

```
# Math 3080 - 1    Paddy Data
# Treiberger
#
# From Devore "Probability and Statistics for Engineers and
# Scientists 5th ed." (Duxbury 199, p. 564)
#
# From "Determination of the Biological Maturity and Effect of Harvesting and
# Drying on Milling Quality of Paddy," (J. Agricultural Eng. & Res. 1975)
#
# Paddy is a grain grown in India. The study compared time to harvest to yield.
# x = number of days after flowering
# y = yield (kg/ha)
#
"x"  "y"
16   2508
18   2518
20   3304
22   3423
24   3057
26   3190
28   3500
30   3883
32   3823
34   3646
36   3708
38   3333
40   3517
42   3241
44   3103
46   2776
```

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

```

>#=====READ PADDY DATA=====
> tt <- read.table("M3081DataPaddy.txt",header=TRUE); tt
  x    y
1 16 2508
2 18 2518
3 20 3304
4 22 3423
5 24 3057
6 26 3190
7 28 3500
8 30 3883
9 32 3823
10 34 3646
11 36 3708
12 38 3333
13 40 3517
14 42 3241
15 44 3103
16 46 2776
> attach(tt)

>#=====FIT SIMPLE REGRESSION=====
> f1 <- lm(y~x); summary(f1); anova(f1)

Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-691.07 -217.65   45.85  271.77  612.14

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2902.96     364.67   7.961 1.45e-06 ***
x              12.26       11.28   1.088  0.295
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 415.8 on 14 degrees of freedom
Multiple R-squared:  0.07791, Adjusted R-squared:  0.01205
F-statistic: 1.183 on 1 and 14 DF,  p-value: 0.2951

Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
x       1  204526   204526   1.1829 0.2951
Residuals 14 2420642   172903

```

```
>#=====FIT QUADRATIC MODEL=====
> x2<- x*x; x3 <-x*x2
> f2 <- lm(y~x+x2); summary(f2); anova(f2)
```

```
Call:
lm(formula = y ~ x + x2)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-303.96 -118.11  13.86  115.67  319.06
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1070.3977   617.2527  -1.734   0.107
x             293.4829    42.1776   6.958 9.94e-06 ***
x2            -4.5358     0.6744  -6.726 1.41e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

```
Residual standard error: 203.9 on 13 degrees of freedom
Multiple R-squared: 0.7942, Adjusted R-squared: 0.7625
F-statistic: 25.08 on 2 and 13 DF, p-value: 3.452e-05
```

Analysis of Variance Table

```
Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
x         1  204526  204526  4.9202  0.04497 *
x2        1 1880253 1880253 45.2328 1.414e-05 ***
Residuals 13  540388   41568
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

```
>#=====PLOT POINTS, FITTED LINE & ADD FITTED QUADRATIC=====
```

```
> xfine<-seq(from=min(x), to=max(x), by=.1)
> y2 <- -1070.3977 + 293.4829*xfine - 4.5358*xfine*xfine
> plot(x,y); abline(f1)
> lines(xfine, y2, lty=2)
```

```
>#=====FIT A CUBIC MODEL=====
```

```
> f3 <- lm(y ~ x + x2 + x3); summary(f3); anova(f3)
```

```
Call:
lm(formula = y ~ x + x2 + x3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-281.970 -113.211  -6.113   97.752  330.924
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-203.60852	2285.13020	-0.089	0.930
x	199.07674	242.92513	0.819	0.428
x2	-1.32071	8.16843	-0.162	0.874
x3	-0.03457	0.08751	-0.395	0.700

Residual standard error: 210.8 on 12 degrees of freedom
Multiple R-squared: 0.7968, Adjusted R-squared: 0.746
F-statistic: 15.68 on 3 and 12 DF, p-value: 0.0001876

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x	1	204526	204526	4.6008	0.05312 .
x2	1	1880253	1880253	42.2964	2.921e-05 ***
x3	1	6937	6937	0.1561	0.69974
Residuals	12	533451	44454		

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

>%=====ADD FITTED CUBIC TO PLOT=====

```
> y3 <- -203.60852 + 199.07674*xfine -1.32071*xfine*xfine-0.03457*xfine^3
> lines(xfine, y3, lty=3)
```

>#=====FITTED VALUES FOR ALL 16 DATA POINTS & THEIR STD. ERRORS=====

