注意：請依題號順序作答，簡短精確地說明理由，可以用中文作答。

Question I (12 points)

Suppose that a military dictator in an unnamed country holds a plebiscite (a yes/no vote of confidence) and claims that he was supported by 65% of the voters. A human rights group suspects foul play and hires you to test the validity of the dictator's claim. You have a budget that allows you to randomly sample 200 voters from the country. You thus collect your sample of 200, and you find that 115 people actually votes yes. You approximate the probability that you would find 115 or fewer yes votes from a random sample of 200 if, in fact, 65% of the entire population voted yes.

1. Explain how and why we could claim that the probability is $P(Z \leq -2.22) \approx 0.013$, where Z is a standard normal random variable.

Question II (20 points, 10 points each)

Let Y denote a Bernoulli(θ) random variable with $0 < \theta < 1$. Suppose we are interested in estimating the odds ratio, $\gamma = \theta/(1-\theta)$, which is the probability of success over the probability of failure. A nature estimator of γ is $G = \bar{Y}/(1-\bar{Y})$, the proportion of successes over the proportion of failures in the sample. Note that $\bar{Y} = (Y_1 + Y_2 + \cdots + Y_n)/n$, the proportion of successes in n trials, given a random sample $\{Y_1, Y_2, \dots, Y_n\}$.

2. Is G an unbiased estimator of γ? Explain.
3. Is G a consistent estimator of γ? Explain.

Question III (20 points, 10 points each)

Consider the case where the true population model has two explanatory variables and an error term:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

and we assume that this model satisfies Gauss-Markov assumptions.

4. Suppose that our primary interest is in β_1, the partial effect of x_1 on y. We thus estimate the equation excluding x_2:
$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1.$$
Note that the variable x_2 may or may not be correlated with x_1. What is the effect, if any, of excluding x_2 from the model on the estimated β_1? Explain.
5. Suppose we specify the model as
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$
However, x_3 has no effect on y after x_1 and x_2 have been controlled for. Because we do not know that $\beta_3 = 0$ in practice, we still estimate the equation including x_3:
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3.$$
Note that the variable x_3 may or may not be correlated with x_1 or x_2. What is the effect, if any, of including x_3 in the model on the estimated β_1? Explain.

見背面

Question IV (48 points, 6 points each)

A problem of interest to health officials is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight; a birth weight that is too low can put an infant at risk for contracting various illnesses. Our data set contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in grams (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). We also take other factors into account and specify the model as

$$\ln(bwght) = \beta_0 + \beta_1 cigs + \beta_2 mage + \beta_3 meduc + \beta_4 feduc + \beta_5 npvis + \beta_6 npvis^2 + u,$$

where

bwght	=	birth weight, in grams
cigs	=	average number of cigarettes the mother smoked per day during pregnancy
mage	=	mother's age
meduc	=	years of schooling for the mother
feduc	=	years of schooling for the father
npvis	=	total number of prenatal visits.

The estimates of this equation, obtained using the method of ordinary least squares (OLS), are given in column (1) of Table 1. Standard errors are listed in parentheses.

6. Interpret the coefficient on *cigs*. In particular, what is the effect on *bwght* from smoking 10 more cigarettes per day?
7. We would like to estimate the partial effect of *npvis* on ln(*bwght*). However, we want a single value to describe this relationship. One intuitively appealing measure is the average partial effect (APE). After computing the partial effect and plugging in the estimated parameters, we average the partial effects for each unit across the sample. Suppose that the sample average of *npvis* is 11.62. What is the APE of *npvis* on ln(*bwght*)?
8. Is *cigs* statistically significant at the 5% level against a two-sided alternative? Construct the hypotheses test in detail. Note that $prob(T > 1.96) = 0.025$ and $prob(T > 1.645) = 0.05$, where $prob(T > t_c)$ is the area to the right of t_c in a *t* distribution with a large enough degree of freedom.
9. Would you say that all the independent variables explain a lot of the variation in birth weight? Explain.

Let us change the unit of measurement of one of the independent variables, *cigs*. Define *packs* to be the number of packs of cigarettes smoked per day, that is, *packs* = *cigs*/20. We can obtain the estimated model:

$$\ln(bwght) = \tilde{\beta}_0 + \tilde{\beta}_1 packs + \tilde{\beta}_2 mage + \tilde{\beta}_3 meduc + \tilde{\beta}_4 feduc + \tilde{\beta}_5 npvis + \tilde{\beta}_6 npvis^2.$$

10. What will be the coefficients and standard errors on *packs* and on *mage*, respectively?
11. Instead of replacing *cigs* with *packs*, we add *packs* to the equation. What is the effect on the coefficients on *cigs* and *packs* if we include both *cigs* and *packs* in the same equation? Explain.

We would like to estimate the effect of parents' education on birth weight, after controlling for *cigs*, *mage*, *npvis*, and $npvis^2$. The estimates of the restricted model are given in column (2) of Table 1.

12. How should we proceed in testing the joint significance of the explanatory variables *meduc* and *feduc*? Construct the hypotheses test in detail.
13. From the given information, why are we unable to compute the *F* statistic for joint significance of *meduc* and *feduc*? Explain.

接次頁

Table 1

Dependent variable: ln(*bwght*)		
Independent variables	(1)	(2)
cigs	-0.0023	-0.0027
	(0.0011)	(0.0012)
mage	0.0008	0.0011
	(0.0011)	(0.0010)
meduc	-0.0018	
	(0.0030)	
feduc	0.0029	
	(0.0026)	
npvis	0.0120	0.0153
	(0.0037)	(0.0037)
$npvis^2$	-0.0002	-0.0003
	(0.0001)	(0.0001)
intercept	7.9752	7.9615
	(0.0012)	(0.0389)
Observations	1,625	1,656
R-squared	0.0173	0.0216
SSR (sum of squared residuals)	62.3072	65.3446

試題隨卷繳回