國立臺灣大學 114 學年度碩士班招生考試試題

第一部份：計算題 (20%)

※ Show the detailed calculation process for all questions.

1. (6%) Find the interval of convergence of $\sum_{n=1}^{\infty} \frac{(x-2)^{2n}}{4^n}$.
2. (6%) Find the slope of the tangent line to the graph of $y^4 + 3y - 4x^3 = 5x + 1$ at the point P(1,-2).
3. (8%) Evaluate $\int_0^4 \int_{x/2}^2 e^{y^2} dy dx$.

第二部份：簡答題 (80%)。請依題號順序作答，每小題作答字數不得超過 5 行，可以用中文作答。

Question I (40 points; 10 points each)

We aim to analyze the effect of a newly constructed domed stadium on housing values. The decision to build the dome was announced in 2001, and construction began in 2005, with the dome expected to be in operation soon after construction. There is a rumor that houses near the dome may be devalued, but wealthier residents might use their political influence to avoid having the dome built in their neighborhoods. We have data on prices of houses that sold in 2001 and another independent random sample on those sold in 2005. All house prices have been adjusted for inflation. Let *rprice* denote the house price in real terms.

4. One hypothesis is that the price of houses located near the dome would fall relative to the price of more distant houses. For illustration, we define a house as "near the dome" if it is within three miles. To estimate the effect of the dome, we use data from 2005 and run the regression:
$$rprice = \alpha_0 + \alpha_1 neardome + u_1,$$
where $neardome$ denotes a dummy variable equal to 1 if the house is near the dome, and 0 otherwise. If the coefficient on $neardome$ is negative and statistically significant, does this imply that the siting of the dome caused the lower housing values? Discuss potential issues with this approach, if any.
5. To examine whether house prices changed over time in response to the dome, we use data on houses located near the dome and run the regression:
$$rprice = \beta_0 + \beta_1 y05 + u_2,$$
where $y05$ is a dummy variable equal to 1 if the house was sold in 2005, and 0 if sold in 2001. If the coefficient on $y05$ is negative and statistically significant, does this imply that building the new dome depressed housing values? Discuss potential issues with this approach, if any.
6. To compare changes in housing values for houses near the dome with those farther away, we use the entire data set and run the regression:
$$\Delta rprice = \gamma_0 + \gamma_1 neardome + u_3,$$
where $\Delta rprice$ is the change in the price of a house between 2001 and 2005. If the coefficient on $neardome$ is negative and statistically significant, does this imply that houses near the dome fall in value more than houses far from the dome? Discuss potential issues with this approach, if any.
7. We propose including various housing characteristics (e.g., house area in square feet, number of rooms, and number of baths) in our analysis of the dome siting. What are the advantages and potential drawbacks of incorporating these variables in our regression models?

Question II (40 points; 10 points each)

A widely used experimental design in business is the single-factor experiment with two levels, where customers in the control group receive the current version of a product or service, and those in the test group receive a new version. If customers are randomly assigned to the two groups and the response variable is quantitative, we can use a two-sample t-test to determine whether the means of the two groups are equal.

8. Suppose now we expand our single-factor experiment to include more levels. What issues might arise when doing pairwise t-tests for all possible treatment means? Explain.

Supermarkets often place similar types of cereal on the same shelf. Suppose a researcher aims to investigate whether the sugar content varies by shelf. The shelf placement for 77 cereals was recorded as their sugar content.

Shelf	Number	Mean	StdDev
1	20	4.80000	4.57223
2	21	9.61905	4.12888
3	36	6.52778	3.83582

The researcher then applies an Analysis of Variance (ANOVA). The test statistic, called F-statistic, compares two quantities that measure variation. The numerator measures the variation *between* the groups (treatments) and is called the Mean Square due to Treatments (MST). The denominator measures the variation *within* the groups, and is called the Mean Square due to Error (MSE). The MST has $k-1$ degrees of freedom (df) because there are k groups, and the MSE has $N-k$ degrees of freedom, where N is the total number of observations. The partial ANOVA table shows the components of the calculation of the F-statistic.

ANOVA					
Source of Variation	Sum of Squares (SS)	df	Mean Square (MS)	F-statistic	P-value
Shelf (Between Groups)	248.4079	#	#	#	0.0012
Error (Within Groups)	1253.1246	#	#		
Total	1501.5326	#			

9. State the null and alternative hypotheses. Compute the F-statistic using the values in the ANOVA table. What does the ANOVA table say about the null hypothesis? Explain your conclusion in terms of sugar content and shelf placement.
10. Explain the conceptual framework of how the ANOVA works. In particular, why does comparing the *variation* between the groups to the *variation* within the groups allow us to determine whether there is a significant difference in the *mean* sugar content?

接次頁

國立臺灣大學 114 學年度碩士班招生考試試題

The researcher would hardly be satisfied with the report since the *F*-test fails to specify which shelves have cereals with higher sugar content and by how much. To address this, multiple comparisons can be performed to compare several pairs of group means. One such method is called the Bonferroni method, which adjusts the tests and confidence intervals to allow for making many comparisons. The result is a wider margin of error (called the minimum significant difference, or MSD) found by replacing the critical *t*-value with a slightly larger number.

Dependent Variable: SUGARS							
	(I) SHELF	(J) SHELF	Mean Difference (I - J)	Std. Error	P-value	95% Confidence Interval	
Bonferroni						Lower Bound	Upper Bound
	1	2	-1.728	1.2857	0.001	-7.969	-1.670
		3	-4.819	1.1476	0.409	-4.539	1.084
	2	1	4.819	1.2857	0.001	1.670	7.969
		3	3.091	1.1299	0.023	0.323	5.859
	3	1	1.728	1.1476	0.409	-1.084	4.539
		2	-3.091	1.1299	0.023	-5.859	-0.323

11. To check for significant differences between the shelf means, we can use a Bonferroni test, the results of which are shown above. For each pair of shelves, the difference is shown along with its standard error and significance level. Identify which pairs of shelves have significant differences in sugar content based on the table. For instance, can we determine that cereals on shelf 1 have a different mean sugar content than cereals on shelf 2? What can we conclude from these results?

試題隨卷繳回