เฉลยข้อสอบชุดที่ 2

ข้อสอบชุดนี้มี 5 ข้อ 100 คะแนน

ข้อ 1. จงเติมคำหรือข้อความให้ถูกต้องและได้ความ

1.1 เราถือว่า Model $Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + e_{n\times 1}$ เป็น

(a) General linear hypothesis model of full rank เมื่อ rank(X) มีค่า เท่ากับ p

หรือ (b) General linear hypothesis model of less than full rank เมื่อ rank(X) มีค่า น้อยกว่า p

1.2 จากข้อ 1.1 model (a) หรือ (b) ที่ x_i (i=1,...,n) มีค่า 0 หรือ 1 เท่านั้น (b)

1.3 จากข้อ 1.1 model (b) มีชื่อเรียกเฉพาะว่า Experimental Design model

1.4 จากข้อ 1.1 สำหรับ model (a) จงเขียน normal equations ในรูป matrix $X'X\hat{\beta} = X'Y$

ดังนั้น least square estimator ของ β คือ $\hat{\beta} = (X'X)^{-1}X'Y$

1.5 ใน model $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + e$

สหสัมพันธ์ที่ใช้วัดความสัมพันธ์ของตัวแปรตาม Y กับตัวแปรอิสระ $X_1, X_2, \ldots, X_k$

มีชื่อเรียกว่า Multiple correlation coefficient

1.6 กำลังสองของสหสัมพันธ์ในข้อ 1.5 ซึ่งมีค่าเท่ากับ $SS(\beta_1,\ldots,\beta_k \mid \beta_0)/SST$

มีชื่อเรียกเฉพาะว่า Coefficient of determination

และใช้เพื่ออธิบาย ส่วนของความแปรปรวนของ Y ซึ่งเป็นผลมาจากอิทธิพลของ $X_1, \ldots, X_k$

1.7 จาก Regression model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$

ถ้า partial $f_c(X_3 | X_1, X_2)$ ไม่มีนัยสำคัญ (n.s.) เราสรุปผลว่า <u>ไม่ควรเพิ่ม X_3 เข้าไปในแบบจำลองที่มี X_1 และ X_2 อยู่แล้ว</u>

1.8 ใน (a x b) factorial design นั้น random model , fixed model และ mixed model แตกต่างกันอย่างไรในการเลือกระดับของ Factor A และ Factor B

<u>Random model: ระดับของทั้งสอง Factor ถูกเลือกมาอย่างสุ่ม</u>

<u>Fixed model: ระดับของทั้งสอง Factor ถูกกำหนดขึ้นตามความสนใจ</u>

<u>Mixed model: ระดับของ Factor หนึ่ง ถูกเลือกมาอย่างสุ่ม</u>

<u>ส่วนระดับของอีก Factor หนึ่ง ถูกกำหนดขึ้นตามความสนใจ</u>

1.9 จงยกตัวอย่าง (4 x 4) standard latin square มา 1 design (ให้วิธีการเป็น A, B, C และ D)

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

1.10 ใน (5 x 5) Greaco latin square design (1 หน่วยทดลอง/cell) เราสามารถทดสอบอิทธิพลหลักได้เป็นจำนวน <u>4</u> อิทธิพล แต่เราใช้หน่วยทดลองสำหรับ design นี้เพียง <u>$5^2 = 25$</u> หน่วยทดลองเท่านั้น

ข้อ 2. ในการทดลองเกี่ยวกับการสึกกร่อนของโลหะ ภายหลังการเคลือบน้ำยาด้วยระดับ Voltage ต่าง ๆ กัน ได้ข้อมูลดังตารางข้างล่าง

x: Voltage	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
y: Corrosion(%)	1.10	1.43	2.11	3.12	2.50	2.21	2.50	5.90

จะหา polynomial model ที่ fit กับข้อมูลข้างต้น (α = .05)

จาก Table of Orthogonal Polynomials Values สำหรับ n = 8 เราได้ linear, quadratic และ cubic polynomial ในตารางต่อไปนี้

y	1.10	1.43	2.11	3.12	2.50	2.21	2.50	5.90	ΣP^2
P_1	-7	-5	-3	-1	1	3	5	7	168
P_2	7	1	-3	-5	-5	-3	1	7	168
P_3	-7	5	7	3	-3	-7	-5	7	264

