There are five questions in total. The table of t distribution is included.

1. (15 points) The standard deviation (or standard error) of the sampling distribution for the sample mean, $\bar{x}$, is equal to $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
   a) As the sample size is increased, what happens to the standard error of $\bar{x}$? Why is this property considered important?
   b) Suppose that a sample statistic has a standard error that is not a function of the sample size. In other words, the standard error remains constant as n changes. What would this imply about the statistic as an estimator of a population parameter?
   c) Suppose another unbiased estimator (call it A) of the population mean is a sample statistic with a standard error equal to $\sigma_A = \frac{\sigma}{\sqrt[3]{n}}$. Which of the sample statistics, $\bar{x}$ or A, is preferable as an estimator of the population mean? Why?

2. (18 points) A random sample of five pairs of observations were selected, one of each pair from a population with mean μ_1, the other from a population with μ_2. The data are shown as below (Note: The conditions that are required for small-sample inferences hold.)

Pair	Value from population 1	Value from population 2
1	28	22
2	31	27
3	24	20
4	30	27
5	22	20

   a) Test the null hypothesis $H_0: \mu_D = 0$ against $H_a: \mu_D \neq 0$, where $\mu_D = \mu_1 - \mu_2$. Use α=0.05.
   b) Another researcher erroneously treats the two samples to be drawn independently from two populations (that is, not pair by pair). Show his test for $H_0: \mu_D = 0$ against $H_a: \mu_D \neq 0$ (Use α=0.05).
   c) The result and statistical inference obtained in (a) and (b) are different. Explain why.

3. Suppose $income = \beta_0 + \beta_1(experience) + \beta_2(sex) + \varepsilon$, where
   income: monthly income (NTD thousand),
   experience: years,
   sex: dummy variable=1 for male, 0 otherwise.

   Based on data for 20 individuals, the OLS estimation results are below:

$$\widehat{income} = 23.96 + 2.93(experience) + 2.50(sex)$$

s.e (2.42) (0.74) (0.96) R^2=0.692

   a) (5 points) Why does this regression model need an error term?
   b) (5 points) How do you interpret the coefficient 2.50 ?
   c) (7 points) Would you reject the hypothesis that there is no difference in income between male and female ? Interpret the economic implication.

4. A random variable X is distributed uniformly between 1 and a, where a is an unknown parameter.
   a) (5 points) Write down the probability density function of X.
   b) (10 points) Compute E(X) and Var(X) as functions of a.
   c) (5 points) With one sample observation of 2.5, test H_0: a=11 against H_a: a<11 at the 10% significance level (i.e. α=0.1).
   d) (5 points) What is the *p-value* in (c)? (Hint: *p-value* is the smallest level of type I error for which the null hypothesis is rejected.)

5. A researcher collects data on household consumption (C), saving (S), and income (Y). It is known that for all observations $Y_i = C_i + S_i$. The researcher tries to estimate the consumption function with the OLS regression:

$$C_i = \alpha + \beta Y_i + e_{1i},$$

where e_{1i} is the regression error. Suppose the OLS result is: $C_i = 2 + 0.6Y_i$.
   a) (8 points) The researcher also uses the same data set to estimate the regression of: $C_i = \gamma + \delta S_i + e_{2i}$, where e_{2i} is the associated regression error. Compute the OLS estimates $\hat{\gamma}$ and $\hat{\delta}$.
   b) (7 points) "*The sum of squared residuals from the first regression* ($\sum \hat{e}_{1i}^2$) *is the same as the sum of squared residuals from the second regression* ($\sum \hat{e}_{2i}^2$)." Is this statement true or false? Explain.
   c) (10 points) Suppose in this sample $\sum Y_i$= 10 and $\sum Y_i^2$=30. A student erroneously assumes that α is zero and estimate β by $\tilde{\beta} = \sum Y_i C_i / \sum Y_i^2$. Prove that $\tilde{\beta}$ is not unbiased. Compute the bias in this sample.

注意：背面有試題

所別:經濟學系碩士班 科目：統計學

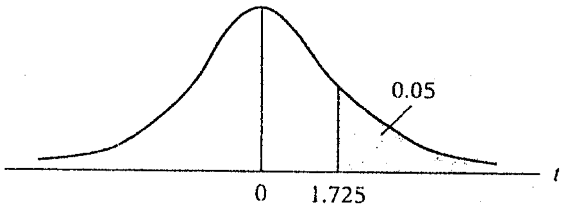
TABLE D.2 PERCENTAGE POINTS OF THE t DISTRIBUTION

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df = 20

$\Pr(|t| > 1.725) = 0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.