<Regression Analysis Lab. Final Exam 2015> Hakbun: _____

(total points : 85 +15 =100) Name in Korean :_____

(괄호 안에 답만 적으시오. 과정생략. 괄호 밖의 내용은 인정하지 않음)

The following output is the summary of p-values from the regressions of FuelCon on Pop, Area, or Gastax variables.

	pop	area	gastax
model fuelcon = pop area gastax ;	0.0296	0.0543	0.1353
model fuelcon = pop area ;	0.0380	0.8240	
model fuelcon = pop gastax;	0.0986		0.0560
model fuelcon = area gastax ;		0.7478	0.0727
model fuelcon = pop ;	0.0359		
model fuelcon = area ;		0.8087	
model fuelcon = gastax ;			0.0718

[1] Using “forward selection with SLE=0.06”, write your selected variables : (5pts)

(Pop Gastax Area)

[2] Using “backward elimination with SLS= 0.08”, write your selected variables : (5pts)

(Pop)

[3] Using “Stepwise Regression with SLE= 0.06 and SLS = 0.08”, write your selected variables (5pts) :

(Gastax)

[4] When you try “Stepwise Regression with SLE=0.20 and SLS=0.20”, complete the following SAS code (5pts) :

```
proc reg ;
    model  fuelcon = pop area gastax / _( Selection=stepwise sle=0.2  sls=0.2 )_ ;
run;
```

[5] Write the selected variables based on Cp-statistic (5pts):

(Pop Gastax)

Number in Model	C(p)	R-Square	Variables in Model
2	2.1458	0.1544	pop GASTAX
1	3.9143	0.0868	pop
3	4.0000	0.1570	pop AREA GASTAX
1	5.1487	0.0646	GASTAX
2	5.8613	0.0877	pop AREA
2	7.0353	0.0667	AREA GASTAX
1	8.6854	0.0012	AREA

[6] Answer the following questions.

6-1 Fill the (a), (b) blanks (10 pts) :

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	397.44586	397.44586	124.41	<.0001
Error	18	57.50414	3.19467		
Lack of Fit	12	(a)	*****	(b)	0.0264
Pure Error	6	5.00000	0.83333		
Corrected Total	19	454.95000			

(a) = (52.50414)

(b) = (5.2506)

6-2 Find the right null hypothesis for the above F-value at (b) (5pts): (3)

(1) The regression model is significant

(2) The regression model is not significant

(3) The Lack of Fit is negligible

(4) The Lack of Fit is not zero

[7] The results for the regression of Salary (y) on Edu (x1), Exper (x2), and Time (x3) as well as the regression of Salary (y) on Edu (x1) are shown below.

model y = x1 x2 x3 ;

Source	DF	Mean Square	F Value	Pr > F
Model	(a)	4663749	12.84	<.0001
Error	89	363281		
Corrected Total				

model y = x1 ;

Source	DF	Mean Square	F Value	Pr > F
Model	(b)	7862534	18.60	<.0001
Error	91	422646		
Corrected Total				

7-1 Write the right degrees of freedom for (a) and (b) : (10 pts)

(a) = (3)

(b) = (1)

7-2 Consider the two different models as above, and conduct the significance test to choose the better one with 0.05 significance level.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

$$y = \beta_0 + \beta_1 x_1 + e$$

a. Write the hypotheses to be tested with beta's. (5pts)

Ho : $\beta_2 = \beta_3 = 0$

Ha : $\beta_2 \neq 0 \text{ or } \beta_3 \neq 0$

b. The decision rule for the test is (5pts)

Reject Ho if Fo > $F_{0.05,\, df_1,\, df_2}$

df_1 = (2)

df_2 = (89)

c. The test statistic is (5pts) Fo = (8.435)

d. The corresponding p-value is 0.0004. Write your selected model equation (5pts)

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

[8] The regression of Sales on ADV and BONUS variables is performed as follows:

```
proc reg ;
        model  Sales = ADV   BONUS /  dwprob ;
    run;
```

Durbin-Watson D	1.191
Pr < DW	0.0050
Pr > DW	0.9950
Number of Observations	22

8-1 This procedure is for the diagnosis of normal distribution assumption (5pts)

(True, False)

8-2 Write your conclusion not beyond the given spaces. (10 pts)

T	h	e		e	r	r	o	r	s		a	r	e		**p**	**o**	**s**	**i**	**t**
i	**v**	**e**	**l**	**y**		**c**	**o**	**r**	**r**	**e**	**l**	**a**	**t**	**e**	**d**	**.**			