

NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
ADVANCED PLACEMENT TEST
(SAMPLE)
ST1131 INTRODUCTION TO STATISTICS
MMM-YYYY Time Allowed: 2 Hours

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FOUR (4)** questions and comprises **EIGHTEEN (18)** printed pages.
2. This is a **CLOSED BOOK** examination.
3. Answer **ALL** questions. The total mark for this paper is 120.
4. Please show working steps and answers in the space provided for each question. **DO NOT** use pencils to write answers. Hand in this booklet at the end of the examination.
5. Only non-programmable and non-graphing calculators without remote communication function may be used.
6. The **Standardized Normal Distribution Table (Z Table)** and the **Student's t -Distribution Table (t Table)** are attached to this test paper for your reference.

Name: _____

Question	Marks scored	Max. marks
1		30
2		28
3		32
4		30

Total
(120)

Question 1

The Land Transport Authority is interested to know the average number of passengers taking a particular bus service upon arrival at a certain bus stop. A sample of 40 buses is selected and the numbers of passengers upon arrival at the particular bus stop on a given day are recorded for this purpose. The data are shown in the stem-and-leaf plot below. A bus is considered overload if there are 55 or more passengers on board. The proportion of overload buses at the particular bus stop was 0.6 based on a similar study done last year.

Stem	Leaf	(Leaf Unit = 1.0)
0	5	
1	2	
1	5 8	
2	4	
2	6 8 9	
3	0 4	
3	5 7 8	
4	2 3 4	
4	5 7 9	
5	3 4	
5	5 6 6 7 8 9 9 9	
6	0 0 1 1 2 3 4 4	
6	5 5 6	

a) Would it be better if a histogram is plotted? Why or why not?

[3 marks]

b) Find the five-number summary for this data set.

[5 marks]

c) Draw a box-plot using the data.

[3 marks]

Question 1 (continued)

- d) What is the likely shape of the population distribution of the numbers of passengers upon arrival at the particular bus stop? Why? [4 marks]
- e) Give an estimate for the proportion of buses that are overload at the particular bus stop. [1 mark]
- f) What is the sample size required to estimate the percentage of overload within 5% with 95% confidence? [2 marks]
- g) Are the assumptions for inference seemed valid? Explain. [4 marks]

Question 1 (continued)

- h) Regardless of your answer in part (g), construct a 90% confidence interval for the proportion of buses that are overload at the particular bus stop. [4 marks]
- i) If the population proportion of overload buses at the particular bus stop has not changed, and a 95% confidence interval constructed for the population proportion is (0.56, 0.74), what is the probability that the proportion of overload buses at the particular bus stop from a sample of 40 will be within the limits of this confidence interval? [4 marks]

Question 2

Sunbucks Coffee sells coffee and food in KR Campus. 90% of its customers orders coffee and 30% of the customers orders only coffee with no food. The manager is keeping a record of the waiting time before each customer is handed the order. Over many observations, the waiting time is known to follow a normal distribution with an average of 5 minutes and a standard deviation of 2.5 minutes. There are 11 customers this morning.

a) What is the probability that a customer orders coffee and food? [3 marks]

b) What is the probability that a customer orders coffee if he has ordered food? [4 marks]

c) Are the orders of coffee and food independent? Explain. [3 marks]

Question 2 (continued)

d) What is the probability that a customer gets the order in 7 minutes? [4 marks]

e) If possible, compute the probability that at most two customers get the orders in 7 minutes, if not, explain what information is needed. [5 marks]

Question 2 (continued)

f) If possible, compute the probability that the average waiting time to get the orders is at most 7 minutes. If not, explain what information is needed. [5 marks]

g) As a promotion strategy, the manager decides to give a 50% discount to the customer if the waiting time is longer than a specific duration. How many seconds should the manager specify so that at most 5% of the customers will get the 50% discount? [4 marks]

Question 3

A researcher has collected the following data from a population.

<u>Family</u>	<u>Daughter's height (cm)</u>	<u>Mother's height (cm)</u>
1	167	172
2	166	162
3	176	157
4	171	159
5	165	157
6	177	177
7	173	174
8	158	150

- a) At 5% significance level, can the researcher conclude that the average height of the daughters in the population is 170cm? [12 marks]

Question 3 (continued)

- b) At 10% significance level, can the researcher conclude that the daughters are taller than their mothers on average? [12 marks]

Question 3 (continued)

c) What type of error you could have made with each of the conclusions above?
[2 marks]

d) At 5% significance level, can we conclude that the proportion of mothers having taller daughters is more than 0.30?
[6 marks]

Question 4

- a) A researcher would like to investigate whether people suffering from insomnia can be helped by taking yoga classes. She designs an experiment as follows. She will obtain a group of volunteers who are suffering from insomnia and who do not currently practice yoga. Each person's degree of insomnia at the start of the experiment will be evaluated. She will investigate which yoga classes, every day or three times per week, is more effective. Each volunteer will choose which group they would like to be part of. At the end of a suitable time period, for example two months, each person's degree of insomnia will again be evaluated. The improvement for the two groups will then be compared. The person evaluating the insomnia levels will be blinded to which group the patients were in. Identify the flaw(s) in this experiment. [7 marks]

Question 4 (continued)

- b) A group of female college students took a test that measured their verbal IQs and also underwent an MRI scan to measure the size of their brains (in 1000s of pixels). Below are the outputs.

