

5. A park district surveyed teenagers in a large metropolitan area regarding participation in a preteen soccer program and subsequent teenage drug use. The results are classified by family income as shown in the following table.

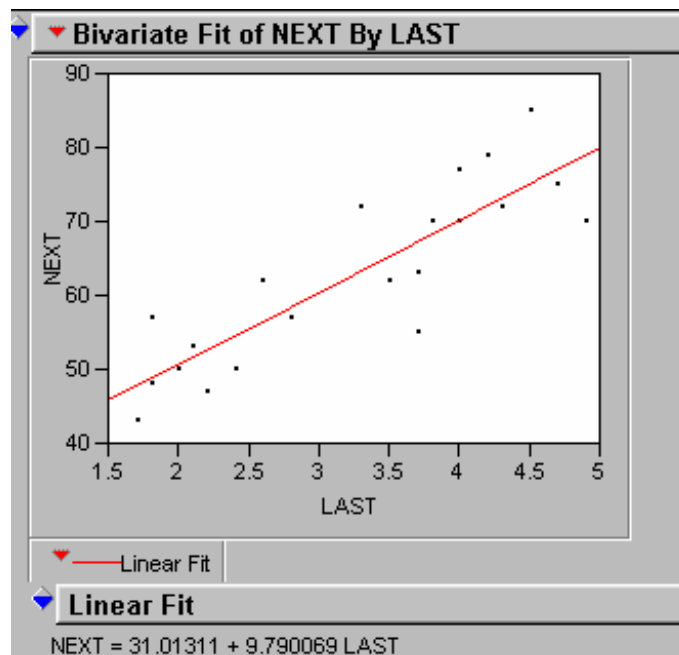
Family Income	Played Soccer	Drug Use	
		Yes	No
Low	Yes	10	40
	No	30	120
Medium	Yes	10	90
	No	10	90
High	Yes	6	144
	No	2	48

- a. Combine the income levels to construct a 2×2 table of drug use (columns) by soccer participation (rows). Compare the proportions of drug users among the teenagers who did and did not participate in soccer. Why would it be misleading to conclude from this table that involvement in soccer lowers the rate of teenage drug use?

Soccer Participation	Drug Use	
	Yes	No
Yes		
No		

- b. For each income level compare the proportion of drug users among teenagers who did and did not play soccer. What do these proportions indicate? Calculate the adjusted (for income) proportion of drug user for soccer players and nonplayers.

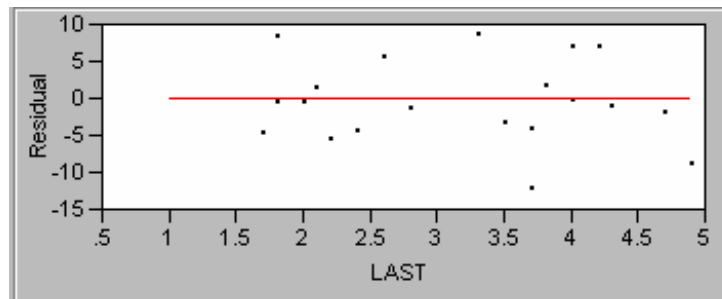
- e. Give a 95 % PI for the time to the next eruption if the last eruption lasted 4 minutes
- f. Give a 95% CI for the *mean* time to the next eruption for a last eruption lasting 4 minutes. Compare this CI with the PI obtained in (e)
- g. Suppose that the last eruption lasted 1 minute. Do you think the PI for 1 minute is reliable? Why or why not?



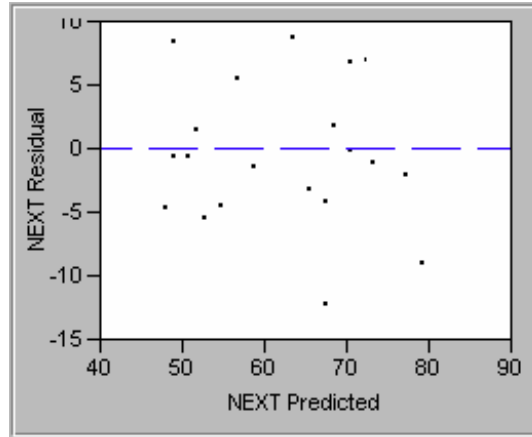
Summary of Fit				
RSquare	0.74908			
RSquare Adj	0.735874			
Root Mean Square Error	6.128823			
Mean of Response	62.71429			
Observations (or Sum Wgts)	21			
Lack Of Fit				
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2130.5987	2130.60	56.7215
Error	19	713.6870	37.56	Prob > F
C. Total	20	2844.2857		<.0001
Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	31.01311	4.416585	7.02	<.0001
LAST	9.790069	1.299906	7.53	<.0001

