※ 注意：請於試卷上「選擇題作答區」依序作答。

Part I. Multiple-Choice Questions: Please select the best answer. (3 points each; 15 points in total)

1. A Pearson's r calculated for the bivariate relationship between X and Y is .40. After controlling for Z, the partial correlation coefficient is .25. This shows that the relationship between X and Y is a(n)
   - (A) direct relationship.
   - (B) spurious relationship.
   - (C) intervening relationship.
   - (D) interactive relationship.
   - (E) spurious or intervening relationship.

2. In the ANOVA table below, which of the following options is CORRECT?

Source of Variation	Sum of Squares	Degree of Freedom	Mean Sum of Square	F
Between	A	D	G	I
Within	B	E	H	
Total	C	F		

   - (A) A + D = G
   - (B) A − B = C
   - (C) D/E = F
   - (D) B×E = F
   - (E) G/H = I

3. Assume that the test scores are normally distributed with a mean of 77 and a standard deviation of 6. What is the probability that a randomly selected case from the sample will have a score less than 65?
   - (A) 0.0228
   - (B) 0.4772
   - (C) 0.5228
   - (D) 0.9772
   - (E) 1

4. Which of the following statements regarding a sampling distribution and a population distribution is CORRECT?
   - (A) The variance of the sampling distribution is usually larger than the variance of the population distribution.
   - (B) The variance of the sampling distribution is usually smaller than the variance of the population distribution.
   - (C) The two distributions will become the same as the sample size increases.
   - (D) The sampling distribution and the population distribution are observable.

5. Which of the following do we need to calculate the standard error of the mean?
   - (A) the standard deviation of the population.
   - (B) the median of the population.
   - (C) the size of the sampling distribution.
   - (D) the correlation between population and sample.

見背面

※ 注意：請於試卷上「非選擇題作答區」標明大題及小題題號，並依序作答。

Part II. Short Answer Questions (85 points in total)

1. **State the differences** between <u>descriptive statistics</u> and <u>inferential statistics</u>. (5 points)
2. Please <u>state the Central Limit Theorem (8 points)</u> and <u>explain how it relates to the confidence interval (7 points)</u>.
3. Pick **two** types of hypothesis tests from the following list. Please create an example for each of the hypothesis tests you picked. State <u>the assumptions,</u> <u>the null and alternative hypotheses,</u> and <u>the steps involved in conducting the test</u>. Also, <u>what conclusion can you make after conducting the test</u>? (24 points)

   a. Two-sample mean test
   b. ANOVA test
   c. Chi-square test

   (A) Hypothesis test 1 selected (1 point):
       1) Example (2 points):
       2) Assumptions (2 points):
       3) Hypotheses (2 points):
       4) Steps (3 points):
       5) Conclusion (2 points):

   (B) Hypothesis test 2 selected (1 point):
       1) Example (2 points):
       2) Assumptions (2 points):
       3) Hypotheses (2 points):
       4) Steps (3 points):
       5) Conclusion (2 points):

4. Please <u>state the assumptions of ordinary least squares (OLS) regression</u>. (12 points)

5. A researcher wanted to know the relationship <u>between socioeconomic status and weight gain</u>. She studied 900 participants over a year and gathered information on the variables below.

   Dependent variable:
       `gain` = weight gain (positive value = weight gain, negative value = weight loss)
   Independent variables:
       `hours`: hours spent exercising
       `ses`: seslow, sesmiddle, seshigh (categorical variable, "high" is the reference group)

接 次 頁

```
OLS Models Predicting Weight Gain
===================================================================================
                                  Dependent Variable: Weight Gain
                         ----------------------------------------------------------
                            Model 1            Model 2              Model 3
                              (1)                (2)                  (3)
-----------------------------------------------------------------------------------
hours                      -2.470**           -3.268***           -6.874***
                           (0.948)            (0.468)             (0.777)

sesmiddle                                      17.785***           18.930***
                                              (0.566)             (2.288)

seslow                                         29.746***            9.933***
                                              (0.567)             (2.177)

hours:sesmiddle                                                   -0.579
                                                                  (1.119)

hours:seslow                                                       9.830***
                                                                  (1.051)

Constant                   -5.076**           -19.320***          -12.149***
                           (1.955)            (1.014)             (1.591)

-----------------------------------------------------------------------------------
Observations                900                900                 900
R2                         0.008              0.759               0.789
Adjusted R2                0.006              0.758               0.787
Residual Std. Error    14.055 (df = 898)    6.937 (df = 896)     6.502 (df = 894)
F Statistic          6.788** (df = 1; 898) 939.491*** (df = 3; 896) 666.752*** (df = 5; 894)
===================================================================================
```

Please refer to the table above for the following questions.

(A) State the OLS regression of Model 1. (3 points)

(B) Interpret the effect of hours spent exercising on weight gain. (3 points)

(C) The researcher added socioeconomic status in Model 2. How would you interpret the effects of "sesmiddle" and "seslow"? (5 points)

(D) The interaction terms between exercise hours and socioeconomic status were added in Model 3. How to interpret the interaction term between hours and seslow (hours:seslow)? (3 points)

(E) Interpret the R-squared of Model 3. (3 points)

(F) Below is the output of Model 1 from R.
  i.   Explain what the "Residuals" mean. (3 points)
  ii. How to calculate the t-value for "hours"? (3 points)
  iii. What does the Pr(>|t|) mean (i.e., 0.00958 and 0.00933)? (3 points)
  iv. What does the F-test do (i.e., F-statistic and its p-value)? (3 points)

```
Call:
lm(formula = gain ~ hours, data = weight)

Residuals:
    Min     1Q  Median     3Q     Max
-42.787 -10.011   2.086  11.388  27.164

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.0757     1.9550  -2.596  0.00958 **
hours        -2.4696     0.9479  -2.605  0.00933 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.06 on 898 degrees of freedom
Multiple R-squared:  0.007502,  Adjusted R-squared:  0.006397
F-statistic: 6.788 on 1 and 898 DF,  p-value: 0.009329
```

試題隨卷繳回