Stat 305: Regression Handout:

Example 1: **Stress / Till-till-fracture data. In the following data,**

x= uniaxial stress applied (kg/mm^2) and y = time till fracture (hours)

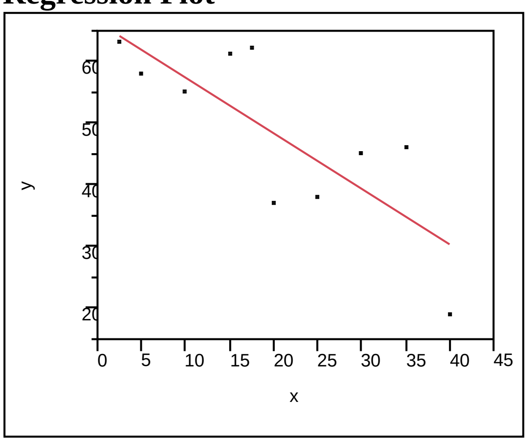
x	y
2.5	63
5	58
10	55
15	61
17.5	62
20	37
25	38
30	45
35	46
40	19

We will fit a linear and quadratic model on this data using least square methods.

JMP Handout:

x	y	Predicted y	Residual y
2.5	63	64.165486726	-1.165486726
5	58	61.913274336	-3.913274336
10	55	57.408849558	-2.408849558
15	61	52.904424779	8.0955752212
17.5	62	50.652212389	11.347787611
20	37	48.4	-11.4
25	38	43.895575221	-5.895575221
30	45	39.391150442	5.6088495575
35	46	34.886725664	11.113274336
40	19	30.382300885	-11.38230088

Response y
Whole Model
Regression Plot



Summary of Fit

RSquare	0.632518
RSquare Adj	0.586583
Root Mean Square Error	9.124307
Mean of Response	48.4
Observations (or Sum Wgts)	10

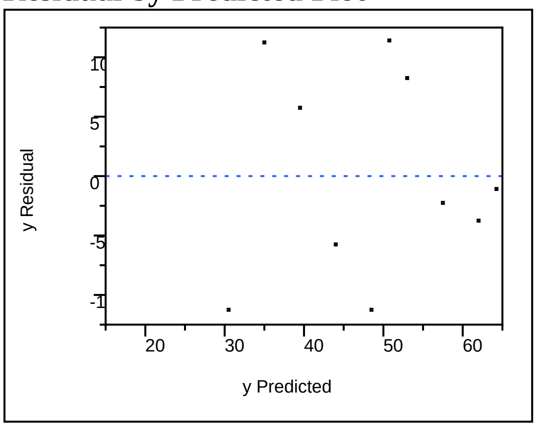
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1146.3761	1146.38	13.7698
Error	8	666.0239	83.25	**Prob > F**
C. Total	9	1812.4000		0.0059*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	66.417699	5.648129	11.76	<.0001*
x	-0.900885	0.242776	-3.71	0.0059*

Residual by Predicted Plot



Prediction Expression

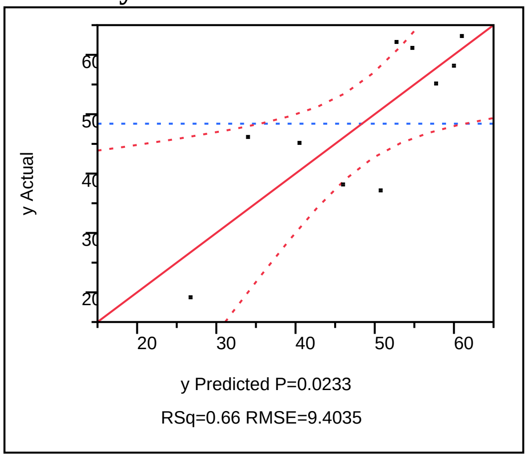
<u>Y=</u>

66.4176991150442 + -0.9008849557522 * x

-

Quadratic Model with x and x^2:

Response y
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.658473
RSquare Adj	0.560894
Root Mean Square Error	9.403518
Mean of Response	48.4
Observations (or Sum Wgts)	10

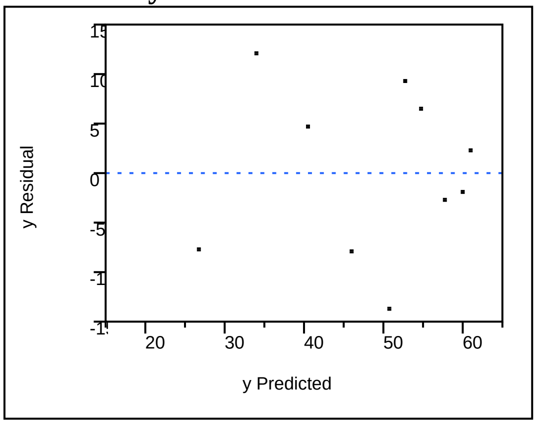
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	1193.4169	596.708	6.7481
Error	7	618.9831	88.426	**Prob > F**
C. Total	9	1812.4000		0.0233*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	61.52298	8.883684	6.93	0.0002*
x	-0.208514	0.981695	-0.21	0.8378
xsq	-0.016541	0.022678	-0.73	0.4895

