is the same as individual t-test for coefficient of sum in full model.

<u>Relationship:</u> t^2=F = $(14.94)^2$ = 223.2036

(3) **<u>Run regression to predict response using all variables, excluding sum. Use SS1 and SS2. Add type I sum of squares, do the same for type II. Do either sum to model sum of squares? Are any predictors for which the two are same? Explain why.</u>**

```
proc reg data=alfalfa2;
model rent = allrent cows pasture liming/ss1 ss2;
run;
```

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	-2.82821	4.67487	-0.60	0.5474	119125	31.72901
allrent	1	0.88327	0.06900	12.80	<.0001	25824	14205
cows	1	0.43176	0.10797	4.00	0.0002	2386.32278	1386.26332
pasture	1	-11.38045	11.89367	-0.96	0.3424	73.86014	79.37040
liming	1	-1.01173	2.84900	-0.36	0.7237	10.93243	10.93243

Type I SS Sum for all predictor variables = 28295.1154

Type II Sum for all predictor variables = 15681.5662

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	28295	7073.84035	81.60	<.0001
Error	62	5374.80658	86.69043		
Corrected Total	66	33670			

Type I SS Sum for all predictor variables = Model Sum of Squares.

Reason:

Type I SSM (cows) = SSM (allrent cows) -SSM(allrent)

Type I SSM (pasture) = SSM (allrent cows pasture) -SSM (allrent cows)

Type I SSM (liming) = SSM (allrent cows pasture liming) – SSM (allrent cows pasture)

Adding these together:

Type I SSM (allrent+cows+pasture+liming) = SSM (allrent cows pasture liming)

Type I SSM = Type II SSM for **variable added last**, in our case: liming.

Reason:

Type I SSM = Extra sum of squares for a variable after all the variables prior to it in the order of addition in model have been included, which is equal to Type II (extra sum of squares for a variable after adding all the other variables in the model) in case of the variable that is added last in the model.

(4) **Run regression to predict the response using a variety of variables, including sum as an explanatory variable, Summarize the result by making a table giving the percentage of variation explained by each model.**

```
proc reg data=alfalfa2;
model rent = allrent sum;
model rent = cows sum;
model rent = pasture sum;
model rent = liming sum;
model rent = allrent cows sum;
model rent = allrent pasture sum;
model rent = allrent liming sum;
model rent = cows pasture sum;
model rent = cows liming sum;
model rent = pasture liming sum;
model rent = allrent cows pasture liming sum;
run;
```

Variables	R squared
rent = allrent sum	0.8379
rent = cows sum	0.8379
rent = pasture sum	0.8155
rent = liming sum	0.7787
rent = allrent cows sum	0.8379
rent = allrent pasture sum	0.84
rent = allrent liming sum	0.838
rent = cows pasture sum	0.84
rent = cows liming sum	0.838
rent = pasture liming sum	0.8168
rent = allrent cows pasture liming sum	0.8404

Part II

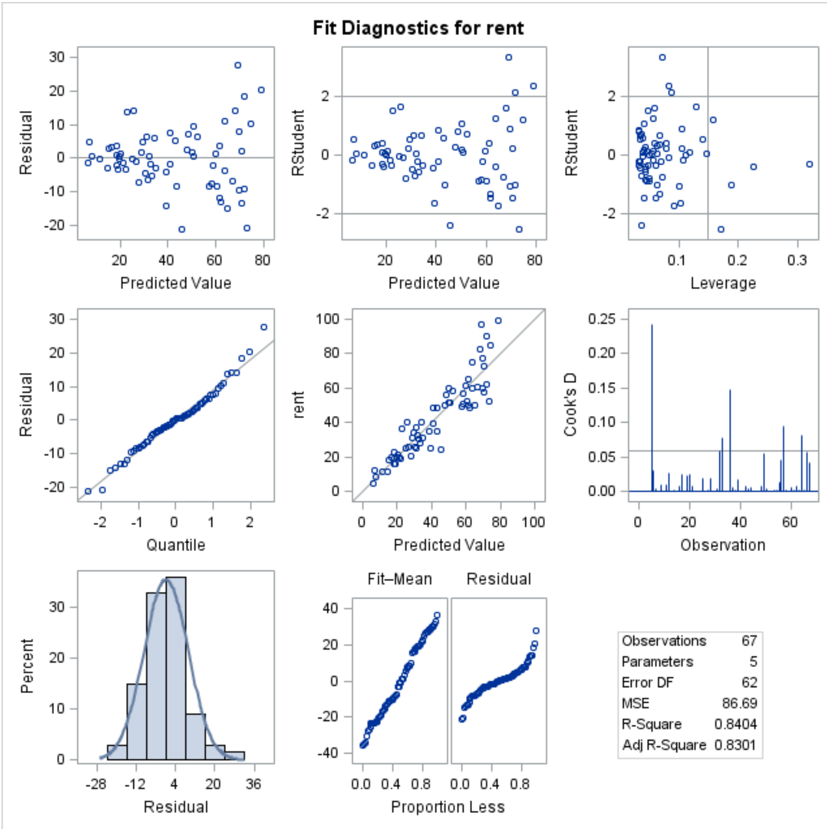
(1) Using techniques learned in class, determine whether the response variable and any of the predictors need to be transformed. Indicate the reasoning for your decision. If a variable need to be transformed, transform it and keep it in the full model for the rest of the questions.

Regression on original model:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	28295	7073.84035	81.60	<.0001
Error	62	5374.80658	86.69043		
Corrected Total	66	33670			

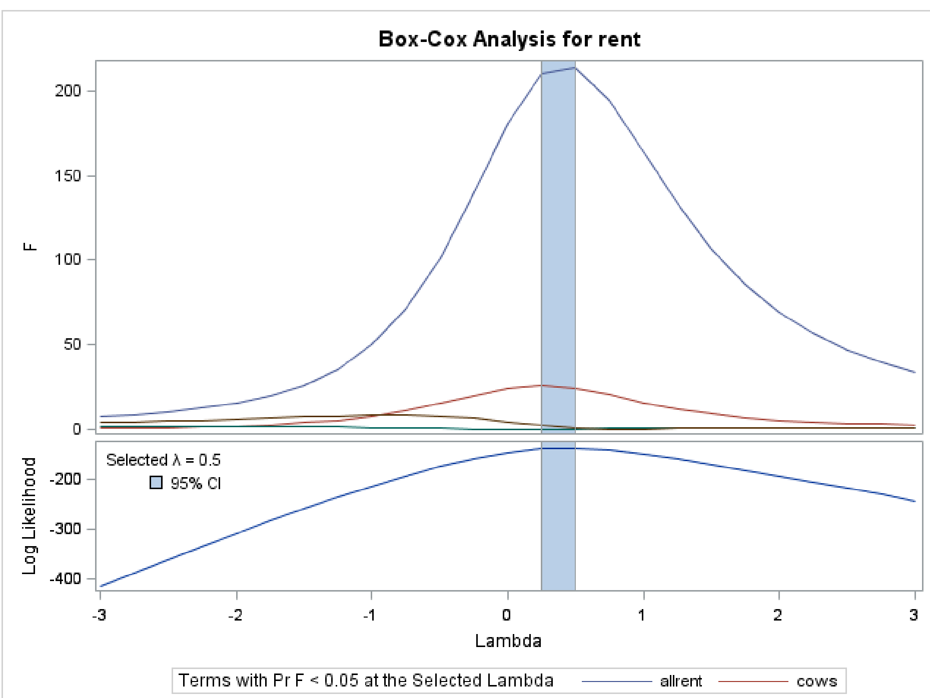
Root MSE	9.31077	R-Square	0.8404
Dependent Mean	42.16612	Adj R-Sq	0.8301
Coeff Var	22.08116		

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS	Variance Inflation
Intercept	1	-2.82821	4.67487	-0.60	0.5474	119125	31.72901	0
allrent	1	0.88327	0.06900	12.80	<.0001	25824	14205	1.62156
cows	1	0.43176	0.10797	4.00	0.0002	2386.32278	1386.26332	2.08740
pasture	1	-11.38045	11.89367	-0.96	0.3424	73.86014	79.37040	2.24858
liming	1	-1.01173	2.84900	-0.36	0.7237	10.93243	10.93243	1.56795



BOXCOX approach

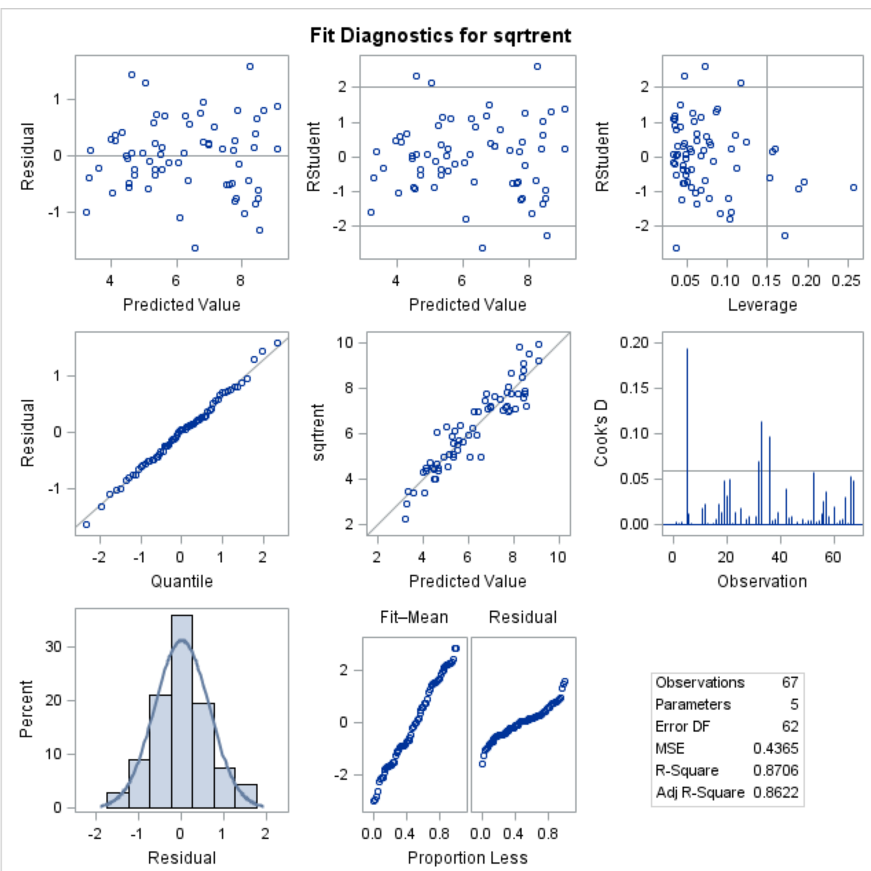
The residuals clearly have a megaphone effect, thus applying box cox to predict a better model. The best lambda is 0.5. Also, since pasture was not linearly related to the response variable, using $e^{-pasture}$ in place of pasture as predictor variable.

