Homework 11 - STAT 231

Suggested Solution [*†]

Due in class: 5 Dec 2019

Problem 1-5 are from the Devore textbook.

1. **(Chapter 13: Problem 35)** Observe taking natural logarithm on both sides of

$$y = \alpha\ e^{\beta x\ +\gamma x^2} \epsilon$$

we arrive

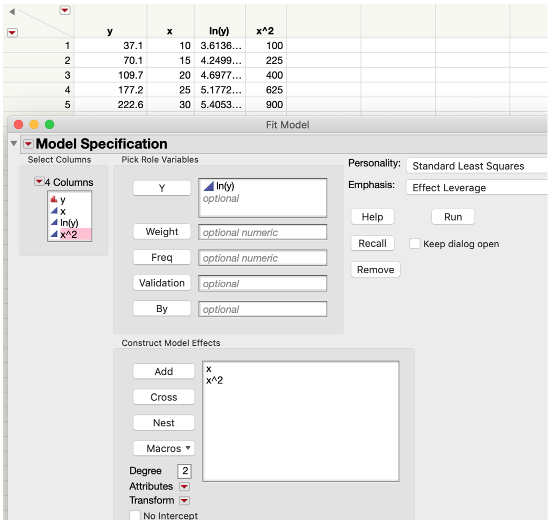
$$\ln(y) = \ln(\alpha)\ + \beta x\ + \gamma x^2 + \ln(\epsilon)$$

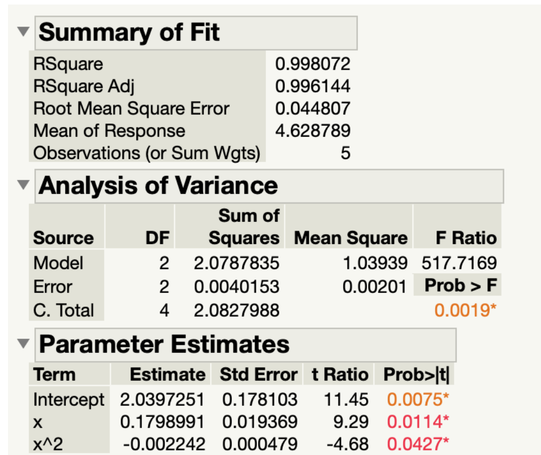
which indicates that we can perform quadratic regression for $\ln(y)$ versus x, or equivalently perform FLM for $\ln(y)$ against x and x^2.

Perform quadratic regression of $\ln(y)$ against x in jmp, we can select *Analyze* $\rightarrow$ *Fit Model*, we have working interface and estimates as following

$$\ln(y) = 2.0397 + 0.1799x - 0.002242x^2$$

[*] Zhiling GU (zlgu@iastate.edu)
[†] Please inform me if there is any error or possible improvement





Summary of Fit

RSquare	0.998072
RSquare Adj	0.996144
Root Mean Square Error	0.044807
Mean of Response	4.628789
Observations (or Sum Wgts)	5

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	2.0787835	1.03939	517.7169
Error	2	0.0040153	0.00201	Prob > F
C. Total	4	2.0827988		0.0019*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.0397251	0.178103	11.45	0.0075*
x	0.1798991	0.019369	9.29	0.0114*
x^2	-0.002242	0.000479	-4.68	0.0427*

2. **(Chapter 13: Problem 38)**

(a) Note

$$Y = 125.0 + 7.75x_1 + .0950x_2 - .0090x_1x_2 + \epsilon$$

$$E(Y) = 125.0 + 7.75x_1 + .0950x_2 - .0090x_1x_2 \implies$$

$$E(Y|x_1 = 40, x_2 = 1100) = 125.0{+}7.75{*}40{+}.0950{*}1100{-}.0090{*}40{*}1100 = 143.5$$

(b) When $x_1 = 30$,

$$E(Y|x_1 = 30, x_2) = 125.0 + 7.75 * 30 + .0950x_2 - .0090 * 30 * x_2$$

$$E(Y|x_1 = 30, x_2 + 1) - E(Y|x_1 = 30, x_2) = .0950 * 1 - .0090 * 30 * 1 = -0.175$$

the mean life will drcrease by 0.175.

Similarly when $x_1 = 40$,

$$E(Y|x_1 = 40, x_2) = 125.0 + 7.75 * 40 + .0950x_2 - .0090 * 40 * x_2$$

then

$$E(Y|x_1 = 40, x_2 + 1) - E(Y|x_1 = 30, x_2) = .0950 * 1 - .0090 * 40 * 1 = -0.265$$

the mean life will drcrease by 0.265.