Residual by Predicted Plot



Prediction Expression:

Y=
61.5229796176377

+ -0.2085138825891 * x

+ -0.0165407888515 * xsq

Example 2: Data collected on a hardness study of a particular alloy. The variables are:

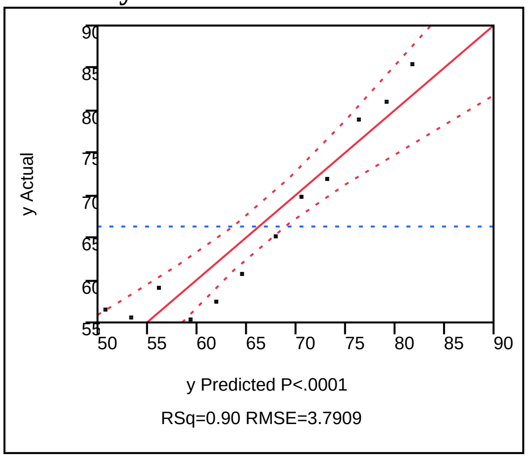
x_1= % of copper in alloy, x_2 = tempering temperature and y = hardness

x_1	x_2	y
0.02	1000	78.9
	1100	65.1
	1200	55.2
	1300	56.4
0.1	1000	80.9
	1100	69.7
	1200	57.4
	1300	55.4
0.18	1000	85.3
	1100	71.8
	1200	60.7
	1300	58.9

We will fit a linear model and some variations on this data using least square methods.

Model with x1 and x2:

Response y
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.899073
RSquare Adj	0.876645
Root Mean Square Error	3.790931
Mean of Response	66.30833
Observations (or Sum Wgts)	12

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	2	1152.1888	576.094	40.0868
Error	9	129.3404	14.371	**Prob > F**
C. Total	11	1281.5292		<.0001*

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	161.33646	11.43285	14.11	<.0001*
x1	32.96875	16.75371	1.97	0.0806
x2	-0.0855	0.009788	-8.74	<.0001*

