

Silence Please

Do Not turn over this page until advised to by the Invigilator

CIT Semester 1 Examinations 2018/19

Note to Candidates:	Check the <u>Programme Title</u> and the <u>Module Description</u> to ensure that you have received the correct examination. If in doubt please contact an Invigilator.
Module Title:	Applied Stats & Probability
Module Code:	STAT8006
Programme Title(s):	HDip Sc Data Science Analytics MSc Data Science & Analytics P1 HDipSc Data Science Analytics PT
Block Code(s):	SDAAN_8_Y5 SDAAN_91Y5 SDAANP8_Y5
External Examiner(s):	Dr. Niall Fitzgerald, Dr. Katarina Domijan
Internal Examiner(s):	Dr. Noreen Quinn, Ms. Sarah Murphy
Instructions:	Answer ALL Questions. All questions carry equal marks.
Duration:	2 Hours
Required Items:	Calculator, Murdoch & Barnes Tables

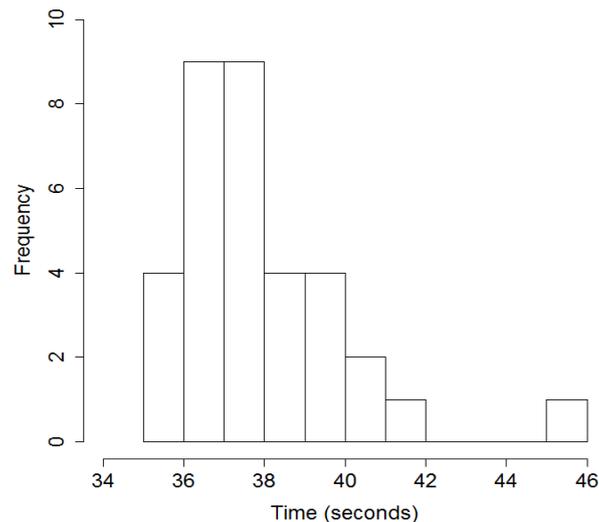
Question 1

(a) Answer each of the following questions:

- i. Outline the difference between a parameter and a statistic.
- ii. What does the coefficient of variation measure?
- iii. Name an advantage and a disadvantage of the range as a measure of spread.

[6 marks]

(b) The histogram below shows the length of time it took 34 people to complete a task.



- i. Comment on the shape of this histogram.
- ii. Based on the shape of the histogram, what is the best measure of centrality and variability for this data?

[3 marks]

(c) The heart rates (beats per minute) of 33 patients admitted to the Accident and Emergency department at a local hospital were measured. The data is given below:

53	60	62	66	69	70	71	76	77	77	78
78	79	80	82	83	87	88	88	89	90	90
91	91	91	92	94	94	95	96	97	99	101

- i. Calculate all three quartiles and the interquartile range for the above dataset.
- ii. Interpret the interquartile range.
- iii. Are there any outliers present in this dataset?
- iv. Construct a boxplot for the above data.
- v. It was discovered that the value of 101 bpm was recorded in error and should have instead been 108 bpm. What effect would this have on the mean and the median? Give reasons for your answer. (Note: you do not need to calculate these values for the updated dataset)
- vi. What other graphical display would be appropriate for this dataset? (Note: you do not need to produce this plot)

[16 marks]

Question 2

(a) There are 7 red balls and 4 black balls in a box. You are to pick three balls from the box without replacement.

- i. What is the probability that you pick three red balls from the box?
- ii. What is the probability that you pick two red balls and one black ball from the box?

[5 marks]

(b) The customer service department for a wholesale electronics outlet claims that 80 percent of all customer complaints are resolved to the satisfaction of the customer. In order to test this claim, a random sample of 6 customers who have filed complaints is selected. Let X be the random variable representing the number of customers whose complaints were resolved to the customer's satisfaction. Assuming the variable X follows a binomial distribution, what is the probability that

- i. Exactly 2 customer complaints are resolved satisfactorily?
- ii. Two or more customer complaints are resolved satisfactorily?

[5 marks]

(c) Suppose that the number of Wi-Fi interruptions on an office network varies, with an average of 1.2 interruptions per day. Assuming that the Poisson model is reasonable for this situation, where X is the random variable representing the daily count of interruptions, calculate the probability of having no more than two interruptions tomorrow.

[5 marks]

(d) It is known that the annual milk yield of a given breed of cow is normally distributed with a mean of $\mu = 4,900$ litres and a standard deviation of $\sigma = 800$ litres.

- i. If one of these cows is selected at random, what is the probability that its annual milk yield is:
 - a. between 4,300 and 5,500 litres?
 - b. less than 4,000 litres?
- ii. Suppose that 60% of the population of the cows yield less than k litres of milk per annum. What is the value of k ?

[10 marks]

Question 3

(a) Answer each of the following questions:

- i. Explain what is meant by a rejection region in the context of statistical hypothesis testing.
- ii. What is a p – value?
- iii. In a given hypothesis test, a p – value of 0.453 was found. In this case, would you reject or fail to reject the null hypothesis at a significance of 5%?
- iv. The following two 95% confidence intervals for the mean were found for a dataset
(33.8, 56.2) and (39.9, 50.1)

Which of these confidence intervals is preferable and why?

[5 marks]

(b) A production line fills ink into bottles. The volume of ink in these bottles is known to follow a Normal distribution with a standard deviation of 0.6ml. A sample of 30 bottles of ink are chosen at random and the mean volume of ink in these bottles was found to be 98ml.

