2018 INIAD stat_A

Exercises

50minutes

On Final Exam

-Scheduled on 11/16

-Paper exam

-You can bring anything, but the internet connection is not allowed.

(Including iniad-WiFi)

PC is OK without the internet connection(Python or R is oK).

Q. 1

- Find below.

$_5P_5 = \boxed{①}$ (i) 60 (ii) 80 (iii) 100 (iv) 120 (v) 160

$_5P_3 = \boxed{②}$ (i) 10 (ii) 15 (iii) 20 (iv) 40 (v) 60

$_7P_2 = \boxed{③}$ (i) 30 (ii) 40 (iii) 42 (iv) 50 (v) 62

$_7C_2 = \boxed{④}$ (i) 5 (ii) 10 (iii) 15 (iv) 21 (v) 30

$_4C_3 = \boxed{⑤}$ (i) 2 (ii) 3 (iii) 4 (iv) 5 (v) 6

$_6C_4 = \boxed{⑥}$ (i) 10 (ii) 15 (iii) 20 (iv) 25 (v) 30

#【A.1】

$_5P_5 =$ ① (i) 60 (ii) 80 (iii) 100 (iv) 120 (v) 160

$_5P_3 =$ ② (i) 10 (ii) 15 (iii) 20 (iv) 40 (v) 60

$_7P_2 =$ ③ (i) 30 (ii) 40 (iii) 42 (iv) 50 (v) 62

$_7C_2 =$ ④ (i) 5 (ii) 10 (iii) 15 (iv) 21 (v) 30

$_4C_3 =$ ⑤ (i) 2 (ii) 3 (iii) 4 (iv) 5 (v) 6

$_6C_4 =$ ⑥ (i) 10 (ii) 15 (iii) 20 (iv) 25 (v) 30

Q. 2

Answer the choice made of only interval scales.

⑦

A: T-score in an exam
B: Room numbers of condos
C:Celsius temperature [℃]
D ： Height[cm]
E ： Answers to a questionnaire with 3 options ;
Good/ normal / bad
F ： Ranking of scores of football players

(i) A and B (ii)A and C (iii) D and E (iv) E and F (v) F

A. 2

Answer the choice made of only interval scales.

⑦

A: T-score in an exam
B: Room numbers of condos
C:Celsius temperature [℃]
D ： Height[cm]
E ： Answers to a questionnaire with 3 options ;
Good/ normal / bad
F ： Ranking of scores of football players

(i) A and B (ii)A and C (iii) D and E (iv) E and F (v) F

問3

The stem-and-leaf display below shows the scores of 40 students. It seems that the median is ⑧

(i) 58 (ii) 59 (iii) 60 (iv) 61 (v) 62

Stem	Leaf
4	1 2 4 4 5
5	1 1 2 3 4 4 4 4 5 6 7 7
6	0 0 1 1 1 2 3 4 4 4 6
7	0 1 2 2 3 3 3 4 7 8
8	3 6

A. 3

The stem-and-leaf display below shows the scores of 40 students. It seems that the median is ⑧

(i) 58 (ii) 59 (iii) 60 (iv) 61 (v) 62

Stem	Leaf
4	1 2 4 4 5 5
5	1 1 2 3 4 4 4 4 5 6 7 7 12
6	0 0 1 1 1 2 3 4 4 4 6 11
7	0 1 2 2 3 3 3 4 7 8 10
8	3 2 2

Q.4

Let a r.v. Z follows the z-distribution.

Then, the probability $P(-4 < Z < 2)$ ⑨

(i) 0.25 (ii) 0.44 (iii) 0.54 (iv) 0.74 (v) 0.98

A.4

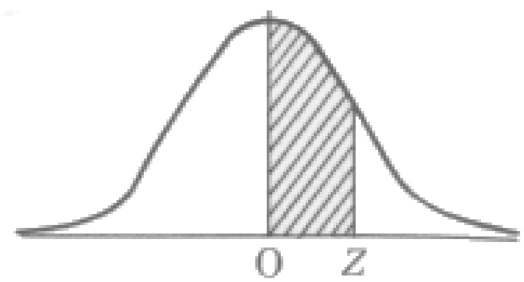
Let a r.v. Z follows the z-distribution.