	No	LAST	NEXT	Lower 95% Mean NEXT	Upper 95% Mean NEXT	Lower 95% Indiv NEXT	Upper 95% Indiv NEXT
20	20	4.2	79	68.2993051	75.9634942	58.743467	85.5193323
21	21	4.3	72	69.08759	77.1332231	59.6666405	86.5541726
22	•	4	•	66.690157	73.6566148	56.8811071	83.4656647
23	•	1	•	34.1013185	47.5050394	26.3302136	55.2761443

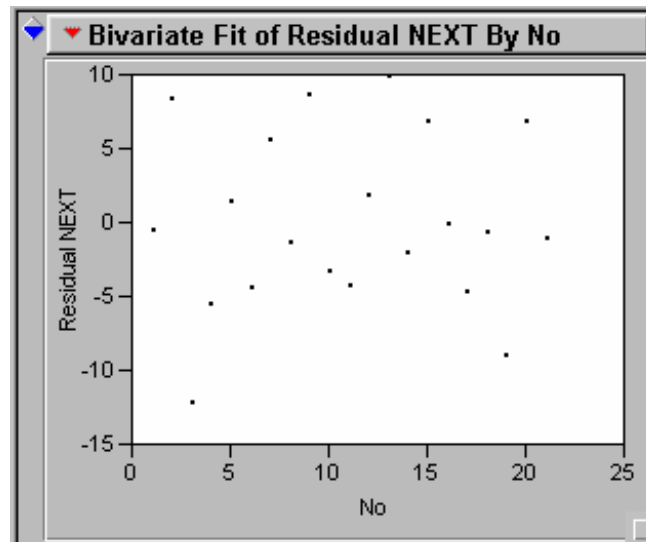
- h. Discuss in detail each of the following residual plots. What do they indicate? Are there any regression assumptions that seem to be violated?



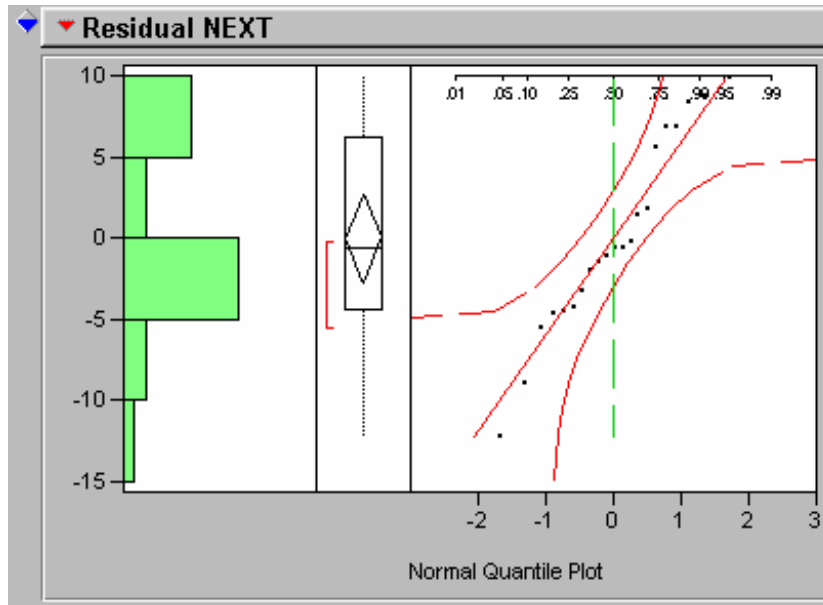
Discussion:



Discussion:



Discussion:



Discussion:

h NEXT	Studentized Resid NEXT
0.11657598	-0.1029851
0.14065379	1.47228747
0.05721691	-2.0562194
0.096097	-0.9526946
0.10588664	0.24636404
0.07921683	-0.7667451
0.06593548	0.93405496
0.05625295	-0.239388
0.04779144	1.45130838
0.05070477	-0.5490069
0.05721691	-0.7118859
0.06182254	0.30062708
0.11925367	1.72669575
0.14375991	-0.3573203
0.07373292	1.15733783
0.07373292	-0.0293947
0.15404224	-0.8260051
0.14065379	-0.1118079
0.17186496	-1.610882
0.08924211	1.17432873
0.09834626	-0.1908029
0.07373292	•
0.27295317	•

i. Look at the column of h_{ii} above. Are there any influential observations? Explain.

j. What do the standardized (studentized) residuals indicate?

