Solution for HW3, MATH3805

1. Exercise 5.7

   (a) (i) first-order;
   (ii) third-order;
   (iii) first-order;
   (iv) second-order.

   (b) (i) $E(y) = \beta_0 + \beta_1 x$
   (ii) $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$
   (iii) $E(y) = \beta_0 + \beta_1 x$
   (iv) $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$

   (c) (i) $\beta_1 > 0$
   (ii) $\beta_3 > 0$
   (iii) $\beta_1 < 0$
   (iv) $\beta_2 < 0$

2. Exercise 5.17

   (a) $E(y) = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_4+\beta_4x_1x_2+\beta_5x_1x_4+\beta_6x_2x_4+\beta_7x_1^2+\beta_8x_2^2+\beta_9x_4^2$

   (b) H_0: $\beta_1 = \beta_2 = \ldots = \beta_9 = 0$ v.s. H_1: at least one $\beta_i \neq 0$
   F-test statistic is 613.27(SAS). P-value $< 0.0001 < \alpha$.
   so reject H_0. we have enough evidence to conclude that the overall model is statistically useful for predicting y.

   (c) compare a reduced model:
   $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_1 x_4 + \beta_5 x_1 x_4 + \beta_6 x_2 x_4$
   H_0: $\beta_7 = \beta_8 = \beta_9 = 0$ v.s. H_1: at least one $\beta_i \neq 0, i = 7, 8, 9$
   $F = \frac{\left(\frac{\text{SSE}_r - \text{SSE}_c}{g}\right)}{\left(\frac{\text{SSE}_c}{n-(k+1)}\right)} = \frac{\left(\frac{3314.94919 - 146.12634}{3}\right)}{\left(\frac{146.12634}{25-10}\right)} = 108.4275$
   so reject H_0. we have enough evidence to conclude that the curvilinear terms is statistically useful for predicting y.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	53769	5974.36818	613.27	<.0001
Error	15	146.12634	9.74176		
Corrected Total	24	53915			

Root MSE	3.12118	R-Square	0.9973
Dependent Mean	88.32000	Adj R-Sq	0.9957
Coeff Var	3.53394		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10283	2335.56968	4.40	0.0005
REDSHIFT	1	276.75689	147.03327	1.88	0.0793
LINEFLUX	1	3325.18799	476.53498	6.98	<.0001
AB1450	1	1301.30580	115.21604	11.29	<.0001
RL	1	41.97757	14.32647	2.93	0.0103
RA	1	15.98078	4.23728	3.77	0.0018
LA	1	207.37869	11.82790	17.53	<.0001
REDSHIFTSQ	1	0.98954	4.30238	0.23	0.8212
LINEFLUXSQ	1	266.55556	24.84031	10.73	<.0001
AB1450SQ	1	40.21654	2.42805	16.56	<.0001

Figure 1: SAS output without intercept for Exercise 5.17

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	50600	8433.41514	45.79	<.0001
Error	18	3314.94919	184.16384		
Corrected Total	24	53915			

Root MSE	13.57070	R-Square	0.9385
Dependent Mean	88.32000	Adj R-Sq	0.9180
Coeff Var	15.36537		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-5590.24207	4472.44114	-1.25	0.2273
REDSHIFT	1	-826.71928	468.64057	-1.76	0.0947
LINEFLUX	1	-216.24176	355.58876	-0.61	0.5507
AB1450	1	625.47619	223.35634	2.80	0.0118
RL	1	-83.23183	41.00448	-2.03	0.0574
RA	1	-15.71527	13.16828	-1.19	0.2482
LA	1	35.57030	17.14111	2.08	0.0526

Figure 2: SAS output without intercept for Exercise 5.17

3. Exercise 5.20

(a) $\bar{x} = 33,\ s_x = 2.16025.$

The coding system equation: $u = (x - 33)/2.16025$

(b) $x = [30, 31, 32, 33, 34, 35, 36]$

$\rightarrow u = [-1.38873, -0.92582, -0.46291, 0, 0.46291, 0.92582, 1.38873]$

(c) $\text{cor}\,(x, x^2) = \frac{\text{cov}(x,x^2)}{\sqrt{\text{var}(x)\,\text{var}(x^2)}} = 0.99966$

(d) $\text{cor}\,(u, u^2) = 0$

(e) $E(y) = 37.57143 - 0.46291u - 5.3333u^2$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	111.00000	55.50000	44.40	0.0019
Error	4	5.00000	1.25000		
Corrected Total	6	116.00000			

