General Linear Model Exercise

Data was collected on the number weeds in a certain area of a plot of grass. Before the experiment was run, each plot of grass had a quality of grass rating recorded (GrassQuality continuous scale from 0-100). Then each plot was randomly selected to either be treated with a weed killer (spray=Y) or not be treated (spray=N). The amount of weeds in the plot was then the response variable (WeedCount 0,1,2,...).

1. What type of variable is Grass Quality? Spray? Weed Count?
2. (Model 1) Ignoring the Grass Quality measurement, what type of model would likely be appropriate here? Write out the model.
3. (Model 2) Ignoring the Spray variable, what type of linear model would likely be appropriate here? Write out the model.
4. (Model 3) Write out a linear model that includes Grass Quality and a different intercept for Spray=N and Spray=Y.
5. (Model 4) Write out a linear model that includes a different slope term for Grass Quality with Spray=N and Spray=Y.
6. (Model 5) Write out a linear model that includes a different intercept and a different slope for Grass Quality with Spray=N and Spray=Y.

   SAS code (available on the website) gives the code for fitting these 5 models in SAS proc reg. The variables are WeedCount, GrassQuality, sprayY=(spray="Y"), and GrassspraY=GrassQuality*sprayY. The output is given below. Answer the following questions:

7. What is the fitted equation of the line for Spray=N and Spray=Y for models 2-5?
8. Using the output, construct the F-statistic for the LOF test as well as the degrees of freedom for the test of each of the following:
   (a) Model 2 with model 3
   (b) Model 2 with model 5
   (c) Model 3 with model 5.
9. Compare these values to the F-stats for these tests in the output to make sure you did them correctly. What conclusions do you make? What model appears to be the best in terms of fit and model assumptions?

The REG Procedure
Model: MODEL1
Dependent Variable: WeedCount

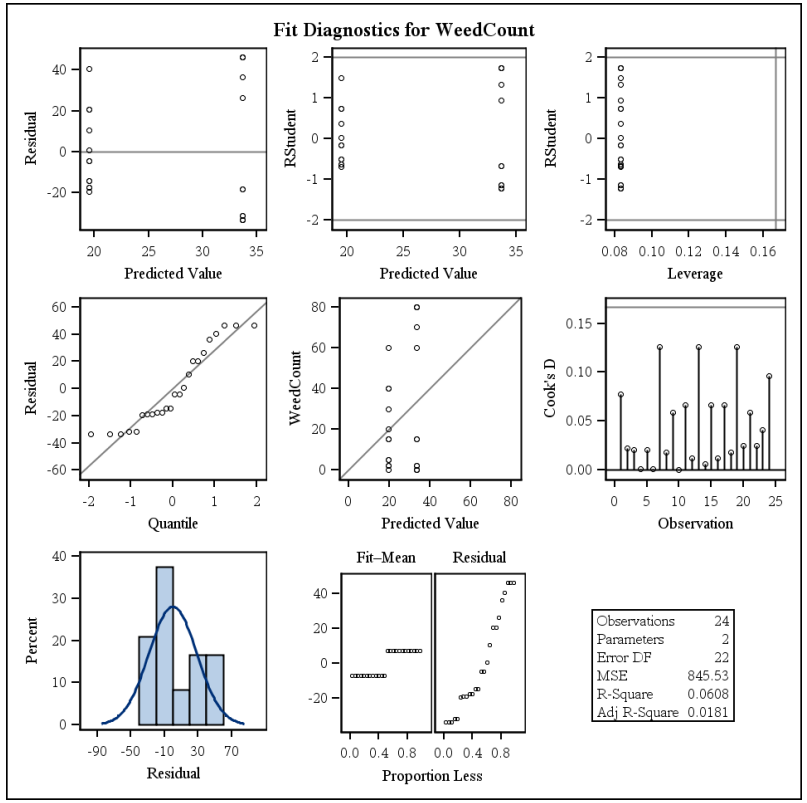
Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1204.16667	1204.16667	1.42	0.2454
Error	22	18602	845.53030		
Corrected Total	23	19806			