Prediction Expression
Y=

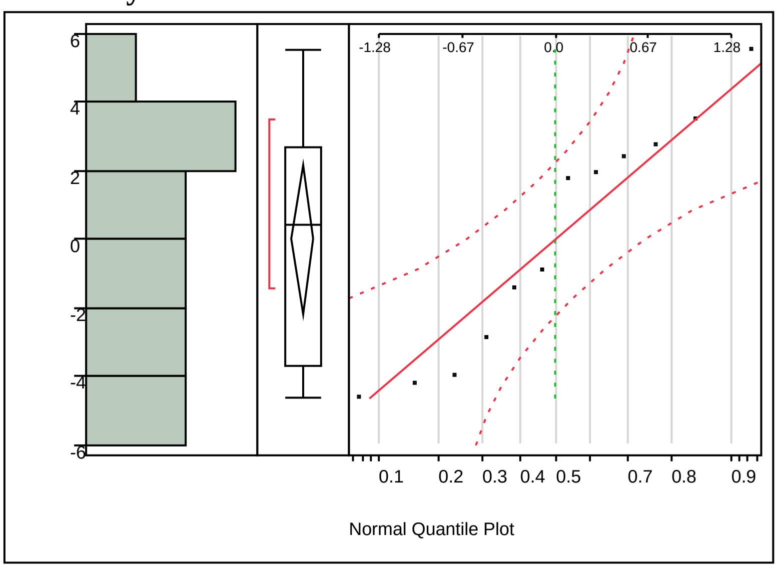
161.336458333333

+ 32.96875 * x1

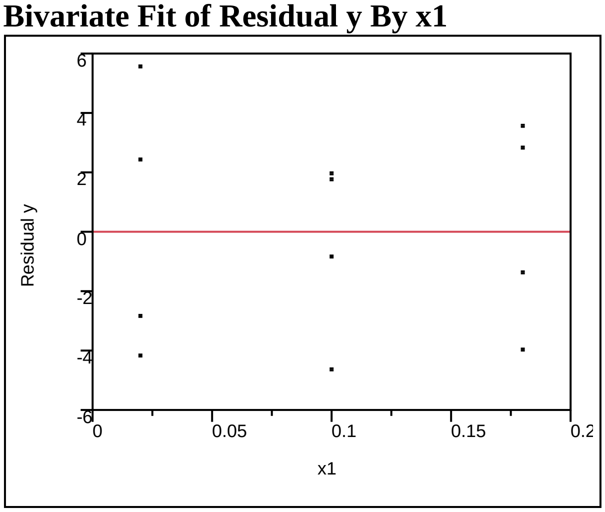
+ -0.0855 * x2

Distributions

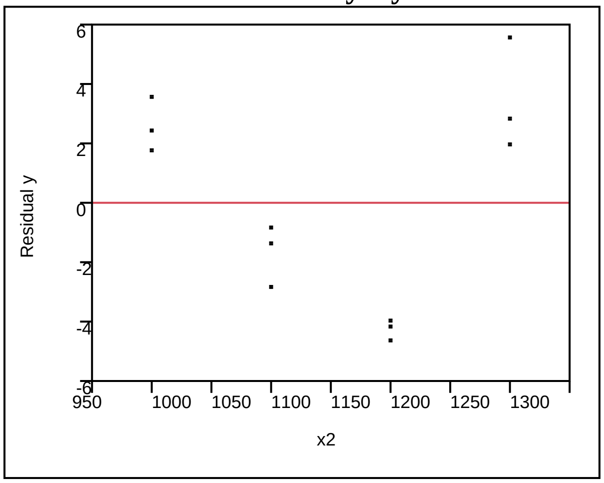
Residual y



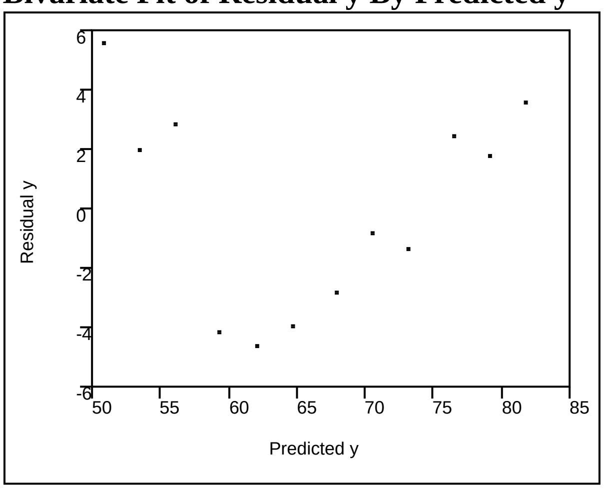
Bivariate Fit of Residual y By x1



Bivariate Fit of Residual y By x2

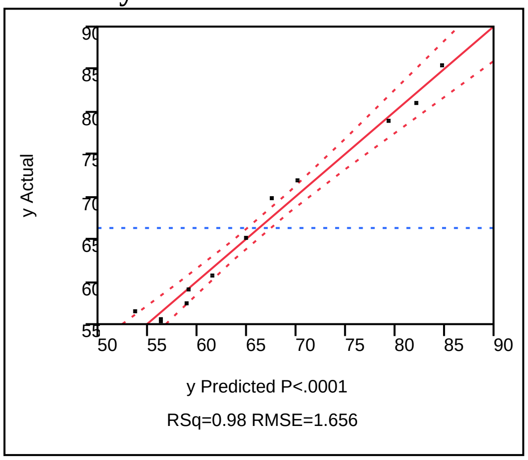


Bivariate Fit of Residual y By Predicted y



Output with x1, x2 and x2^2 in the model.

Response y
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.98288	← R^2
RSquare Adj	0.97646	← modified R^2, adjusts for overfitting
Root Mean Square Error	1.656034	← s_{SF} (or s_{LF})
Mean of Response	66.30833	← $\bar{y}$
Observations (or Sum Wgts)	12	← n

Analysis of Variance (ANOVA table)

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1259.5896	419.863	153.0980
Error	8	21.9396	2.742	**Prob > F**
C. Total	11	1281.5292		<.0001*

For testing $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ vs. H_a: not H_0

p-value for the F-test.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept (β_0) →	553.24479	62.82414	8.81	<.0001*
x1 (β_1)	32.96875	7.318705	4.50	0.0020*
x2 (β_2)	-0.773583	0.110036	-7.03	0.0001*
x2-sq (β_3)	0.0002992	4.781e-5	6.26	0.0002*

β_ℓ

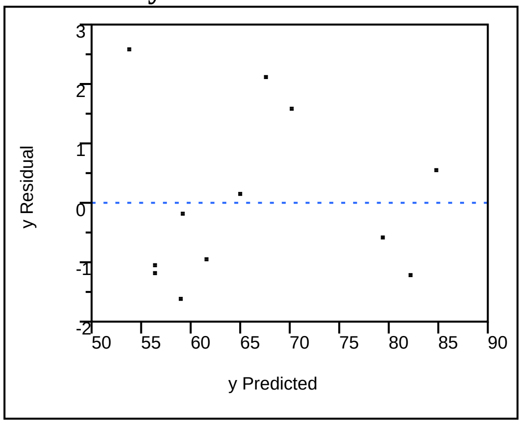
$se(\beta_\ell) = s_{SF}\sqrt{d_\ell}$, $\ell = 0, 1, 2, 3$