Root MSE	1.11803	R-Square	0.9569
Dependent Mean	33.00000	Adj R-Sq	0.9353
Coeff Var	3.38798		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	37.57143	0.64550	58.21	<.0001
u	1	-0.46291	0.45644	-1.01	0.3679
usq	1	-5.33333	0.56928	-9.37	0.0007

Figure 3: SAS output without intercept for Exercise 5.20

4. Exercise 5.24

   (a) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
   where $x_1 = \begin{cases} 1, & \text{male ;} \\ 0, & \text{female.} \end{cases}$ and $x_2 = \begin{cases} 1, & \text{expert testimont} = \text{yes} \\ 0, & \text{otherwise} \end{cases}$
   Interpretation of model parameters:
   $\beta_0 = u_{11}$ mean at gender = female and expert testimony = yes
   $\beta_1 = u_{2j} - u_{1j}$, for any level of expert testimony
   $\beta_2 = u_{i2} - u_{i1}$, for any level of gender

   (b) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$
   Interpretation of model parameters:
   $\beta_0 = u_{11}$ mean at gender -female and expert testimony = yes
   $\beta_1 = u_{2j} - u_{1j}$, for any level of expert testimony
   $\beta_2 = u_{i2} - u_{i1}$, for any level of gender
   $\beta_3 = u_{22} - u_{12} - u_{21} + u_{11}$,

   (c) Model in part (b).
   when $x_1 = 0, x_2 = 1, E(y) = \beta_0 + \beta_2$
   when $x_1 = 1, x_2 = 1, E(y) = \beta_0 + \beta_1 + \beta_2 + \beta_3$
   when $x_1 = 0, x_2 = 0, E(y) = \beta_0$
   when $x_1 = 1, x_2 = 0, E(y) = \beta_0 + \beta_1$
   and we assume that

$$u_{12} > u_{22} \Rightarrow \beta_0 + \beta_2 > \beta_0 + \beta_1 + \beta_2 + \beta_3$$
$$u_{21} > u_{21} \Rightarrow \beta_0 < \beta_0 + \beta_1$$

   Thus,

$$\beta_1 > 0$$
$$\beta_1 + \beta_3 < 0$$

5. Exercise 5.34

   (a)

$$\begin{aligned} E(y) = &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ &+ \beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_7 x_1 x_4 + \beta_8 x_2 x_3 + \beta_9 x_2 x_4 \\ &+ \beta_{10} x_1^2 + \beta_{11} x_2^2 \end{aligned}$$

   where x_1 is cycle speed, x_2 is cycle pressure ratio, take traditional level as base level.
   $x_3 = \begin{cases} 1, & \text{at advanced level;} \\ 0, & \text{not at advanced level.} \end{cases}$ and $x_4 = \begin{cases} 1, & \text{at aeroderivative level;} \\ 0, & \text{not at aeroderivative level.} \end{cases}$

   (b) for traditional level, $x_3 = x_4 = 0$, model is

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_1 x_2 + \beta_{10} x_1^2 + \beta_{11} x_2^2$$

for advanced level, $x_3 = 1, x_4 = 0$, model is

$$\begin{aligned} E(y) = &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &+ \beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_8 x_2 x_3 \\ &+ \beta_{10} x_1^2 + \beta_{11} x_2^2 \end{aligned}$$

for aeroderivative level, $x_3 = 0, x_4 = 1$, model is

$$\begin{aligned} E(y) = &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 \\ &+ \beta_5 x_1 x_2 + \beta_7 x_1 x_4 + \beta_9 x_2 x_4 \\ &+ \beta_{10} x_1^2 + \beta_{11} x_2^2 \end{aligned}$$

(c)

$$\begin{aligned} E(y) = &14485 + 0.11816 x_1 - 370.44605 x_2 - 2662.13217 x_3 - 1763.69024 x_4 \\ &+ 0.00054906 x_1 x_2 + 0.17531 x_1 x_3 + 0.03417 x_1 x_4 \\ &+ 96.17827 x_2 x_3 + 78.05405 x_2 x_4 - 2.05897\text{E-}7 x_1^2 + 4.68457 x_2^2 \end{aligned}$$

(d) H_0: $\beta_1 = \beta_2 = \ldots = \beta_{11}$ v.s. H_1: at least one is not. F-value is 44.22, and p-value is less than 0.0001

we have enough evidence to conclude that the overall model is useful.

