

2021 Statistics and data analysis I

Week 6
"Review"

Takashi Sano and Hirotada Honda

INIAD

Lecture plan

Week1: Introduction of the course and some mathematical preliminaries

Week2: Overview of statistics, One dimensional data(1): frequency and histogram

Week3: One dimensional data(2): basic statistical measures

Week4: Two dimensional data(1): scatter plot and contingency table

Week5: Two dimensional data(2): correlation coefficients, simple linear regression and concepts of Probability /

Probability(1):randomness and probability, sample space and probabilistic events

Week6:Probability(2): definition of probability, additive theorem, conditional probability and independency

Week7:Review and exam(i)

Week8: Random variable(1): random variable and expectation

Week9: Random variable(2): Chebyshev's inequality, Probability distribution(1):binomial and

Poisson distributions

Week10: Probability distribution(2): normal and exponential distributions

Week11: From descriptive statistics to inferential statistics -z-table and confidwncw interval-

Week12: Hypothesis test(1) -Introduction, and distributions of test statistic (t-distribution)-

Week13: Hypothesis test(2) -Test for mean-

Week14: Hypothesis test(3) -Test for difference of mean-

Week15: Review and exam(2)





Agenda

Today we introduce the past test of this course (the mid-term exam in 2018) as an exercise. We'll give an explanation after each question.



Exam

Choose the correct statistical measures for the items below.

- Amount of money spent for hobby in one year.
- 2

(1)

The ranking of football players.

3

- (3) T-scores of students in the exams of physics.
- (4) The reputation of a song evaluated with five choices ("very good/good/normal/bad/ definitely bad").
- (5) Birth places of students.

(5)

4

(ii) Interval scale (iii) Ordinal scale (iv) Nominal scale (i)Ratio scale



Q.1 (Answers)

Choose the correct statistical measures for the items below.

- (1) Amount of money spent for hobby in one year. (i
- (2) The ranking of football players. (iii)
 - (3) T-scores of students in the exams of physics. (ii)
 - (4) The reputation of a song evaluated with five choices ("very good/good/normal/bad/ definitely bad"). (iii
 - (5) Birth places of students.

(i)Ratio scale (ii) Interval scale (iii) Ordinal scale (iv) Nominal scale



Choose the correct answers.

(1)
$${}_{5}C_{3} = \frac{6}{2}$$
 5 (ii) 10 (iii) 15 (iv) 25

(2)
$$_{3}P_{3} = \frac{0}{100}$$
 1 (ii) 3 (iii) 6 (iv) 10

(3)
$$_{4}C_{2} = \frac{8}{100}$$
 i) 1 (ii) 4 (iii) 6 (iv) 12

(4)
$$_{4}^{2}P_{2}^{2}=$$
 9 1 (ii) 4 (iii) 6 (iv) 12

(5)
$$_{6}C_{2} =$$
 (i) 1 (ii) 6 (iii) 12 (iv) 15

Q.2[Answer]

- (1) ₅C₃ (i) 5 (ii) 10 (iii) 15 (iv) 25
- (2) $_{3}P_{3}$ (i) 1 (ii) 3 (iii) 6 (iv) 10
- (3) $_{4}C_{2}$ (i) 1 (ii) 4 (iii) 6 (iv) 12
- (4) $_{4}P_{2}$ (i) 1 (ii) 4 (iii) 6 (iv) 12
- (5) $_{6}C_{2}$ (i) 1 (ii) 6 (iii) 12 (iv) 15

The following table shows the height of 10 students.

Then, the mean of them is . ①

Students	1	2	3	4	5	6	7	8	9	10
Height [cm]	178	165	168	175	163	155	164	170	162	171

(i) 167.1cm (ii) 169.3cm (iii) 165.3cm (iv) 167.2cm

(i) 167.1cm (ii) 169.3cm (iii) 165.3cm (iv) 167.2cm



Find the number of ways for each item below.

- (1) The ways of exhibiting 3 cars chosen from 4 in a showroom.
 - (i) 24 (ii) 48 (iii) 56 (iv) 96



(i) 124 (ii) 168 (iii) 256 (iv) 336



(i) 64 (ii) 126 (iii) 256 (iv) 512



- (i) 12 (ii) 16 (iii) 18 (iv) 24
- (5) The ways of making a group of 2 students chosen from 6 students.
 - (i) 9 (ii) 12 (iii) 15 (iv) 20











(15)



Q.4[Answer]

Find the number of ways for each item below.