Then, the probability $P(-4 < Z < 2)$ ⑨

(i) 0.25 (ii) 0.44 (iii) 0.54 (iv) 0.74 (v) 0.98

```
> pnorm(2)-pnorm(-4)
[1] 0.9772182
```

Z-table



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.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

```
]: #Q. 4
   from scipy.stats import norm
   norm.cdf(2)-norm.cdf(-4)
```

```
]: 0.9772181968099877
```

Q.5

Let X be a r.v. following $N(5, 5^2)$.

Then, $P(-1<X<5)=$ ⑩

(i) 0.264 (ii) 0.385 (iii) 0.481 (iv) 0.545 (v) 0.652

A.5

Let X be a r.v. following N（5，5^2）.

Then, P(-1<X<5)= $\boxed{⑩}$

(i) 0.264 (ii) 0.385 (iii) 0.481 (iv) 0.545 (v) 0.652

Translate X into Z.

Z=(X-5)/5

X=-1 ⇒ Z= -6/5 = -1.2

X=5 ⇒ Z= (5-5)/5 = 0

Thus, P(-1<X<5) = P(-1.2<Z<0).

= 0.385

```
> pnorm(1.2)-0.5
[1] 0.3849303
```

```python
from scipy.stats import norm
norm.cdf(5, loc=5,scale=5)-norm.cdf(-1, loc=5,scale=5)
```

0.3849303297782918

Q.6

In a certain lottery, we can get a winning piece with
The probability of p in one trial.

Now, as a result of 3 trials, you won at the first trial, and then got the losing pieces twice continuously.
Then, the maximum likelihood estimator of p is $\boxed{⑪}$.

(i) 0.1 (ii) 0.33 (iii) 0.4 (iv) 0.5 (v) 0.6

A.6

In a certain lottery, we can get a winning piece with The probability of p in one trial.

Now, as a result of 3 trials, you won at the first trial, and then got the losing pieces twice continuously.

Then, the maximum likelihood estimator of p is $\boxed{⑪}$.

(i) 0.1 (ii) 0.33 (iii) 0.4 (iv) 0.5 (v) 0.6

Q.7

① Find the number of ways to choose 2 students out of 5 students. ⑫

(i) 6 (ii) 8 (iii) 10 (iv) 999

② The number of ways to choose 4 students out of 7 is ⑬

(i) 28 (ii) 35 (iii) 42 (iv) 999

A.7

① Find the number of ways to choose 2 students out of 5 students.

⑫

(i) 6 (ii) 8 (iii) 10 (iv) 999

$$\frac{5!}{(5-3)! \times 3!} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