Σy_i = 20.87, Σy_i^2 = 69.6355

$\Sigma P_1 y_i$ = 38.63, $\Sigma P_2 y_i$ = 11.87, $\Sigma P_3 y_i$ = 29.41

$\Sigma P_1 = \Sigma P_2 = \Sigma P_3 = \underline{0}$

$\Sigma P_1 P_2 = \Sigma P_1 P_3 = \Sigma P_2 P_3 = \underline{0}$

โดยที่ Σ คือ $\sum_{i=1}^{8}$

จากการ fit linear model, quadratic model และ cubic model กับข้อมูลข้างต้นได้ผลสรุปบางส่วน ตามตาราง ANOVA ต่อไปนี้

ANOVA

S.V.	d.f.	SS.	MS.	f_c	$Pr[F>f_c]$
Total (uncorrected)	8	69.6355			
Reduction for mean($n\bar{y}^2$)	1	54.4446			
Linear	1	8.8826	8.8826	8.45	< 5%
Error for linear	6	6.3083	1.0514		
Quadratic	1	0.8387	0.8387	0.77	> 5%
Error for quadratic	5	5.4696	0.5483		
Cubic	1	3.2763	3.2763	5.98	> 5%
Error for cubic	4	2.1933	2.1933		

SST(corrected) = $\underline{69.6355 - 54.4446}$

= $\underline{15.1909 \qquad , \text{d.f.} = 7}$

จงเติมตาราง ANOVA ให้สมบูรณ์ และกำหนด $\alpha = .05$ จงหา polynomial model ที่ fit กับข้อมูลที่กำหนดให้

(ในการเติมตาราง ANOVA ให้ใช้ทศนิยม 4 ตำแหน่ง)

หมายเหตุ ในการเติมตาราง ANOVA นักศึกษาอาจใช้ค่าจากการ fit cubic model ต่อไปนี้ช่วยได้

Cubic model: $Y = X\gamma + e$ หรือ $y = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \gamma_3 x^3 + e$

$\gamma' = [\ \gamma_0 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4]$

$p = 4$

$$X'X = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 168 & 0 & 0 \\ 0 & 0 & 168 & 0 \\ 0 & 0 & 0 & 264 \end{bmatrix} \quad , \; X'Y = \begin{bmatrix} 20.87 \\ 38.63 \\ 11.87 \\ 29.41 \end{bmatrix}$$

$$\hat{\gamma}' = (X'X)^{-1}X'Y = \begin{bmatrix} 20.87/8 \\ 38.63/168 \\ 11.87/168 \\ 29.41/264 \end{bmatrix} = \begin{bmatrix} 2.60875 \\ 0.2299404 \\ 0.0706455 \\ 0.1114015 \end{bmatrix}$$

$$\begin{aligned} R(\gamma) &= \hat{\gamma}'X'Y = R(\gamma_0, \gamma_1, \gamma_2, \gamma_o) \\ &= 54.444612 + 8.8826005 + 0.838672 + 3.2763185 \\ &= 63.3272120 + 0.838672 + 3.2763185 \\ &= 64.165884 + 3.2763185 \\ &= 64.4422020 \end{aligned}$$

