1. (10 points) Consider the following back-to-back stem-and-leaf plot for two groups:

A		B
	0	348
	1	01256
843	2	29
65210	3	2557
92	4	
7552	5	6
	6	1458
6	7	09
8541	8	
90	9	

Group	mean	Standard deviation
A	34.857	25.031
B	54.857	

(a) (3 points) Which of the following are true statements?
   (1) The distributions have the same median.
   (2) The distributions have the same IQR (interquartile range).
   (3) The distributions have the same standard deviation.

(b) (7 points) The table provides summary statistics on the data. Use these statistics to calculate a 95% confidence interval for the difference between average score of group A and Group B, and interpret this interval in the context of the data.

2. (7 points) Suppose the summary statistics for a sample of 100 observations is given by

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00125	0.13000	0.27800	0.41100	0.56200	2.11000

(a) (5 points) Are there outliers in this sample? Justify your answer.
(b) (2 points) Would you expect the distribution in part (a) to be left-skewed, right-skewed or symmetric? Justify your answer.

3. (5 points) The water from a lake is tested and is found to contain on average four bacteria per litre (L) of water. It is assumed that the number of bacteria contained in the lake follows a Poisson distribution. Now, a sample of 250 ml is collected from the lake. Calculate the probability that the 250 ml water sample will contain

at least two bacteria.

4. (10 points) Fasting blood glucose levels for normal non-diabetic individuals are normally distributed in the population, with mean $\mu = 85$ mg/dL and standard deviation $\sigma = 7.5$ mg/dL.
   (a) (5 points) What is the probability that a randomly chosen member of the population has a fasting glucose level higher than 100 mg/dL?
   (b) (5 points) What is the probability of observing 30 or fewer individuals with a fasting glucose level higher than 100 mg/dL in a random sample of 1000 members from the population?

5. (13 points) Suppose the table below represents a cohort of 100,000 women age 40 - 49 in whom mammogram screening and breast cancer behaves just like the larger population.

Breast Cancer present	Mammogram Result		Total
	Positive	Negative	
Yes	3,296	37	3,333
No	8,313	88,354	96,667
Total	11,609	88,391	100,000

   (a) (3 points) Using this data, calculate the sensitivity and specificity of the mammogram test.
   (b) (5 points) If the prevalence of breast cancer in the population is 10%, please calculate the PPV (positive predictive value) and NPV (negative predictive value) of the test.
   (c) (5 points) Many studies have shown that breast cancer is almost twice as prevalent in women with family history. Would you expect the positive predictive value of the mammogram test to be larger or smaller than what you calculated in part (b).

6. (5 points) A university newspaper is conducting a survey to determine what fraction of students support a $200 per year increase in annual fees to pay for a new gymnasium. How big of a sample is required to ensure the error of the sample proportion estimate is smaller than 0.04 using a 95% confidence level?

7. (24 points) When we have a data set $(x_i, y_i), i = 1, 2, .., 27$. Fit simple linear regression:
   Model: $Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon$
   The summary of results shown below

接次頁

Call: lm(formula = Y1 ~ X1)				
Coefficients:				
	Estimate.	Std. Error.	t-statistic	Pr(>|t|)
(Intercept)	10.4838	0.4391	23.878	<2e-16
X1	1.5863	0.7603	2.08	0.0473
$R^2 = 0.1483$				

(a) (2 points) What is the degree of freedom of t-statistic for testing $H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$?

(b) (2 points) For testing $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$, what is the F-statistic?

(c) (2 points) For testing $H_0: \beta_1 \geq 0$ vs. $H_1: \beta_1 < 0$ with significant level α=0.05. What is the p-value?

(d) (2 points) Fine $\sum_{i=1}^{n}(y_i - \hat{y}_i)\hat{y}_i$ =?

When fit model: $Y_1 = \beta_0 + \beta_1 X_2 + \varepsilon, where\ X_2 = 10X_1$, and $\varepsilon \sim N(0, \sigma^2)$

(e) (2 points) What is the t-statistic for testing $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$?

(f) (2 points) What is the $\hat{\beta}_0$?

When fit model: $X_1 = \beta_0 + \beta_1 Y_1 + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$

(g) (2 points) What is the t-statistic for testing $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$?

(h) (2 points) R^2 = 0.1487. True or False?

When fit model: $Y_1 = \beta_0 + \beta_1 X_3 + \varepsilon,, where\ X_3 = X_1 + 2$ and $\varepsilon \sim N(0, \sigma^2)$

(i) (2 points) What is the t-statistic for testing $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$?

(j) (2 points) R^2 < 0.1487. True or False?

When fit model: $Y_2 = \beta_0 + \beta_1 X_4 + \varepsilon, where\ Y_2$ and X_4 are separately the normalizations of Y_1 and X_1, and $\varepsilon \sim N(0, \sigma^2)$.

(k) (2 points) R^2 > 0.1487. True or False?

(l) (2 points) $\hat{\beta}_1$ is equal to the correlation coefficient of Y_1 and X_1. True or False?