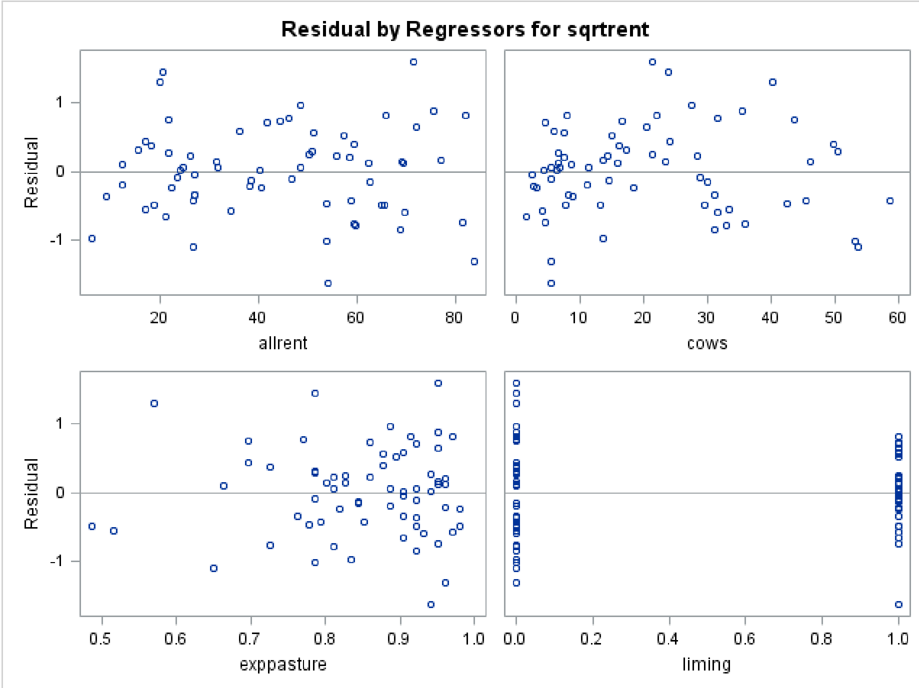


Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	182.03370	45.50843	104.26	<.0001
Error	62	27.06251	0.43649		
Corrected Total	66	209.09621			

Root MSE	0.66068	R-Square	0.8706
Dependent Mean	6.24862	Adj R-Sq	0.8622
Coeff Var	10.57314		

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS	Variance Inflation
Intercept	1	1.76025	0.98957	1.78	0.0802	2616.03379	1.38111	0
allrent	1	0.07167	0.00510	14.06	<.0001	165.75980	86.25617	1.75804
cows	1	0.03759	0.00801	4.69	<.0001	15.84868	9.60719	2.28294
exppasture	1	0.59682	1.18939	0.50	0.6176	0.14624	0.10990	2.55382
liming	1	0.16132	0.20178	0.80	0.4271	0.27898	0.27898	1.56206





Residual by Regressors for sqrtrent
Residual
1
0
-1
20
40
60
80
allrent
0
10
20
30
40
50
60
cows
0.5
0.6
0.7
0.8
0.9
1.0
exppasture
0.0
0.2
0.4
0.6
0.8
1.0
liming

(2) Use C_p criterion to select the best subset of variables for your data. Use original and transformed variables, not SUM. Summarize the results and explain your choice of best model.

With modified response variables and modified predictor variables: Cp =1.9742

```
proc reg data=transalfalfa;
model sqrtrent = allrent cows exppasture liming/ selection = cp b;
run;
```

```
Result: sqrtrent = 2.37024 + 0.07380*allrent + 0.03199*cows
```

Number in Model	C(p)	R-Square	Parameter Estimates				
			Intercept	allrent	cows	exppasture	liming
2	1.9742	0.8685	2.37024	0.07380	0.03199	.	.
3	3.2518	0.8700	2.23907	0.07332	0.03530	.	0.17075
3	3.6391	0.8692	1.81198	0.07187	0.03483	0.68545	.
4	5.0000	0.8706	1.76025	0.07167	0.03759	0.59682	0.16132
2	24.8287	0.8208	5.32356	0.08343	.	-3.18798	.
3	25.0100	0.8246	4.97954	0.08235	.	-2.58662	-0.24569
2	30.0216	0.8100	3.16515	0.07592	.	.	-0.46590
1	36.2834	0.7927	2.97876	0.07493	.	.	.
3	200.6122	0.4581	-5.42145	.	0.08789	11.40198	0.30233
2	200.8627	0.4534	-5.36241	.	0.08299	11.62590	.
2	348.9506	0.1442	0.85526	.	.	6.80777	-0.82516
2	356.4162	0.1286	4.85168	.	0.05072	.	0.71864
1	367.4855	0.1014	5.48882	.	0.03695	.	.
1	368.1231	0.1000	1.85956	.	.	5.15187	.
1	414.7626	0.0027	6.33842	.	.	.	-0.18232

With original response and predictor variables, Output: 1.9781

```
proc reg data=alfalfa;
model rent = allrent cows pasture liming/selection = cp b;
run;
```

```
rent = -6.11433 + .92137*allrent + .39255*cows
```

Number in Model	C(p)	R-Square	Parameter Estimates				
			Intercept	allrent	cows	pasture	liming
2	1.9781	0.8379	-6.11433	0.92137	0.39255	.	.
3	3.1261	0.8400	-3.70912	0.88212	0.44890	-10.90999	.
3	3.9156	0.8380	-5.57040	0.92335	0.37885	.	-0.70808
4	5.0000	0.8404	-2.82821	0.88327	0.43176	-11.38045	-1.01173
2	18.6049	0.7950	4.36983	0.95121	.	.	-7.54168
3	18.9910	0.7992	-0.34036	0.99249	.	12.97780	-6.10707
2	22.7348	0.7844	-6.24168	1.01660	.	23.82764	.
1	27.5051	0.7670	1.35258	0.93524	.	.	.
2	164.9157	0.4183	39.95497	.	0.96761	-104.21877	.
3	166.8570	0.4185	39.31634	.	0.97883	-103.81592	0.68922
2	262.3174	0.1676	60.01473	.	.	-69.06150	-12.44327
1	284.7796	0.1046	50.74431	.	.	-50.54868	.
2	285.5289	0.1078	27.33146	.	0.57311	.	6.19179
1	288.4139	0.0952	32.82105	.	0.45445	.	.
1	322.3225	0.0079	44.13059	.	.	.	-3.98847

With modified response variables and original predictor variables, No Cp is good enough

```
proc reg data=transalfalfa;
model sqrtrent= allrent cows pasture liming/selection = cp b;
run;
```

No model is good enough

Number in Model	C(p)	R-Square	Parameter Estimates				
			Intercept	allrent	cows	pasture	liming
2	2.0912	0.8685	2.37024	0.07380	0.03199	.	.
3	3.3675	0.8700	2.23907	0.07332	0.03530	.	0.17075
3	3.6051	0.8696	2.49903	0.07170	0.03501	-0.58421	.
4	5.0000	0.8708	2.36223	0.07152	0.03767	-0.51115	0.15711
3	27.2229	0.8203	2.57930	0.08105	.	1.61416	-0.28746
2	27.7551	0.8151	2.30153	0.08218	.	2.12486	.

Now, since Cp came out to be less than p in both cases when response variable is rent and sqrrent, we are looking at the residual plot to decide out model.

Model with modified rent and allrent and cows: Residual plot is good

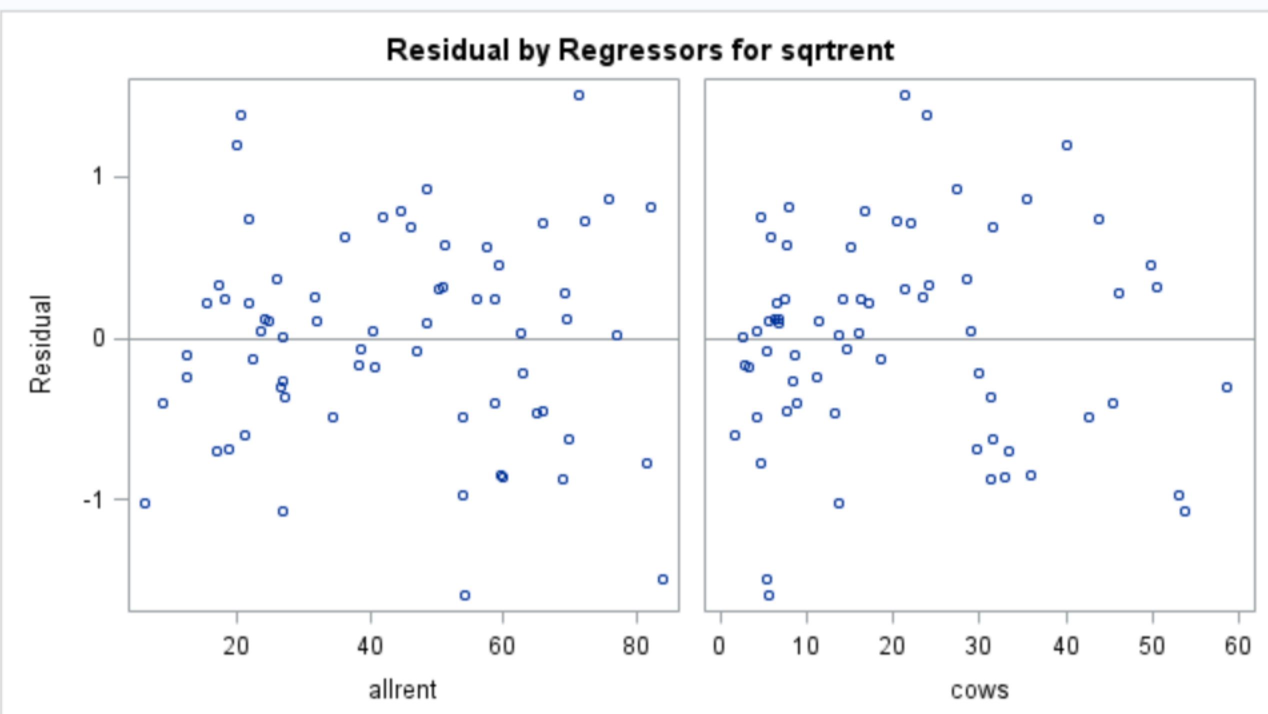
Model: MODEL1
Dependent Variable: sqrtrent

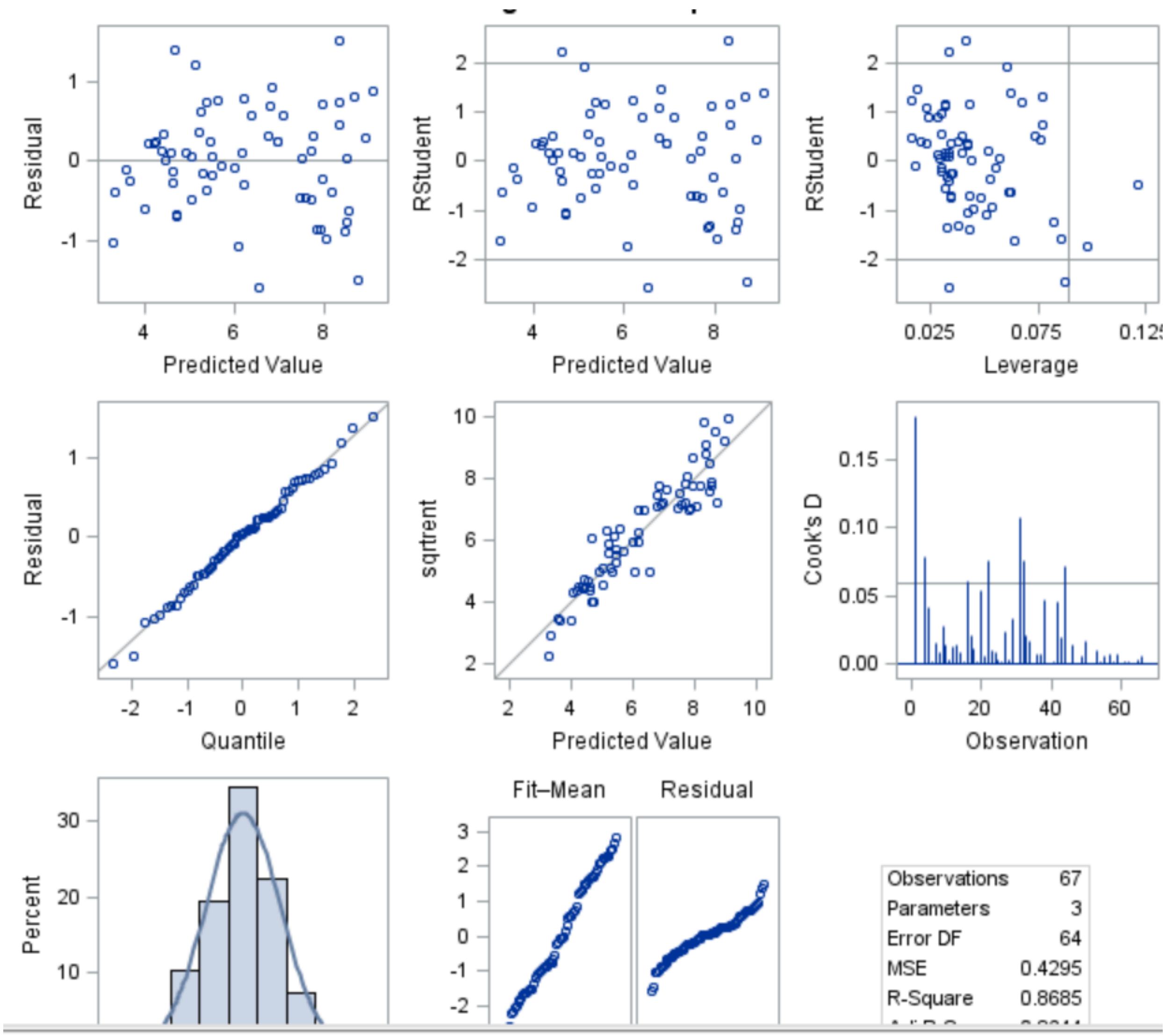
Number of Observations Read	67
Number of Observations Used	67

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	181.60848	90.80424	211.42	<.0001
Error	64	27.48772	0.42950		
Corrected Total	66	209.09621			

Root MSE	0.65536	R-Square	0.8685
Dependent Mean	6.24862	Adj R-Sq	0.8644
Coeff Var	10.48806		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	2.37024	0.21012	11.28	<.0001	0
allrent	1	0.07380	0.00382	19.33	<.0001	1.00238
cows	1	0.03199	0.00527	6.07	<.0001	1.00238




Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
sqrtrent
Predicted Value
Cook's D
Observation
Percent
Fit–Mean
Residual
Observations 67
Parameters 3
Error DF 64
MSE 0.4295
R-Square 0.8685

Model with rent allrent cows: Residual Plot shows megaphone effect.