② The number of ways to choose 4 students out of 7 is ⑬

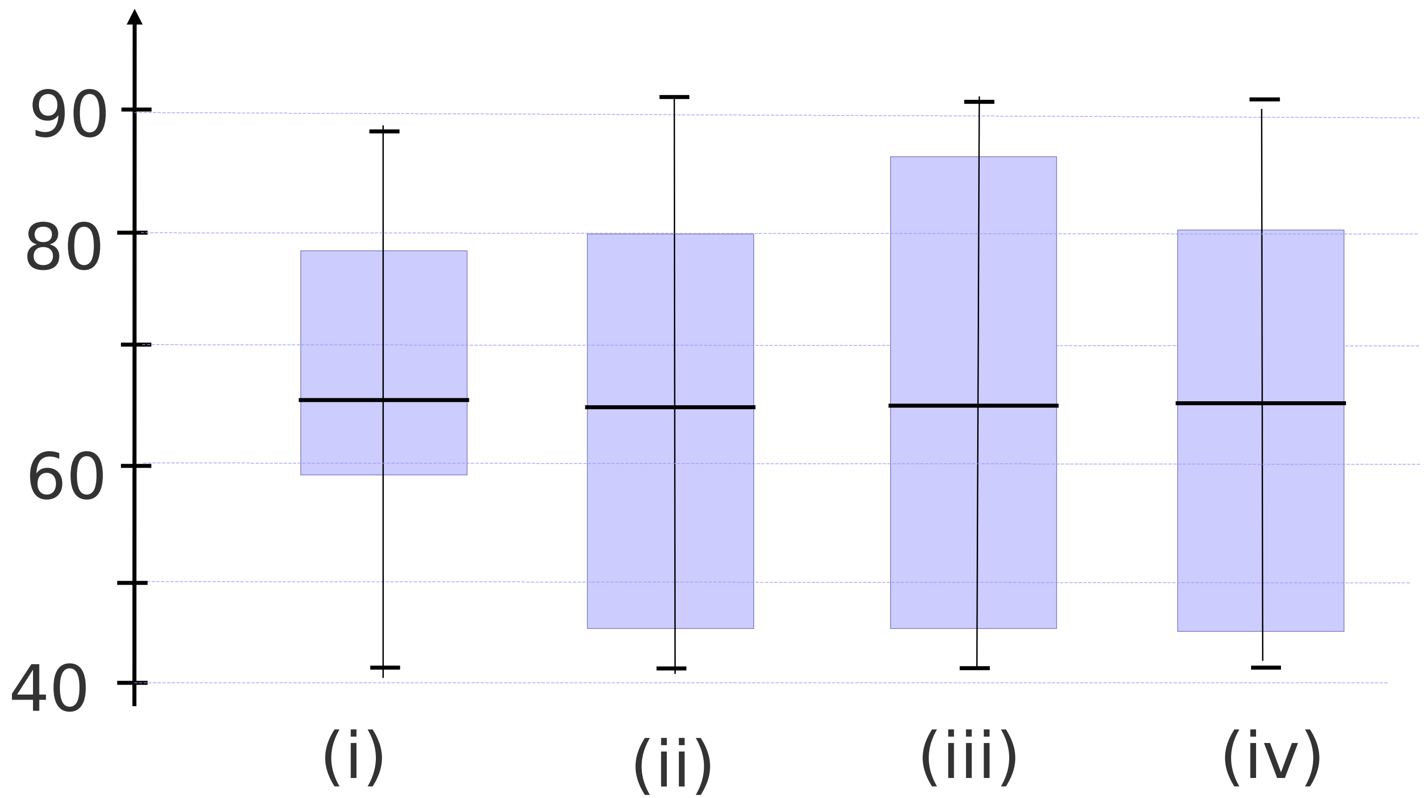
(i) 28 (ii) 35 (iii) 42 (iv) 999

$$\frac{7!}{(7-4)! \times 4!} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{4 \cdot 3 \cdot 2 \cdot 1} = 35$$