3. **(Chapter 13: Problem 42)**

(a) Model utility test. The test:

$$H_0 : \beta_1 = \beta_2 = 0 \quad H_1 : \text{at least one of } \beta_1, \beta_2 \neq 0$$

The full model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad k = 2$$

The reduced model:

$$Y = \beta_0 + \epsilon \quad m = 0$$

F-statistic:

$$F = \frac{\frac{SSR(full) - SSR(reduced)}{k-m}}{\frac{SSE(full)}{(n-k-1)}}, \quad f = \frac{\frac{715.50-0}{2-0}}{\frac{6.72}{9-2-1}} = 319.42$$

You can also see directly from the minitab output that $F = 319.31$. Compare it with the critical value $F_{k-m,n-k-1,\alpha} = F_{2,6,.001} = 27.00 < f$, which means $P(F_{2,6,.001} > f) < .001$. There is sufficient evidence to reject H_0 and conclude that there is a relationship between Y (temperature difference) and the regressor variables X_1, X_2 (furnace temparture and the temperature difference on the die surface).

(b) Calculate and interprete a 95% confidence interval for β_2. Note

$$\frac{\hat{\beta}_2 - \beta_2}{\widehat{SE}(\beta_2)} \sim t(df = n - k - 1 = 6)$$

by which we can construct 95%

$$\hat{\beta}_2 \pm t(df = 6, \alpha/2)\widehat{SE}(\beta_2) = 3.00 \pm 2.447 * 0.4321 = (1.942651, 4.057349) \quad \square$$

Use JMP to construct the confidence interval for (c) (d) . See "hw11q3 Fit Model.jmp".

(c) The theoretical CI

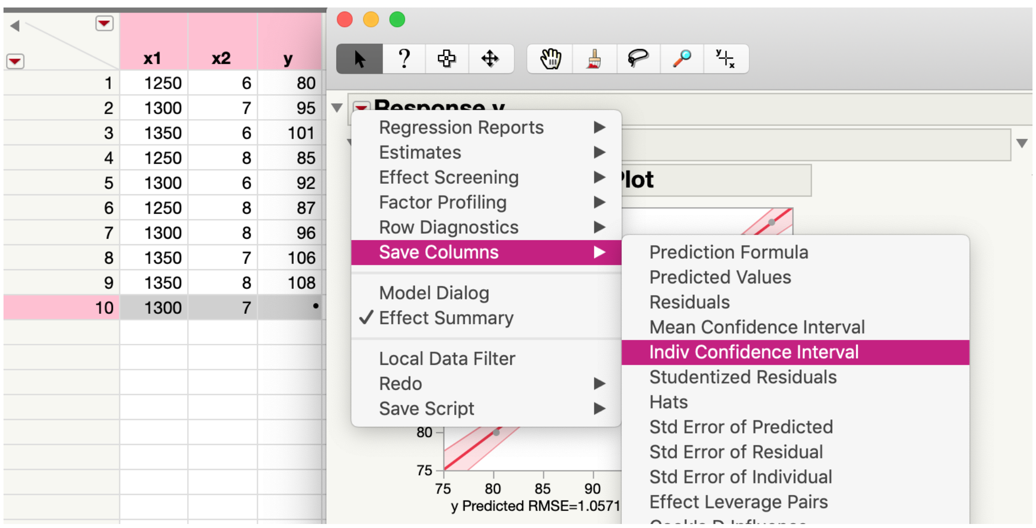
$$\hat{Y} \pm t_{n-k-1,\alpha/2}\hat{\sigma}\sqrt{Est.SDp(\hat{y})}, \quad \hat{\sigma}^2 = SSE/(n-k-1) = MSE$$

JMP: '*Save Columns* $\rightarrow$ *Mean Confidence Interval*'. From the output we have the 95% CI for $Y|x_1 = 1300, x_2 = 7$ is $(93.2601, 94.9988)$

(d) The theoretical CI

$$\hat{Y} \pm t_{n-k-1,\alpha/2}\hat{\sigma}\sqrt{1 + Est.SDp(\hat{y})}, \quad \hat{\sigma}^2 = SSE/(n-k-1) = MSE$$

JMP: '*Save Columns* $\rightarrow$ *Indiv Confidence Interval*' to find the prediction interval. From the graph we have the 95% CI for $Y|x_1 = 1300, x_2 = 7$ is $(91.4006, 96.8583)$.