Root MSE	29.07800	R-Square	0.0608
Dependent Mean	26.58333	Adj R-Sq	0.0181
Coeff Var	109.38434		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	33.66667	8.39410	4.01	0.0006
sprayY	1	-14.16667	11.87105	-1.19	0.2454

The REG Procedure
Model: MODEL1
Dependent Variable: WeedCount



The REG Procedure
Model: MODEL2
Dependent Variable: WeedCount

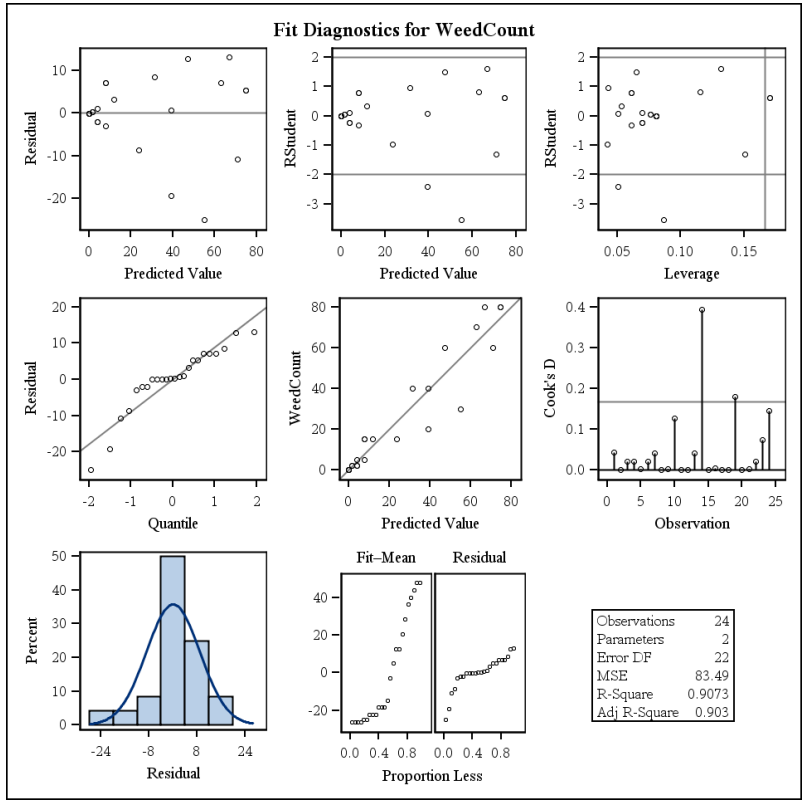
Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	17969	17969	215.22	<.0001
Error	22	1836.78361	83.49016		
Corrected Total	23	19806			

Root MSE	9.13730	R-Square	0.9073
Dependent Mean	26.58333	Adj R-Sq	0.9030
Coeff Var	34.37227		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	78.66421	4.01018	19.62	<.0001
GrassQuality	1	-0.78563	0.05355	-14.67	<.0001