- (1) The ways of exhibiting 3 cars chosen from 4 in a showroom.
 - (i) 24 (ii) 48 (iii) 56 (iv) 96
- (2) The ways of exhibiting 3 cars chosen from 8 in a showroom.
 - (i) 124 (ii) 168 (iii) 256 (iv) 336
- (3) The ways of dividing 10 students into 2 groups.
 - (i) 64 (ii) 126 (iii) 256 (iv) 512
- (4) There is a group made of 4 boys and 3 girls. Then, when you choose 2 students from this group, find the number of situations in which at least one boy is included.
 - (i) 12 (ii) 16 (iii) **(**8 (iv) **2**4
- (5) The ways of making a group of 2 students chosen from 6 students.
 - (i) 9 (ii) 12 (iii) 15 (iv) 20





The table below shows the numbers of correct answers of 5 students for a quiz made of 10 questions.

Students	A	В	C	D	Е
Number of correct answers	8	3	2	6	1



- ②Find the median. (i) 3 (ii) 4 (iii) 5 (iv) 6 (v) 7
- ③Find the variance. (i) 3.2 (ii) 4.3 (iii) 5.5 (iv) 6.8 (v) 8.5
- 4) Find the z-score of student A.
- (i) 1.12 (ii) 1.36 (iii) 1.54 (iv) 1.68 (v) 1.78
- (5) Find the T-score of student C.
- (i) -0.77 (ii) 30.34 (iii) 42.31 (iv) 43.86 (v) 55.68











Q.5[Answer]

Students	Α	В	С	D	Е	
Correct	8	3	2	6	1	4
Deviation	4	-1	-2	2	-3	681=Variance
Deviation ²	16	1	4	4	9	6.8(=Variance) S.D. About 2.6

The table below shows the numbers of correct answers of 5 students for a quiz made of 10 questions.

- 1) Find the mean. (i) 3 (ii) 4 (iii) 5 (iv) 6 (v) 7
- ②Find the median. (i) 3 (ii) 4 (iii) 5 (iv) 6 (v) 7
- (3) Find the variance. (i) 3.2 (ii) 4.3 (iii) 5.5 (iv) 6.8 (v) 8.5
- 4) Find the z-score of student A.
- (i) 1.12 (ii) 1.36 (iii) 1.54 (iv) 1.88 (v) 1.78
- 5 Find the T-score of student C.
- (i) -0.77 (ii) 30.34 (iii) 42.31 (iv) 43.86 (v) 55.68





Student	Α	В	С	D	Е	
Correct answers	8	3	2	6	1	4
Deviation	4	-1	-2	2	-3	A (O () (auxi aux a a)
Deviation ²	16	1	4	4	9	6.8(=Variance) S.D. About 2.6
Z	1.538		-0.769			3.D. ADOUT 2.0
Т			42.307			

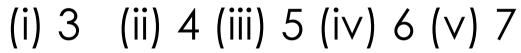
問5[Answer]



The table below shows the numbers of correct answers of 5 students for a quiz made of 10 questions.

Student	Α	В	С	D	Е
Number of correct answers	7	3	ŚŚ	6	9

1)The mean was known to be 6. Then, find the number of C.





- 2) Find the first quartile.
- (i) 3 (ii) 4 (iii) 5 (iv) 6 (v) 7



- 3Afterward, the score of a certain student was modified, whose rank was improved. However, the median and first quartile remained unchanged after the modification. Answer the student whose score was modified.
- (i) A (ii) B (iii) C (iv) D (v) E



Q.6[Answer]

The table below shows the numbers of correct answers of 5 students for a quiz made of 10 questions.

Student	Α	В	С	D	Е
Number of correct answers	7	3	ŚŚ	6	9

- 1)The mean was known to be 6. Then, find the number of C.
- (i) 3 (ii) 4 (iii) 5 (iv) 6 (v) 7
- 2Find the first quartile.
- (i) 3 ((ii) 4 (iii) 5 (iv) 6 (v) 7
- 3 Afterward, the score of a certain student was modified, whose rank was improved. However, the median and first quartile remained unchanged after the modification. Answer the student whose score was modelified.
- (i) A (ii) B (iii) C (iv) D (v) E





The table below shows the result of exams of math and physics in a certain class. The scores are integer, and if you answer completely, your score will be one-hundred.

	Mean	Median	Minimum	Maximum	Standard deviation
Math	65.8	66.0	56	77	5.45
Physics	65.9	65.0	54	77	6.06
Covariano	e:29.013				

(1) Find the correlation coefficient.