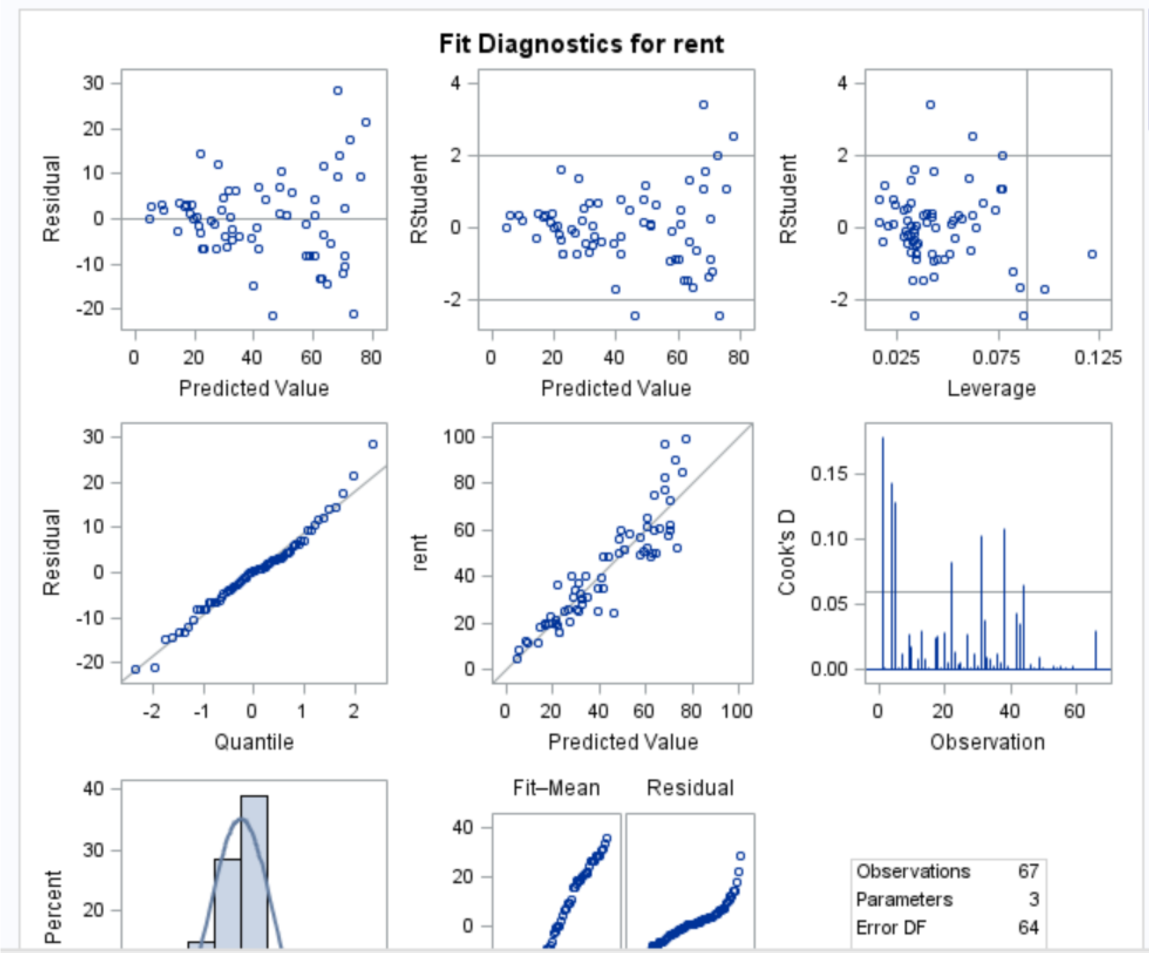
The REG Procedure
Model: MODEL1
Dependent Variable: rent

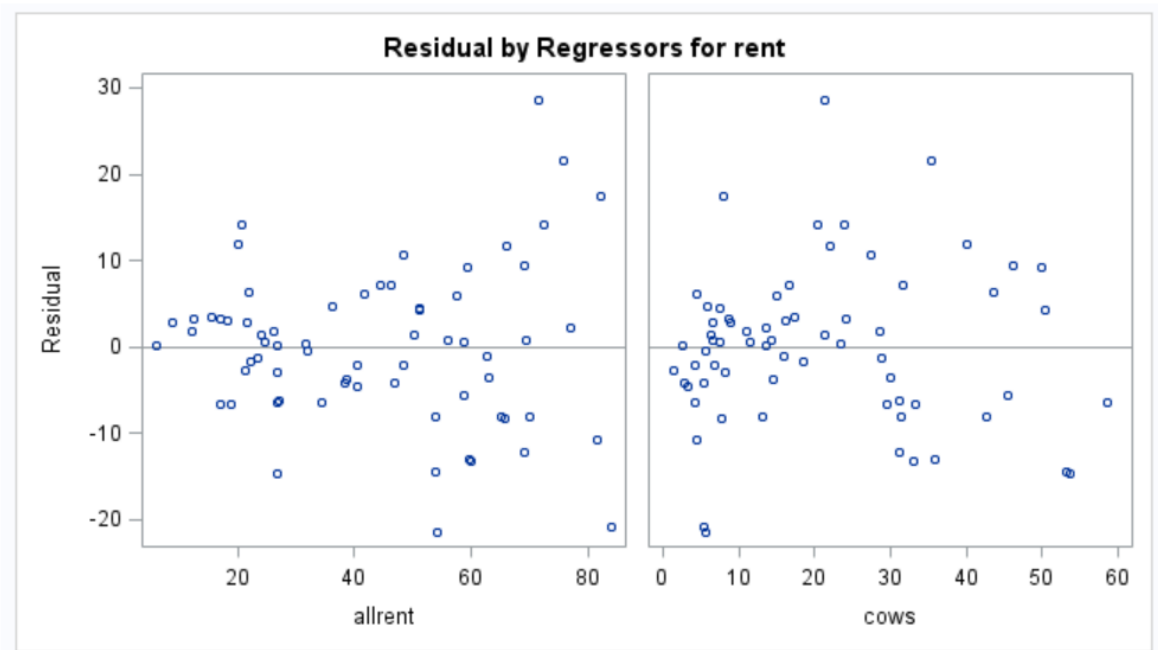
Number of Observations Read	67
Number of Observations Used	67

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	28211	14105	165.35	<.0001
Error	64	5459.59915	85.30624		
Corrected Total	66	33670			

Root MSE	9.23614	R-Square	0.8379
Dependent Mean	42.16612	Adj R-Sq	0.8328
Coeff Var	21.90417		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-6.11433	2.96123	-2.06	0.0430	0
allrent	1	0.92137	0.05382	17.12	<.0001	1.00238
cows	1	0.39255	0.07422	5.29	<.0001	1.00238





Since , the residuals are random and P value is less than α for sqrtrent and allrent and cows taking that as a model.

Result: sqrtrent = 2.37024 + 0.07380*allrent + 0.03199*cows

(3) Use selection = stepwise criterion to report the best model

We changed the significance level to 0.05 .SAS by default for stepwise selection uses 0.15 significance level.

```
proc reg data= std;
model sqrtrent= allrent cows epasture liming/slentry=0.05 slstay=0.05 selection =
stepwise;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: sqrtrent

Number of Observations Read	67
Number of Observations Used	67

Stepwise Selection: Step 1

Variable allrent Entered: R-Square = 0.7927 and C(p) = 36.6649

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	165.75980	165.75980	248.62	<.0001
Error	65	43.33641	0.66671		
Corrected Total	66	209.09621			

Stepwise Selection: Step 1

Variable allrent Entered: R-Square = 0.7927 and C(p) = 36.6649

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	165.75980	165.75980	248.62	<.0001
Error	65	43.33641	0.66671		
Corrected Total	66	209.09621			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.97876	0.23012	111.71036	167.55	<.0001
allrent	0.07493	0.00475	165.75980	248.62	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	181.60848	90.80424	211.42	<.0001
Error	64	27.48772	0.42950		
Corrected Total	66	209.09621			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.37024	0.21012	54.65393	127.25	<.0001
allrent	0.07380	0.00382	160.41575	373.50	<.0001
cows	0.03199	0.00527	15.84868	36.90	<.0001

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.97876	0.23012	111.71036	167.55	<.0001
allrent	0.07493	0.00475	165.75980	248.62	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable cows Entered: R-Square = 0.8685 and C(p) = 2.2161

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.37024	0.21012	54.65393	127.25	<.0001
allrent	0.07380	0.00382	160.41575	373.50	<.0001
cows	0.03199	0.00527	15.84868	36.90	<.0001

Bounds on condition number: 1.0024, 4.0095

All variables left in the model are significant at the 0.0500 level.

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

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	allrent		1	0.7927	0.7927	36.6649	248.62	<.0001
2	cows		2	0.0758	0.8685	2.2161	36.90	<.0001

Thus the stepwise selection method gives us the best model as

$$sqrtrent = \beta_0 + \beta_1(allrent) + \beta_2\,(cows) + \varepsilon$$

$$sqrtrent = \beta_0 + \beta_1(allrent) + \beta_2\,(cows) + \beta_2\,(epasture) + \beta_2\,(liming) + \varepsilon_i$$

(4) Check the assumptions of this best model using all the usual plots. Explain in detail whether for not each assumption appears to be substantially violated.

The best model is as follows:
Response variable= Square-root of rent
Predictors= Allrent and Cows

$\widehat{sqrtrent}$ = 2.37024 + 0.07380*allrent + 0.03199*cows

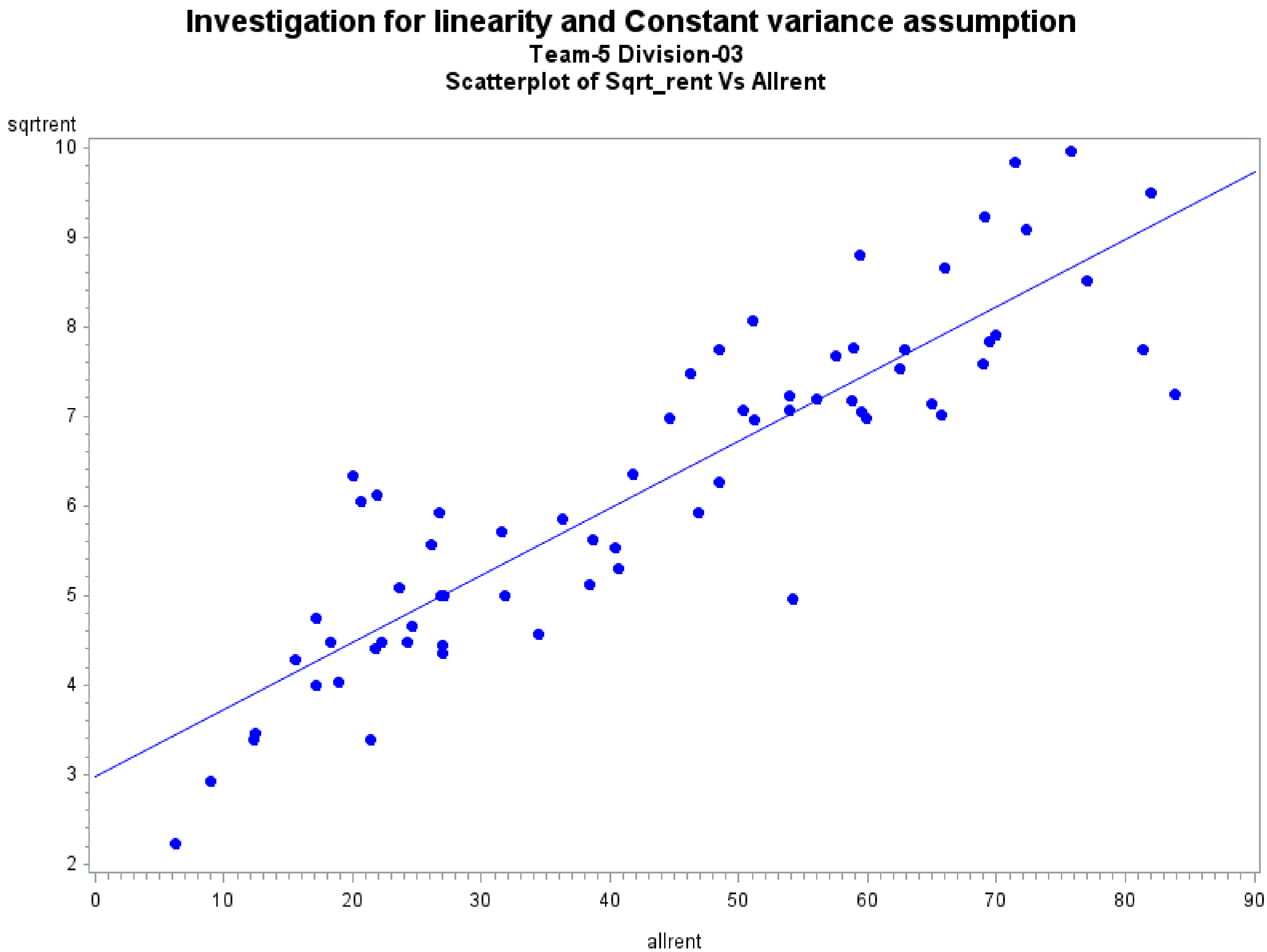
```
title1 ' Initial Investigation of the predictors ';
title2 ' Team-5 Division-03 ';
goptions colors=(blue);
symbol v=dot ;