Q.8

Choose a suitable boxplot to the data below: ⑭

「42, 43.5, 46.9, 56.9, 63.8, 70.2, 78.2, 86.6, 88.5, 91.0」

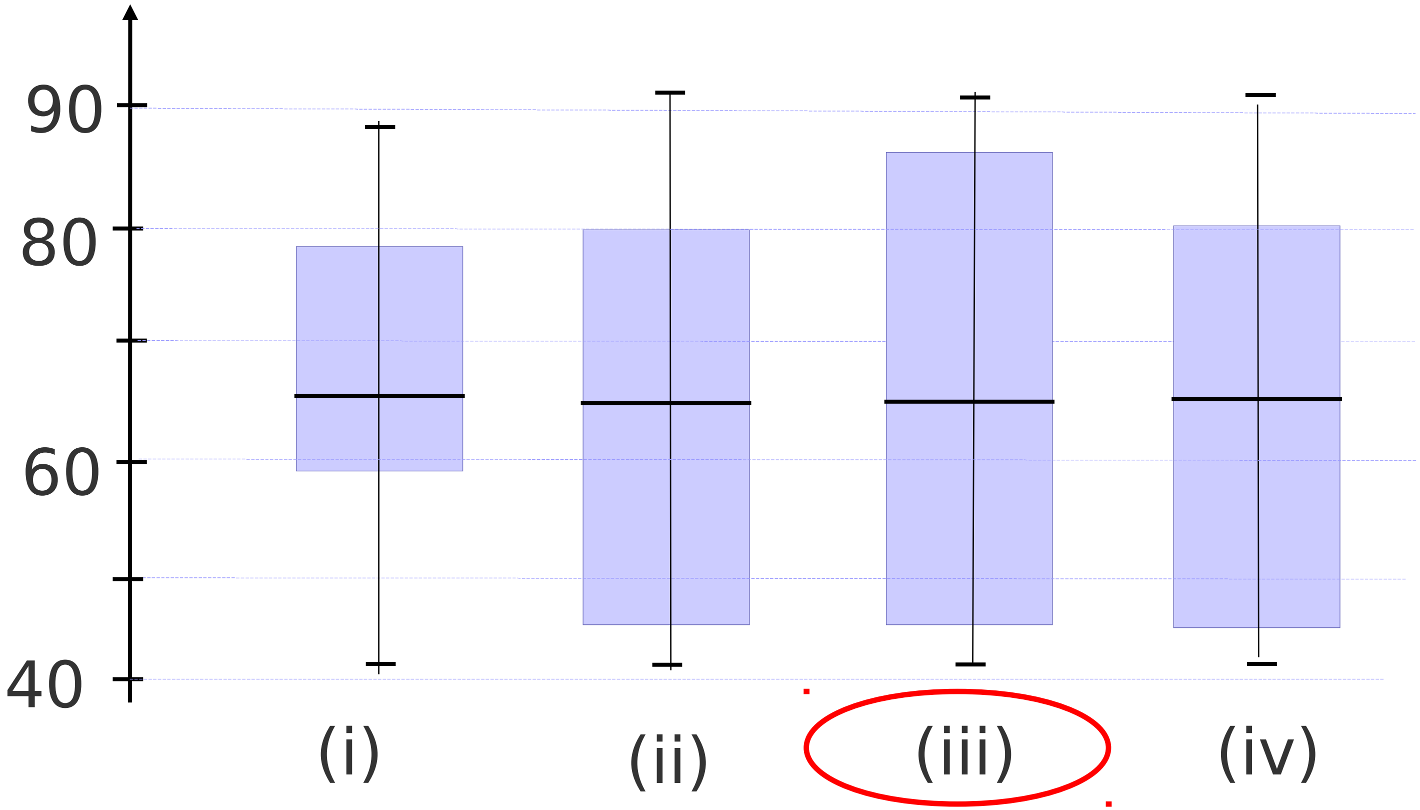


A.8

Choose a suitable boxplot to the data below: ⑭

「42, 43.5, 46.9, 56.9, 63.8, 70.2, 78.2, 86.6, 88.5, 91.0」

Min ： 42, 1st quartile ： 46.9 Median ： 67 3rd quartile ： 86.6



Q.9

The table below shows the heights of 10 students.

①The mean is ⑮

(i) 167.1cm (ii) 168.3cm (iii) 169.3cm (iv) 170.2cm

Student	1	2	3	4	5	6	7	8	9	10
Height [cm]	173	162	168	175	183	155	164	170	162	171

②The variance is ⑯

(i) 57.1 (ii) 56.8 (iii) 69.3 (iv) 70.2

③The median ⑰

(i) 167.5 (ii) 168cm (iii) 168.5cm (iv) 169cm

A.9

The table below shows the heights of 10 students.

① The mean is ⑮

(i) 167.1cm (ii) 168.3cm (iii) 169.3cm (iv) 170.2cm

Student	1	2	3	4	5	6	7	8	9	10
Height [cm]	173	162	168	175	183	155	164	170	162	171

② The variance is ⑯

(i) 57.1 (ii) 56.8 (iii) 69.3 (iv) 70.2

③ The median is ⑰

(i) 167.5 (ii) 168cm (iii) 168.5cm (iv) 169cm

155 162 162 164 168 170 171 173 175 183

A.9

(i) 167.1cm (ii) 168.3cm (iii) 169.3cm (iv) 170.2cm

学生	1	2	3	4	5	6	7	8	9	10
身長	173	162	168	175	183	155	164	170	162	171
偏差	4.7	-6.3	-0.3	6.7	14.7	-13.3	-4.3	1.7	-6.3	2.7
偏差 ^{2}	22.09	39.69	0.09	44.89	216.09	176.89	18.49	2.89	39.69	7.29

Mean:

56.81(=Variance
)

(i) 57.1 (ii) 56.8 (iii) 69.3 (iv) 70.2

(i) 167.5 (ii) 168cm (iii) 168.5cm (iv) 169cm

昇順に並べると

155 162 162 164 168 170 171 173 175 183

Q.10

A certain lottery has one winning piece out of 40.

The distribution of winning pieces follows the Poisson distribution. Then, find the probability that

3 winning pieces are included

in 10 you take out.

⑱

(i) 0.002 (ii) 0.004 (iii) 0.006 (iv) 0.008 (v) 0.01

A.10

A certain lottery has one winning piece out of 40.

The distribution of winning pieces follows the Poisson distribution. Then, find the probability that

3 winning pieces are included

in 10 you take out.

⑱

(i) 0.002 (ii) 0.004 (iii) 0.006 (iv) 0.008 (v) 0.01

Find λ first !

$\lambda = 10 * 1/40 = 0.25$

$P(X=3) = 0.25^3 \times e^{-0.25} / 3! = 0.002$

Q.11

In a certain pond, there 10 marked fish and 200 non-marked fish. Now, suppose you take out 5 fish from this pond.