- i. Construct a 95% confidence interval for the mean volume of ink in the bottles.
- ii. What sample size is needed in order for upper and lower bound of the 95% confidence interval for the mean volume of ink to be within 0.1ml of the mean value.

[7 marks]

(c) A company is conducting a one day orientation program and wishes to test the effect it has on the work performance of their employees. Sixteen employees are selected at random of which 8 workers attend the orientation program and the other 8 employees do not attend the orientation program. After one month, an on-the-job evaluation was conducted for all 16 participants. The descriptive statistics for both groups are shown below.

	Sample Size	Mean	Standard deviation
Received Orientation	8	26.25	4.367
No Orientation	8	19.875	3.044

- i. List and explain two methods which you could use to test a dataset for normality.
- ii. Is there evidence of a significant difference between the work performance of employees who attend this orientation program and those who do not attend the program? Conduct a suitable test of hypothesis using a 1% significance level and you may assume equal variances. Clearly state your hypotheses and conclusions.

[13 marks]

Question 4

- (a) The probability that Bob will like a new action movie is 0.8 and the probability that his girlfriend Ann will like it is 0.7. Based on previous dates, the probability is 0.18 that he will like the movie and she will dislike it. What is the probability that he will like the movie given that she is not going to like it?

[5 marks]

- (b) Let A be the event that a person takes an overseas summer vacation and the probability of this event is 0.75. Let B be the event that a person takes an overseas winter vacation and the probability of this event is 0.32. The probability that a person takes an overseas winter vacation and takes an overseas summer vacation is 0.15.

- i. What is the probability that the person does not take an overseas summer vacation?
- ii. What is the probability that a person takes an overseas winter vacation given that person takes an overseas summer vacation?
- iii. Are the events A and B mutually exclusive? Give a reason for your answer using probabilities to demonstrate.
- iv. Are the events A and B independent? Give a reason for your answer using probabilities to demonstrate.

[10 marks]

- (c) A student was asked to investigate the amount of credit given by a bank to customers based on whether
- the customer had a telephone or not, and
 - how long the customer lived at their particular residence (This is recorded as a categorical nominal variable with 4 levels/categories).

The following output was generated in R. Answer the following questions:

- i. Explain the reasons why the student would choose the methods below to investigate these problems.
- ii. Comment on the findings of the student.

```
wilcoxon rank sum test with continuity correction
data: Amount by Telephone
W = 6078, p-value = 0.001179
alternative hypothesis: true location shift is not equal to 0
```

```
Kruskal-wallis rank sum test
data: Amount by ResidenceDuration
Kruskal-wallis chi-squared = 1.7832, df = 3, p-value = 0.6186
```

[10 marks]

Statistical Formulae

Mean:

$$\mu = \frac{\sum x}{N} \qquad \bar{x} = \frac{\sum x}{n} \qquad \mu = \frac{\sum(fx)}{\sum f}$$

Standard Deviation:

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}} \qquad s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \qquad \sigma = \sqrt{\frac{\sum(fx^2)}{\sum f} - \left(\frac{\sum(fx)}{\sum f}\right)^2}$$

Coefficient of Skewness: $\frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$

Expected Value: $E(X) = \sum xP(x)$

Variance: $V(X) = E(X^2) - (E(X))^2$

Distributions:

Binomial Distribution:

$$P(X = r) = {}^n C_r p^r q^{n-r}$$

Standard Units:

$$z = \frac{x - \mu}{\sigma}$$

Poisson Distribution:

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

Confidence Intervals and Hypothesis Testing:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$$

$$\bar{d} \pm t_{\frac{\alpha}{2}, n-1} \frac{s_d}{\sqrt{n}}$$

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$t_0 = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

$$Z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

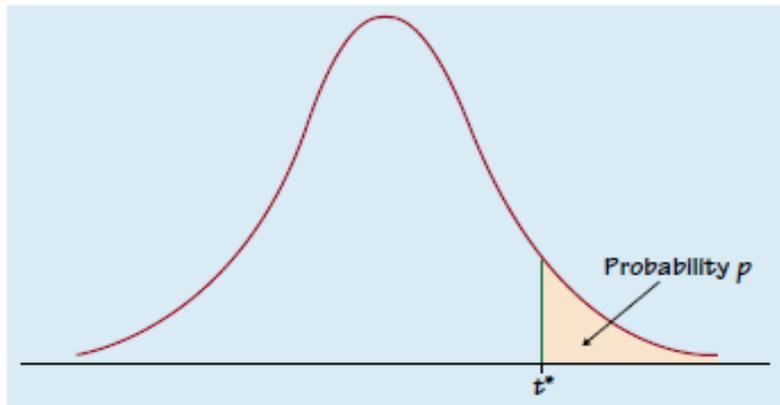
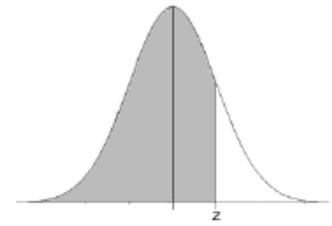


TABLE D t distribution critical values

df	Tail probability p											
	.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	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	0.674	0.841	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

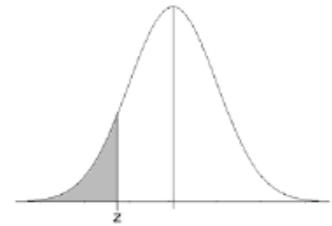
Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

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