proc reg data=transalfalfa;
model sqrtrent= allrent cows;
output out=output r=resid1 p= pred1 ;
run;
```

1. Linearity and Constant variance assumption :

Let us observe the scatter plot of each predictor against the response variable to examine the linearity and constant variation assumption.

a) Scatterplot of sqrtrent vs allrent

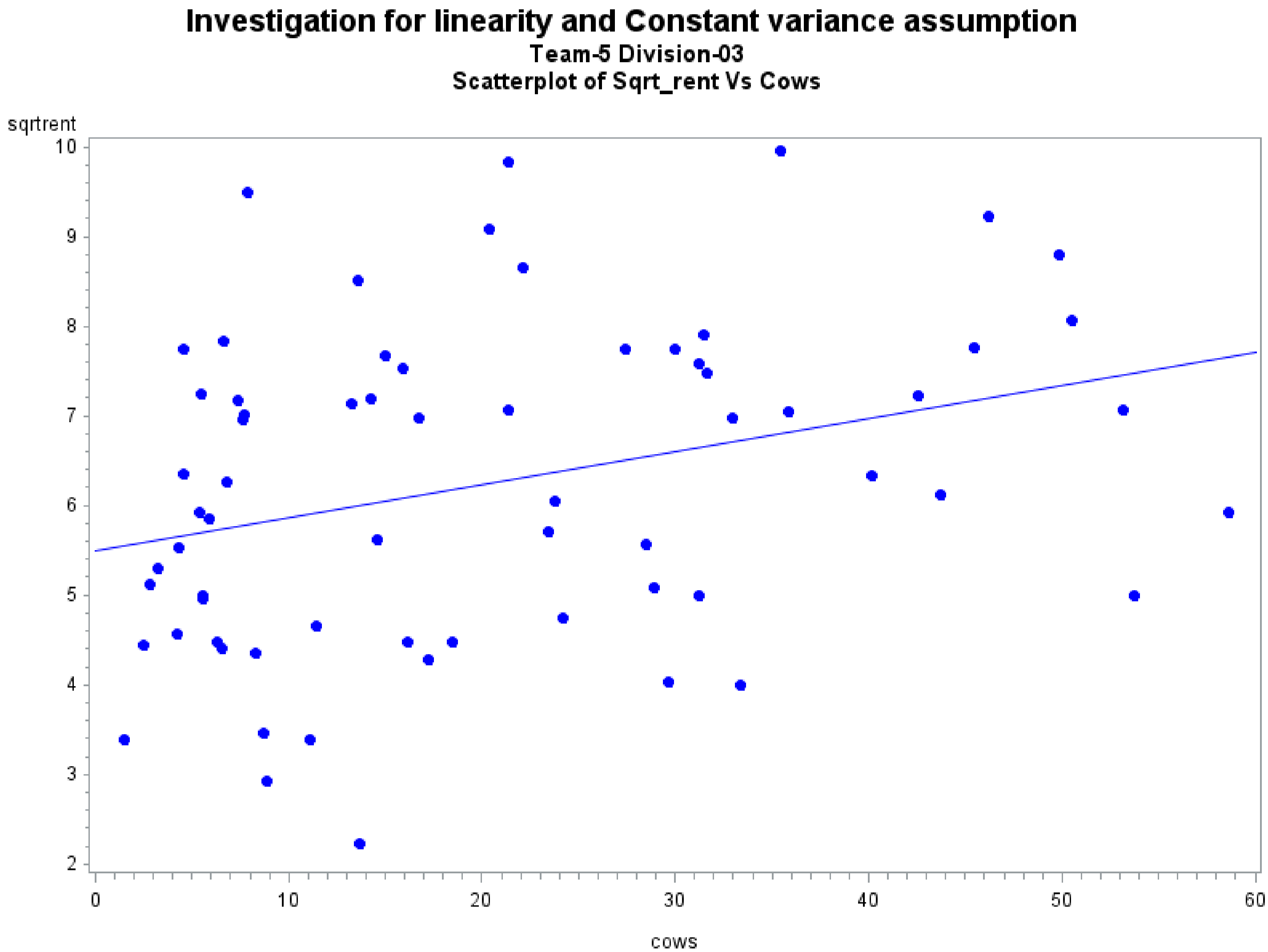
```
title1 'Investigation for linearity and Constant variance assumption ';
title2 ' Team-5 Division-03 ';
title3  'Scatterplot of Sqrt_rent Vs Allrent';
goptions colors=(blue);
symbol v=dot ;
/*------------------Q4 Linearity and Constant Variance assumption-------  */
proc gplot data= transalfalfa;  /* Scatterplot*/
plot sqrtrent* allrent ;
run;
```



Conclusion- The scatterplot does not show any striking deviation from linearity. In fact the data points are pretty close to least square line. Thus the allrent predictor follows a linear relationship with the response variable.

b) Scatterplot of sqrtrent vs cows

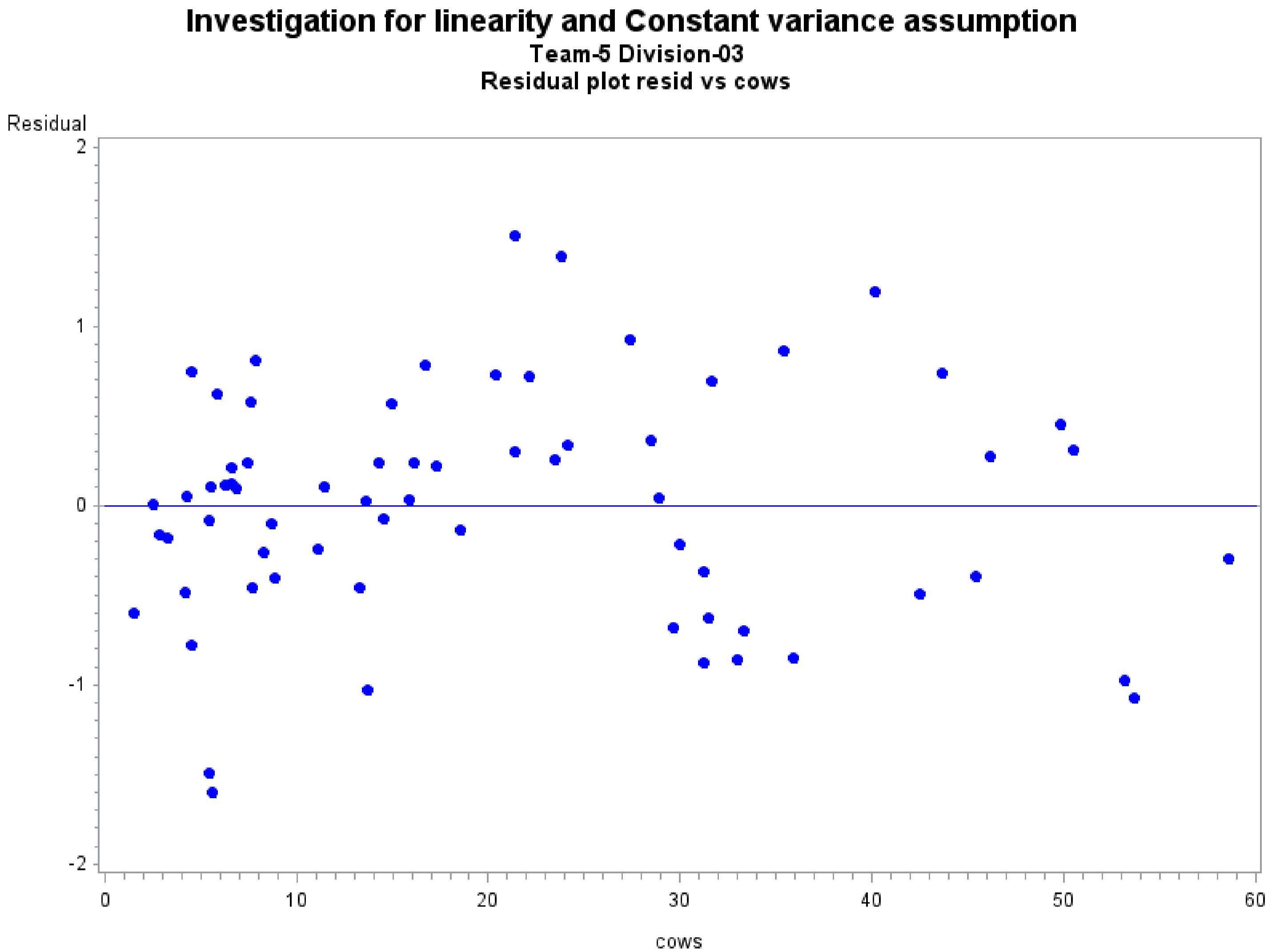
```
title1 'Investigation for linearity and Constant variance assumption ';
title2 ' Team-5 Division-03 ';
title3  'Scatterplot of Sqrt_rent Vs Cows';
goptions colors=(blue);
symbol v=dot ;
proc gplot data= transalfalfa;  /* Scatterplot*/
plot sqrtrent* cows;
run;
```



Conclusion- The scatterplot does not show any striking deviation from linearity. In fact the data points are pretty close to least square line. Thus the cows predictor follows a linear relationship with the response variable.

c) The residual plot for Sqrtrent vs cows:

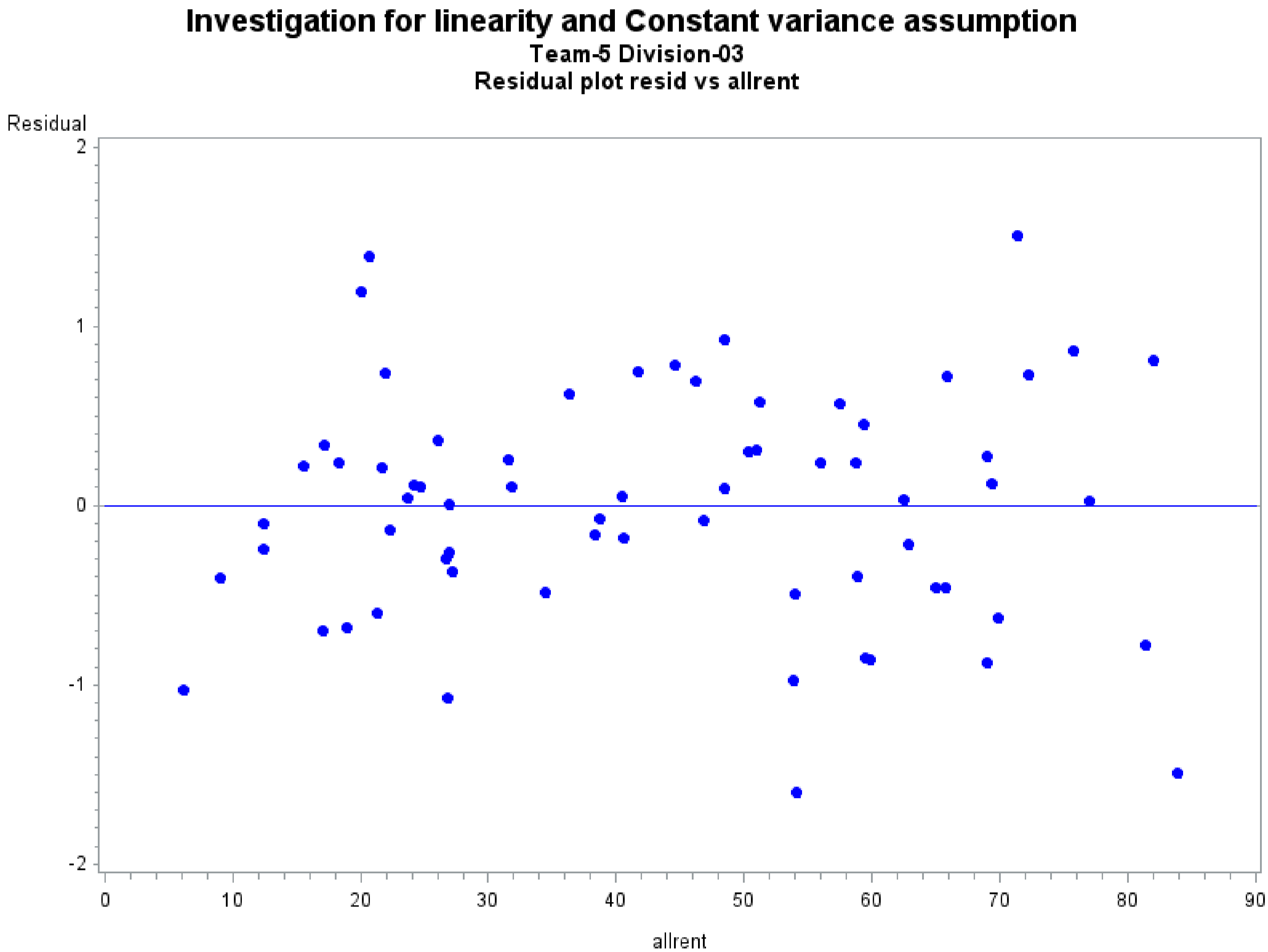
```
title1 'Investigation for linearity and Constant variance assumption ';
title2 ' Team-5 Division-03 ';
title3  'Residual plot resid vs cows';
proc gplot data= output;           /* Residual plot for residual vs Cows*/
plot resid1*cows /vref=0;
run;
```



Conclusion- The Residual plot does not show any striking deviation from random pattern. Thus the cows predictor does not violate the constant variance assumption.

d) The residual plot for Sqrtrent vs allrent:

