國立臺北大學114學年度碩士班一般入學考試試題

系（所）組別：經濟學系
科　　　目：統計學

☑可　☐不可使用計算機

Part I

※**Reference tables are at the end of the document for your reference.**

1. A factory produces 60% of products from Machine A and 40% from Machine B. Defect rates are 5% for A and 10% for B. If a randomly chosen product is defective, what is the probability that product is from Machine A? (5%)

2. Given X=[1, 2, 3] and Y=[2, 4, 3], find the correlation coefficient. (5%)

3. A sample of n=25 has mean=50 and standard deviation=10. Find a 95% confidence interval for the mean. (Assume normal distribution) (5%)

4. A Die is rolled, if the number is SIX, then roll a second die and add two numbers together. If the second die is SIX again, then roll a third die and add all numbers together. If the third die is SIX again, then roll another die, and so on. Until the final roll is not SIX, then stop and add all numbers together as the result. What is the expected value of the result? (5%)

5. A die is rolled 120 times, and the following frequencies are observed for each face:
Face 1: 15, Face 2: 18, Face 3: 20, Face 4: 22, Face 5: 25, Face 6: 20. Perform a chi-square goodness-of-fit test at the 0.05 significance level to determine if the die is fair. (5%)

6. Our basketball team is against "zone defense strategy" now.
When the basketball player is unguarded and shooting within the three-point line, there is an 90% chance of scoring 2 points.
When the basketball player is guarded and shooting within the three-point line, the chance of scoring 2 points drops to 30%.
Within the three-point line, the player will be guarded 70% of the time and unguarded 30% of the time.
When the basketball player is unguarded and shooting outside the three-point line, there is an 60% chance of scoring 3 points.
When the basketball player is guarded and shooting outside the three-point line, the chance of scoring 3 points drops to 20%.
Outside the three-point line, the player will be guarded 50% of the time and unguarded 50% of the time.
Calculate the expected points scored when the player attempts 2-point shoot or 3-point shoot. Which one is a better offense choice given "zone defense strategy" faced? (5%)

7. Given data X=[1, 2, 3, 4] and Y=[2, 3, 5, 7], find the regression line Y = a + b X. What are the values of a and b? (10%)

8. A survey claims that 50% of adults prefer coffee over tea. To test this claim, a random sample of 100 adults is selected, and 60 of them indicate a preference for coffee. Perform a hypothesis test at the 0.05 significance level to determine if the true proportion differs from the claimed value. (10%)

Appendix 1　Standard Normal Distribution

Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786

Appendix 2　Chi-square Distribution Table (right-tail probabilities)

	Probability					
d.f.	0.5	0.10	0.05	0.02	0.01	0.001
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

Part II

選擇題（每題5%，共50%）

1. In the Ordinary Least Squares (OLS) regression model:
   $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \epsilon_i$,
   where $i = 1, \ldots, n$, the OLS estimator is the Best Linear Unbiased Estimator (BLUE) if:
   (A) $E(\epsilon_i \mid x_{1i}, x_{2i}, \ldots, x_{ki}) = 0$ and $Var(\epsilon_i) = \sigma^2$ only
   (B) $x_{1i}, x_{2i}, \ldots, x_{ki}$ are linearly independent
   (C) The regressors$x_{1i}, x_{2i}, \ldots, x_{ki}$ are fixed, and ϵ_i satisfies the Gauss-Markov assumptions
   (D) $E(x_i\epsilon_i) = 0$ for all i

2. Consider the model:
   $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$, and $E(x_{1i}\epsilon_i) \neq 0$,
   where x_{1i} is correlated with the error term. This correlation arises could be due to:
   (A) Measurement error in y_i
   (B) Multicollinearity among regressors
   (C) Endogeneity in x_{1i}
   (D) Nonlinearity in the relationship between x_{1i} and y_i

3. Under the setup in Question 2, the OLS estimator for β_1 has the following property:
   (A) $E(\hat{\beta}_1) = \beta_1 + Cov(x_{1i}, \varepsilon) / Var(x_{1i})$
   (B) $E(\hat{\beta}_1) = \beta_1 + Cov(\epsilon_i, y_i) / Var(y_i)$
   (C) $E(\hat{\beta}_1) = \beta_1 + Cov(x_{1i}, y_i) / Var(y_i)$
   (D) $E(\hat{\beta}_1) = \beta_1$

