國立成功大學 109 學年度碩士班招生考試試題

系　　所：交通管理科學系

考試科目：統計學

考試日期：0211，節次：2

※ 考生請注意：本試題不可使用計算機。　請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

Show all work in order to facilitate partial credit. Be sure to clearly indicate your final answer. When doing hypothesis tests be sure to state the reasoning. Simple reject or accept answers with no work will receive no credit.

I. Long questions.

1. For 114 restaurants in NYC we have Zagat ratings on food as well as information on the price of a meal. We want to understand if there is a relationship between quality (as summarized by the Zagat ratings, the overall ratings are on a 30 point scale.) and price. We a run a simple regression of price on quality and obtain the following output

	Estimate	Std error	t-ratio	P-value
Intercept	-18.154	6.553	2.77	0.007
slope	2.6253	0.3315	7.92	0.000

In addition, we know that s_{pred} = 8.93 and R-square = 35.9%

(a - 2pts) Give an economic interpretation for the sign of the slope.

(b - 3pts) Find an approximate 95% confidence interval for the true slope.

(c - 3pts) Test the hypothesis that the slope is equal to zero at the 5% level. (You should be very precise here.)

(d - 3pts) Test the hypothesis that the slope is equal to 2 at the 5% level. (You should use our usual rule-of-thumb here.)

(e - 2pts) You are planning to have dinner in Soho at the restaurant YumYum. You know that the Zagat rating is 24. How much do you expect to pay?

(f - 4pts) Find a 95% prediction interval for the total price given a rating equal to 24.

(g - 4pts) Please interpret the meaning of the prediction interval in (f).

(h - 4pts) Do you think quality of food is sufficient to explain the price of a meal? Briefly explain your answer.

2. A student analyzed data for a one-way analysis of variance situation for which there were 3 levels of the treatment, and 21 people measured at each level. Unfortunately, after running the analysis, the student lost the computer output. She said "All I remember is that one of the mean squares was 100 and the other one was 500, but I can't remember which was which. Oh, and I remember that the p-value for the test was about .01."

(a - 13pts) Based on this information, can you construct the analysis of variance table? (The headings of the table structure are shown below to remind you.) If so, fill it in. If not, explain why not. If you think you can partially fill it in, do that.

Source	SS	df	MS	F*	p-value

(b - 6pts) In the statement of the question, it wasn't specified whether the treatment was fixed or random. Write the null and alternative hypotheses being tested in each of the two cases. Make sure you define any symbols you use.

(c - 6pts) We suppose that analysis of variance test is significant. Describe the procedure to determine which pairs of treatment are significantly different at 5% level. Justify your choice.

3. A course can be taken for credit either by attending lecture sessions at fixed times and days, or by doing online session that can be done at the student's own pace and at those times the student chooses. The course coordinator wants to determine if these two ways of taking the course resulted in a significant difference in achievement as measured by the final exam for the course. The table below gives the scores on an examination with 45 possible points for one group of $n_1 = 9$ students and a second group of $n_2 = 9$ students who took the course with conventional lectures. The sample standard deviations of the two groups are $s_1 = 4.94$ and $s_2 = 4.47$, respectively. We want to determine, via a one-sided hypothesis test, whether these data present sufficient evidence to indicate that the average grade for students who take the course online is significantly higher than those who attend a conventional class. (Given $\sqrt{4.94} \cong 2.22$, $\sqrt{9.41} \cong 3.07$)

Group 1: online	32	37	35	28	41	44	35	31	34
Group 2: classroom	35	31	29	25	34	40	27	32	31

(a - 4pts) Develop appropriate hypotheses for the test.

(b - 4pts) Compute the appropriate test statistic.

(c - 4pts) What is the rejection region for the null hypothesis, with a significance level $\alpha = 0.05$? With the above test statistic, do we reject or not reject H_0?

(d - 4pts) What is the one-sided confidence interval with a significance level $\alpha = 0.05$? Do we reject or not reject H_0?

