(101)輔仁大學碩士班招生考試試題	考試日期：101 年 3 月 9 日第 3 節
	本試題共 2 頁 (本頁為第 1 頁)

科目：迴歸分析　　系所組：統計資訊與金融統計碩士班

1. A study about modeling the length of stay in hospital was conducted and the corresponding data set consists of 80 random selected samples (hospitals) was observed. The details about the data set in listed in Table A (Note that LS is employed as the dependent variable and the rest variables are the independent variables in the model). If all the variables can be directly used in building the model, the corresponding output from package R is listed in Output 1. Please answer the following questions:
   a. Write down the estimated model using the variable names in Table A. (10%)
   b. Write down the meaning of each coefficient in your model in a (i.e., explain the model). (10%)
   c. What is the value of the multiple R-squared in this model and what does it means? (10%)
   d. What is the value of the adjusted R-squared in this model and what is the difference between multiple R-squared and adjusted R-squared? (10%)
   e. According to Output 1, IR has the largest coefficient with it. Does this means IR is the most important independent variable in the model (Explain your answer in detail)? (10%)
   f. The F-statistic and associated p-value are printed at the end of Output 1. Write down the null hypothesis of this test. (10%)
   g. What conclusion can you obtain according to the test in f.? (10%)
   h. If a new important categorical variable "Region" with four levels (N,W,E, and C) will be added in the model. Write down the formula of the indicator variables that you can use to code "Region". (10%) What is the reference level in your coding? (10%) What is the meaning of each associated coefficient? (10%)

科目：迴歸分析　　系所組：統計資訊學系應用統計碩士班

Table A

Variable Name	Description
LS (Length of stay)	Average length of stay in hospital (in days)
IR (Infection risk)	Average probability of acquiring infection (in percent)
ADC (daily census)	Average number of patients in hospital
NN (Number of nurses)	Average number of nurses

Output 1

Call:

lm(formula = LS ~ IR + ADC + NN)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.941030	0.610021	9.739	5.19e-15 ***
IR	0.672785	0.141677	4.749	9.44e-06 ***
ADC	0.013716	0.002653	5.170	1.84e-06 ***
NN	-0.011046	0.002802	-3.943	0.000178 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.51 on 76 degrees of freedom

Multiple R-squared: 0.4954, Adjusted R-squared: 0.4755

F-statistic: 24.87 on 3 and 76 DF, p-value: 2.58e-11