```
> predict(f2 , se.fit=TRUE)
```

\$fit

1	2	3	4	5	6	7	8	9
2464.164	2742.696	2984.941	3190.899	3360.571	3493.957	3591.056	3651.869	3676.396
10	11	12	13	14	15	16		
3664.636	3616.589	3532.257	3411.637	3254.732	3061.540	2832.061		

\$se.fit

[1]	135.60941	104.75122	82.98101	71.28155	68.44831	70.70752	74.16358	76.40619
[9]	76.40619	74.16358	70.70752	68.44831	71.28155	82.98101	104.75122	135.60941

\$df

[1] 13

\$residual.scale

[1] 203.8831

```
>#=====RESIDUALS, STUDENTIZED RESIDUALS & STANDARDIZED RESIDUALS=====
```

```
> residuals(f2)
```

```
      1      2      3      4      5      6      7
43.83578 -224.69559 319.05945 232.10091 -303.57122 -303.95693 -91.05623
      8      9     10     11     12     13     14
231.13088 146.60441 -18.63564  91.41071 -199.25651 105.36268 -13.73172
     15     16
41.46029 -56.06127
```

```
> rstudent(f2)
```

```
      1      2      3      4      5      6      7
0.27752047 -1.32087516 1.87069873 1.23994717 -1.68971352 -1.70137801 -0.46477326
      8      9     10     11     12     13     14
1.24878749 0.76301743 -0.09431106 0.46335248 -1.04084582 0.53626262 -0.07085650
     15     16
0.22822484 -0.35564507
```

```
> rstandard(f2)
```

```
      1      2      3      4      5      6      7
0.28792995 -1.28459327 1.71323323 1.21508349 -1.58068925 -1.58948634 -0.47945521
      8      9     10     11     12     13     14
1.22275333 0.77558235 -0.09812570 0.47801536 -1.03752467 0.55158959 -0.07373436
     15     16
0.23703009 -0.36823158
```

```
>#====FOR LIN, QUAD, CUBIC FITTED VALUES, STD.ERRORS, CI'S PI'S FOR GIVEN X VALUES=====
```

```
> predict(f1 , data.frame(x=c(22,33,44)), se.fit=TRUE)
```

```
$fit
```

```
      1      2      3
3172.756 3307.651 3442.547
```

```
$se.fit
```

```
      1      2      3
145.2733 106.3719 179.7002
```

```
$df
```

```
[1] 14
```

```
$residual.scale
```

```
[1] 415.816
```

```
> predict(f2 , data.frame(x=c(22,33,44),x2=c(484,1089,1936)), se.fit=TRUE)
```

```
$fit
```

```
      1      2      3
3190.899 3675.051 3061.540
```

```
$se.fit
```

```
      1      2      3
71.28155 75.52786 104.75122
```

```
$df
```

```
[1] 13
```

```
$residual.scale
```

```
[1] 203.8831
```

```

> xstar <- c(22,33,44); xstar2 <- xstar*xstar; xstar3 <- xstar*xstar2

> predict(f2 , data.frame(x=xstar,x2=xstar2), se.fit=TRUE,interval="confidence")
$fit
      fit      lwr      upr
1 3190.899 3036.905 3344.894
2 3675.051 3511.883 3838.219
3 3061.540 2835.238 3287.841

$se.fit
      1      2      3
71.28155 75.52786 104.75122

$df
[1] 13

$residual.scale
[1] 203.8831

> predict(f2 , data.frame(x=xstar,x2=xstar2), se.fit=TRUE,interval="prediction")
$fit
      fit      lwr      upr
1 3190.899 2724.292 3657.506
2 3675.051 3205.337 4144.765
3 3061.540 2566.343 3556.736

$se.fit
      1      2      3
71.28155 75.52786 104.75122

$df
[1] 13

$residual.scale
[1] 203.8831
>

```

```
> predict(f3 , data.frame(x=xstar,x2=xstar2,x3=xstar3), se.fit=TRUE,interval="confidence")
$fit
```

	fit	lwr	upr
1	3168.746	2966.944	3370.548
2	3685.298	3505.982	3864.615
3	3053.989	2814.321	3293.658