SSE (cubic) = 63.6355 - 67.4422 = 2.1933 , d.f.= 4

MSE (cubic) = 0.5483

Quadratic model: $Y = X\lambda + e$

$\lambda' = [\lambda_0 \quad \lambda_1 \quad \lambda_2]$, $R(\lambda) = 64.1659$

SSE (quadratic) = 69.6355 - 64.1659 = 5.4696 , d.f. = 5

Linear model: $Y = X\alpha + e$

$\alpha' = [\alpha_0 \quad \alpha_1]$, $R(\alpha) = 63.3272$

SSE (linear) = 69.6355 - 63.3272 = 6.3083 , d.f.= 6

จากตาราง ANOVA

เมื่อทดสอบ H_0: $\alpha_1 = 0$, $f_c = 8.45$, $Pr[F>f_c] < 0.05$

เราปฏิเสธ H_0 แสดงว่า $\alpha_1 \neq 0$

ส่วนการทดสอบ H_0: $\lambda_2 = 0$, $f_c = 0.77$, $Pr[F>f_c] > 0.05$

เราไม่ปฏิเสธ H_0 แสดงว่า $\lambda_2 = 0$

ส่วนการทดสอบ H_0: $\gamma_3 = 0$, $f_c = 5.98$, $Pr[F>f_c] > 0.05$

เราไม่ปฏิเสธ H_0 แสดงว่า $\gamma_3 = 0$

สรุป Model ที่ fit กับข้อมูลข้างต้นคือ **Linear model**

ข้อ 3. สมมุติว่าเราต้องการสำรวจดูว่าน้ำหนัก (Y) ของเด็กที่เป็นโรคขาดอาหารชนิดหนึ่ง ว่าเปลี่ยนแปลงไปตามความสูง (X_1) และอายุ (X_2) อย่างไร ถ้าตัวอย่างสุ่มประกอบด้วยเด็ก 12 คนที่คลีนิคแห่งหนึ่ง น้ำหนัก ความสูง และอายุของเด็กแต่ละคน แสดงในตารางต่อไปนี้

คนที่	1	2	3	4	5	6	7	8	9	10	11	12
Y	64	71	53	67	55	58	77	57	56	51	76	68
X_1	57	59	49	62	51	50	55	48	42	42	61	57
X_2	8	10	6	11	8	7	10	9	10	6	12	9

ในการศึกษาถึงความสัมพันธ์แบบ Multiple regression ของ Y กับ X_1, X_2 และ $X_3 = X_2^2$

ได้พิจารณา 3 models คือ

Model 1: Y on X_1 นั่นคือ $y = \beta_0 + \beta_1 X_1 + e$

Model 2: Y on X_1 และ X_2 นั่นคือ $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$

Model 3: Y on X_1, X_2 และ X_3

นั่นคือ $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$

จาก Computer printouts ที่กำหนดให้

จงเติมตาราง (ให้ใช้ทศนิยม 2 ตำแหน่ง) และตอบปัญหาต่อไปนี้

Least square estimates

Model	ตัวแปรอิสระใน Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$s_{\hat{\beta}_1}$	$s_{\hat{\beta}_2}$	$s_{\hat{\beta}_3}$
1	X_1	6.19	1.07	-	-	0.24	-	-
2	X_1 และ X_2	6.55	0.72	2.05	-	0.26	0.94	-
3	X_1, X_2 และ X_3	3.44	0.72	2.78	-0.04	0.28	7.43	0.42

ANOVA

Model 1			Model 2			Model 3		
S.V.	d.f.	SS.	S.V.	d.f.	SS.	S.V.	d.f.	SS.
X_1	1	588.92	X_1, X_2	2	692.82	X_1,X_2,X_3	3	693.06
Residual	10	299.33	Residual	9	195.43	Residual	8	195.19

3.1 จาก Model 3

Prediction equation คือ $\hat{y} = 3.44 + 0.72x_1 + 2.78x_2 - 0.04x_2^2$

จงทำนายน้ำหนัก (Y) ของเด็กที่มีความสูง (X_1) 50 และมีอายุ (X_2) 10

$\hat{y} = 3.44 + 0.72(50) + 2.78(10) - 0.04(10)^2 = 63.24$

3.2 จงหา R^2 ของ model ทั้ง 3

Model 1: $R^2 = 0.66301$

Model 2: $R^2 = 0.77999$

Model 3: $R^2 = 0.78025$

3.3 จงทำ Overall F test ของแต่ละ model (α=.10)โดยให้ระบุ H_o ที่ทดสอบให้ชัดเจน

Model 1: H_o: $\beta_1=0$, f_c = 19.67486**,signif F = .0013<.10

Model 2: H_o: $\beta_1=\beta_2=0$, f_c = 15.95325**,signif F = .0011<.10

Model 3: H_o: $\beta_1=\beta_2=\beta_3=0$, f_c = 9.46855**,signif F = .0052<.10

3.4 จงทำ Partial F test (α=.10) สำหรับการ

1) เพิ่ม X_2 เข้าไปใน model ที่มี X_1 อยู่แล้ว

และ 2) เพิ่ม X_3 เข้าไปใน model ที่มี X_1 และ X_2 อยู่แล้ว

3.4 1) $H_0: \beta_2 = 0$, $f_c = f(x_2 | x_1) = T^2(x_2 | x_1) = (2.187)^2 = 4.7830$