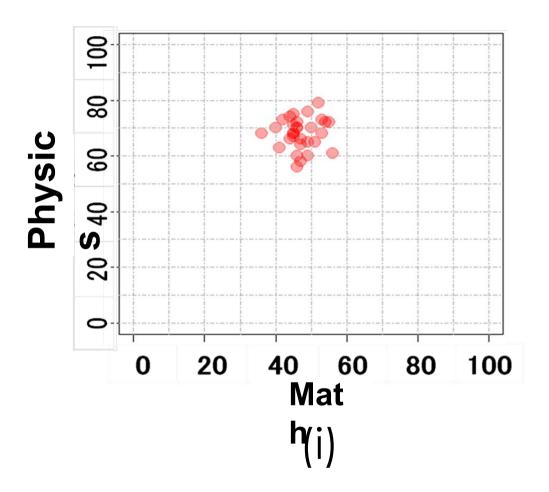


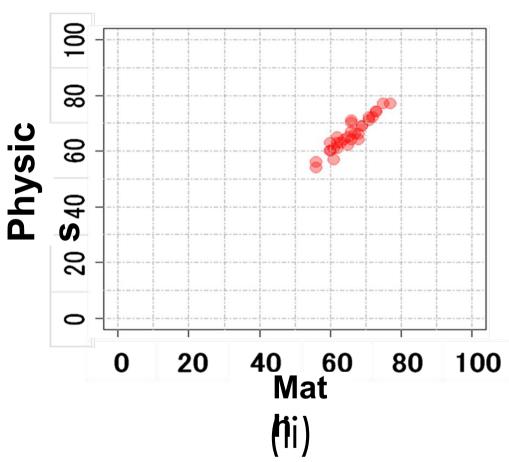
(i) 0.64 (ii) 0.72 (iii) 0.88 (iv) 0.91 (v) 0.96

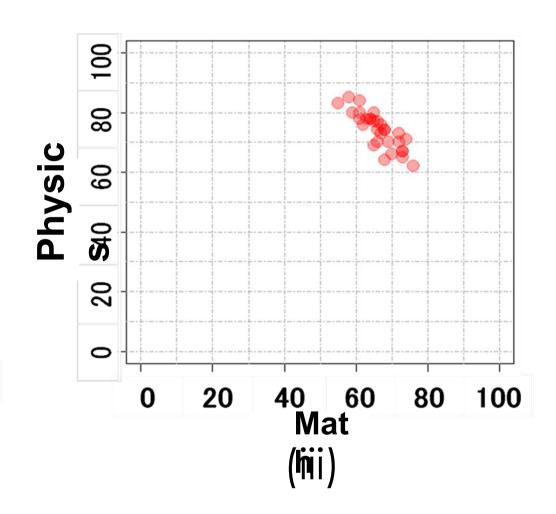
Q.7(Continued)

(2) Find a correct scatter plot.









Q.7(Continued)

(3) Afterward, the scores of physics exam were made twice. (The one who answered completely got the score of two hundred).

Then, choose all the correct choices.

- (i) The correlation coefficient will decrease.
- (ii) The correlation coefficient will increase.
- (iii) The correlation coefficient will remain unchanged.



Q.7[Answer]

The table below shows the result of exams of math and physics in a certain class. The scores are integer, and if you answer completely, your score will be one-hundred.

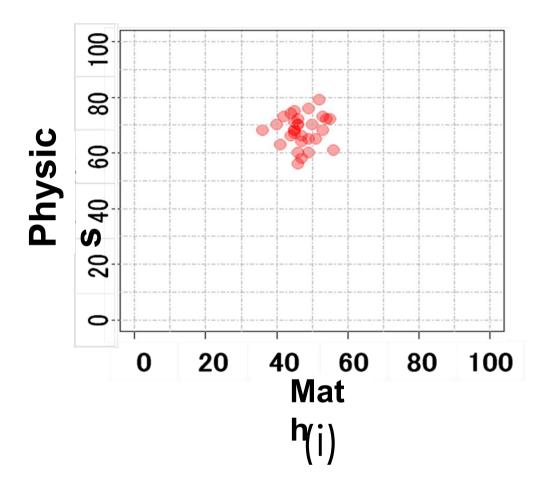
	Mean	Median	Minimum	Maximum	Standard deviation
Math	65.8	66.0	56	77	5.45
Physics	65.9	65.0	54	77	6.06
Covarianc	e:29.013				

