※ 考生請注意：本試題 ☑可 ☐不可 使用計算機

1. (15%) Consider a random variable X which has p.d.f. (or p.m.f.) $f_1(x)$ with probability p_1 and p.d.f. (or p.m.f.) $f_2(x)$ with probability p_2, where $p_1 + p_2 = 1$. We can think of first observing a Bernoulli random variable Y which equals 1 with probability p_1 and 2 with probability p_2. If Y=1, then $X = X_1 \sim f_1(x)$, and if Y=2, then $X = X_2 \sim f_2(x)$. We say that X has a **mixture distribution.**

a. (5%) Show that the p.d.f. (or p.m.f.) of X is

$$f(x) = p_1 f_1(x) + p_2 f_2(x)$$

b. (5%) Let μ_1 and μ_2 be the means of $f_1(x)$ and $f_2(x)$, respectively. Show that

$$E(X) = \mu = p_1\mu_1 + p_2\mu_2$$

c. (5%) Let σ_1^2 and σ_2^2 be the variances of $f_1(x)$ and $f_2(x)$, respectively. Show that

$$\mathrm{Var}(X) = \sigma^2 = p_1\sigma_1^2 + p_2\sigma_2^2 + p_1\mu_1^2 + p_2\mu_2^2 - (p_1\mu_1 + p_2\mu_2)^2$$

2. (35%) Imagine there is a help line for the BELL telephone company, and we want to study its service time. To collect data, we simply call the help line at the busiest time of day and measure how long it takes until the service is finished. We speculated that the service times would be lognormally distributed. We will assume that this is true, and thus the natural log of the service times would have a normal distribution. Below are the service times and their natural logs.

The service times in minutes:
0.21, 0.22, 14.40, 4.79, 11.75, 3.13, 9.03, 0.13, 7.69, 2.63

The natural log of the service times:
-1.561, -1.514, 2.667, 1.567, 2.464, 1.141, 2.201, -2.040, 2.040, 0.967

Descriptive Statistics: ln (service times)

Variable	N	Mean	StDev
ln (service times)	10	0.793	1.810

※ 考生請注意：本試題 ☑可 ☐不可 使用計算機

a. (2%) If we define $\bar{x}$ as the sample mean of ten logged service times (independent and identically distributed with mean μ and variance σ^2), then what is the sampling distribution of $\bar{x}$ (with its expected value and variance)?

b. (3%) If we define s, as the sample standard deviation of ten logged service times, then what is the sampling distribution of s ? Describe the detail as much as you can.

c. (5%) Suppose that the telephone company wants to study the 90^{th} percentile of their service time. To ease the calculation they start to study the estimator 90^{th} percentile of the **natural log of the service time**. What estimator should they use for this?

d. (2%) What is the expected value of the estimator in (c) as a function of μ and σ^2? Explain how you derive it.

e. (3%) What is the variance of the estimator in (c) as a function of μ and σ^2?

f. (5%) Do you know the form of the sampling distribution of the estimator in (c) or do you need to use a bootstrap method to approximate the form of the distribution of the 90^{th} percentile estimator? Explain.

g. (2%) Estimate the 90^{th} percentile of the service time using the data above (a single number).

h. (3%) Make a 95% confidence interval for the mean of the logged service time, μ.

i. (3%) Make a 95% confidence interval for the standard deviation of the logged service time, σ.

j. (7%) Suppose that the telephone company wants to be confident that the 90^{th} percentile of their service time is less than 30 minutes. Use some or all of oyur answers in (a) through (i) above to tell the telephone company whether you are confident that the 90^{th} percentile for the service time is less than 30 minutes. Explain.

※ 考生請注意：本試題 ☑可 ☐不可 使用計算機

3. The table below gives beverage preferences for random samples of teens and adults.

	Teens	**Adults**	**Total**
Coffee	50	200	250
Tea	100	150	250
Soft Drink	200	200	400
Other	50	50	100
	400	600	1,000

We are asked to test for independence between age (i.e., adult and teen) and drink preferences.

(a) (6%) Formulate null hypothesis and calculate test statistic.

(b) (6%) In an independence test, if an expected frequency is less than five, what will you do, and how to know the degrees of freedom of the critical value?

4. There are two possible ways for collecting samples in performing the analysis of variance about one factor. When the sample for each population is collected independently, it is called a completely randomized design. If every chosen experimental unit will generate an observation for each population, it is called a randomized block design.

(a) (6%) Which design will be better and why?

(b) (6%) What are the random processes involved in completely randomized design and randomized block design?

(c) (6%) In a randomized block design, discuss whether it is appropriate to consider the experimental unit as a second factor for significance test, and explain.