7. Explain how one uses variance inflation factors (VIF)

8. Cork deposits were collected on the north, east, south, and west sides of each of 28 trees. Trees are a blocking factor in this RB design. The weights of cork deposits collected for each tree are given below.

Tree	Direction				Tree	Direction			
	N	E	S	W		N	E	S	W
1	72	66	76	77	15	91	79	100	75
2	60	53	66	63	16	56	68	47	50
3	56	57	64	58	17	79	65	70	61
4	41	29	36	38	18	81	80	68	58
5	32	32	35	36	19	78	55	67	60
6	30	35	34	26	20	46	38	37	38
7	39	39	31	27	21	39	35	34	37
8	42	43	31	25	22	32	30	30	32
9	37	40	31	25	23	60	50	67	54
10	33	29	27	36	24	35	37	48	39
11	32	30	34	28	25	39	36	39	31
12	63	45	74	63	26	50	34	37	40
13	54	46	60	52	27	43	37	39	50
14	47	51	52	43	28	48	54	57	43

Based on the JMP output that follows answer the following questions:

- a. It is of interest to determine if the weight of the cork deposit is influenced by direction. Is the direction effect significant?

- b. Does blocking by tree seem to help in this experiment? Justify your answer.

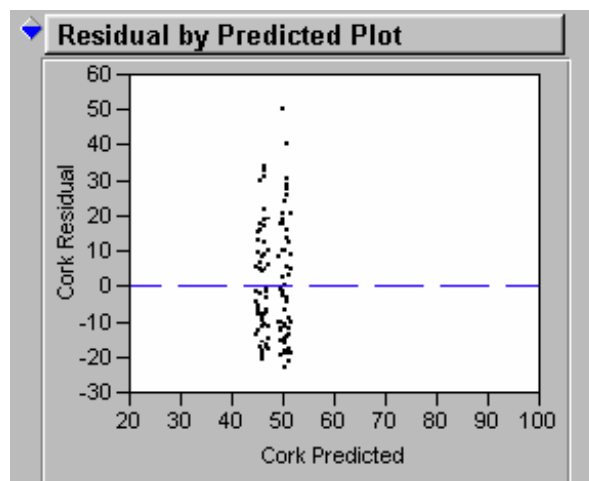
- c. Does the constant variance assumption seem to hold? Explain.

Summary of Fit	
RSquare	0.019922
RSquare Adj	-0.01672
Root Mean Square Error	16.55039
Mean of Response	47.89286
Observations (or Sum Wgts)	112

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	595.762	148.941	0.5437
Error	107	29308.952	273.915	Prob > F
C. Total	111	29904.714		0.7039

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	48.694444	3.213437	15.15	<.0001
Side [E]	-1.714286	2.708694	-0.63	0.5282
Side [N]	2.6428571	2.708694	0.98	0.3314
Side [S]	1.7857143	2.708694	0.66	0.5111
Tree	-0.055282	0.193602	-0.29	0.7758

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Side	3	3	573.42857	0.6978	0.5554
Tree	1	1	22.33388	0.0815	0.7758



	Liquid	Texture	Score
1	Low	Course	35
2	Low	Course	39
3	Low	Course	77
4	Low	Course	16
5	Low	Fine	104
6	Low	Fine	129
7	Low	Fine	97
8	Low	Fine	84
9	High	Course	24
10	High	Course	21
11	High	Course	39
12	High	Course	60
13	High	Fine	65
14	High	Fine	94
15	High	Fine	86
16	High	Fine	64

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	12053.250	4017.75	10.3273
Error	12	4668.500	389.04	Prob > F
C. Total	15	16721.750		0.0012

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Liquid	1	1	1024.000	2.6321	0.1307
Texture	1	1	10609.000	27.2696	0.0002
Liquid*Texture	1	1	420.250	1.0802	0.3191