(e - 4pts) Find the p-value of the test statistic. With a significance level $\alpha = 0.05$, do we reject or not reject H_0?

II. Short questions. (5 points each, 30 points total.)

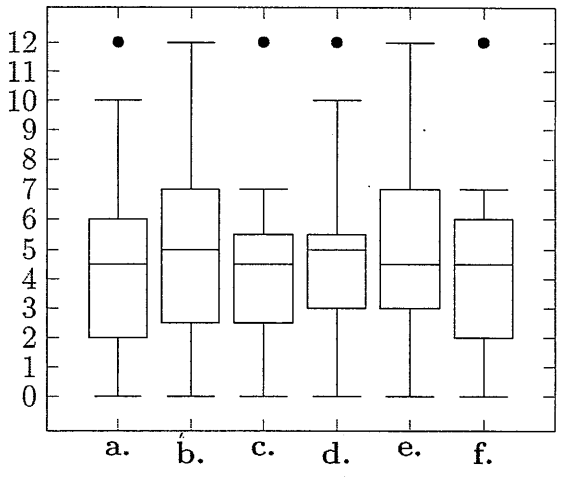
1. State, as precisely as you can remember, the Central Limit Theorem (CLT).

2. True or false? "A 90% confidence interval for the average number of children per household based on a simple random sample is found to be $I_{90\%} = [.7, 2.1]$. Under these assumptions, we conclude that 90% of households have between .7 and 2.1 children." (Justify your answer.)

3. Consider the two-sided hypotheses $H_0 : \mu = 0$ and $H_A : \mu \neq 0$, and suppose we are performing a t-test with significance level $\alpha = 0.05$. If the sample size n increases, does the rejection region R_0 for H_0 become bigger or smaller? (Justify your answer.)

4. Let T_1 and T_2 be two unbiased estimators for an unknown parameter θ. It is given that $\text{Var}(T_1) = 1$ and $\text{Var}(T_2) = 3$. Furthermore, T_1 and T_2 are independent. Next, consider the estimator $T_3 = \alpha T_1 + (1 - \alpha)T_2$, where $\alpha \in [0, 1]$. For which value of α has T_3 the smallest mean squared error (MSE)? (Provide the process of calculation.)

5. Consider the following ordered dataset: 0, 1, 3, 3, 4, 5, 5, 5, 7, 12
Which boxplot below corresponds to the given dataset? (Justify your answer.)



6. Let X have a *Exp*(2) distribution. Using Chebyshev's inequality it holds that:
P (|X − "A" | < 2) "____" "B" , please find the number of A and B and also determine the inequality symbols to use ($<, \leq, >, \geq$) in the underline blank of the inequality. (Do NOT only show the results, it is necessary to provide the process of calculation.)

國立成功大學 109 學年度碩士班招生考試試題

系　　所：交通管理科學系

考試科目：統計學

考試日期：0211，節次：2

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.

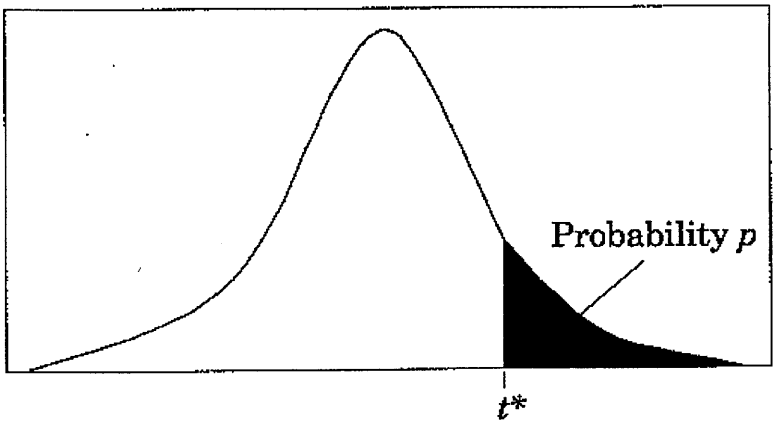


Table B t distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											