國立臺灣大學97學年度碩士班招生考試試題

※ 注意：請於試卷上「非選擇題作答區」依序作答，並應註明作答之部份及題號。

I.單選題，每題 5 分。

1. A recent article in the paper claims that business ethics are at an all-time low. Reporting on a recent sample, the paper claims that 41% of all employees believe their company president possesses low ethical standards. Suppose 20 of a company's employees are randomly and independently sampled and asked if they thought their company president had low ethical standards and their years of experience at the company. Could the probability distribution for the number of years of experience be modelled by a binomial probability distribution?
   A) No, the employees would not be considered independent in the present sample.
   B) Yes, the sample size is n = 20.
   C) No, a binomial distribution requires only two possible outcomes for each experimental unit sampled.
   D) Yes, the sample is a random and independent sample.

2. Which of the following statements is not a property of the normal curve?
   A) mound-shaped (or bell shaped)
   B) $P(\mu - \sigma < x < \mu + \sigma) \approx .95$
   C) symmetric about μ
   D) $P(\mu - 3\sigma < x < \mu + 3\sigma) \approx .997$

3. The registrar's office at State University would like to estimate the average commute time and determine a 95% confidence interval for the average commute time of evening university students from their usual starting point to campus. A member of the staff randomly chooses a parking lot and selects the first 100 evening students who park in the chosen lot starting at 5:00 p.m. The confidence interval is
   A) not meaningful because the sampling distribution of the sample mean is not normal.
   B) meaningful because the sample size exceeds 30 and the central limit theorem ensures normality of the sampling distribution of the sample mean.
   C) not meaningful because of the lack of random sampling.
   D) meaningful because the sample is representative of the population.

4. We never conclude "Accept H_0" in a test of hypothesis. This is because:
   A) The rejection region is not known.
   B) α is the probability of a Type I error.
   C) The p-value is not small enough.
   D) β = p(Type II error) is not known.

見背面

5. Data was collected from CEO's of companies within the consumer products industry sector. The following ASP printout compares the mean return-to-pay ratios between CEO's in the low tech industry with CEO's in the consumer products industry.

HYPOTHESIS: MEAN X = MEAN Y

SAMPLES SELECTED FROM RETURN

industry 1	(low tech)	(NUMBER = 15)
industry 3	(consumer products)	(NUMBER = 15)

X = industry1
Y = industry3

SAMPLE MEAN OF X =	157.286
SAMPLE VARIANCE OF X =	1563.45
SAMPLE SIZE OF X =	14
SAMPLE MEAN OF Y =	217.583
SAMPLE VARIANCE OF Y =	1601.54
SAMPLE SIZE OF Y =	12
MEAN X - MEAN Y =	-60.2976
t =	-4.23468
P-VALUE =	0.000290753
P-VALUE/2 =	0.000145377
SD. ERROR =	14.239

Using the printout, which assumption is not necessary for the test above to be valid?

A) The population variances are equal.

B) The population means are equal.

C) The samples were randomly and independently selected.

D) Both populations have approximate normal distributions.

6. Which of the following probabilities for the sample points A, B, and C could be true if A, B, and C are the only sample points in an experiment?

A) P(A) = 1/10, P(B) = 1/6, P(C) = 1/2

B) P(A) = 1/8, P(B) = 1/8, P(C) = 1/8

C) P(A) = -1/4, P(B) = 1/2, P(C) = 3/4

D) P(A) = 0, P(B) = 1/10, P(C) = 9/10

7. Suppose a 95% confidence interval for μ turns out to be (190, 260). To make more useful inferences from the data, it is desired to reduce the width of the confidence interval. Which of the following will result in a reduced interval width?

A) Increase the confidence level.

B) Increase the sample size.

C) Decrease the sample size and decrease the confidence level.

D) All of the choices will result in a reduced interval width.

8. As an aid in the establishment of personnel requirements, the director of a hospital wishes to estimate the mean number of people who are admitted to the emergency room during a 24-hour period. The director randomly selects 81 different 24-hour periods and determines the number of admissions for each. For this sample, $\bar{x} = 18.2$ and $s^2 = 25$. Estimate the mean number of admissions per 24-hour period with a 95% confidence interval.

A) 18.2 ± 1.089

B) 18.2 ± .528

C) 18.2 ± .121

D) 18.2 ± 5.444

接次頁

9. A computer package was used to generate the following printout for estimating the sale price of homes in a particular neighborhood.

X = sale_price

SAMPLE MEAN OF X =	46,600
SAMPLE STANDARD DEV =	13,747
SAMPLE SIZE OF X =	15
CONFIDENCE =	98
UPPER LIMIT =	55,913.8
SAMPLE MEAN OF X =	46,600
LOWER LIMIT =	37,286.2

Which of the following is a practical interpretation of the interval above?

A) We are 98% confident that the mean sale price of all homes in this neighborhood fall between $37,286.20 and $55,913.80.

B) 98% of the homes in this neighborhood have sale prices that fall between $37,286.20 and $55,913.80.

C) We are 98% confident that the true sale price of all homes in this neighborhood fall between $37,286.20 and $55,913.80.