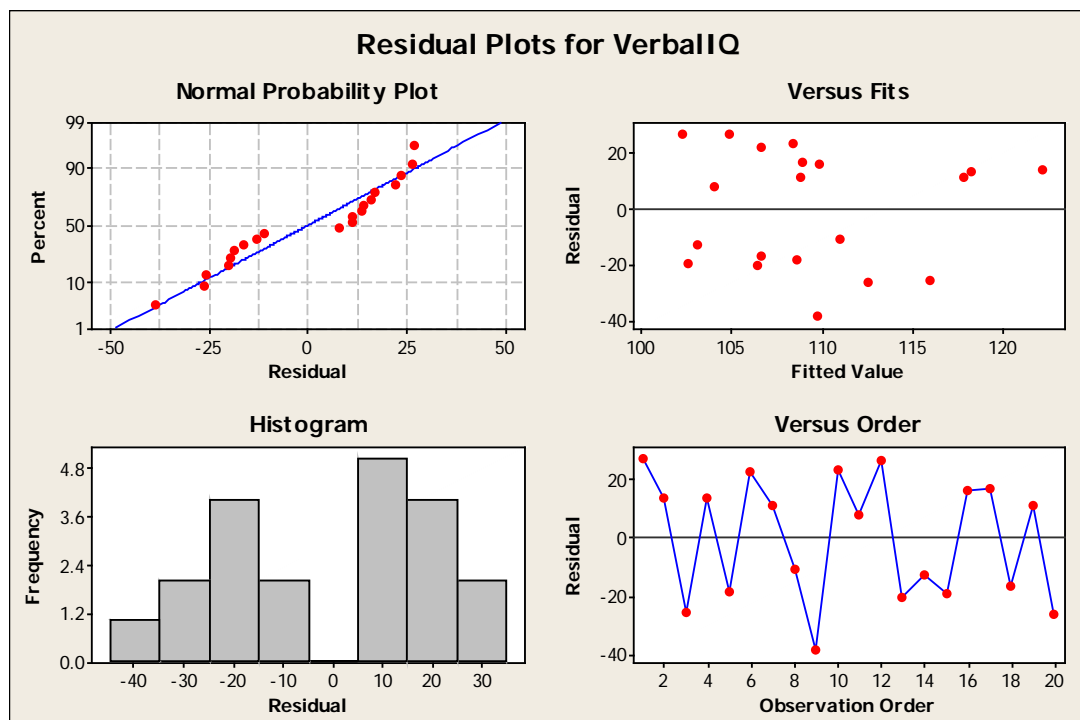
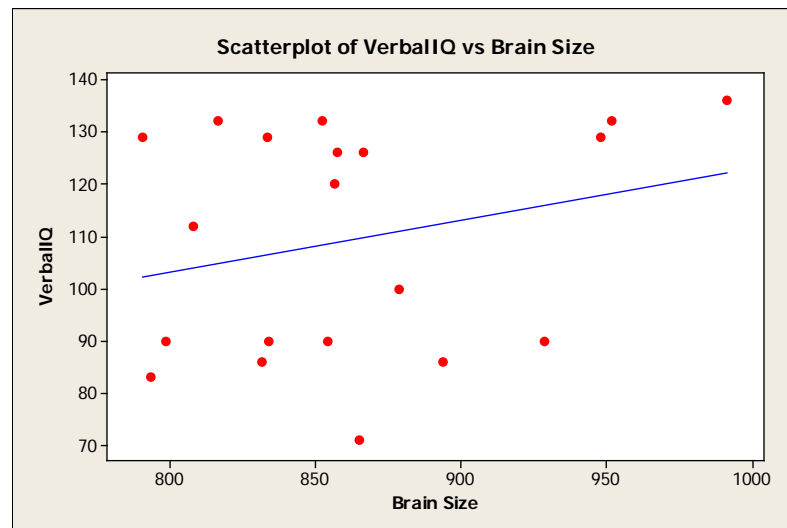
Regression Analysis: Verbal IQ versus Brain Size

Predictor	Coef	SE Coef
Constant	24.18	76.38
Brain Size	0.09884	0.08837

Dependent variable: IQ_verbal

R-squared = 6.5%

DF = 18



Question 4 (continued)

- iv. Is there evidence of an association between the IQ levels and brain size? Test an appropriate hypothesis and state your conclusion. [6 marks]

- v. Are the conditions for inference satisfied? Explain. [6 marks]

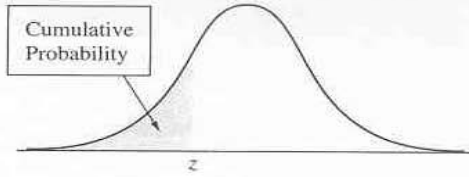
Question 4 (continued)

vi. Construct a 90% confidence interval for the slope of the true line. [3 marks]

vii. Interpret the interval in context. [2 marks]

***** END OF QUESTION *****

Appendix A: Standard Normal Distribution Table (Z Table)



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

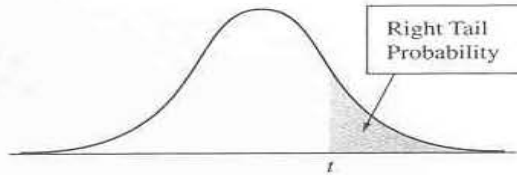
Appendix B: Student's t -Table (t Table)

TABLE B t Distribution Critical Values						
df	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
	Right-Tail Probability					
	$\alpha_{.100}$	$\alpha_{.050}$	$\alpha_{.025}$	$\alpha_{.010}$	$\alpha_{.005}$	$\alpha_{.001}$
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.091