```
title1 'Investigation for linearity and Constant variance assumption ';
title2 ' Team-5 Division-03 ';
title3  'Residual plot resid vs allrent';
proc gplot data= output;          /* Residual plot for residual vs AllRent*/
plot resid1* AllRent /vref=0;

run;
```

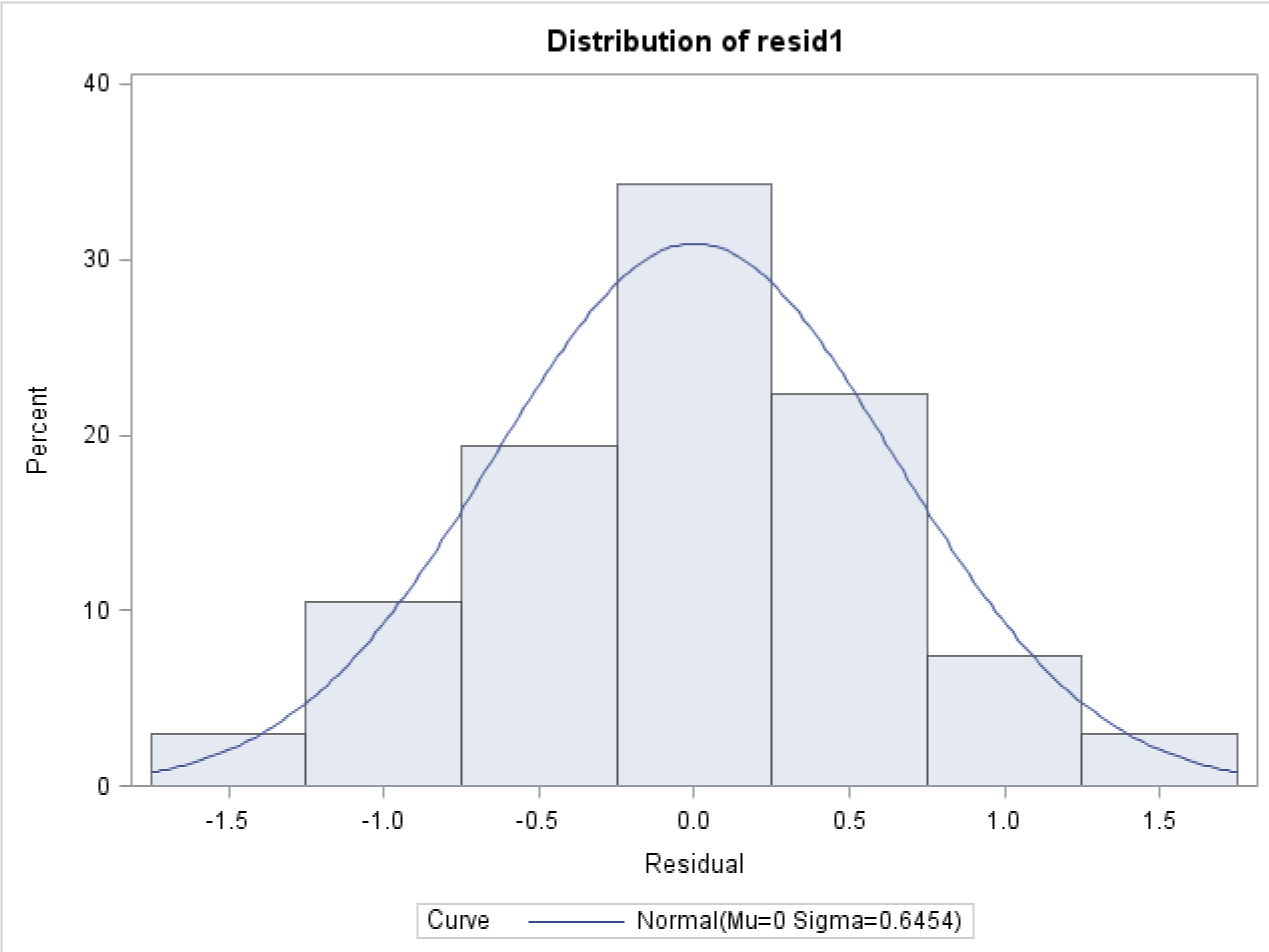


Conclusion- The Residual plot does not show any striking deviation from random pattern. Thus the allrent predictor does not violate the constant variance assumption.

2) Normality

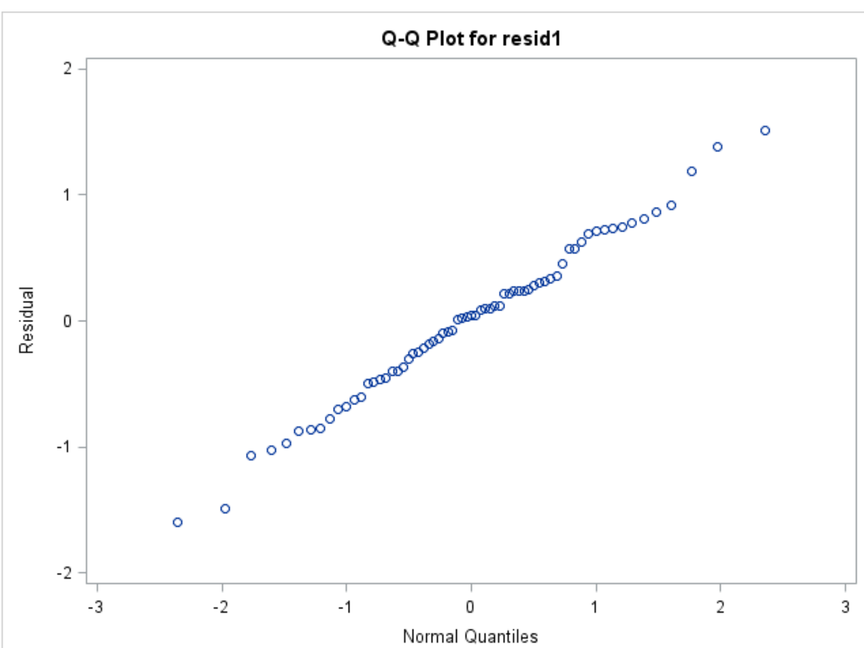
a) Histogram of the residuals