D) All are correct practical interpretations of this interval.

10. A statistician in a mail order house wished to estimate how many mail order parcels prepared last Friday by the wrapping and packaging department were improperly packaged. He took a random sample of 2% of all the parcels prepared that day, had the sample parcels unwrapped and inspected, and found that 26 were improperly packaged. He reported that 1300 mail order parcels were improperly packaged on Friday. The statistician's report utilized

A) a two-sided confidence interval.

B) a point estimate.

C) a one-sided confidence interval.

D) a population parameter.

II. 單選題， 每題 5 分

Each year U.S. News & World Report conducts its "Survey of America's Best Graduate and Professional Schools." For the top 25 business schools in 1991, three variables were measured: (1) GMAT score for the typical incoming student; (2) student acceptance rate; and (3) starting salary of the typical graduating student. The academic advisor wants to predict the typical starting salary of a graduate at a top business school using GMAT score of the school as a predictor variable. A simple regression of SALARY versus GMAT using the 25 data points are shown below:

	School	GMAT	Acc. Rate	Salary
1.	Harvard	644	15.0%	$ 63,000
2.	Stanford	665	10.2	60,000
3.	Penn	644	19.4	55,000
4.	Northwestern	640	22.6	54,000
5.	MIT	650	21.3	57,000
6.	Chicago	632	30.0	55,269
7.	Duke	630	18.2	53,300
8.	Dartmouth	649	13.4	52,000
9.	Virginia	630	23.0	55,269
10.	Michigan	620	32.4	53.300
11.	Columbia	635	37.1	52,000
12.	Cornell	648	14.9	50,700
13.	CMU	630	31.2	52,050
14.	UNC	625	15.4	50,800
15.	Cal-Berkeley	634	24.7	50,000
16.	UCLA	640	20.7	51,494
17.	Texas	612	28.1	43,985
18.	Indiana	600	29.0	44,119
19.	NYU	610	35.0	53,161
20.	Purdue	595	26.8	43,500
21.	USC	610	31.9	49,080
22.	Pittsburgh	605	33.0	43,500
23.	Georgetown	617	31.7	45,156
24.	Maryland	593	28.1	42,925
25.	Rochester	605	35.9	44,499

The academic advisor wants to predict the typical starting salary of a graduate at a top business school using GMAT score of the school as a predictor variable. A simple linear regression of SALARY versus GMAT using the 25 data points in the table are shown below.

$\beta_0 = -92040$ $\hat{\beta}_1 = 228$ $s = 3213$ $R^2 = .66$ $r = .81$ $df = 23$ $t = 6.67$

1. For the situation above, give a practical interpretation of $\hat{\beta}_0 = -92040$.

A) The value has no practical interpretation since a GMAT of 0 is nonsensical and outside the range of the sample data.

B) We estimate the base SALARY of graduates of a top business school to be $ -92,040.

C) We expect to predict SALARY to within 2(92040) = $184,080 of its true value using GMAT in a straight-line model.

D) We estimate SALARY to decrease $92,040 for every 1-point increase in GMAT.

接次頁

2. For the situation above, give a practical interpretation of $s = 3213$.
   A) Our predicted value of SALARY will equal 2(3213) = $6,426 for any value of GMAT.
   B) We estimate SALARY to increase $3,213 for every 1-point increase in GMAT.
   C) We expect the predicted SALARY to deviate from actual SALARY by at least 2(3213) = $6,426 using GMAT in a straight-line model.
   D) We expect to predict SALARY to within 2(3213) = $6,426 of its true value using GMAT in a straight-line model.

3. For the situation above, give a practical interpretation of $R^2 = .66$.
   A) 66% of the sample variation in SALARY can be explained by using GMAT in a straight-line model.
   B) We expect to predict SALARY to within $2(\sqrt{.66})$ = $1,620 of its true value using GMAT in a straight-line model.
   C) We estimate SALARY to increase $.66 for every 1-point increase in GMAT.
   D) We can predict SALARY correctly 66% of the time using GMAT in a straight-line model.

4. A 95% prediction interval for SALARY when GMAT = 600 is ($37,915, $51,948). Interpret this interval for the situation above.
   A) We are 95% confident that the SALARY of a top business school graduate will fall between $37,915 and $51,984.
   B) We are 95% confident that the SALARY of a top business school graduate with a GMAT of 600 will fall between $37,915 and $51,984.
   C) We are 95% confident that the mean SALARY of all top business school graduates with GMATs of 600 will fall between $37,915 and $51,984.
   D) We are 95% confident that the increase in SALARY for a 600-point increase in GMAT will fall between $37,915 and $51,984.

5. For the situation above, which of the following is *not* an assumption required for the simple linear regression analysis to be valid?
   A) The errors of predicting SALARY have a mean of 0.
   B) The errors of predicting SALARY have a variance that is constant for any given value of GMAT.
   C) The errors of predicting SALARY are normally distributed.
   D) SALARY is independent of GMAT.

