中國醫藥大學九十九學年度研究所碩士班暨碩士在職專班入學招生考試試題

所別：生物統計研究所碩士班

科目：統計學　【此科考生可攜帶電子計算機應試】

考生注意：答案必須寫在答案卷上，否則不予計分。

1. (A) What is the standard deviation of a sample mean? Please compare this standard deviation with the standard deviation of the population. (5 pts)
   (B) What happens to the amount of sampling variability among a set of sample means $\{\bar{x}_1, \bar{x}_2, \ldots\}$ as the size of the samples increases? (5 pts)

2. A study of patients with diabetes was conducted to investigate the effects of cigarette smoking on renal complication. The basline measures of systolic blood pressure across four different subgroups, nonsmokers, current smokers, ex-smokers, and tobacco chewers, are measured. The revelant data are shown below with means and standard deviations expressed in mmHg. Assume that systolic blood pressure is normally distributed.

	n	$\bar{x}$	s
Nonsmokers	20	115	13.4
Current Smokers	50	114	10.1
Ex-smokers	25	118	11.6
Tobacco Chewers	9	126	12.2

(A) Please complete the ANOVA table: (10 pts)

Sourece	**SS**	**df**	**MS**	**F**	**P**
Between groups					
Within groups					
Total					

(B) At the 0.05 level of significance, test the null hypothesis of identical mean systolic blood pressures. What do you conclude? (5 pts)
(C) What is your next step if you find that the population means are not all equal? (3 pts)

3. In a study of low birth weight infant, systolic blood pressure (sbp) and gestational age (gestage, in weeks) measurements are recorded for a sample of 12 infants in the following table.

Infant	1	2	3	4	5	6	7	8	9	10	11	12
Sbp	43	51	31	31	49	51	33	64	36	39	53	34
Gestage	29	31	25	27	29	28	27	33	28	28	31	29

(A) Please use a measure to show whether there is any evidence of a linear relationship between sbp and gestational age. What do you conclude? (4 pts)
(B) Using sbp as the response and gestational age as the explanatory variable, the SAS output displaying the simple linear regression is shown below. At the 0.05 level of significance, test the null hypothesis that the true population slope b is equal to 0. What do you conclude? (3 pts)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	917.76244	917.76244	28.58	0.0003
Error	10	321.15423	32.11542		
Corrected Total	11	1238.91667			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-79.95025	23.04220	-3.47	0.0060
gestage	1	4.27363	0.79945	5.35	0.0003

(C) Construct a 95% confidence interval for the population slope. (4 pts)
(D) Compute the coefficients of determination (R^2). (4 pts)
(E) What is the estimated mean systolic blood pressure for the population of low birth weight infants whose gestational age is 30 weeks? (2 pts)

4. In a study of observer variability in the assessment of cervical smears (子宮頸抹片), 3325 slides were screened for the presence or absence of abnormal cells. Each slide was screened by a particular observer and then rescreened six months later by the same observer. The results can be displayed in the following two manners.
(A) Please select the appropriate table to analyze these data and explain why you use it. (5 pts)
(B) At the 0.05 level of significance, does there appear to be an association between time of screening and diagnosis? (10 pts)

First Screening	Second Screening		Total
	Present	Absence	
Present	1763	489	2252
Absence	403	670	1073
Total	2166	1159	3325

Abnormal Cells	Screening		Total
	First	Second	
Present	2252	2166	4418
Absence	1073	1159	2232
Total	3325	3325	6650

5. In Taiwan, the distribution of birth weights for full term infants whose gestational age is 40 weeks is approximately normal with mean m=3500 grams and standard deviation s=430 grams. An investigator plans to conduct a study to determine whether the birth weights of full-term babies whose mothers smoked throughout pregnancy have the same mean. If the true mean birth weight for the infants whose mothers smoked is as low as 3200 grams (or as high as 3800 grams), the investigator wants to risk only a 10% chance of failing to detect this difference. A two-sided test conducted as the 0.05 level of significance will be used. What sample size is needed for this study? (10 pts)

6. The following data come from a smoking data. Seven subjects were each required to smoke a single cigarette, and then cotinine levels were recorded for all individuals 12, and 24 hours after smoking. It is believed that the mean cotinine level 24 hours after smoking (m_{24}) must be lower than the mean cotinine level 12 hours after smoking (m_{12}).

Subject	Cotinine Levels (nmol/l)	
	After 12 Hours	After 24 Hours
1	73	24
2	58	27
3	67	49
4	93	59

5	33	0
6	18	11
7	147	43

(A) What are the appropriate null and alternative hypotheses for testing this belief? (5 pts)
(B) Conduct the test at the 0.05 level of significance. What do you conclude? (5 pts)
(C) Construct a one-sided 95% confidence interval for the true difference in population means m_{12}- m_{24}. (5 pts)

7. A study was conducted to investigate the effectiveness of the drug (RU486) for terminating early pregnancy. 488 women were administered RU486 followed 48 hours later by a single dose of a second drug. In 473 of these women, the pregnancy was terminated.
(A) Estimate the proportion of successfully terminated early pregnancies among women using the described treatment regimen. (5 pts)
(B) Construct a 95% confidence interval for the true population proportion p. (5 pts)
(C) Interpret this confidence interval. (5 pts)

Appendix:

1. Percentiles of standard normal distribution for area in the upper tail

Area in the upper tail					
0.1	0.05	0.025	0.01	0.005	0.0005
1.28	1.64	1.96	2.33	2.58	3.29

2. Percentiles of the t distribution for area in the upper tail

Area in the upper tail				
df	**0.05**	**0.025**	**0.01**	**0.005**
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169

3. Percentiles of the chi-square distribution

Area in the upper tail				
df	**0.05**	**0.025**	**0.01**	**0.001**
1	3.84	5.02	6.63	10.83
2	5.99	7.38	9.21	13.82
3	7.81	9.35	11.34	16.27
4	9.49	11.14	13.28	18.47

4. Percentiles of the F distribution in upper tail =0.05

	Numerator Degrees of Freedom (df)					
Denominator df		**3**	**4**	**5**	**6**	**7**
	60	2.76	2.53	2.37	2.25	2.17
	80	2.72	2.49	2.33	2.21	2.13
	100	2.70	2.46	2.31	2.19	2.10
	104	2.69	2.46	2.30	2.19	2.10
	120	2.68	2.45	2.29	2.18	2.09