① The probability that just one marked fish is in the 5 is ⑲

(i) 0.2 (ii) 0.3 (iii) 0.4 (iv) 0.5 (v) 0.6

② If you repeat this trial in the catch-and-release manner, the expected value of the marked fish in the 5 is ⑳

(i) 0.13 (ii) 0.24 (iii) 0.67 (iv) 0.98 (v) 2.07

A.11

In a certain pond, there 10 marked fish and 200 non-marked fish. Now, suppose you take out 5 fish from this pond.

① The probability that just one marked fish is in the 5 is ⑲

(i) 0.2 (ii) 0.3 (iii) 0.4 (iv) 0.5 (v) 0.6

$$f(x) = {}_{M}C_{x} \cdot {}_{N\text{-}M}C_{n\text{-}x} / {}_{N}C_{n}$$

$$f(1) = {}_{10}C_{1} \cdot {}_{200}C_{4} / {}_{210}C_{5} \fallingdotseq 0.2$$

② If you repeat this trial in the catch-and-release manner, the expected value of the marked fish in the 5 is ⑳

np = 5*10/210 = 0.24

(i) 0.13 (ii) 0.24 (iii) 0.67 (iv) 0.98 (v) 2.07

Q.12

- Suppose you roll a dice 4 times. Then, the probability that the pip of 4 appears just twice is ㉑

- (i) 0.025 (ii) 0.116 (iii) 0.135 (iv) 0.141 (v) 0.225

- Suppose you roll a dice 4 times. Then, the probability that the pip of 4 appears once or more is ㉒

(i) 0.48 (ii) 0.52 (iii) 0.65 (iv) 0.15 (v) 0.35

A.12

- Suppose you roll a dice 4 times. Then, the probability that the pip of 4 appears just twice is ㉑
- (i) 0.025 (ii) 0.116 (iii) 0.135 (iv) 0.141 (v) 0.225

$f(x) = {}_nC_x\, p^x (1-p)^{n-x}$

$f(2) = {}_4C_2 (1/6)^2 (1-1/6)^2 = 6(1/36)(25/36) \fallingdotseq 0.116$

- Suppose you roll a dice 4 times. Then, the probability that the pip of 4 appears once or more is ㉒

(i) 0.48 (ii) 0.52 (iii) 0.65 (iv) 0.15 (v) 0.35

$P(X \geqq 1) = 1 - P(X = 0)$

$= 1 - P(0) = 1 - (5/6)^4$

$= 0.52$

Q.13

In a certain call center, they receive 4 calls per 30 minutes on average. Then, the probability that they receive 5 calls within one hour is ㉓ .

Apply the Poisson dist.

(i) 0.011 (ii) 0.045 (iii) 0.63 (iv) 0.092 (v) 0.385

A.13

In a certain call center, they receive 4 calls per 30 minutes on average. Then, the probability that They receive 5 calls within one hour is $\boxed{㉓}$.

Apply the Poisson dist.

(i) 0.011 (ii) 0.045 (iii) 0.63 (iv) 0.092 (v) 0.385

Find λ and its unit !

Set “one hour” as the time unit here. Then, λ=8[calls/unit time]

$P(x=5) = 8^5 \times e^{-8}/5! = 0.092$

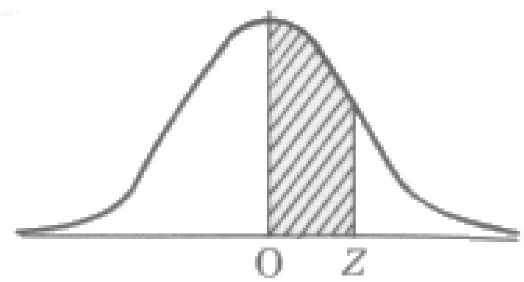
Q.14

Let a r.v. z follows the z-dist. Then,

$P\ (0 < Z < 1$ ㉔

(i) 0.1179 (ii) 0.3413 (iii) 0.2514 (iv) 0.3372 (v) 0.433

Z-table



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.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

A.14

Let a r.v. z follows the z-dist. Then,

$$P\ (0 < Z < 1 \quad ㉔$$

(i) 0.1179 (ii) 0.3413 (iii) 0.2514 (iv) 0.3372 (v) 0.433

Q.15

In a certain store, 2 customers arrive per hour on average. Now, if you assume that the customer intervals follow the exponential distribution, the probability that a customer interval is 5 minutes or less is $\boxed{㉕}$

(i) 0.153 (ii) 0.23 (iii) 0.34 (iv) 0.48 (v) 0.52

A.15

In a certain store, 2 customers arrive per hour on average. Now, if you assume that the customer intervals follow the exponential distribution, the probability that a customer interval is 5 minutes or less is ㉕

(i) 0.153 (ii) 0.23 (iii) 0.34 (iv) 0.48 (v) 0.52

λ＝ 2 ［customers／hour］， x＝1/12 ［hour］，

$P(X \leqq 1/12) = 1 - e^{-\lambda x} = 1 - e^{-2/12}$

$= 0.153$

Q.16

The measured customer arrival intervals of a certain Amusement park were as below [sec].