(e) H_0:$\beta_3 = \beta_4 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$ v.s. H_1: at least one is not. $F = \frac{\left(\frac{\text{SSE}_r - \text{SSE}_c}{g}\right)}{\left(\frac{\text{SSE}_c}{n-(k+1)}\right)} = \frac{\left(\frac{19370350-17056349}{6}\right)}{\left(\frac{17056349}{67-12}\right)} = 1.243623$

we do not have enough evidence to conclude that the surfaces are not identical.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	148526859	29705372	93.55	<.0001
Error	61	19370350	317547		
Corrected Total	66	167897208			

Root MSE	563.51284	R-Square	0.8846
Dependent Mean	11066	Adj R-Sq	0.8752
Coeff Var	5.09209		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15583	1142.85985	13.63	<.0001
X1	1	0.07823	0.11044	0.71	0.4814
X2	1	-523.13391	103.37571	-5.06	<.0001
X1X2	1	0.00445	0.00558	0.80	0.4282
X12	1	-1.80598E-7	0.00000197	-0.09	0.9272
X22	1	8.84007	2.16320	4.09	0.0001

Figure 4: SAS output for Exercise 5.34

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	150840860	13712805	44.22	<.0001
Error	55	17056349	310115		
Corrected Total	66	167897208			

Root MSE	556.88009	R-Square	0.8984
Dependent Mean	11066	Adj R-Sq	0.8781
Coeff Var	5.03216		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	14485	1897.87728	7.63	<.0001
X1	1	0.11816	0.14123	0.84	0.4064
X2	1	-370.44605	213.99588	-1.73	0.0890
X3	1	-2662.13217	1458.64130	-1.83	0.0734
X4	1	-1763.69024	3249.94564	-0.54	0.5895
X1X2	1	0.00054906	0.00727	0.08	0.9400
X12	1	-2.05897E-7	0.00000267	-0.08	0.9388
X22	1	4.68457	6.37404	0.73	0.4655
X1X3	1	0.17531	0.09531	1.84	0.0713
X2X3	1	96.14827	84.09666	1.14	0.2579
X1X4	1	0.03417	0.07590	0.45	0.6543
X2X4	1	78.05405	161.90057	0.48	0.6316

Figure 5: SAS output for Exercise 5.34

6. Exercise 5.35

(a) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

(b) when$x_2 = 0, E(y) = \beta_0 + \beta_1 x_1$

when $x_2 = 1, E(y) = \beta_0 + \beta_2 + (\beta_1 + \beta_3)\, x_1$

(c) when $x_2 = 1$, the change of y for every one foot increse in elevation for moss specimens is $\beta_1 + \beta_3$.

(d) $E(y) = 2.38487 + 0.00181 x_1 + 3.20146 x_2 - 0.00133 x_1 x_2$

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ v.s. H_1: at least one is not.

F-value is 0.26, and p-value is $0.8567 > 0.1$

we do not have enough evidence to conclude that the model is useful.

(e) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_1 x_2$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	20.23864	6.74621	0.26	0.8567
Error	66	1738.08227	26.33458		
Corrected Total	69	1758.32091			

Root MSE	5.13172	R-Square	0.0115
Dependent Mean	6.83994	Adj R-Sq	-0.0334
Coeff Var	75.02582		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.38487	5.39314	0.44	0.6598
ELEVATION	1	0.00181	0.00214	0.84	0.4014
SLOPE	1	3.20146	7.66988	0.42	0.6777
ES	1	-0.00133	0.00303	-0.44	0.6629

Figure 6: SAS output for Exercise 5.35

7. Exercise 5.37

(a) $E(y) = \beta_0 + +\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

(b) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$

(c) For level AL, the slope of $E(y)$ is β_1
For level 3A, the slope of $E(y)$ is $\beta_1 + \beta_4$
For level FE, the slope of $E(y)$ is $\beta_1 + \beta_5$

(d) H_0: $\beta_4 = \beta_5 = 0$ v.s. H_1: at least one is not.
fit complete model and reduced model then use F test.

8. Exercise 5.45

(a) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
where $x_1 = \begin{cases} 1, & \text{boy ;} \\ 0, & \text{girl.} \end{cases}$, $x_2 = \begin{cases} 1, & \text{youngest;} \\ 0, & \text{not.} \end{cases}$, $x_3 = \begin{cases} 1, & \text{middle} \\ 0, & \text{mnot.} \end{cases}$

(b) β_0: mean at level: girl, oldest.
β_1: mean difference between girl and boy.
β_2: mean difference between youngest and oldest.
β_3: mean difference between middle and oldest.