見背面

III. 單選題， 每題5分

1. In any production process in which one or more workers are engaged in a variety of tasks, the total time spent in production varies as a function of the size of the workpool and the level of output of the various activities. In a large metropolitan department store, it is believed that the number of man-hours worked (y) per day by the clerical staff depends on the number of pieces of mail processed per day (x_2) and the number of checks cashed per day (x_1). Data collected for n = 20 working days were used to fit the model:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Suppose a 95% confidence interval for β_1 is (-.02, -.01). Interpret this result.

A) We are 95% confident that the mean number of man-hours worked per day falls between .01 and .02.

B) We are 95% confident that the daily number of man-hours worked decreases by an amount between .01 and .02 man-hour for each extra piece of mail sorted, holding number of checks cashed constant.

C) We are 95% confident that the daily number of man-hours worked decreases by an amount between .01 and .02 man-hour for each extra check cashed, holding number of pieces of mail sorted constant.

D) We expect to predict most (approximately 95%) of the daily number of man-hours worked to within a range of .01 to .02 of its true value.

2. A magazine for car enthusiasts conducts comprehensive road tests on all new car models and reports the results in its latest issue. One variable measured is the time it takes a car to accelerate from 0 to 60 miles per hour. To model acceleration time, a regression analysis is conducted on the following data collected for a random sample of 129 new cars:

TIME60: y = Elapsed time (in seconds) from 0 mph to 60 mph
MAX x_1 = Maximum speed attained (miles per hour)
GRIP: x_2 = 1 good road-holding grip during cornering, 0 if poor grip

Write a model which proposes a quadratic relationship between elapsed acceleration time and the maximum speed attained, but allows different curves for good and poor road-holding grips.

A) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

B) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 (x_1)^2 + \beta_5 (x_2)^2 x_2$

C) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 (x_1)^2 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 (x_1)^2 x_2$

D) $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 (x_1)^2 + \beta_3 x_2$

接次頁

3. A firm has developed a new type of light bulb and is interested in evaluating its performance in order to help decide whether to market the bulb. It is known that the level of light output of the bulb depends on the cleanliness of the bulb's surface area. The data below gives the drop in light output y (measured as a percent of original output), cleanliness of the bulb surface (clean or dirty), and color (white, yellow, or blue) for n = 10 randomly selected bulbs which had been in use at least 500 hours.

Bulb	Drop in Light Output y (%)	Bulb Surface	Color
1	20	Clean	White
2	30	Clean	Blue
3	50	Dirty	White
4	25	Clean	Yellow
5	90	Dirty	Yellow
6	80	Dirty	White
7	30	Clean	White
8	50	Dirty	Blue
9	35	Clean	Yellow
10	40	Clean	Blue

The following model is proposed to relate the mean drop in light output to bulb color:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where $x_1 = 1$ if white, 0 otherwise;
$x_2 = 1$ if yellow, 0 otherwise

Use the data in the table to estimate the value of β_0. (Hint: First, interpret the value of β_0. Then, remember that the sample mean is used to estimate a population mean.)

A) 40 B) 50 C) 5 D) 45

4. The Bureau of Census collects and summarizes data on voters during each presidential election. Suppose the Bureau is interested in predicting the percentage of eligible voters in the state who will vote in the next presidential election, y, as a function of the following independent variable: (The levels of the qualitative variable are given in parentheses.)

Location of State (North, South, East, or West)

Consider the model relating mean voter turnout, E(y), to location:

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x_2 + \beta_3 x_3$$

where $x_1 = 1$ if North, 0 if not
$x_2 = 1$ if South, 0 if not
$x_3 = 1$ if East, 0 if not (Base level = West)

The p-value for testing $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ is .14. Interpret this result.

A) Reject the model since it only explains 14% of the variation.

B) Reject H_0 at $\alpha = .10$; the model is useful for predicting voter turnout.

C) Do not reject H_0 at $\alpha = .10$; there is no evidence of differences among the mean voter turnouts for states in the North, South, East, or West.

D) Reject H_0 at $\alpha = .10$; there is evidence of differences among the mean voter turnouts for states in the North, South, East, or West.

見背面

5、 In Hawaii, condemnation proceedings are under way to enable private citizens to own the property that their homes are built on. Prior to 1980, only estates were permitted to own land, and homeowners leased the land from the estate. In order to comply with the new law, a large Hawaiian estate wants to use regression analysis to estimate the fair market value of the land. Each of the following three models were fit to data collected for n = 20 properties (10 of which are located near a cove):

Model 1: $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 (x_1)^2 + \beta_5 (x_1)^2 x_2$
Model 2: $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$
Model 3: $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

where y = Sale price of property ($ thousands)
x_1 = Size of property in square feet (thousands)
x_2 = 1 if property near cove, 0 if not

The SSE's and MSE's for the three models are given below. Use this information to calculate the test statistic appropriate for determining whether the relationship between sale price (y) and size (x_1) is linear or curvilinear.

	SSE	MSE
Model 1	1280	91
Model 2	3140	196
Model 3	7225	425

A) 56.3　　B) 10.2　　C) .4　　D) 21.8

試題隨卷繳回