	x1	x2	y	Lower 95% Mean y	Upper 95% Mean y	Lower 95% Indiv y	Upper 95% Indiv y
1	1250	6	80	78.455817192	82.188810908	77.132496082	83.512132017
2	1300	7	95	93.260139333	94.998813835	91.400580214	96.858372954
3	1350	6	101	100.60996203	103.92447323	99.195145973	105.33928929
4	1250	8	85	84.449711242	87.533759832	82.980262962	89.003208112
5	1300	6	92	89.890374154	92.699157527	88.351392919	94.238138761
6	1250	8	87	84.449711242	87.533759832	82.980262962	89.003208112
7	1300	8	96	95.727284999	98.201089657	94.09694902	99.831425636
8	1350	7	106	103.73748019	106.46637656	102.17740329	108.02645346
9	1350	8	108	106.21394053	109.65933771	104.82877513	111.04450311
10	1300	7	•	93.260139333	94.998813835	91.400580214	96.858372954

4. **(Chapter 13: Problem 44 (a) - (e))**

(a) $n = 80 + 1 = 81$

(b) $R^2 = 88.6\%$ The MLR

$$Ra = \beta_0 + \beta_1 a + \beta_2 d + \beta_3 f + \beta_4 v$$

can explain 88.6% of the variation in Roughness (Ra)

(c)

$$H_0 : \beta_i = 0, i = 1, 2, 3, 4, \quad H_1 : \text{ at least one of } \beta_i \text{ is nonzero}$$

F-test:

$$F = \frac{SSR/(k-1)}{SSE/(n-k-1)} = \frac{MSR}{MSE} \sim F_{n-1,n-k-1}$$

Note from the output, we have

$$f = \frac{100.25}{.68} = 147.4$$

P-value $P(F_{4,76} > f)$ is less than .001 therefore there is sufficient evidence to reject the null hypothesis and conclude that there is a relationship between Ra and a,d,f,v.

(d) The average value of Ra increases by 18.2602 if variable f increases by 1 unit (mm/rev) given all other variables (a,d,f) are unchanged.

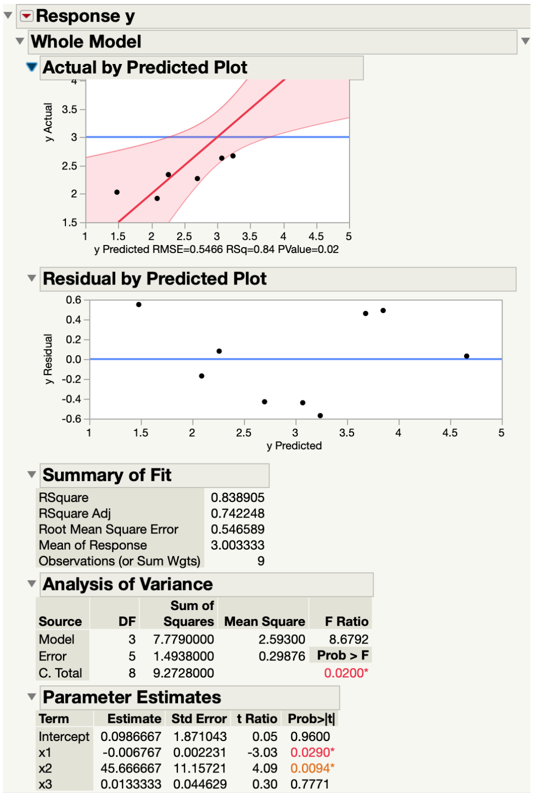
(e) The P in the minitab output represents the P-value for

$$H_0 : \beta_i = 0, \quad H_1 : \beta_i \neq 0$$

given that *all the other variables remain in the model!* See here for detailed explanation. So we can look at the variables with P larger than desired α to decide whether to eliminate that variable from the current model. We can only eliminate one variable at a time. After one-time elimination, we need to do the multiple linear regression again for the reduced model to decide whether to eliminate more variables that are not significant.

In our case, at $\alpha = .1$ we will elimite v from current model given every other variable remains in the model.

5. **(Chapter 13: Problem 51 with (b) - (e) adjusted)** See the output from JMP:



