<Regression Analysis Mid-term, 2015> 학번:_____

(총 120점) 이름 (한글) :_____

[1] The following SAS output (1) is obtained from the regression between annual incomes (INCOME) and consumptions (CONS). Fill the blanks
(10 pts.)

SAS-output (1)

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: CONS

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2603912710	2603912710		<.0001
Error					
Corrected Total		2796433722			

Root MSE	4387.72164	R-Square	
Dependent Mean	29189	Adj R-Sq	0.9243
Coeff Var	15.03224		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2521.37891	2619.58763	0.96	0.3585
INCOME	1	0.82690	0.07110	11.63	<.0001

1.1 Write the corresponding SAS code in the blank. (5 pts)

proc reg data=a;

_____

run;

1.2 R-Square is “Coefficient of D_____” (5 pts)

1.3 Write the estimated regression equation (5 pts)

1.4 Calculate the average INCOME based on the SAS Output. (5 pts)

1.5 Construct the approximate 95% Confidence Interval for the slope using (Z_0.025 = 1.96, Z_0.05=1.64). (5 pts)

1.6 For the hypotheses,

$$H_0 : \beta_0 = 0, \quad H_a : \beta_0 > 0$$

i) Write the corresponding p-value based on Output (5 pts) :

ii) Given the answer of (i), if someone has no income, how much is he going to consume ? (5pts)

[2] In hypothesis testing, the null hypothesis is rejected if … (5 pts.) ( )

A. P-value > the significance level

B. P-value < the significance level

C. P-value > (the significance level) / 2

D. P-value > (the significance level) x 2

[3] The ABX company sells winter sports merchandise. The time (TIME) period represents the first quarter of 1994 through the ends in 2003. Dependent variable is SALES. The following is the SAS output.

SAS-output (2)

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: SALES

Number of Observations Read	40
Number of Observations Used	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	34818	34818	137.50	<.0001
Error	38	9622.06053	253.21212		
Corrected Total	39	44440			

Root MSE	15.91264	R-Square	0.7835
Dependent Mean	251.41250	Adj R-Sq	0.7778
Coeff Var	6.32930		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	199.01731	5.12788	38.81	<.0001
TIME	1	2.55586	0.21796	11.73	<.0001

3.1 Can we say there is a linear trend about SALES ? (5 pts)

(Yes or No)

3.2 Test whether the linear trend is 3.0 or not with 5% level of significance. Fill the blanks. (10 pts)

$$t_0 = \frac{(\quad (1) \quad) - (\quad (2) \quad)}{(\quad (3) \quad)} = (\quad (4) \quad)$$

(1) = ( )
(2) = ( )
(3) = ( )
(4) = ( )

$$t_{0.05,38} = 1.685, \quad t_{0.02538} = 2.024$$

Therefore we (can or can't) say that the linear trend is 3.0.

3.3 Fill the blanks in the SAS code for the above regression output. (10 pts)

```
proc  reg  data=a ;
      model   ( (1) ) = ( (2) )   ;
run;
```

(1) = ( )

(2) = ( )

3.4 What percentage of the variation in SALES has been explained by the regression? (5 pts)

( )

[4] Translate the following english term in Korean, or translate the Korean term into English one. (each 2 pts)

4.1 Significance Level ( )

4.2 통계학 ( )

4.3 L.S.E. (L S E)

4.4 Two-Tailed Test ( )

4.5 상관계수 (C C)

4.6 Confidence Interval ( )

4.7 $n \geq 100$ (n is 100)

4.8 i.i.d. (i and i d)

4.9 잔차 ( )

4.10 y를 x에 회귀시키다 ( )

[5] Fill the blanks (2 pts each / total 20 pts)

Recall that a s_____ is any value calculated from a sample.

Because statistics are r_____, they have probability distributions called s_____.

The variance of the estimate of the conditional mean of y given x = x_m is equal

$$\sigma_m^2 = \sigma_e^2 \left(\frac{1}{n} + \frac{(x_m - \bar{x})^2}{(n-1)s_x^2} \right)$$

where

$$s_x^2 = \frac{1}{(\quad)} \sum_{i=1}^{n} (\qquad)$$

The standard _____ of the estimate of the point on the regression line is affected by the distance of the value of x_m from _____ .

The (closer or farther) the value of x_m to the mean value, the smaller the variance of the estimate, so the (less or more) accurate the estimate is expected to be. Because all least-squares lines pass through the point (,).