4. When consider the instrumental variable estimator, a valid instrumental variable z_i must satisfy the following conditions:
   (A) $Cov(z_i, x_{1i}) \neq 0$, and $Cov(z_i, \epsilon_i) = 0$
   (B) $Cov(z_i, \epsilon_i) = 0$, and $Var(z_i) \neq 0$
   (C) $Cov(z_i\ x_{1i}) \neq 0$, and $Cov(z_i, x_{1i}) = 0$
   (D) $Cov(z_i, \epsilon_i) = 0$, and $E(z_i) = 0$

5. In a linear regression model:
   $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$,
   under the assumptions:
   a. $E(\epsilon_i \mid x_{1i}, x_{2i}) = 0$,
   b. $Var(\epsilon_i \mid x_{1i}, x_{2i}) = \sigma^2$,
   c. $Cov(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$,
   the OLS estimator minimizes:
   (A) The standard error among all estimators
   (B) The variance among all linear unbiased estimators
   (C) The variance among all nonlinear estimators
   (D) The bias among all consistent estimators

6. The coefficient of determination R^2 is defined as:
   $R^2 = 1 - \Sigma_i \hat{\epsilon}_i^2 / \Sigma_i (y_i - \bar{y})^2$,
   where $\hat{\epsilon}_i$ are the residuals and $\bar{y}$ is the sample mean of y_i.
   When additional regressors are included in the model, R^2 will:
   (A) Decrease if the regressors are irrelevant
   (B) Increase or stay constant, as $\Sigma_i \hat{\epsilon}_i^2$ decreases
   (C) Increase or decrease depending on the sample size
   (D) Decrease due to the penalty for overfitting

國立臺北大學 114 學年度碩士班一般入學考試試題

系（所）組別：經濟學系
科　　　目：統計學

☑可 ☐不可使用計算機

7. If the regressors $x_{1i}, x_{2i}, \dots, x_{ki}$ in the model:
   $y_i = \beta^0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i$,
   are perfectly collinear, there exists a non-zero vector such that:
   $\sum_{j=1}^{k} \lambda_j x_{ji} = 0$ for all i.
   This causes:
   (A) The OLS estimator is undefined
   (B) The OLS estimator to be unbiased but inefficient
   (C) The OLS estimator to have infinite variance
   (D) No effect on the estimator, but it affects hypothesis testing

8. In the true model:
   $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$,
   if x_{2i} is omitted, the estimated model becomes:
   $y_i = \beta_0 + \beta_1 x_{1i} + u_i$,
   where $u_i = \beta_2 x_{2i} + \epsilon_i$.
   If x_{1i} and x_{2i} are correlated, the bias in $\hat{\beta}_1$ from OLS is given by:
   (A) $\beta_2 Cov(x_{1i}, x_{2i}) Var(x_{1i})$
   (B) $\beta_2 Cov(x_{2i}, u_i) Var(x_{2i})$
   (C) $Cov(x_{1i}, u_i)$
   (D) $E(u_i^2)$

9. In the model:
   $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$,
   if $Var(\epsilon_i | x_{1i}) = \sigma_i^2 \neq \sigma^2$, then:
   (A) The OLS estimator is unbiased, but its standard errors are inconsistent
   (B) The OLS estimator is biased and inefficient
   (C) The OLS estimator is unbiased but inefficient, and standard errors are inconsistent
   (D) The OLS estimator is consistent, but hypothesis tests are invalid

10. Consider the following simple linear regression model:
    $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$,
    where $n = 10$. After using the OLS, the following results are obtained: $\hat{\beta}_1 = 1.22$ and the standard error of $\hat{\beta}_1$ is 0.62. Please conclude that: (Critical values for standard Normal distribution: $z_{0.05/2} = 1.96$, $z_{0.1/2} = 1.645, z_{0.2/2} = 1.280$, and for t-distribution: $t_{0.05/2,8} = 2.306$, $t_{0.1/2,8} = 1.860, t_{0.2/2,8} = 1.397$)
    (A) The t-statistic is 3.17, and $\hat{\beta}_1$ is statistically significantly different from zero at 20%
    (B) The t-statistic is 1.97, and $\hat{\beta}_1$ is statistically significantly different from zero at 5% significance level
    (C) The t-statistic is 3.17, and $\hat{\beta}_1$ is statistically significantly different from zero at 1% significance level
    (D) The t-statistic is 1.97, and $\hat{\beta}_1$ is statistically significantly different from zero at 10% significance level