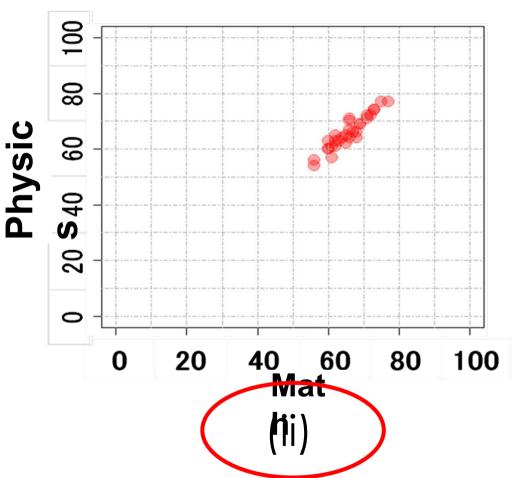
(1) Find the correlation coefficient.

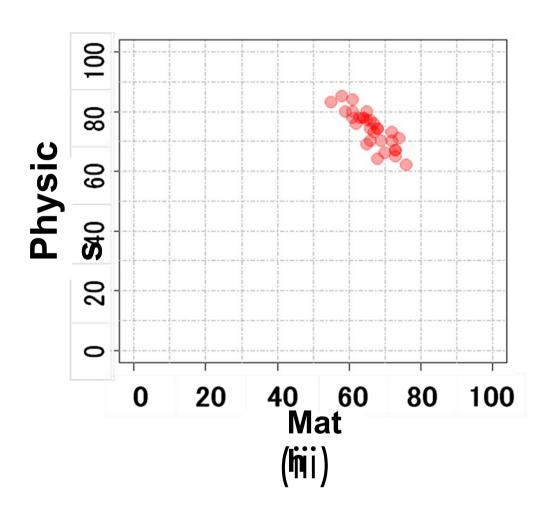
(i) 0.64 (ii) 0.72 (iii) 0.88 (iv) 0.91 (v) 0.96

Q.7(Continued) [Answer]

(2) Find a correct scatter plot.







Q.7(Continued) [Answer]

(3) Afterward, the scores of physics exam were made twice. (The one who answered completely got the score of two hundred). Then, choose all the correct choices.

- (i) The correlation coefficient will decrease.
- (ii) The correlation coefficient will increase.
- (iii) The correlation coefficient will remain unchanged.





The table below shows the numbers of correct answers of 8 students in a certain quiz made of 10 questions.

Students	A	В	С	D	E	F	G	Н
Number of correct	2	5	4	4	10	6	7	3
answers								

- (1) Find the mean.
- (i) 3.332 (ii) 4.825 (iii) 5.125 (iv) 5.334 (v) 6.012



- (2) Find the standard deviation. Use the supplementary table below if needed.
- (i) 1.639 (ii) 2.368(iii) 3.224 (iv) 4.634 (v) 5.892

Students	Α	В	С	D	E	F	G	Н
Number of correct answers	2	5	4	4	10	6	7	3
Deviation								
Deviation ²								



Q.8(continued)



The table below shows the numbers of correct answers of 8 students in a certain quiz made of 10 questions.

Students	A	В	С	D	Е	F	G	Н
Number of correct	2	5	4	4	10	6	7	3
answers								

- (3) Find the z-score of student E.
- (i) 1.332 (ii) 2.058 (iii) 3.223 (iv) 4.412 (v) 5.824
- (4) Afterward, the result of another quiz, made of 100 questions for the same students was conducted. Its mean and standard deviation were 5 and 2.0, respectively. Then, answer which result has larger variation.
- (i) Quiz of 10 questions (ii) Quiz of 100 questions (iii) Cannot be determined

Q.8 (Answer)



The table below shows the numbers of correct answers of 8 students in a certain quiz made of 10 questions.

Students	Α	В	С	D	E	F	G	Н
Number of correct	2	5	4	4	10	6	7	3
answers								

- (1) Find the mean.
- (i) 3.332 (ii) 4.825 (ii) 5.125 (iv) 5.334 (v) 6.012
- (2) Find the standard deviation. Use the supplementary table below if needed.
- (i) 1.639 (ii) 2.368(iii) 3.224 (iv) 4.634 (v) 5.892

Student	Α	В	С	D	E	F	G	Н
Score	2	5	4	4	10	6	7	3
Dev.	-3.125	-0.125	-1.125	-1.125	4.875	0.875	1.875	-2.125
Dev. ²								

Var.: 5.609

S.D.: 2.368





Student	A	В	С	D	E	F	G	Н
Score	2	5	4	4	10	6	7	3

(3) Find the Z-score of student-E.