15.5 28.4 58.4 39.2

Assuming that the intervals follow the exponential distribution, find the MLE of λ. ㉖

(i) 0.0015 (ii) 0.028 (iii) 0.375 (iv) 35.375 (v) 55.88

Next, by using the MLE above, the probability that The next customer arrivals within on minute is ㉗

(i) 0.75 (ii) 0.79 (iii) 0.81 (iv) 0.86 (v) 0.92

A.16

The measured customer arrival intervals of a certain Amusement park were as below [sec].

15.5 28.4 58.4 39.2

Assuming that the intervals follow the exponential distribution, find the MLE of λ. ㉖

(i) 0.0015 (ii) 0.028 (iii) 0.375 (iv) 35.375 (v) 55.88

(15.5+28.4+58.4+39.2)/4 = 35.375

λ = 1/35.375 = 0.0283[customers/sec]

Next, by using the MLE above, the probability that The next customer arrivals within on minute is ㉗

(i) 0.75 (ii) 0.79 (iii) 0.81 (iv) 0.86 (v) 0.92

$P(X \leqq 60) = 1-e^{-\lambda x} = 1-e^{-0.028*60} = 0.813$

Q.17

The table below shows the socres of 8 students in the math exam.

The mean is ㉘, and the median is ㉙.

(i) 63.375 (ii) 66 (iii) 66.125 (iv) 65.625 (v) 64.5

SD is ㉚

(i) 12.15 (ii)13.96 (iii) 14.93 (iv) 15.88 (v) 24.37

Students	A	B	C	D	E	F	G	H
Score	62	52	34	81	76	65	73	64

A.17

The table below shows the socres of 8 students in the math exam.

The mean is ㉘ , and the median is ㉙ .

(i) 63.375 (ii) 66 (iii) 66.125 (iv) 65.625 (v) 64.5

SD is ㉚

(i) 12.15 (ii)13.96 (iii) 14.93 (iv) 15.88 (v) 24.37

Students	A	B	C	D	E	F	G	H
Score	62	52	34	81	76	65	73	64

34, 52, 62, 64, 65, 73, 76, 81

A.17

Students	A	B	C	D	E	F	G	H
Score	62	52	34	81	76	65	73	64
Dev.	-1.375	-11.375	-29.375	17.625	12.625	1.625	9.625	0.625
Dev.^{2}	1.890625	129.390625	862.890625	310.640625	159.390625	2.640625	92.640625	0.390625

63.375

Mean194.984

Q.18

The diameter of a certain product of a certain factory is specified as 1.54 inches. Now, as a result of a sample survey, the measured one sample’s diameter was 1.5475 inches. The population variance is known to be 0.0001. Then, check if the hypothesis “The spec is satisfied.” is Rejected or not. Set the significance level at 5%.

㉛

(i) Not rejected

(ii) Rejected (the spec is not met.)

A.18

The diameter of a certain product of a certain factory is specified as 1.54 inches. Now, as a result of a sample survey, the measured one sample's diameter was 1.5475 inches. The population variance is known to be 0.0001. Then, check if the hypothesis "The spec is satisfied." is Rejected or not. Set the significance level at 5%.

(i) Not rejected

(ii) Rejected (the spec is not met.)

SD=0.01. Thus,

$(X-\mu)/\sigma =$ （1.5475-1.54） $/0.01 = 0.75 < 1.96$.

Not rejected.

Q.19

We like to estimate the average height of Japanese students, and prepared 5 samples. The population variance is known to be 121.

[Samples] 150cm, 155cm, 160cm, 165cm, 175cm

①The sample mean is ㉜

(i) 159 (ii) 161 (iii) 163 (iv) 165 (v) 170

②The 95% CI is ㉝

(i) 151.4～170.6 (iii) 161.2～170.8

(ii) 151.2～172.5 (iv) 161.2～172.5

③In order to make the width of CI 10cm or less, at least how many samples are needed? ㉞

(i) 15 (ii) 16 (iii) 19 (iv) 20 (v) 21

A.19

We like to estimate the average height of Japanese students, and prepared 5 samples. The population variance is known to be 121.