sig T = .0565 < .10 เราปฏิเสธ H_0 ที่ $\alpha = .10$

นั่นคือควรเพิ่ม X_2 เข้าไปใน model ที่มี X_1 อยู่แล้ว

2) $H_0: \beta_3 = 0$, $f_c = f(x_3 | x_1, x_2) = T^2(x_3 | x_1, x_2)$

$= (-0.099)^2 = 0.0098$

sig T = .9238 > .10 เราไม่ปฏิเสธ H_0 ที่ $\alpha = .10$

นั่นคือไม่ควรเพิ่ม X_3 เข้าไปใน model ที่มี X_1 และ X_2 อยู่แล้ว

[$T(x_2 | x_1)$ และ $T(x_3 | x_1, x_2)$ จากหน้า 15 ของ printout]

ข้อ 4. จากโจทย์ และ computer printouts ในข้อ 3 จงสรุปผลการเลือก regression model ที่ดีที่สุดสำหรับทำนาย Y จากแต่ละวิธีที่กำหนดให้ในข้อนี้ จากนั้น **ให้เลือกอธิบายขั้นตอนอย่างละเอียด** 1 **วิธี** จาก 3 วิธีที่กำหนดให้

3 วิธีที่กำหนดให้คือ 1) Forward selection method

2) Backward elimination method

3) Stepwise method

ทั้งนี้ต้องใช้ค่าจาก printouts ประกอบคำอธิบายทุกขั้นตอนด้วย

สรุปผลการเลือก Regression model จากแต่ละวิธี

1) Forward selection method เลือก model ที่ 2

2) Backward elimination method เลือก model ที่ 2

3) Stepwise method เลือก model ที่ 2

<u>เฉลย</u> ดูหัวข้อ 6.8.1 ในหนังสือ ST 313

ข้อ 5.1 การทดลองได้จำแนกระดับสติปัญญา (IQ) และชนิดของโรงเรียนที่นักเรียนเรียน นักเรียน 4 คน ถูกเลือกอย่างสุ่มจากกลุ่มที่จำแนกโดย IQ-SCHOOL นักเรียนทั้งหมดสอบวิชาภาษาอังกฤษ ข้อมูลดังแสดงในตารางต่อไปนี้

IQ	School		Total
	B_1:Spectial	B_2:Public	
A_1: IQ < 60	40.1,41.1,40.9,39.4	63.0,61.9,61.6,64.0	
รวม	161.5	250.5	412.0
A_2:60<IQ<70	52.8,53.6,53.9,53.8	68.6,70.7,69.1,73.3	
รวม	214.1	281.7	495.8
A_3: IQ > 70	41.6,37.7,43.2,42.0	62.8,56.6,63.3,60.9	
รวม	165.5	243.6	408.1
รวม	540.1	775.8	1315.9

จงเติมตาราง ANOVA ที่กำหนดให้ ให้สมบูรณ์ แล้วทดสอบอิทธิพลร่วม และอิทธิพลหลักที่ α = .05 ทั้งนี้ให้ตั้งสมมติฐาน และสรุปผลการทดสอบให้ชัดเจน

ANOVA

S.V.	d.f.	SS.	MS.	f_c
IQ	2	613.706	306.853	$f_1=85.637^{**}$
School	1	2314.770	2314.770	$f_2=646.007^{**}$
IQ x School	2	28.676	14.338	$f_3=4.001^{*}$
Residual	18	64.498	3.583	
Total	23	3021.65		

