（考試時間 2 小時）

1. (10%) A new test has been devised for detecting a particular type of cancer. If the test is applied to a person who has this type of cancer, the probability that the person will have a positive reaction is 0.95 and the probability that the person will have a negative reaction is 0.05. If the test is applied to a person who does not have this type of cancer, the probability that this person will have a positive reaction is 0.05 and the probability that the person will have a negative reaction is 0.95. Suppose that in the general population, one person out of every 100,000 people has this type of cancer. If a person selected at random has a positive reaction to the test, what is the probability that he has this type of cancer?

2. (10%) Bus tickets in a certain city contain four digits, U, V, W, and X. Each of these numbers is equally likely to be any of the 10 digits 0, 1, …, 9, and the four numbers are chosen independently. A bus rider is said to be lucky if $U + V = W + X$. What proportion of the riders is lucky?

3. (10%) Explain the famous "Simpson's Paradox". Give an example to explain how it can happen.

4. (10%) This question may have one or two incorrect statements. Identify them and state your reasons why they are incorrect.
   About multiple linear regression models:
   (a) R^2 never decreases as more explanatory variables are added to a model.
   (b) The multiple correlation coefficient, multiple R, measures the strength and direction of linear association between the observed Y-values and the fitted Y-values from the model.
   (c) If an explanatory variable has the largest coefficient value, then this explanatory variable has the largest impact on the response variable.
   (d) When comparing two multiple linear regression models on the same data set, if the residual standard error of model A is larger than that of model B, then the R^2 of model A is smaller.
   (e) Adjusted R^2 will increase as R^2 increases.

試題請隨卷繳回

5. (10%) A company constructs towers for power lines. The project manager for the company has selected a random sample of 10 completed tower installations at various locations and has recorded the time required to complete the construction of each tower.
   (a) In order to use a t-based confidence interval for the population mean time required to complete the construction of a tower, what must be true concerning the distribution of the population of all construction time from which this sample of times was selected?
   (b) The manager decides to construct a confidence interval for the median of the population of construction times using the interval determined by $x_{(2)}$ and $x_{(9)}$; that is, the lowest and highest construction times in the sample will be dropped. Determine the coverage probability of this interval. Round your answer to 4 decimal places.

6. (10%) Let X and Y be continuous random variables with density function f(x,y) = xy for $0 \leqq x \leqq y \leqq 1$ and f(x,y)=0 otherwise. Please show $P(XY \geqq 1/2)$.

7. (10%) Let X be uniformly distributed on the interval [1, 10]. Please find
   (a) 1/E(X)
   (b) E(1/X)

8. (10%) Assume the time to failure for a type of unit follows an exponentially distribution with mean of ten hours.
   (a) Find the probability that one unit lasts less than 4 hours.
   (b) If the time to failure follows the same distribution but the mean reduces to two hours. Let w be the first failure in the pair. Find E(w).

9. (10%) The computer output for the multiple regression model
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$
is shown below. However, because of a printer malfunction some of the results are not shown. These are indicated by the boldface letters a to i. Fill in the missing results (up to three decimal places).

Predictor	*Coef*	*StDev*	*T*
Constant	***a***	6.15	4.11
x_1	3.51	***b***	1.25
x_2	–0.71	0.30	***c***

S = ***d*** R–Sq = ***e***

ANALYSIS OF VARIANCE

Source of Variation	*df*	*SS*	*MS*	*F*
Regression	2	412	***g***	***i***
Error	37	***f***	***h***	
Total	39	974		

10. (10%) Ten business people who fly frequently from Chicago to New York were asked to rank four airlines in terms of the quality of service. The people assigned scores using a 5-point Likert scale where: 1 = bad, 2 = poor, 3 = average, 4 = good, and 5 = excellent. The results are shown below:

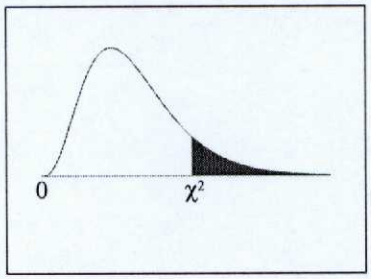
	Airline			
Person	A	B	C	D
1	1	3	5	2
2	5	4	2	1
3	2	5	3	2
4	4	2	4	1
5	3	3	1	5
6	4	4	5	3
7	3	4	1	4
8	2	5	2	1
9	5	3	4	2
10	5	5	4	3

(a) Which test is appropriate if you want to compare the quality of service of the four airlines?

(b) Can we conclude at the 5% significance level that there are differences in service quality among the four airlines?

(c) Using the appropriate statistical table, what statement can be made about the p-value for the test in the previous question? Explain how to use the p-value for testing the hypotheses.

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169