

<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)	x	P(x)
	1	(0.5)
	2	(0.5)

(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).

SAS 시스템							
The REG Procedure							
Model: MODEL1							
Dependent Variable: son							
Number of Observations Read				5			
Number of Observations Used				5			
Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F		
Model	1	28.90000	28.90000	4.11	0.1367		
Error	3	21.10000	7.03333				
Corrected Total	4	50.00000					
Root MSE		2.65204	R-Square	0.5780			
Dependent Mean		180.00000	Adj R-Sq	0.4373			
Coeff Var		1.47336					
Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Intercept	1	151.10000	14.30629	10.56	0.0018	105.57099	196.62901
fa	1	0.17000	0.08386	2.03	0.1367	-0.09690	0.43690

(2 pts) Write the regression equation :

$$(\hat{S}on = 151.1 + 0.17 Fa)$$

The following hypotheses are tested :

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0$$

(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 :

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 > 0$$

(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) $H_0 : \mu \geq 27.5$ vs. $H_1 : \mu < 27.5$

3) $H_0 : \mu \leq 27.5$ 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:

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

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

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

$$(2) = (\quad 147 \quad)$$

(2 pts) Write the test statistic value : $t_0 = (\quad 1.205 \quad)$

(5 pts) We could say that 2003 cars meet the CAFE standards. (alpha=0.05)

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

[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.

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

(5 pts) Can we say there is a linear trend about SALES ? (Yes or No)

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

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

$$t_{0.05,38} = 1.685, \quad t_{0.025,38} = 2.024$$

$$(1) = (2.555)$$

$$(2) = (3.0)$$

$$(3) = (0.217)$$

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

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.

```
proc reg data=a ;
```

```
model ( (1) ) = ( (2) ) ;
```

```
( (3) )
```

$$(1) = (\text{Sales})$$

$$(2) = (\text{Time})$$

$$(3) = (\text{run ;})$$

(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)	
Total	81	(5)		

$$(1) = (1)$$

$$(2) = (1000)$$

$$(3) = (100)$$

$$(4) = (10)$$

$$(5) = (1800)$$