```
title1 'Rent for land planted to Alfalfa ';
title2 ' Team-5 Division-03 ';
title3  'Qqplot';


proc univariate data=output plot;
var resid1;
histogram resid1 / normal ;   /* The histogram has a perfect bell shape!! :)*/
qqplot resid1;
run;
```



Conclusion- The Histogram depicts almost a perfect bell shape, thus the normality assumption does not seen to be violated.

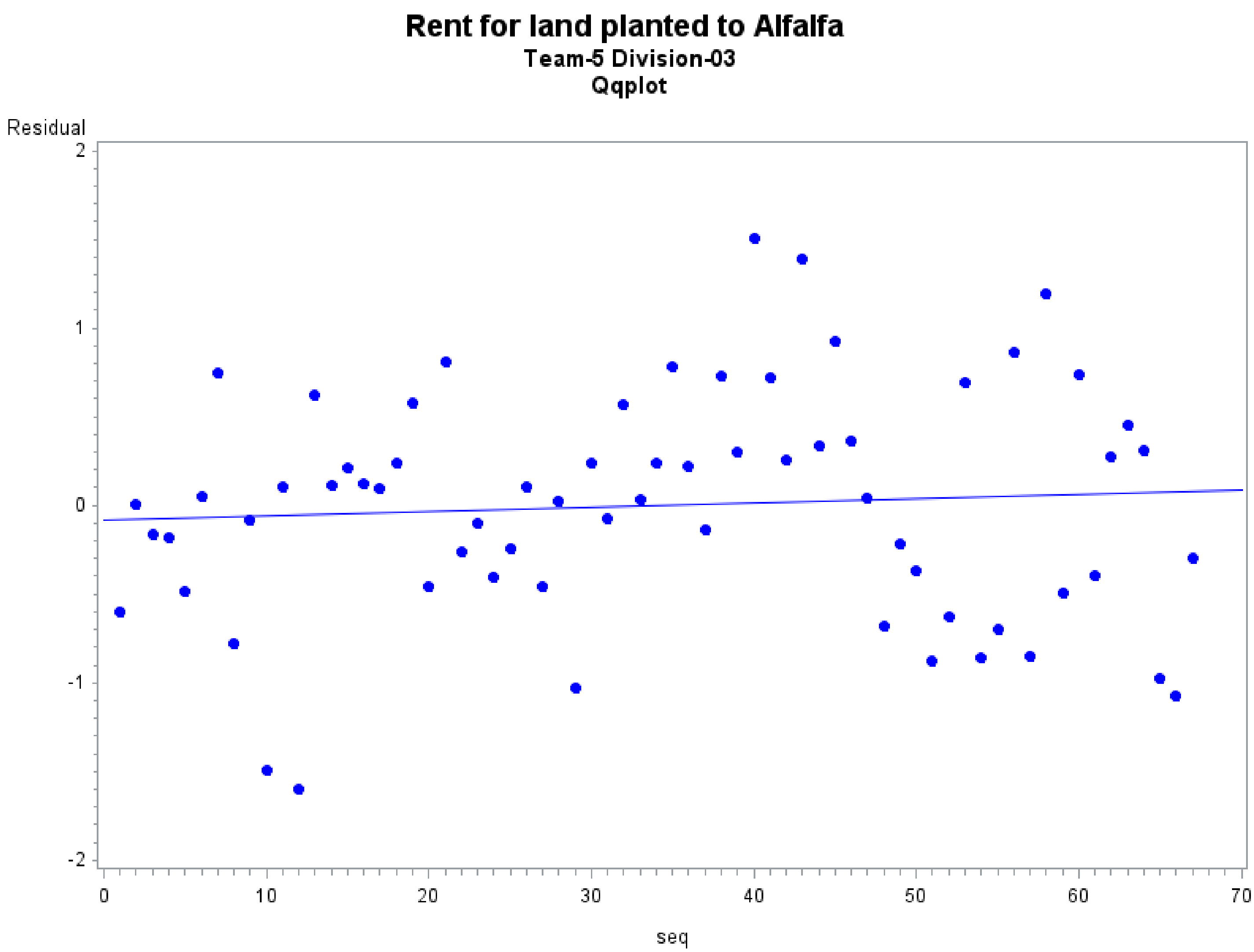
b) Qqplot/ Normal Quantile plot



Conclusion- The QQplot does not show any striking deviation from the 45 degree line through origin. Thus it is safe to assume the normality of response variable.

3) Independence

```
data output;
set output;
seq=_n_;
proc print data= output;
run;

proc gplot data= output;            /* Residual plot for residual vs AllRent*/
plot resid1* seq ;

run;
```



ConclusionThe Sequence plot does not show any striking deviation from random patterm. Thus it is safe to assume Independence of the residuals.

(5) Use the best model to predict the response variable. Examine other diagnositics such as studentized, Cook's etc. Explain any problems such as outliers, highly influential observations or multicollinearity that these diagnostics point out.

<u>Examination of Diagnostics</u>

Examinations of diagnostics are done for :
1) Outliers:

Can be diagnosed by observing ***Studentized residuals, semi Studentized residuals*** and ***Studentized deleted residuals***

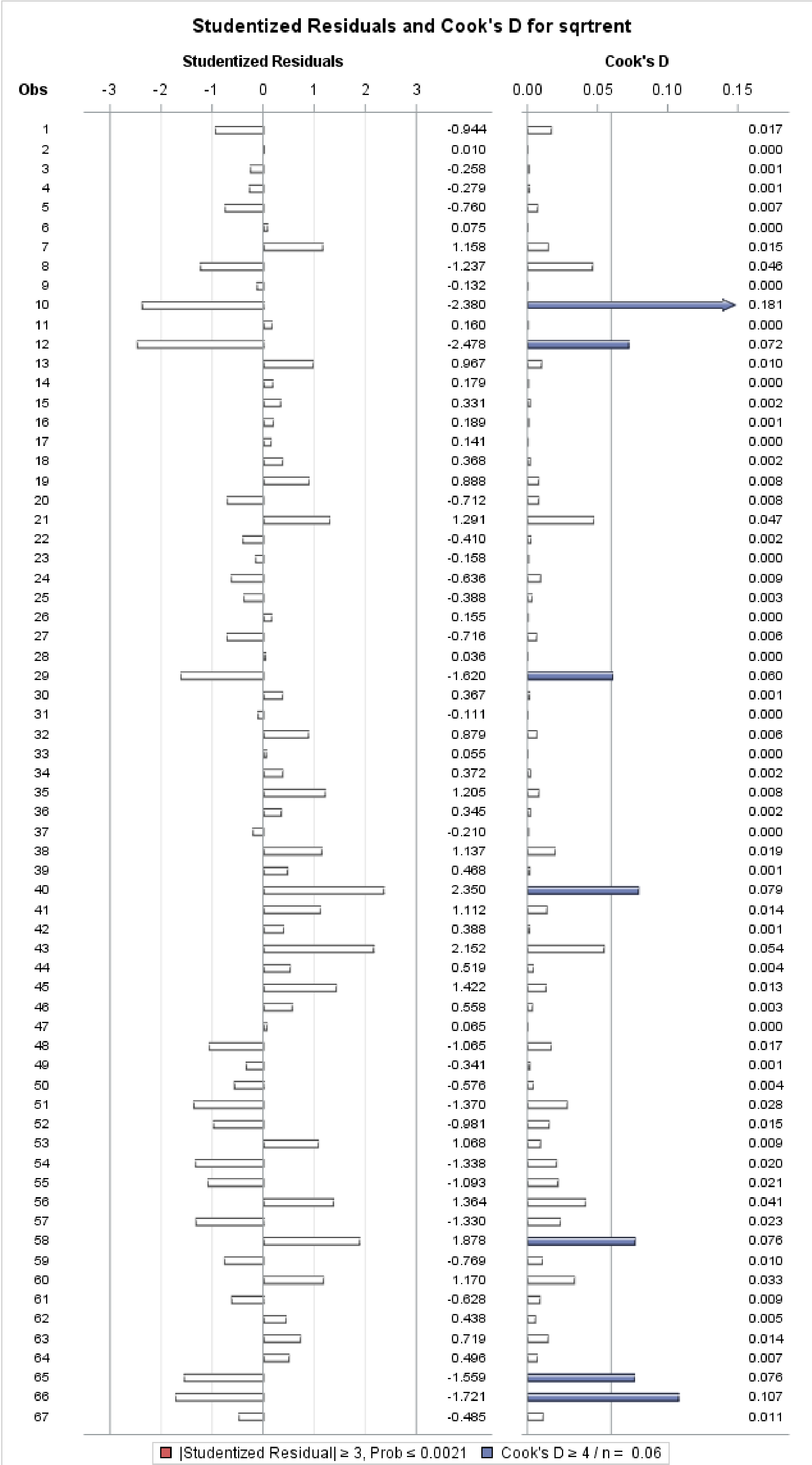
2) Influencial observations:

Can be diagnosed by observing ***Cook' distance*** **,** ***hat matrix*** **,** ***DFFITS , DFBETAS***

3) Multicollinearity issue:

Can be examined by metrics such as tolerance and VIF along with the partial residual plots.

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows/r influence alpha= 0.05;
plot r.*(allrent cows);
output out=datatrans p=sqrt_renthat r=resid;
run;
```



Studentized Residuals and Cook's D for sqrtrent
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.05 0.10 0.15
1 -0.944 0.017
2 0.010 0.000
3 -0.258 0.001
4 -0.279 0.001
5 -0.760 0.007
6 0.075 0.000
7 1.158 0.015
8 -1.237 0.046
9 -0.132 0.000
10 -2.380 0.181
11 0.160 0.000
12 -2.478 0.072
13 0.967 0.010
14 0.179 0.000
15 0.331 0.002
16 0.189 0.001
17 0.141 0.000
18 0.368 0.002
19 0.888 0.008
20 -0.712 0.008
21 1.291 0.047
22 -0.410 0.002
23 -0.158 0.000
24 -0.636 0.009
25 -0.388 0.003
26 0.155 0.000
27 -0.716 0.006
28 0.036 0.000
29 -1.620 0.060
30 0.367 0.001
31 -0.111 0.000
32 0.879 0.006
33 0.055 0.000
34 0.372 0.002
35 1.205 0.008
36 0.345 0.002
37 -0.210 0.000
38 1.137 0.019
39 0.468 0.001
40 2.350 0.079
41 1.112 0.014
42 0.388 0.001
43 2.152 0.054
44 0.519 0.004
45 1.422 0.013
46 0.558 0.003
47 0.065 0.000
48 -1.065 0.017
49 -0.341 0.001
50 -0.576 0.004
51 -1.370 0.028
52 -0.981 0.015
53 1.068 0.009
54 -1.338 0.020
55 -1.093 0.021
56 1.364 0.041
57 -1.330 0.023
58 1.878 0.076
59 -0.769 0.010
60 1.170 0.033
61 -0.628 0.009
62 0.438 0.005
63 0.719 0.014
64 0.496 0.007
65 -1.559 0.076
66 -1.721 0.107
67 -0.485 0.011
|Studentized Residual| ≥ 3, Prob ≤ 0.0021
Cook's D ≥ 4 / n = 0.06

a) Studentized Residuals-

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /r influence alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid student=Studentized;
run;

proc univariate data= datatrans;
var studentized;
run;
```

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.47822	67	1.36436	57
-2.38049	5	1.42153	11
-1.72128	32	1.87838	33
-1.62040	52	2.15226	19
-1.55865	66	2.35018	36

After observing the student residual column in the output statistics , we notice that there exists no

Studentized residual < -3 or Studentized residual >3, thereby implying there is no striking outlier based on this criteria.

It is also very clear from the chart shown just above. None of the horizontal bars have crossed the 3/-3 boundaries.

Thus, according to the ***Studentized residual*** criterion, ***we have no possible outlier in our data set***.

b) Studentized deleted Residuals-

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /r influence alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid rstudent= deleted_resid;
run;

proc print data=datatrans;
run;

proc univariate data= datatrans;
var deleted_resid;
run;
```

Here, t_c (n-p-1, 1-$\frac{\alpha}{2n}$) = t(n-p-1, 1-$\frac{0.05}{134}$)

From our data set, n= 67, p=3

= t(63, 1-(0.0003731))

= -3.551

$|t_c|$ = 3.545

Thus, after examining the R student column of the output statistics,

We observe that there is no $|t_i| > |t_c|$;

Thus, according to the ***Studentized deleted residual*** criterion, ***we have no possible outlier in our data set***.

Also, we verified the same by executing a proc univariate over the regression output for the variable deleted_residual .

The table shown below delineates the extreme values obtained for the deleted residuals, as we can clearly see that:

the highest extreme value(|2.43|) <| -3.545| and

the lowest extreme value(|-2.58|) < |3.545| .

Thus it verifies our claim that there exists no possible outlier according to the Studentized deleted residual criterion.