$T = \dfrac{b_\ell - 0}{se(b_\ell)}$, $\ell = 0, 1, 2, 3$

p-values for these tests

For testing $H_0: \beta_\ell = 0$ vs $H_a: \beta_\ell \neq 0$

Residual by Predicted Plot



Prediction Expression

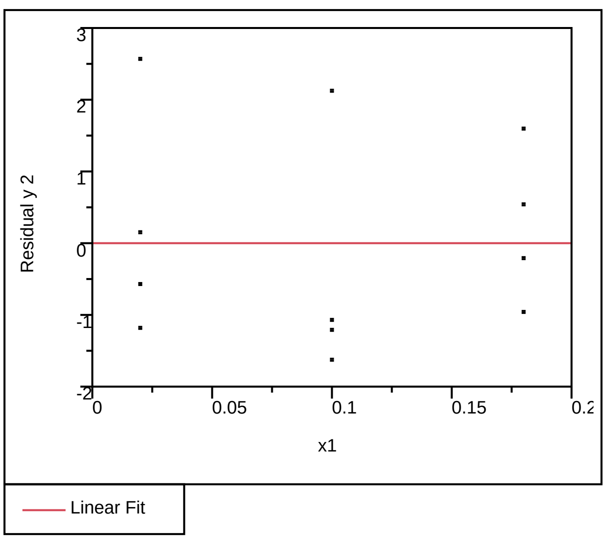
553.244791666667

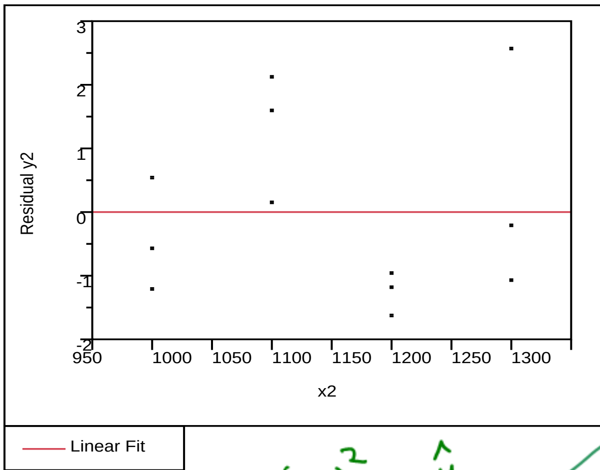
+ 32.96875 * x1

+ -0.7735833333333 * x2

+ 0.00029916666667 * x2-sq

Residuals:





does the same job, except one deals w/ new x values

$s_{SF}A$ // $se(\hat{y})$

x1	x2	y	x2-sq	Predicted y	Pred Formula y	StdErr Pred y	Residual y	Studentized Resid y
x_1	x_2	y	$(x_2)^2$	$\hat{y}$		$se(\hat{y})$	e_i	e_i^*
0.02	1000	78.9	1000000	79.4875	79.4875	1.1005670492	-0.5875	-0.474779434
0.02	1100	65.1	1210000	64.954166667	64.954166667	0.9195586483	0.1458333333	0.1058861567
0.02	1200	55.2	1440000	56.404166667	56.404166667	0.9195586483	-1.204166667	-0.874317123
0.02	1300	56.4	1690000	53.8375	53.8375	1.1005670492	2.5625	2.0708464676
0.1	1000	80.9	1000000	82.125	82.125	0.9319022697	-1.225	-0.894850719
0.1	1100	69.7	1210000	67.591666667	67.591666667	0.7090713067	2.1083333333	1.4087945605
0.1	1200	57.4	1440000	59.041666667	59.041666667	0.7090713068	-1.641666667	-1.096966515
0.1	1300	55.4	1690000	56.475	56.475	0.9319022697	-1.075	-0.785277161
0.18	1000	85.3	1000000	84.7625	84.7625	1.1005670492	0.5375	0.4343726737
0.18	1100	71.8	1210000	70.229166667	70.229166667	0.9195586483	1.5708333333	1.1405451738
0.18	1200	60.7	1440000	61.679166667	61.679166667	0.9195586483	-0.979166667	-0.710949909
0.18	1300	58.9	1690000	59.1125	59.1125	1.1005670492	-0.2125	-0.171728731
.	.	.	.	.	.	.	.	.
0.2	1150	.	1322500	.	65.865625	.	.	.

new x-values for prediction of y

predicted y