[Samples] 150cm, 155cm, 160cm, 165cm, 175cm

① The sample mean is ㉜

(i) 159 (ii) 161 (iii) 163 (iv) 165 (v) 170

② The 95% CI is ㉝

(i) 151.4 ～ 170.6 (iii) 161.2 ～ 170.8

(ii) 151.2

161 – 1.96*11/2.236=151.36

161 + 1.96*11/2.236=170.64

③ In order to make the width of CI 10cm or less, at least how many samples are needed? ㉞

(i) 15 (ii) 16 (iii) 19 (iv) 20 (v) 21

1.96*2*11/√ n ≦ 10 √ n ≧ 1.96*22/10 = 1.96*2.2

```python
import numpy as np
import scipy.stats as st

x=np.array([150,155,160, 165,175])
#sample size.
n=x.size

#sample mean
x_mean = x.mean()
print(x_mean)
print(n)
#Known S.D.
x_sd=11#np.std(x,ddof=1)

#significance level.
alpha_r=0.05

z=abs(st.norm.ppf(alpha_r/2))
x_min = x_mean - z*x_sd/np.sqrt(n)
x_max = x_mean + z*x_sd/np.sqrt(n)
print("C.I. : {0}≦μ≦{1}".format(x_min,x_max))

#another way.
st.norm.interval(alpha=0.95,loc=x_mean,scale=x_sd/np.sqrt(n))
```

```
C.I. : 151.3582520536576≦μ≦170.6417479463424

(151.3582520536576, 170.6417479463424)
```

Q.20

- In a certain experiment, we measured the pH –value of a

Solution 5 times. The results were:

6.7, 7.0, 7.2 7.3, 7.5

① The unbiased variance i ㉟

(i) 0.057 (ii)0.677 (iii) 0.085 (iv) 0.093 (v) 1.053

② The 95% CI is ㊱

(i) 5.43～7.76 (iii) 6.14～7.69

(ii) 5.12～7.27 (iv) 6.76～7.52

A.20

- In a certain experiment, we measured the pH –value of a

Solution 5 times. The results were:

6.7, 7.0, 7.2 7.3, 7.5

① The unbiased variance i ㉟

(i) 0.057 (ii)0.677 (iii) 0.085 (iv) 0.093 (v) 1.053

② The 95% CI is ㊱

(i) 5.43～7.76 (iii) 6.14～7.69

(ii) 5.12～7.27 (iv) 6.76～7.52

A.20

(i) 0.057 (ii)0.677 (iii) 0.085 (iv) 0.093 (v) 1.053

	6.7	7	7.2	7.3	7.5
偏差	-0.44	-0.14	0.06	0.16	0.36
偏差 ^{2}	0.1936	0.0196	0.0036	0.0256	0.1296

Mean:7.14

S^2=0.093

⇒S=0.305

$$t_4(0.025) = 2.78$$

(i) 5.43～7.76

(ii) 5.12～7.27

(iii) 6.14～7.69

(iv) 6.76～7.52

7.14 – S*t/2.236 = 6.76

7.14 + S*t/2.236 = 7.52

$$\boxed{\bar{x} - \frac{St_{\frac{\alpha}{2}}}{\sqrt{n}} < \mu < \bar{x} + \frac{St_{\frac{\alpha}{2}}}{\sqrt{n}}}$$

```
import numpy as np
import scipy.stats as st

x=np.array([6.7, 7.0, 7.2, 7.3, 7.5])

#sample size.
n=x.size

#sample mean
x_mean = x.mean()
print(x_mean)
print(n)
#Unknown S.D.
x_sd=np.std(x,ddof=1)
print(x_sd)

#significance level.
alpha_r=0.05

stat_t = abs(st.t.ppf(alpha_r/2,df=n-1))

x_min = x_mean - stat_t * x_sd/np.sqrt(n)
x_max = x_mean + stat_t * x_sd/np.sqrt(n)
print("信頼区間：{0}≦μ≦{1}".format(x_min,x_max))

#another way.
st.t.interval(alpha=0.95,df=n-1,loc=x_mean,scale=x_sd/np.sqrt(n))
```

```
信頼区間：6.761343371836054５≦μ≦7.518656628163947

(6.7613433718360545, 7.518656628163947)
```