(c) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$

(d) $\beta_0 = 0.21$
$\beta_1 = 0.16 - 0.21 = -0.05$
$\beta_2 = 0.27 - 0.21 = 0.06$
$\beta_3 = 0.18 - 0.21 = -0.03$
$\beta_4 = 0.33 - 0.27 - 0.16 + 0.21 = 0.11$
$\beta_5 = 0.33 - 0.18 - 0.16 + 0.21 = 0.20$

(e) H_0: $\beta_4 = \beta_5 = 0$ v.s. H_1: at least one is not.

9. Exercise 5.46

(a) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

(b) β_0 : mean at level: female, 1880.
β_1 : the slope for winning time.
β_2 : the mean difference of male and female.

(c) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$
for male: $E(y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)\, x_1$
for female: $E(y) = \beta_0 + \beta_1 x_1$

(d) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3 + \beta_4 x_1^2 + \beta_5 x_3^2$

(e) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3 + \beta_4 x_1^2 + \beta_5 x_3^2 + \beta_6 x_2$

(f)

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3 + \beta_4 x_1^2 + \beta_5 x_3^2 + \beta_6 x_2 + \beta_7 x_1 x_2 + \beta_8 x_3 x_2 + \beta_9 x_1 x_3 x_2 + \beta_{10} x_1^2 x_2 + \beta_{11} x_3^2 x_2$$

(g)

$$E(y) = \left(\beta_0 + \beta_1 x_1 + \beta_4 x_1^2\right) + \left(\beta_2 + \beta_3 x_1\right) x_3 + \beta_5 x_3^2$$
$$+ \left(\beta_6 + \beta_7 x_1 + \beta_{10} x_1^2\right) x_2 + \left(\beta_8 + \beta_9 x_1\right) x_2 x_3 + \beta_{11} x_3^2 x_2$$

The interaction of x_2 and x_3 should be deleted. So $\beta_8 + \beta_9 x_1 = 0, \beta_{11} = 0$ and $\beta_6 + \beta_7 x_1 + \beta_{10} x_1^2 \neq 0$

(h) $\beta_8 + \beta_9 x_1 = 0, \beta_{11} = 0$

(i) $\beta_8 + \beta_9 x_1 = 0, \beta_{11} = 0$ and $\beta_6 + \beta_7 x_1 + \beta_{10} x_1^2 = 0$

10. Exercise 5.47

(a) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$

where $x_1 = \begin{cases} 1, & \text{Pl;} \\ 0, & \text{P2} \end{cases}, x_2 = \begin{cases} 1, & \text{L1;} \\ 0, & \text{not.} \end{cases}, x_3 = \begin{cases} 1, & \text{L2;} \\ 0, & \text{not.} \end{cases}, x_4 = \begin{cases} 1, & \text{L4} \\ 0, & \text{not.} \end{cases}$

(b) 8 parameters, $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_7 x_1 x_4$

(c) H_0: $\beta_5 = \beta_6 = \beta_7 = 0$ v.s. H_1: at least one is not.

F-value $= \frac{(422.336 - 346.65)/3}{346.63/32} = 2.3297 < F^{-1}_{(3,32)}(0.95) = 2.9011$

We don't have enough evidence to state that for different location, type of packaging influence the total weekly sales.

11. Exercise 6.1

(a) F-Statistic is equivalent to T-Statistic for one variable case. So we compare the absolute value of T-Statistic $\left|\frac{\hat{\beta}}{S_\beta}\right|$

$|t(x_1)| = 3.81, |t(x_2)| = 90, |t(x_3)| = 2.98$

$|t(x_4)| = 1.21, |t(x_5)| = 6.03, |t(x_6)| = 0.86$

x_2 is the best candidate.

(b) Yes. Reject region is $|t(x)| > 2.0106$. x_2 should be included in model.

(c) (i) Fit all two-variable model as: $y = \beta_0 + \beta_1 x_2 + \beta_2 x_i$ where $i = 1, 3, 4, 5, 6$.

(ii) Check all fitted two-variable model with a reduced model: $y = \beta_0 + \beta_1 x_2$ compare F-value.

(iii) keep the two-variable model with largest F-value, check all included variable.

12. Exercise 6.10

take the Traditional level as base level, represent the variables engine as two-valued variable.

$$\text{ADVANCED} = \begin{cases} 1, & \text{at advanced level;} \\ 0, & \text{not at advanced level.} \end{cases}$$

$$\text{AERODERIVATIVE} = \begin{cases} 1, & \text{at aeroderivative level;} \\ 0, & \text{not at aeroderivative level.} \end{cases}$$