$f_{(2,18),.05} = 3.55$, $f_{(1,18),.05} = 4.41$

ทดสอบอิทธิพลร่วม

H_o: ไม่มีอิทธิพลร่วมระหว่าง IQ และ School

$f_c = 4.001 > f_{(2,18),.05} = 3.55$

เราปฏิเสธ H_o นั่นคือมีอิทธิพลร่วมระหว่าง IQ และ School ต่อคะแนนสอบภาษาอังกฤษ

ทดสอบอิทธิพลหลัก

IQ H_o: ไม่มีอิทธิพลของ IQ

$f_c = 85.637 > f_{(2,18),.05} = 3.55$

เราปฏิเสธ H_o นั่นคือมีอิทธิพลของ IQ ต่อคะแนนสอบภาษาอังกฤษ

School H_o: ไม่มีอิทธิพลของ School

$f_c = 646.007 > f_{(1,18),.05} = 4.41$

เราปฏิเสธ H_o นั่นคือมีอิทธิพลของ School ต่อคะแนนสอบภาษาอังกฤษ

จากการวิเคราะห์ข้างต้น เป็นการวิเคราะห์ตาม Fixed หรือ Random หรือ Mixed model

Fixed model

ข้อ 5.2 ในการศึกษาเกี่ยวกับอิทธิพลของดนตรี ซึ่งเปิดในขณะที่พนักงานธนาคารกำลังปฏิบัติงานว่ามีผลต่อการทำงานหรือไม่ ดนตรี 5 ชนิดที่นำมาเปรียบเทียบคือ

วิธีการที่	Latin letter	Tempo and Style of Music
1	A	Slow, instrumental and vocal
2	B	Medium, instrumental and vocal
3	C	Fast, instrumental and vocal
4	D	Medium, instrumental only
5	E	Fast, instrumental only

หน่วยทดลองคือพนักงานที่รับ-จ่ายเงินของธนาคารในวันทำงานในแต่ละสัปดาห์ จากข้อมูลที่รวบรวมมาได้ในตารางต่อไปนี้ **จงเติมตาราง ANOVA ให้สมบูรณ์ และทดสอบอิทธิพลหลักทั้งสาม ที่ α = .05 ทั้งนี้ให้ตั้งสมมติฐาน และสรุปผลการทดสอบให้ชัดเจน**

Week	Day					Total
	M	T	W	Th	F	
1	D 18	C 17	A 14	B 21	E 17	87
2	C 13	B 34	E 21	A 16	D 15	99
3	A 7	D 29	B 32	E 27	C 13	108
4	E 17	A 13	C 24	D 31	B 25	110
5	B 21	E 26	D 26	C 31	A 7	111
Total	76	119	117	126	77	515

จงเติมตารางต่อไปนี้

Latin letter	A	B	C	D	E	Total
Total	57	133	98	119	108	515

Model: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk}$, $i = 1,\ldots,5;$

$j = 1,\ldots,5;$

$k = 1,\ldots,5$

ANOVA

S.V.	d.f.	SS.	MS.	f_c
Week	4	82.00	20.5	f_1=1.306(n.s.)
Day	4	477.20	119.3	f_2=7.599*
Latin letter	4	664.40	116.1	f_3=10.58*
Error	12	188.40	15.7	
Total	24	1412.00		$f_{(4,12),.05}$=3.26

Week $H_0: \alpha_i = 0$, $i = 1,\ldots,5$

$f_c = 1.306 < f_{(4,12),.05} = 3.26$ เราไม่ปฏิเสธ H_0

นั่นคือไม่มีอิทธิพลของสัปดาห์ต่อผลการปฏิบัติงาน

Day $H_0: \beta_j = 0$, $j = 1,\ldots,5$

$f_c = 7.599 > f_{(4,12),.05} = 3.26$ เราปฏิเสธ H_0

นั่นคือมีอิทธิพลของวันในสัปดาห์ต่อผลการปฏิบัติงาน

วิธีการ หรือ Latin letter $H_0: \gamma_k = 0$, $k = 1,\ldots,5$

$f_c = 10.58 > f_{(4,12),.05} = 3.26$ เราปฏิเสธ H_0

นั่นคือมีอิทธิพลของดนตรีต่อผลการปฏิบัติงาน