```
$se.fit
```

	1	2	3
	92.62025	82.30005	109.99956

```
$df
```

```
[1] 12
```

```
$residual.scale
```

```
[1] 210.8417
```

```
> predict(f3 , data.frame(x=xstar,x2=xstar2,x3=xstar3), se.fit=TRUE,interval="prediction")
$fit
```

	fit	lwr	upr
1	3168.746	2666.991	3670.501
2	3685.298	3192.157	4178.440
3	3053.989	2535.843	3572.135

```
$se.fit
```

	1	2	3
	92.62025	82.30005	109.99956

```
$df
```

```
[1] 12
```

```
$residual.scale
```

```
[1] 210.8417
```

```
> predict(f1 , data.frame(x=xstar), se.fit=TRUE,interval="confidence")
$fit
```

	fit	lwr	upr
1	3172.756	2861.176	3484.336
2	3307.651	3079.507	3535.796
3	3442.547	3057.128	3827.966

```
$se.fit
```

	1	2	3
	145.2733	106.3719	179.7002

```
$df
```

```
[1] 14
```

```
$residual.scale
```

```
[1] 415.816
```

```
>#=====QUADRATIC MODEL USING CENTERED VARIABLES=====
```

```
> xmean <- mean(x); xmean
[1] 31
> xc <- x-xmean
> f4 <- lm(y~ xc + I(xc*xc))
> summary(f4); anova(f4)
```

Call:

```
lm(formula = y ~ xc + I(xc * xc))
```

Residuals:

Min	1Q	Median	3Q	Max
-303.96	-118.11	13.86	115.67	319.06

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3668.6682	76.7086	47.826	5.33e-16	***
xc	12.2632	5.5286	2.218	0.045	*
I(xc * xc)	-4.5358	0.6744	-6.726	1.41e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 203.9 on 13 degrees of freedom

Multiple R-squared: 0.7942, Adjusted R-squared: 0.7625

F-statistic: 25.08 on 2 and 13 DF, p-value: 3.452e-05

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
xc	1	204526	204526	4.9202	0.04497 *
I(xc * xc)	1	1880253	1880253	45.2328	1.414e-05 ***
Residuals	13	540388	41568		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
>#=====CUBIC MODEL USING NORMALIZED VARIABLES=====
```

```
> stdev <- sd(x); stdev
[1] 9.521905
> xs <- xc/stdev
> f5 <- lm(y~ xs + I(xs^2) + I(xs^3))
> summary(f5); anova(f5)
```

Call:

```
lm(formula = y ~ xs + I(xs^2) + I(xs^3))
```

Residuals:

Min	1Q	Median	3Q	Max
-281.970	-113.211	-6.113	97.752	330.924

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3668.67	79.33	46.248	6.82e-15	***
xs	166.87	138.02	1.209	0.25	
I(xs^2)	-411.25	63.23	-6.504	2.92e-05	***
I(xs^3)	-29.85	75.55	-0.395	0.70	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 210.8 on 12 degrees of freedom

Multiple R-squared: 0.7968, Adjusted R-squared: 0.746

F-statistic: 15.68 on 3 and 12 DF, p-value: 0.0001876

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
xs	1	204526	204526	4.6008	0.05312	.
I(xs^2)	1	1880253	1880253	42.2964	2.921e-05	***
I(xs^3)	1	6937	6937	0.1561	0.69974	
Residuals	12	533451	44454			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>#=====DERIVING COEFFICIENTS FOR CUBIC MODEL USING COEFS OF NORMALIZED MODEL=====

```
> beta0 <- 3668.67+166.87*(-xmean/stdev)-411.25*(-xmean/stdev)^2-29.85*(-xmean/stdev)^3;beta0
[1] -203.499
```

```
> beta1 <- 166.87/stdev-2*411.25*(-xmean/stdev^2)-3*29.85*xmean^2/stdev^3;beta1
[1] 199.0651
```

```
> beta2 <- -411.25/stdev^2-3*29.85*(-xmean/stdev^3); beta2
[1] -1.320292
```

```
> beta3 <- -29.85/stdev^3; beta3
[1] -0.03457585
```

>#=====DIAGNOSTICS FOR QUADRATIC FIT=====

```
> layout(matrix(1:4,ncol=2))
> plot(rstudent(f2)~fitted(f2))
> abline(h=0,lty=2)
> plot(rstudent(f2)~x)
> plot(fitted(f2)~y)
> abline(a=0,b=1,lty=5)
> qqnorm(rstudent(f2))
> qqline(rstudent(f2))
```