The REG Procedure
Model: MODEL2
Dependent Variable: WeedCount



The REG Procedure
Model: MODEL3
Dependent Variable: WeedCount

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

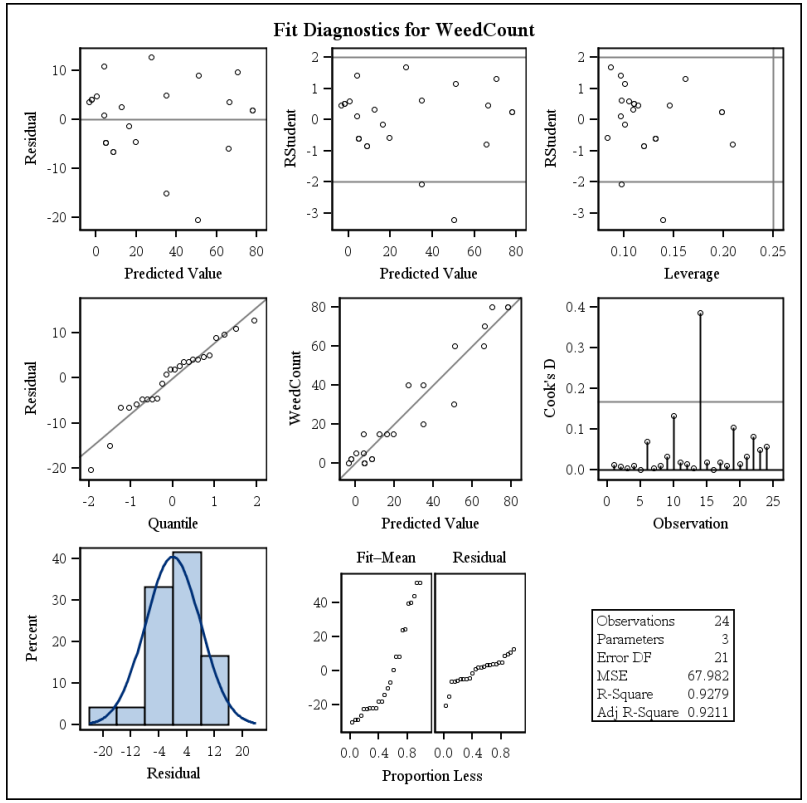
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	18378	9189.11037	135.17	<.0001
Error	21	1427.61260	67.98155		
Corrected Total	23	19806			

Root MSE	8.24509	R-Square	0.9279
Dependent Mean	26.58333	Adj R-Sq	0.9211
Coeff Var	31.01602		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	81.95722	3.85953	21.24	<.0001
GrassQuality	1	-0.77265	0.04861	-15.89	<.0001
sprayY	1	-8.30741	3.38617	-2.45	0.0230

The REG Procedure
Model: MODEL3
Dependent Variable: WeedCount



The REG Procedure
Model: MODEL3

Test 1 Results for Dependent Variable WeedCount				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	409.17101	6.02	0.0230
Denominator	21	67.98155		

The REG Procedure
Model: MODEL4
Dependent Variable: WeedCount

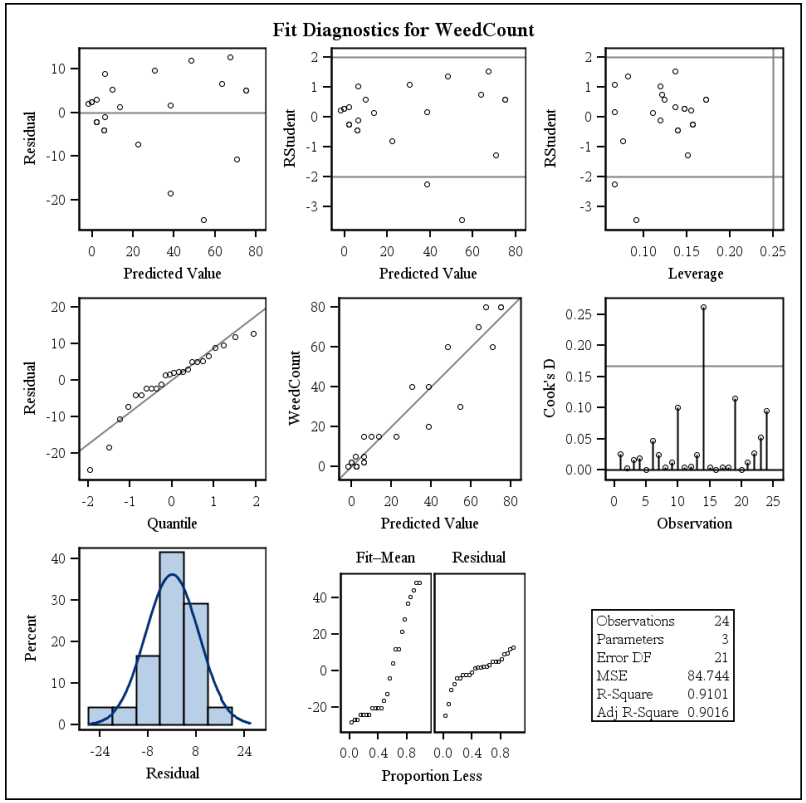
Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	18026	9013.10212	106.36	<.0001
Error	21	1779.62909	84.74424		
Corrected Total	23	19806			