8. (12 points) Let X be the number of insects collected during a fixed time period in a farm, and follow a Poisson distribution. Poisson distribution with *pmf*:

$P(X = x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, 2, \ldots$. To test $H_0: \lambda \geq 3\ vs.\ H_A: \lambda < 3$ and we have

X	0	1	2	3	4	5	6	7	8
$P(X = x, \lambda = 4)$	0.018	0.073	0.147	0.195	0.195	0.156	0.104	0.060	0.030
$P(X = x, \lambda = 3)$	0.050	0.149	0.224	0.224	0.168	0.101	0.050	0.022	0.008
$P(X = x, \lambda = 2)$	0.135	0.271	0.271	0.180	0.090	0.036	0.012	0.003	0.001

(a) (2 points) If the rejection region of X is $\{0, 1\}$, what is the probability of type I error?

(b) (2 points) When significant level is 0.05, what is the rejection region of X.

(c) (2 points) If significant level is 0.05 and $\lambda = 4$, what is the probability of type I error?

(d) (2 points) If significant level is 0.05 and $\lambda = 2$, what is the probability of type II error?

(e) (2 points) If $X = 2$, what is the p-value?

(f) (2 points) When testing result is rejecting H_0, that also imply rejecting H_0: $\lambda \geq 4$. True or False?

9. (6 points) A dice was tossed 30 times to test whether the dice is fair. The outcome of the 30 trials is shown in the table.

Dice point	1	2	3	4	5	6
The frequency of dice point	2	3	6	7	10	2

Note: $P(\chi^2_{df=4} > 9.5) = 0.05$; $P(\chi^2_{df=5} > 11.1) = 0.05$;
$P(\chi^2_{df=6} > 12.6) = 0.05$;

Q: What is the testing result when significant level is 0.05? Please show the detail. of your calculation.

10. (8 points) To test whether an insect is randomly spatially distributed on the farm. 20 plots with a fixed size were randomly selected from the farm and counted the number of individuals in each selected plot. The sampling outcome was shown in the table.

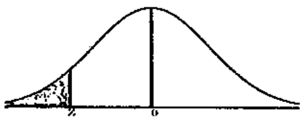
The number of insects	0	1	2	3	4
The number of plots	5	2	4	6	3

Note: $P(\chi^2_{df=3} > 7.8) = 0.05$; $P(\chi^2_{df=4} > 9.5) = 0.05$;
$P(\chi^2_{df=5} > 11.1) = 0.05$;

Q: What is the testing result when significant level is 0.05? Please show the detail of your calculation.

接次頁

表一、標準常態分佈的累積機率表



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

見背面

表二、卜瓦松分布 (Poission distribution) 的累積機率表

Tables of the Poisson Cumulative Distribution

The table below gives the probability of that a Poisson random variable X with mean = λ is less than or equal to x. That is, the table gives

$$P(X \le x) = \sum_{r=0}^{x} \lambda^r \frac{e^{-\lambda}}{r!}$$

λ=	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8
x= 0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679	0.3012	0.2466	0.2019	0.1653
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358	0.6626	0.5918	0.5249	0.4628
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197	0.8795	0.8335	0.7834	0.7306
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810	0.9662	0.9463	0.9212	0.8913
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963	0.9923	0.9857	0.9763	0.9636
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994	0.9985	0.9968	0.9940	0.9896
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9974
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9994
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
λ=	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0	4.5	5.0	5.5
x= 0	0.1353	0.1108	0.0907	0.0743	0.0608	0.0498	0.0408	0.0334	0.0273	0.0224	0.0183	0.0111	0.0067	0.0041
1	0.4060	0.3546	0.3084	0.2674	0.2311	0.1991	0.1712	0.1468	0.1257	0.1074	0.0916	0.0611	0.0404	0.0266
2	0.6767	0.6227	0.5697	0.5184	0.4695	0.4232	0.3799	0.3397	0.3027	0.2689	0.2381	0.1736	0.1247	0.0884
3	0.8571	0.8194	0.7787	0.7360	0.6919	0.6472	0.6025	0.5584	0.5152	0.4735	0.4335	0.3423	0.2650	0.2017
4	0.9473	0.9275	0.9041	0.8774	0.8477	0.8153	0.7806	0.7442	0.7064	0.6678	0.6288	0.5321	0.4405	0.3575
5	0.9834	0.9751	0.9643	0.9510	0.9349	0.9161	0.8946	0.8705	0.8441	0.8156	0.7851	0.7029	0.6160	0.5289
6	0.9955	0.9925	0.9884	0.9828	0.9756	0.9665	0.9554	0.9421	0.9267	0.9091	0.8893	0.8311	0.7622	0.6860
7	0.9989	0.9980	0.9967	0.9947	0.9919	0.9881	0.9832	0.9769	0.9692	0.9599	0.9489	0.9134	0.8666	0.8095
8	0.9998	0.9995	0.9991	0.9985	0.9976	0.9962	0.9943	0.9917	0.9883	0.9840	0.9786	0.9597	0.9319	0.8944
9	1.0000	0.9999	0.9998	0.9996	0.9993	0.9989	0.9982	0.9973	0.9960	0.9942	0.9919	0.9829	0.9682	0.9462
10	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992	0.9987	0.9981	0.9972	0.9933	0.9863	0.9747
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9996	0.9994	0.9991	0.9976	0.9945	0.9890
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9992	0.9980	0.9955
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9983
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

接次頁

表三、t 分佈的累積機率表

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

試題隨卷繳回