(i) 1.332 (ii) 2.058 (iii) 3.223 (iv) 4.412 (v) 5.824

(4) Compare the CVs.

(i) 10Q-quiz (ii) 100q-quiz (iii) Not determined 10-q quiz: 2.368/5.125=0.462 100-q quiz: 2/5=0.4





There are 3 cards, A, B and C, whose sides are painted with red or blue as follows.

Card A: Both sides are painted with blue.

Card B: One side is painted with red, and another with blue.

Card C: Both sides are painted with red.

Now, suppose that you randomly take out one of these 3 cards from a bag, and then place it on a desk. Then, you see the blue side. Now, find the probability of the case that red side will appear

when you turn it over.

(i) 1/4 (ii) 1/3 (iii) 1/2 (iv) 2/3 (v) 3/4





There are 3 cards, A, B and C, whose sides are painted with red or blue as follows.

Card A: Both sides are painted with blue.

Card B: One side is painted with red, and another with blue.

Card C: Both sides are painted with red.

Now, suppose that you randomly take out one of these 3 cards from a bag, and then place it on a desk. Then, you see the blue side. Now, find the probability of the case that red side will appear

when you turn it over.

(i) 1/4 (ii) 1/3 (iii) 1/2 (iv) 2/3 (v) 3/4

A:Turned over, then red B:Blue at first

ANB={Card-B is chosen, and in addition, the blue side is upward}

 $P(A \mid B) = P(A \cap B) / P(B) = 1/6 / \frac{1}{2} = 1/3$

Mr. Ichiro, Jiro, Saburo and Siro get a strike with the probability of 60%, 70%, 90% and 98%, respectively.

Now, suppose somebody among them gets a strike. Then, find the probability that he is Ichiro.

(i)0.153 (ii)0.164 (iii) 0.189 (iv) 0.254



Q.10[Answer]

Mr. Ichiro, Jiro, Saburo and Siro get a strike with the probability of 60%, 70%, 90% and 98%, respectively.

Now, suppose somebody among them gets a strike. Then, find the probability that he is Ichiro.

Apply the Bayes' theorem.

 $P(H_1)$: Ichiro throws.

P(A): Gets a strike.

(i)0.153 (ii)0.164 (iii) 0.189 (iv) 0.254

- > ans1=1/4*0.6
- > ans2=1/4*0.6+1/4*0.7+1/4*0.9+1/4*0.98
- > ans1/ans2

[1] 0.1886792



The following frequency table shows the results of a certain exam of a class made of 50 students. If you answer completely at the exam, you get the score of 100. Now, fill in the blanks.

Class interval	Class value	Frequency	Cumulative frequency
0-19	10	2	35)
20-39	30	8	36
40-59	50	<u>34</u>)	37)
60-79	70	12	38
80-100	90	10	39





- 34 (i) 14 (ii) 16 (iii) 18 (iv) 20
- 35 (i) 2 (ii) 3 (iii) 4 (iv) 5
- 36 (i) 2 (ii) 6 (iii) 8 (iv) 10
- ③ (i) 20 (ii) 24(iii) 26 (iv) 28

- 38 (i) 20 (ii) 30 (iii) 40 (iv) 50
- 39 (i) 20 (ii) 30 (iii) 40 (iv) 50

- (1) Find the number of events that the sum of pips of two dice is equal to 7.(i) 5 (ii) 6 (iii) 7 (iv) 8
 - (2) Suppose you roll a dice, and let us define an event that the pip is even as "event A", and an event that the pip of 4 or more appear as "event B". Then, answer the event A^C\nB.
 - (i) $\{5\}$ (ii) $\{3,5\}$ (iii) $\{4,6\}$ (iv) $\{4,5,6\}$



Q12[Ans]

(1) Find the number of events that the sum of pips of two dice is equal to 7.

```
6
(1,6)(2,5) (3,4)(4,3)(5,2)(6,1)
(2)
(i) {5} {ii) {3,5} (iii) {4,6} (iv) {4,5,6}
```