Summary of Fit

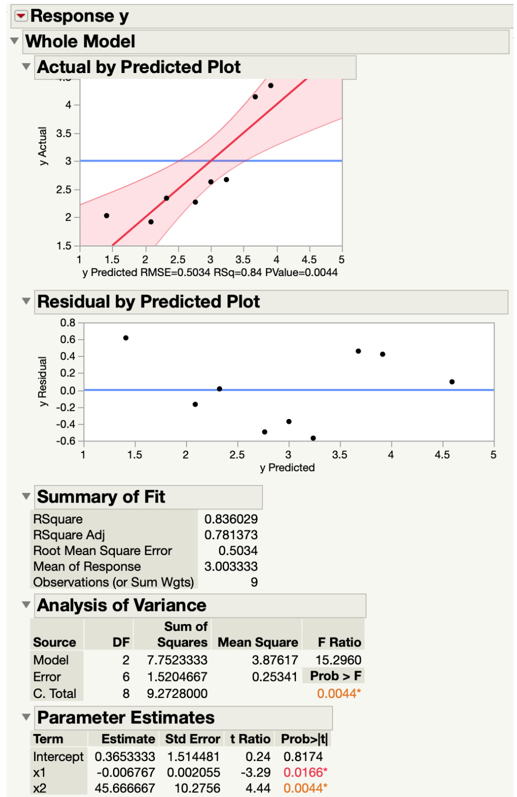
RSquare	0.838905
RSquare Adj	0.742248
Root Mean Square Error	0.546589
Mean of Response	3.003333
Observations (or Sum Wgts)	9

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	7.7790000	2.59300	8.6792
Error	5	1.4938000	0.29876	Prob > F
C. Total	8	9.2728000		0.0200*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0986667	1.871043	0.05	0.9600
x1	-0.006767	0.002231	-3.03	0.0290*
x2	45.666667	11.15721	4.09	0.0094*
x3	0.0133333	0.044629	0.30	0.7771



Summary of Fit

RSquare	0.836029
RSquare Adj	0.781373
Root Mean Square Error	0.5034
Mean of Response	3.003333
Observations (or Sum Wgts)	9

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	7.7523333	3.87617	15.2960
Error	6	1.5204667	0.25341	Prob > F
C. Total	8	9.2728000		0.0044*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.3653333	1.514481	0.24	0.8174
x1	-0.006767	0.002055	-3.29	0.0166*
x2	45.666667	10.2756	4.44	0.0044*

(a) Note the P-value of

$$H_0 : \beta_3 = 0, \quad \beta_3 \neq 0$$

is .7771, sigifinicantly larger than $\alpha = .05$. There is no sufficient evidence to reject H_0, i.e. drilling depth (X_3) does not provide useful information about roughness (Y) given X_1, X_2 are in the model.

(b)

$$H_0 : \beta_1 = \beta_2 = 0, \quad \text{at least one of } \beta_1, \beta_2 \text{ is not zero}$$

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{(SSR/SST)/k}{(SSE/SST)/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1}$$

In this problem, the F statistic

$$f = \frac{.836029/2}{(1-.836029)/6} = 15.29592$$

P-value: $P(F_{k,n-k-1} > f) = .0044 < .05 = \alpha$, therefore there is sufficient evidence to reject H_0 and conclude that the relationship of the reduced model

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ is significant.

(c) The 95% CI for β_1 is

$$\hat{\beta}_1 \pm t_{n-k-1,\alpha/2} \hat{SE}(\hat{\beta}_1) = -.006767 \pm 2.447 * .002055 = (-0.01179559, -0.001738415)$$

This indicates we are 95% confident that the decrease in Y, when X_1 increases for 1 unit and X_2 is unchanged, is between 0.001738415 and 0.01179559.

(d) Again use JMP, we have the following result

	x1	x2	x3	y	e*	Lower 95% Mean y	Upper 95% Mean y
1	320	0.1	15	2.27	-1.32	1.9454837767	3.5878495566
2	320	0.12	20	4.14	1.08	3.0307979231	4.3292020769
3	320	0.14	25	4.69	0.26	3.7721504434	5.4145162233
4	420	0.1	20	1.92	-0.4	1.4407979231	2.7392020769
5	420	0.12	25	2.63	-0.79	2.5927418884	3.4139247783
6	420	0.14	15	4.34	0.99	3.2674645897	4.5658687436
7	520	0.1	25	2.03	1.64	0.5921504434	2.2345162233
8	520	0.12	15	2.34	0.03	1.6774645897	2.9758687436
9	520	0.14	20	2.67	-1.52	2.4188171101	4.0611828899
10	400	0.125	•	•	•	2.925972416	3.808027584

The confidence level for $Y|x_1 = 400, x_2 = .125$ is $(2.9260, 3.8080)$

(e) e^* are essentially the regression errors for each observation, we can also see the residual plot in the JMP output. They seems to scatter randomly around zero, which indicates the model is capturing most of the behaviour of Y.