The SAS System

The UNIVARIATE Procedure
Variable: deleted_resid (Studentized Residual without Current Obs)

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.58599	12	1.37379	56
-2.47388	10	1.43318	45
-1.74874	66	1.91724	58
-1.64172	29	2.21713	43
-1.57664	65	2.43938	40

c) Cook's Distance

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /r influence alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid rstudent= deleted_resid cookd=cook_distance;
run;

proc print data=datatrans;
run;

proc univariate data= datatrans;
var cook_distance;
run;
```

From our data set, n= 67, p=3

F critical (p-1, n-p) = F(2, 64)at 50th percentile= 0.7

Thus, after examining the Cook's D column of the output statistics,

We observe that there is no distance > F_c;

Thus, according to the ***Cook's Distance*** criterion, ***we have no possible influential observation in our data set***.

Also, we verified the same by executing a proc univariate over the regression output for the variable cook_distance .

The table shown below delineates the extreme values obtained for the deleted residuals, as we can clearly see that:

the highest extreme value(0.18) <F_c (0.7),

Thus it verifies our claim that there exists no possible influential observation according to the Cook's distance criterion.

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1.53261E-06	2	0.0757894	65
2.58842E-05	28	0.0762199	58
2.99145E-05	33	0.0788073	40
4.90551E-05	47	0.1073695	66
6.23269E-05	6	0.1811329	10

c) Hat Matrix:

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /r influence alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid rstudent= deleted_resid cookd=cook_distance
h=hat_matrix;
run;

proc print data=datatrans;
run;

proc univariate data= datatrans;
var hat_matrix;
run;
```

$h_c = \frac{2p}{n} = \frac{6}{67} = 0.089$

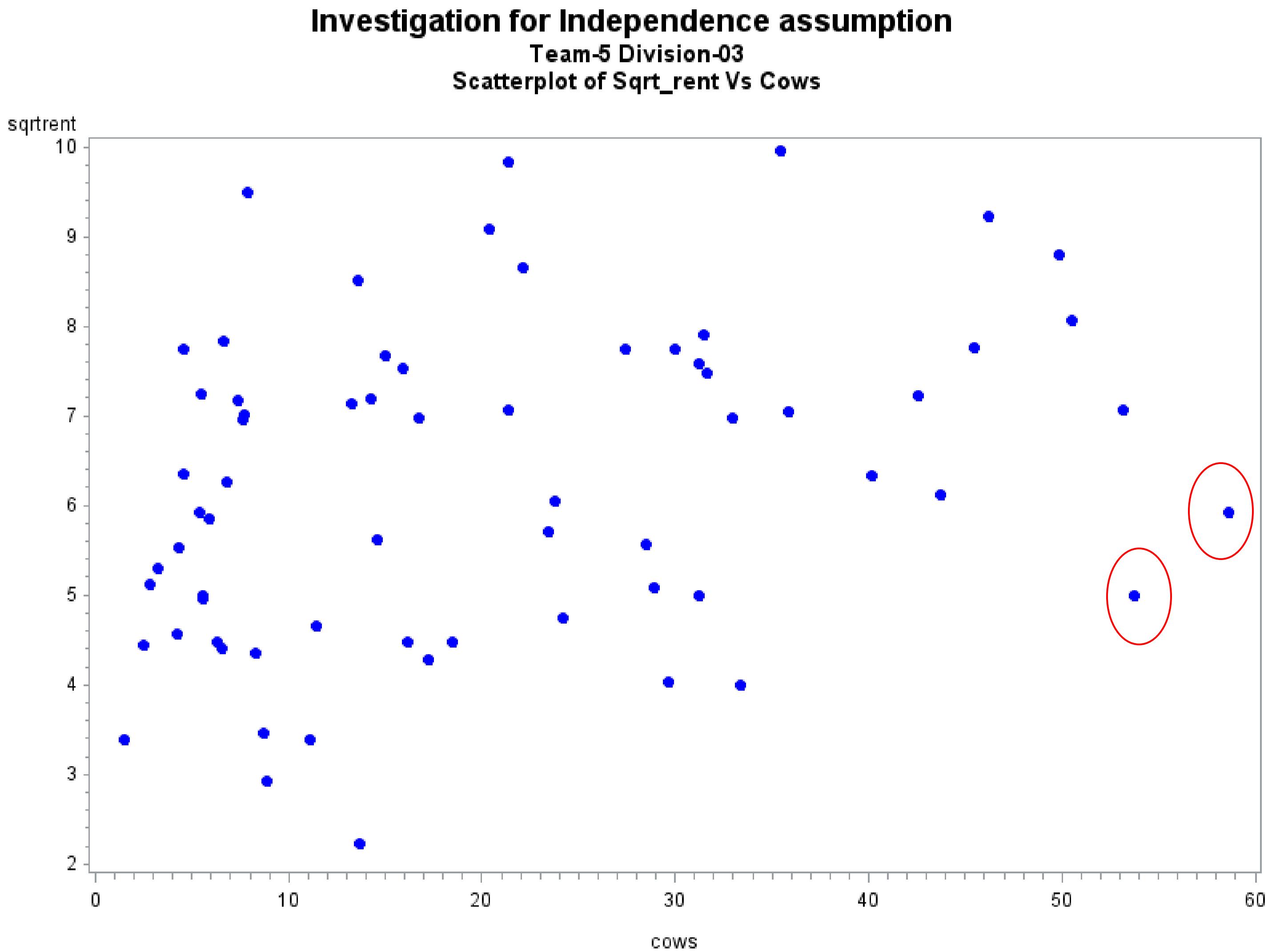
Thus, values lager than 0.089 would be considered potential influential observations

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0159341	43	0.0826723	20
0.0164571	22	0.0855806	66
0.0179594	26	0.0875017	5
0.0185820	11	0.0980567	32
0.0205936	45	0.1210585	56

We can observe from the table that the two observations 66 and 67 have the hat matrix diagonal values greater than 0.089. Thus these two data points are influential.

Thus, according to the ***Hat matrix diagonal*** criterion, ***we two (66 , 67) possible influential observation in our data set***.

From the Scatterplot (shown below) of cows vs SqrtRent, we can see that the observations 66 (corresponding cows value= 53) and 67 corresponding cows value= 58) are slightly influential.



c) DFFITS: Measure of influence of case i on its $\widehat{Y}_i$

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /r influence alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid rstudent= deleted_resid cookd=cook_distance
h=hat_matrix dffits=df_fit;
run;

proc univariate data= datatrans;
var df_fit;
run;
```

Our dataset has n=67 observations, thus we consider it as a medium sized data set.

In the DFFITS column of the output statistics, there exists no entry whose value is larger than 1. Thus,

Thus, according to the ***DFFITS*** criterion, ***there are no possible influential observations in our data set***.

d) DFBETAS: Measure of influence of case i on each of the regression coefficients.

Our dataset has n=67 observations, thus we consider it as a medium sized data set.

In the DFFITS columns for intercept, allrent and cows of the output statistics, there exists no entry whose value is larger than 1. Thus,

Thus, according to the ***DFBETAS*** criterion, ***there are no possible influential observations in our data set***.

e) VIF (Variation Inflation Factor)

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /vif tol alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid rstudent= deleted_resid cookd=cook_distance
h=hat_matrix;
run;
```

The VIF for both the predictors viz. allrent and cows is approximately 1.0 , which is < 10.

Thus, according to the ***VIF*** criterion, ***there exists no excessive multicollinearity issue***.

f) Tolerance

```
proc reg data= transalfalfa;
model sqrtrent= allrent cows /vif tol alpha= 0.05;
output out= datatrans p=sqrt_renthat r=resid rstudent= deleted_resid cookd=cook_distance
h=hat_matrix;
run;
```

The tolerance values for both the predictors is 0.99763 which is <0.1.

Thus, according to the ***Tolerance*** criterion, ***there exists no excessive multicollinearity issue***.

Parameter Estimates

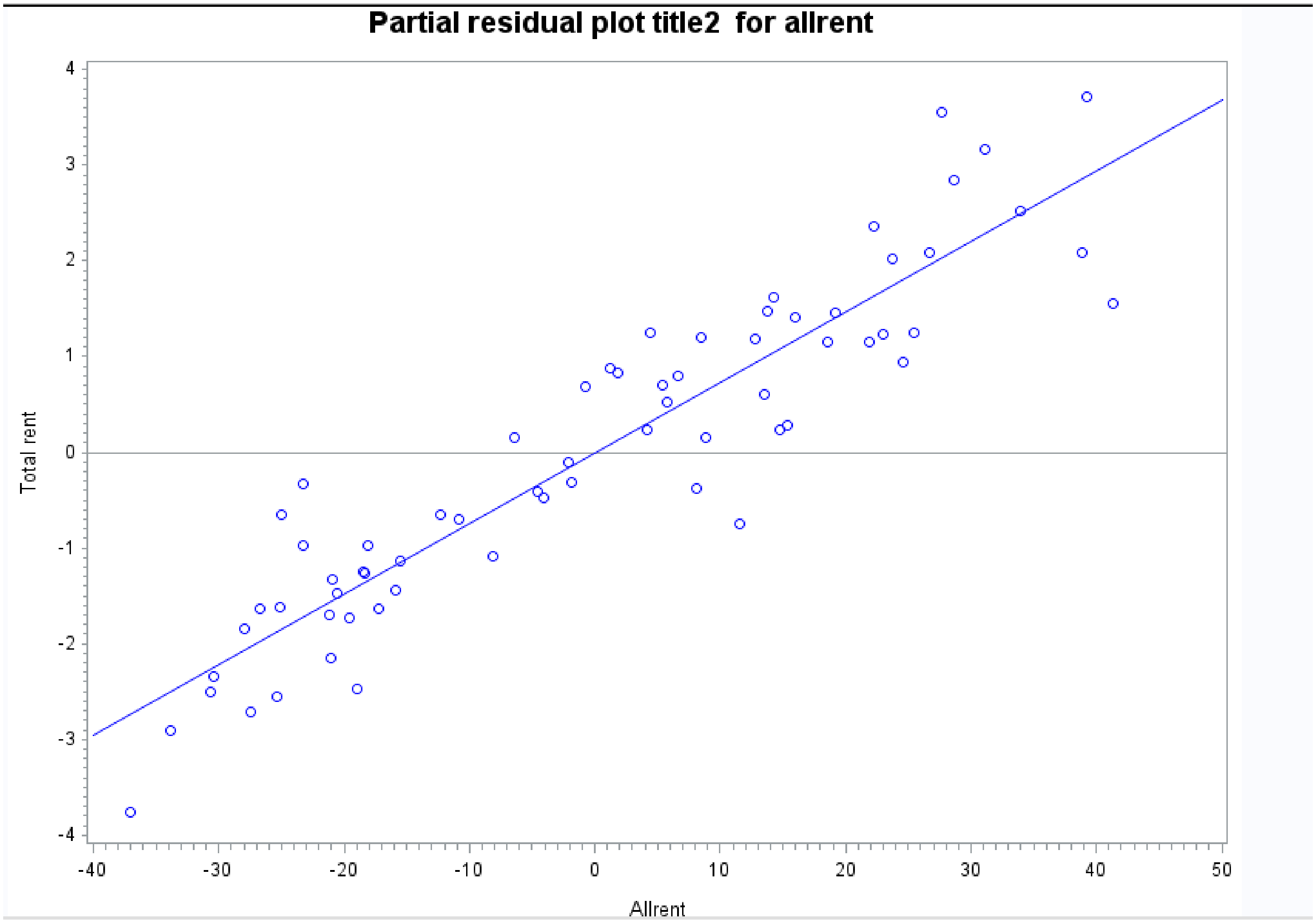
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	2.37024	0.21012	11.28	<.0001	.	0
allrent	1	0.07380	0.00382	19.33	<.0001	0.99763	1.00238
cows	1	0.03199	0.00527	6.07	<.0001	0.99763	1.00238

g) Partial residual plots

Analyzing the worth of Allrent in the model.

```
proc reg data= transalfalfa;
model sqrtrent allrent= cows / alpha= 0.05;
output out= partialAllrent p=sqrt_renthat r=resid_sqrt_rent resid_allrent;
run;
```