T-table

	有意確率								
	0.10	0.05	0.01	0.001	両側	0.10	0.05	0.01	0.001
df	0.05	0.025	0.005	0.0005	片側	0.05	0.025	0.005	0.0005
1	6.3138	12.706	63.657	636.62	18	1.7341	2.1009	2.8784	3.922
2	2.9200	4.3027	9.9248	31.598	19	1.7291	2.0930	2.8609	3.883
3	2.3534	3.1825	5.8409	12.941	20	1.7247	2.0860	2.8453	3.850
4	2.1318	2.7764	4.6041	8.610	21	1.7207	2.0796	2.8314	3.819
5	2.0150	2.5706	4.0321	6.859	22	1.7171	2.0739	2.8188	3.792
6	1.9432	2.4469	3.7074	5.959	23	1.7139	2.0687	2.8073	3.767
7	1.8946	2.3646	3.4995	5.405	24	1.7109	2.0639	2.7969	3.745
8	1.8595	2.3060	3.3554	5.041	25	1.7081	2.0595	2.7874	3.725
9	1.8331	2.2622	3.2498	4.781	26	1.7056	2.0555	2.7787	3.707
10	1.8125	2.2281	3.1693	4.587	27	1.7033	2.0518	2.7707	3.690
11	1.7959	2.2010	3.1058	4.437	28	1.7011	2.0484	2.7633	3.674
12	1.7823	2.1788	3.0545	4.318	29	1.6991	2.0452	2.7564	3.659
13	1.7709	2.1604	3.0123	4.221	30	1.6973	2.0423	2.7500	3.646

Q.21

As a result of an exam, we got the following table.

Classes A and B are made of 50 and 40 students, resp.

	Mean	**SD**
Class-A	70	9
Class-B	80	12

i) Find the total mean.

ii) Find the total SD.

iii) Afterward, the result of Class-C, made of 45 students was found. The mean and SD of Class-C were 72 and 5, resp. Then, find the total SD of these 3 classes.

A.21

	Mean	SD
Class-A	70	9
Class-B	80	12

i) Find the total mean.

(70*50+80*40)/90

[1] **74.44444**

ii)Find the total SD.

$$V[X] = E[X^2] - \bar{x}^2 = \frac{n_A\left(S_A^2 + \bar{x}_A^2\right) + n_B\left(S_B^2 + \bar{x}_B^2\right)}{n} - \bar{x}^2$$

```
> var_total = ( 50*(9^2 + 70^2) + 40*(12^2+80^2) )/90 - 74.44444^2
> sqrt(var_total)
[1] 11.56253
```

```
import numpy as np
import scipy.stats as st

M_ab = (70*50+80*40)/90
print("M_ab is", M_ab)

V_ab = ( 50*( pow(9,2) + pow(70,2)) + 40*( pow(12,2) + pow(80,2)) )/90 - pow(M_ab,2)
print("V_ab is", V_ab)
sd_ab = np.sqrt(V_ab)
print("sd_ab is", sd_ab)
```

```
M_ab is 74.44444444444444
V_ab is 133.6913580246919
sd_ab is 11.562497914581083
```

A.21

i) Afterward, the result of Class-C, made of 45 students was found. The mean and SD of Class-C were 72 and 5, resp. Then, find the total SD of these 3 classes.

Find the new total mean first.

```
> mean_3classes = (70*50 + 80*40 + 72*45)/(50+40+45)
> mean_3classes
[1] 73.62963
```

Then, by using

$$V[X] = E[X^2] - \bar{x}^2 = \frac{\sum_{l=1}^{k} n_l (S_l^2 + x_l^2)}{n} - \bar{x}^2$$

```
> var_3classes = ( 50*(9^2 + 70^2) + 40*(12^2+80^2) + 45* (5^2+72^2) )/(50+40+45) - mean_3classes^2
> sqrt(var_3classes)
[1] 9.939253
>
```

```
#######(iii)

M_abc = (70*50+80*40 + 72*45)/(50+40+45)
print("M_abc is", M_abc)

V_abc = ( 50*( pow(9,2) + pow(70,2)) + 40*( pow(12,2) + pow(80,2)) + 45*( pow(5,2) + pow(72,2)) )/(50+40+45) - pow(M_abc,2)
print("V_abc is", V_abc)
sd_abc = np.sqrt(V_abc)
print("sd_abc is", sd_abc)
```

```
M_abc is 73.62962962962963
V_abc is 98.78875171467735
sd_abc is 9.939253076296898
```