6. Let Y denote the High Price, X_1 denote Length, X_2 denote Weight, X_3 denote QtrMile. The model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

(a) i. Overall F test / Model utility test:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad H_1 : \text{at least one of } \beta_1, \beta_2, \beta_3 \neq 0$$

From the JMP output, the P-value $< .0001$ therefore there is sufficient evidence to reject H_0, i.e. there is relationship between Y and X_1, X_2, X_3.

ii. For β_0, we observe P-value less than .05. Therefore there is sufficient evidence to reject $H_0 : \beta_0 = 0$ against $H_1 : \beta_0 \neq 0$, i.e. the intercept (level) is significant in predicting High Price given X_1, X_2, X_3 are in the model, i.e.

$$E(Y) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

.

iii. For β_1, β_3, we observe P-value less than .05. Therefore there is sufficient evidence to reject $H_0 : \beta_i = 0, i = 1, 3$ against $H_1 : \beta_i \neq 0$ given that the other two variables remain in the model. Therefore there is sufficient evidence that Length and QtrMile are significant in predicting High Price given that the other two variables remain in the model.

iv. For β_2, we observe P-value $= .1408 > .05$. Therefore there is no sufficient evidence to reject $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ given that Length and QtrMile remain in the model, i.e. Weight add little information to predict High Price given that Length and QtrMile is already in the model.

(b) Let X_4 denote HwyMPG, X_5 denote FuelCap and X_6 denote UTurn. Then the full model can be expressed as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon, \quad k = 6$$

$$H_0 : E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$H_1 : E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

While the former model is reduced model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon, \quad m = 3$$

Apply partial F-test,

$$F = \frac{\frac{SSR(full) - SSR(reduced)}{k-m}}{\frac{SSE(full)}{(n-k-1)}} \sim F_{k-m, n-k-1}, \quad f = \frac{\frac{2288.2326 - 2144.6717}{6-3}}{\frac{434.2391}{22-6-1}} = 1.65$$

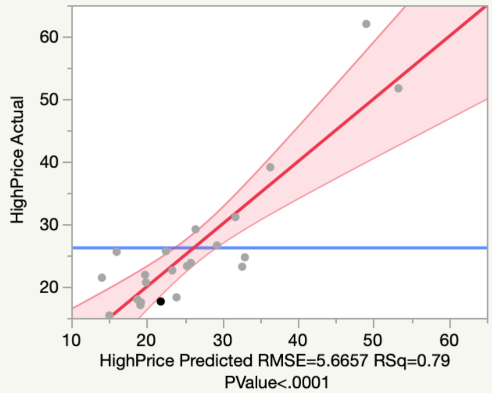
Compare with $F_{\nu_1=5, \nu_2=15, \alpha=.001} = 9.34 > f$. Therefore there is sufficient evidence to reject H_0 with 99.9% confidence, i.e. the additional three variables do not help improve predicitng Y.

JMP output:

Response HighPrice

Whole Model

Actual by Predicted Plot



Residual by Predicted Plot

Summary of Fit

RSquare	0.787766
RSquare Adj	0.752394
Root Mean Square Error	5.665686
Mean of Response	26.19473
Observations (or Sum Wgts)	22

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	2144.6717	714.891	22.2707
Error	18	577.8000	32.100	Prob > F
C. Total	21	2722.4717		<.0001*

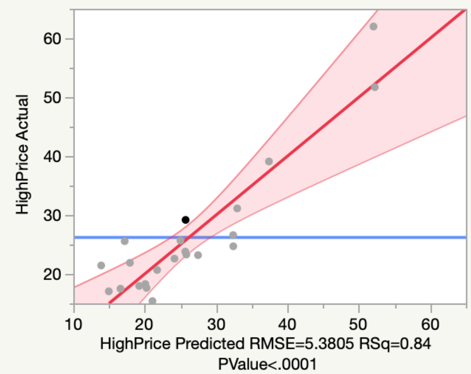
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	235.85816	55.17824	4.27	0.0005*
Length	-0.484625	0.146701	-3.30	0.0040*
Weight	0.0099513	0.00646	1.54	0.1408
QtrMile	-8.913568	2.084961	-4.28	0.0005*

Response HighPrice

Whole Model

Actual by Predicted Plot



Residual by Predicted Plot

Summary of Fit

RSquare	0.840498
RSquare Adj	0.776697
Root Mean Square Error	5.380453
Mean of Response	26.19473
Observations (or Sum Wgts)	22

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	2288.2326	381.372	13.1738
Error	15	434.2391	28.949	Prob > F
C. Total	21	2722.4717		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	205.96479	60.49056	3.40	0.0039*
Length	-0.592865	0.274743	-2.16	0.0476*
Weight	0.0146843	0.008918	1.65	0.1204
QtrMile	-8.489935	2.014196	-4.22	0.0007*
HwyMPG	1.0170523	0.561969	1.81	0.0904
FuelCap	1.2066698	2.088989	0.58	0.5721
UTurn	-0.678282	1.375365	-0.49	0.6290