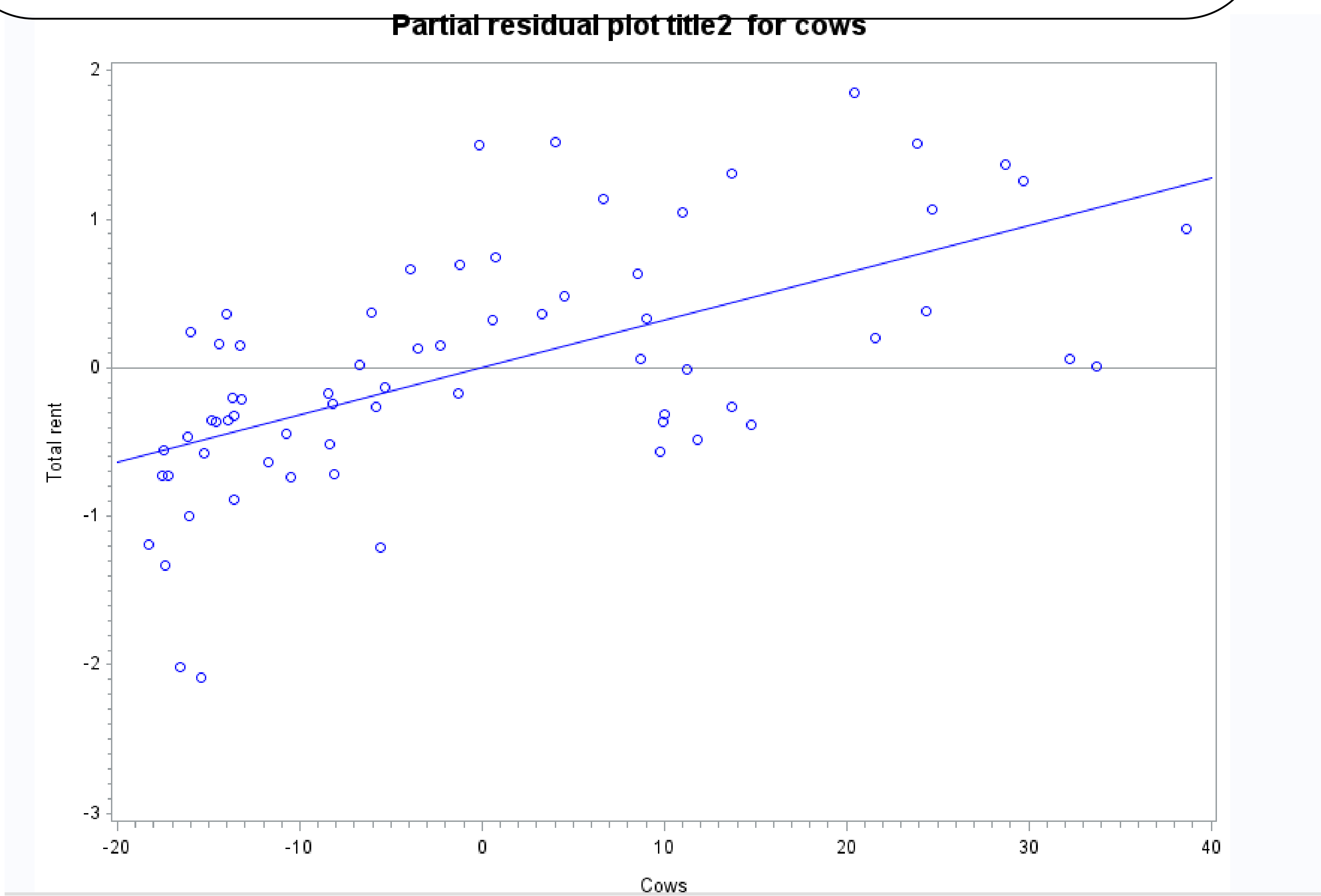
```
title1 'Partial residual plot'
title2 ' for allrent ';
symbol v=circle i=rl;
axis1 label= ('Allrent');
axis2 label= (angle=90 'Total rent');


proc gplot data= partialAllrent;
plot resid_sqrt_rent * resid_allrent / haxis=axis1 vaxis=axis2 vref = 0;
run;
```



As we can see a high value of r, thus indicating there is much to be gained by including allrent in the model.

Analyzing the worth of cows in the model.

```
proc reg data= transalfalfa;
model sqrtrent cows= allrent / alpha= 0.05;
output out= partialAllrent p=sqrt_renthat r=resid_sqrt_rent resid_cows;
run;

title1 'Partial residual plot';
title2 ' for cows ';
symbol v=circle i=rl;
axis1 label= ('Cows');
axis2 label= (angle=90 'Total rent');


proc gplot data= partialAllrent;
plot resid_sqrt_rent * resid_cows / haxis=axis1 vaxis=axis2 vref = 0;
run;
```



As we can see a fairly good value of r, thus indicating there is much to be gained by including cows in the model.

In both the scatterplots, we can see the points are closer to the regression line than to the X –axis. Thus, there is much to be gained from adding both the predictors.

(6) **For the best model report the following:**

(a) **Equation of the regression model**

```
Result: sqrtrent = 2.37024 + 0.07380*allrent + 0.03199*cows
```

(b) **90% confidence interval for the mean of the response variable**

(c) **90% prediction interval for individual observations**

(d) **90% confidence intervals for regression coefficients**

```
proc reg data=transalfalfa;
model sqrtrent = allrent cows/ clm cli clb alpha=0.1;
run;
```

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	90% CL Mean		90% CL Predict		Residuals
1	4.2872	4.066	0.1345	3.8415	4.2904	2.9494	5.1825	0.2212
2	4.4721	4.6074	0.1144	4.4164	4.7983	3.497	5.7177	-0.1352
3	3.3912	3.6384	0.1502	3.3877	3.8892	2.5163	4.7606	-0.2473
4	5.000	4.8972	0.1198	4.6973	5.0972	3.7853	6.0091	0.1028
5	7.2457	8.7359	0.1939	8.4124	9.0595	7.5953	9.8766	-1.4903
6	9.083	8.3538	0.1355	8.1277	8.58	7.2369	9.4708	0.7291
7	5.000	5.3712	0.1177	5.1747	5.5677	4.2599	6.4826	-0.3712
8	5.5381	5.4897	0.1175	5.2936	5.6858	4.3784	6.6009	0.0484
9	3.4641	3.5648	0.1543	3.3073	3.8223	2.4411	4.6885	-0.1007
10	7.8262	7.7054	0.149	7.4568	7.954	6.5837	8.8271	0.1208
11	7.746	6.8231	0.0893	6.674	6.9722	5.7191	7.927	0.9229
12	7.5829	8.4614	0.1357	8.2349	8.6879	7.3444	9.5784	-0.8785
13	5.5678	5.2074	0.1137	5.0177	5.3971	4.0972	6.3175	0.3604
14	7.746	7.9661	0.1178	7.7694	8.1627	6.8547	9.0774	-0.2201
15	8.5147	8.4919	0.1565	8.2306	8.7531	7.3673	9.6165	0.0228
16	7.7672	8.1661	0.1619	7.8958	8.4364	7.0394	9.2928	-0.3989
17	7.0534	7.9082	0.127	7.6963	8.1201	6.7941	9.0224	-0.8549
18	2.9155	3.3188	0.164	3.0451	3.5925	2.1913	4.4464	-0.4033
19	6.0415	4.6551	0.1207	4.4537	4.8565	3.5429	5.7673	1.3864
20	7.746	8.5227	0.1884	8.2082	8.8372	7.3845	9.6608	-0.7767
21	4.0311	4.7141	0.1343	4.4899	4.9382	3.5975	5.8306	-0.6829

22	7.0711	6.7671	0.0841	6.6268	6.9074	5.6643	7.8699	0.304
23	3.3912	3.9933	0.1513	3.7408	4.2458	2.8707	5.1159	-0.6021
24	5.9161	6.0011	0.1141	5.8107	6.1915	4.8908	7.1113	-0.085
25	8.6603	7.9435	0.1169	7.7484	8.1386	6.8324	9.0546	0.7168
26	5.6178	5.6902	0.0878	5.5436	5.8368	4.5866	6.7938	-0.0724
27	6.9642	6.3908	0.11	6.2072	6.5744	5.2817	7.4999	0.5734
28	8.8034	8.3504	0.1815	8.0475	8.6533	7.2154	9.4854	0.453
29	4.6551	4.5552	0.1168	4.3604	4.7501	3.4442	5.6663	0.0999
30	4.4441	4.4377	0.1377	4.2079	4.6675	3.32	5.5554	0.006394
31	7.4833	6.7913	0.0992	6.6257	6.9568	5.685	7.8975	0.692
32	5.000	6.0713	0.2052	5.7288	6.4138	4.9251	7.2175	-1.0713
33	6.3246	5.1316	0.1617	4.8617	5.4014	4.005	6.2582	1.193
34	7.5279	7.4924	0.1113	7.3067	7.6782	6.383	8.6019	0.0355
35	7.1965	6.9588	0.0995	6.7928	7.1248	5.8525	8.0651	0.2377
36	9.8321	8.3238	0.1328	8.1022	8.5454	7.2078	9.4398	1.5083
37	7.1295	7.5907	0.1219	7.3873	7.7941	6.4781	8.7032	-0.4612
38	5.8592	5.2348	0.114	5.0446	5.425	4.1246	6.345	0.6244
39	6.9821	7.8447	0.1189	7.6462	8.0431	6.733	8.9563	-0.8625
40	5.0794	5.0376	0.1204	4.8366	5.2386	3.9254	6.1497	0.0418
41	4.4721	4.3574	0.1305	4.1396	4.5752	3.2421	5.4727	0.1148
42	4.000	4.698	0.1479	4.4511	4.9449	3.5767	5.8193	-0.698
43	6.9764	6.1929	0.0827	6.0549	6.331	5.0905	7.2954	0.7835
44	4.5585	5.0477	0.1215	4.8448	5.2505	3.9352	6.1601	-0.4892
45	5.7009	5.4494	0.094	5.2924	5.6064	4.3444	6.5544	0.2515
46	4.3589	4.6232	0.1194	4.4239	4.8226	3.5114	5.7351	-0.2644
47	7.1764	6.9397	0.1221	6.7358	7.1435	5.827	8.0523	0.2367
48	7.0121	7.4684	0.1366	7.2403	7.6964	6.3511	8.5857	-0.4563
49	9.2195	8.9433	0.181	8.6413	9.2454	7.8086	10.0781	0.2762
50	7.6649	7.0958	0.1012	6.9269	7.2647	5.989	8.2026	0.569
51	4.3966	4.1844	0.135	3.959	4.4097	3.0676	5.3012	0.2122
52	2.2361	3.2632	0.1664	2.9855	3.5409	2.1347	4.3917	-1.0271

53	8.0623	7.7495	0.1778	7.4527	8.0463	6.6161	8.8828	0.3128
54	4.4721	4.2327	0.127	4.0207	4.4448	3.1186	5.3469	0.2394
55	7.9057	8.5343	0.1385	8.3031	8.7656	7.4164	9.6523	-0.6286
56	5.9161	6.2138	0.228	5.8333	6.5944	5.0557	7.372	-0.2978
57	9.9584	9.0924	0.1632	8.8201	9.3647	7.9652	10.2196	0.866
58	6.3443	5.5977	0.1163	5.4036	5.7919	4.4868	6.7086	0.7466
59	6.2586	6.1676	0.1101	5.9839	6.3514	5.0585	7.2768	0.091
60	6.1237	5.3837	0.1707	5.0988	5.6686	4.2534	6.514	0.74
61	5.1235	5.2895	0.1239	5.0826	5.4963	4.1763	6.4027	-0.166
62	7.2208	7.7125	0.1446	7.4712	7.9539	6.5924	8.8327	-0.4917
63	4.7434	4.4103	0.131	4.1915	4.629	3.2948	5.5257	0.3332
64	9.4868	8.6741	0.1824	8.3697	8.9786	7.5387	9.8095	0.8127
65	5.2915	5.4711	0.1214	5.2685	5.6737	4.3587	6.5835	-0.1796
66	7.0711	8.0479	0.1917	7.7279	8.3678	6.9082	9.1875	-0.9768
67	4.9497	6.5461	0.1207	6.3446	6.7476	5.4339	7.6583	-1.5963

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	90% Confidence Limits	
Intercept	1	2.37024	0.21012	11.28	<.0001	2.01955	2.72093
allrent	1	0.07380	0.00382	19.33	<.0001	0.06743	0.08017
cows	1	0.03199	0.00527	6.07	<.0001	0.02320	0.04078