5. Consider the simple linear regression.

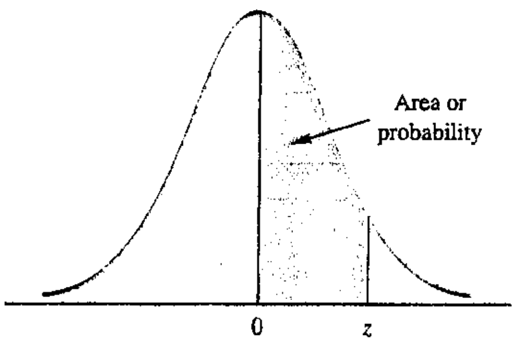
(a) (10%) Let the estimated linear regression equation for data (x_j, y_j) for j = 1, 2, …, 10 be $\hat{y} = 60 + 5x$. Suppose that $\bar{x} = 14$ and $\sum_{j=1}^{10}(x_j - \bar{x})^2 = 568$. Re-estimate the linear regression model when two new observations (10, 95) and (18, 165) are observed. ($b_1 = \dfrac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}, b_0 = \bar{y} - b_1\bar{x}$)

(b) (10%) Let $SST_j = \sum_i (y_i - \bar{y})^2$ and $SSR_j = \sum_i (\hat{y}_i - \bar{y})^2$ be the total sum of squares and the sum of squares due to regression for data set j, respectively. Draw figures for two data sets 1 and 2 such that $SST_1 > SST_2$ and $SSR_1 = SSR_2$, and justify your answer.

系所組別： 工業與資訊管理學系甲、乙、丙組
考試科目： 統計學
考試日期：0308，節次：3

※ 考生請注意：本試題 ☑可 □不可 使用計算機

STANDARD NORMAL DISTRIBUTION

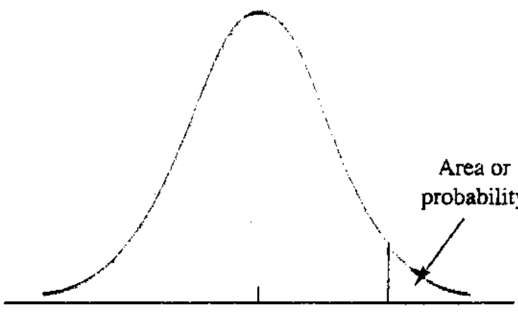


Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean and z is .3944.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

※ 考生請注意：本試題 ☑可 □不可 使用計算機

t DISTRIBUTION



Entries in the table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
20	.860	1.325	1.725	2.086	2.528	2.845
21	.859	1.323	1.721	2.080	2.518	2.831
22	.858	1.321	1.717	2.074	2.508	2.819
23	.858	1.319	1.714	2.069	2.500	2.807
24	.857	1.318	1.711	2.064	2.492	2.797
25	.856	1.316	1.708	2.060	2.485	2.787
26	.856	1.315	1.706	2.056	2.479	2.779
27	.855	1.314	1.703	2.052	2.473	2.771
28	.855	1.313	1.701	2.048	2.467	2.763
29	.854	1.311	1.699	2.045	2.462	2.756
30	.854	1.310	1.697	2.042	2.457	2.750
40	.851	1.303	1.684	2.021	2.423	2.704
50	.849	1.299	1.676	2.009	2.403	2.678
60	.848	1.296	1.671	2.000	2.390	2.660
80	.846	1.292	1.664	1.990	2.374	2.639
100	.845	1.290	1.660	1.984	2.364	2.626
∞	.842	1.282	1.645	1.960	2.326	2.576

Note: A more extensive table is provided as Table 2 of Appendix B.

(背面仍有題目,請繼續作答)

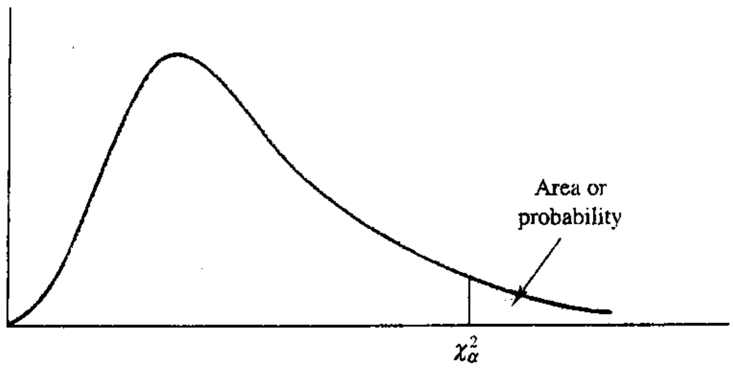
系所組別： 工業與資訊管理學系甲、乙、丙組

考試科目： 統計學

考試日期：0308，節次：3

※ 考生請注意：本試題 ☑可 ☐不可 使用計算機

CHI-SQUARE DISTRIBUTION



Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	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	7.041	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	45.558
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	12.878	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	50.994
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335