Root MSE	9.20566	**R-Square**	0.9101
Dependent Mean	26.58333	**Adj R-Sq**	0.9016
Coeff Var	34.62946		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	78.87184	4.04809	19.48	<.0001
GrassQuality	1	-0.76693	0.05856	-13.10	<.0001
GrassprayY	1	-0.04131	0.05030	-0.82	0.4207

The REG Procedure
Model: MODEL4
Dependent Variable: WeedCount



The REG Procedure
Model: MODEL5
Dependent Variable: WeedCount

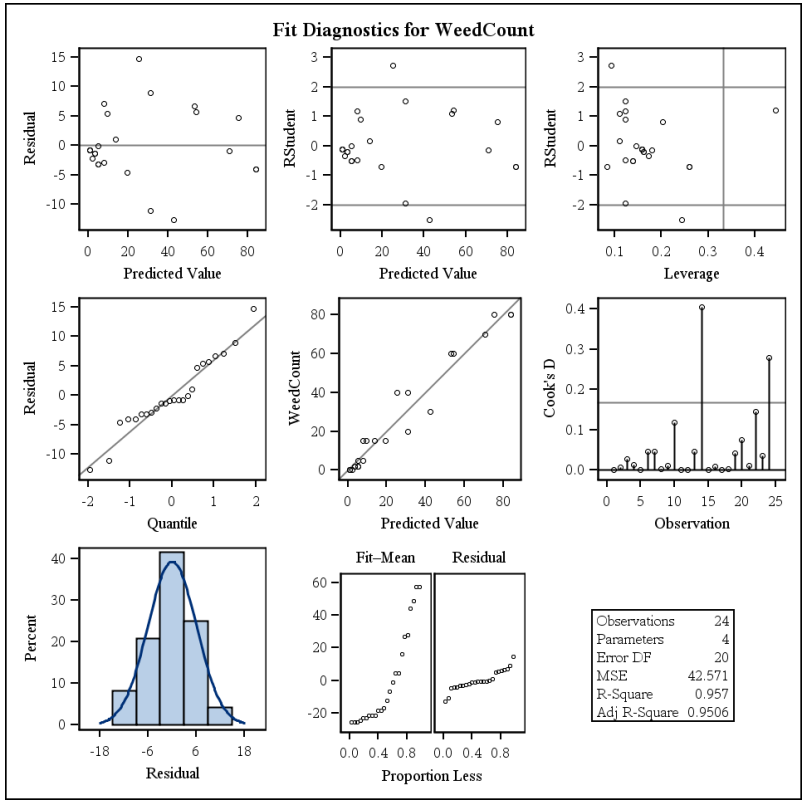
Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	18954	6318.14108	148.42	<.0001
Error	20	851.41009	42.57050		
Corrected Total	23	19806			

Root MSE	6.52461	R-Square	0.9570
Dependent Mean	26.58333	Adj R-Sq	0.9506
Coeff Var	24.54398		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	88.41034	3.52201	25.10	<.0001
GrassQuality	1	-0.87590	0.04762	-18.39	<.0001
sprayY	1	-28.35593	6.07258	-4.67	0.0001
GrassprayY	1	0.29724	0.08079	3.68	0.0015

The REG Procedure
Model: MODEL5
Dependent Variable: WeedCount



The REG Procedure
Model: MODEL5

Test 2 Results for Dependent Variable WeedCount				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	2	492.68676	11.57	0.0005
Denominator	20	42.57050		

The REG Procedure
Model: MODEL5

Test 3 Results for Dependent Variable WeedCount				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	576.20250	13.54	0.0015
Denominator	20	42.57050		