<Regression Analysis Lab. Mid-term, 2015> Hakbun: \_\_\_\_\_\_

(total points : 90 + 10 = 100)

Name in Korean :\_\_\_\_\_

## [1] Let X be a random variable defined as

- X=1 if an even number (짝수) of dots appears on the roll of a single dice (주사위)
- X=2 if an odd number (홀수) of dots appears on the roll of a single dice

Construct the probability distribution of X by completing the following table :

(5 pts) Show that the expected value of X is 1.5

$$E(X) = (1 \times 0.5 + 2 \times 0.5) = 1.5$$

(5 pts) Find the standard deviation of X

( 0.5 )

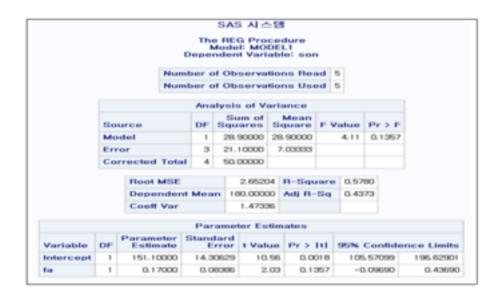
(5 pts) As the expected value is 1.5, we can expect the value of X to be 1.5 if we roll a dice

(True or False)

[2] If a null hypothesis is rejected at the 5% level of significance, what decision would have been made at the 10% level?

(5 pts) ( 1 )

- 1. Reject the null hypothesis
- 2. Do not reject the null hypothesis
- 3. We don't know
- [3] The following SAS output is obtained by regressing the son's height (Son) on the father's height (Fa).



(2 pts) Write the regression equation : (  $\hat{\mathrm{Son}} = 151.1 + 0.17\,\mathrm{Fa}$  )

The following hypotheses are tested:

$$H_0: \beta_1 = 0$$
 vs.  $H_1: \beta_1 \neq 0$ 

(2 pts) Write the p-value for this test ( 0.1357 )

(2 pts) We reject the null hypothesis (alpha=0.05) ( True or False )

The following hypotheses are tested:

$$H_0: \beta_1 = 0$$
 vs.  $H_1: \beta_1 > 0$ 

(2 pts) Write the p-value for this test ( 0.06785 )
(2 pts) We reject the null hypothesis (alpha=0.10) ( True or False )

[4] Highway mileages of 147 cars are in a file named CARS2. Assume these cars represent a random sample of all new cars produced in 2003. The corporate average fuel economy (CAFE) standards set by the government require that average fuel economy for cars be more than 27.5 miles per gallon.

(5 pts) To examine whether the CAFE standard is being met, which hypotheses are correct?

( 3 )

- 1)  $H_0: \mu = 27.5$  vs.  $H_1: \mu \neq 27.5$
- 2)  $H_0: \mu \ge 27.5$  vs.  $H_1: \mu < 27.5$
- 3)  $H_0: \mu \le 27.5$  vs.  $H_1: \mu > 27.5$

We can t-test using SAS as shown below:

The TTEST Procedure

Variable: HWYMPG

N Mean Std Dev Std Err Minimum Maximum 147 28.1497 6.5337 0.5389 13.0000 68.0000

Mean 95% CL Mean Std Dev 95% CL Std Dev 28.1497 27.2576 Infty 6.5337 5.8626 7.3799 (8 pts) Fill the blanks:

$$\bar{y} = \frac{1}{147} \sum_{i=1}^{147} y_i =$$
 ( 28.1497 )

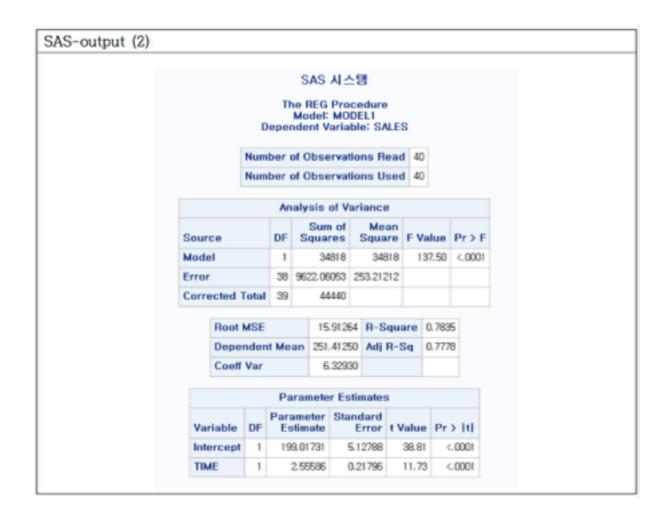
$$s_y^2 = \frac{1}{146} \sum_{i=1}^{147} (y_i - \bar{y})^2 =$$
 ( 42.689

$$SE(\bar{y}) = \sqrt{\frac{(\quad (1) \quad )}{(\quad (2) \quad )}} = 0.5389$$
 (1) = ( 42.689 (2) = ( 147 )

- (2 pts) Write the test statistic value :  $t_0 =$  ( 1.205
- (5 pts) We could say that 2003 cars meet the CAFE standards. (alpha=0.05)

$$t_{0.05,146} = 1.655, \quad t_{0.025,146} = 1.976$$
 ( True or False )

[5] 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.



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

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

$$(1) = ( 2.555 )$$

$$(2) = ( 3.0 )$$

$$(3) = ( 0.217 )$$

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

$$(4) = ( -2.050 )$$

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

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

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

( 78.3 )

[6] Fill the missing blanks on the following ANOVA table from simple linear regression: (10 pts)

Source	DF	SS	MS	F0
Regression	(1)	(2)	1000	(3)
Error	80	800	(4)	
<u>Total</u>	81	(5)		

$$(1) = (1)$$
  $(2) = (1000)$   $(3) = (100)$   $(4) = (10)$ 

(5) = (1800)