- (1) Suppose that you are rolling 5 dice. Then, find the probability of an event that three of them shows the pip one, and other two show other pips.
 - (i) 0.012 (ii) 0.024 (iii) 0.032 (iv) 0.048 (v) 0.053
 - (2) Suppose that you roll 4 dice. Then, find the probability that two of them show the pip of 2, and others show other than 2. (i) 0.115 (ii) 0.224 (iii) 0.232 (iv) 0.348 (v) 0.432
 - (3) A certain exam is made of 5 questions, each of which are 3 choice question. Then, find the probability of the event you correctly answer 3 questions if you answer randomly.
 - (i) 0.122 (ii) 0.142 (iii) 0.132 (iv) 0.148 (v) 0.165



- (1)Suppose that you are rolling 5 dice. Then, find the probability of an event that three of them shows the pip one, and other two show other pips.
- (i) 0.012 (ii) 0.024 (iii) 0.032 (iv) 0.048 (v) 0.053
 - (2)Suppose that you roll 4 dice. Then, find the probability that two of them show the pip of 2, and others show other than
 - 6*1/36*25/36 = 0.115
 - (i) 0.115 (ii) 0.224 (iii) 0.232 (iv) 0.348 (v) 0.432
 - (3) A certain exam is made of 5 questions, each of which are 3 choice question. Then, find the probability of the event you correctly answer 3 questions if you answer randomly.
 - (i) 0.122 (ii) 0.142 (iii) 0.132 (iv) 0.148 (v) 0.165



There are 3 boxes in which red and white balls are placed as below.

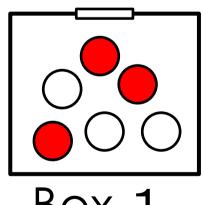
Now, suppose that you choose a box randomly and take a ball out of it. Then, under the situation that the ball is red, consider the most probable box from which you take the ball out. Let us call the event that the red ball is taken out as "event-A".

Below, the probability you choose Box-1 is assumed to be 1/3, and the probability you take a red ball out from Box-1 is denoted as p1, and the probability of an event-A is P(A). Then, under the situation a red ball is taken out, the probability that it was taken out from Box-1 is

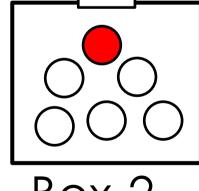


(ii) 1/3*p1/P(A)

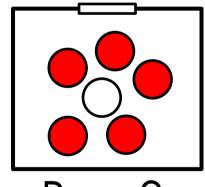
(iii) p1/P(A)-1/3(iv) 1/3+p1/P(A)



Box-1



Box-2



Box-3

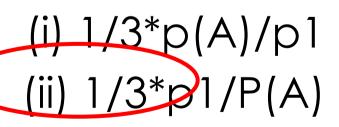


Q.14 (Answer)

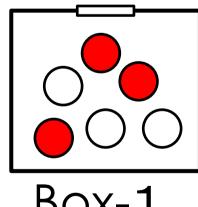
There are 3 boxes in which red and white balls are placed as below.

Now, suppose that you choose a box randomly and take a ball out of it. Then, under the situation that the ball is red, consider the most probable box from which you take the ball out. Let us call the event that the red ball is taken out as "event-A".

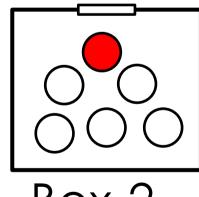
Below, the probability you choose Box-1 is assumed to be 1/3, and the probability you take a red ball out from Box-1 is denoted as p1, and the probability of an event-A is P(A). Then, under the situation a red ball is taken out, the probability that it was taken out from Box-1 is



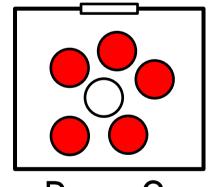
(iii) p1/P(A)-1/3(iv) 1/3+p1/P(A)



Box-1



Box-2



Box-3

As a result of a certain exam, both Mr. Yamamoto in Class-A and Mr. Takahashi in Class-B got the score of 65. It is known that the mean of classes A and B are 60 and 55, respectively, and the standard deviation, 5 and 7, respectively.

Then, compare the relative location of Mr. Yamamoto and Mr. Takahashi in each class.

- (i) Mr. Yamamoto is higher.
- (ii) Mr. Takahashi is higher.
- (iii) It is not determined.

As a result of a certain exam, both Mr. Yamamoto in Class-A and Mr. Takahashi in Class-B got the score of 65. It is known that the mean of classes A and B are 60 and 55, respectively, and the standard deviation, 5 and 7, respectively.

Then, compare the relative location of Mr. Yamamoto and Mr. Takahashi in each class.

Yamamoto: (65-60)/5 = 1Takahashi: (65-55)/7 = 1.43

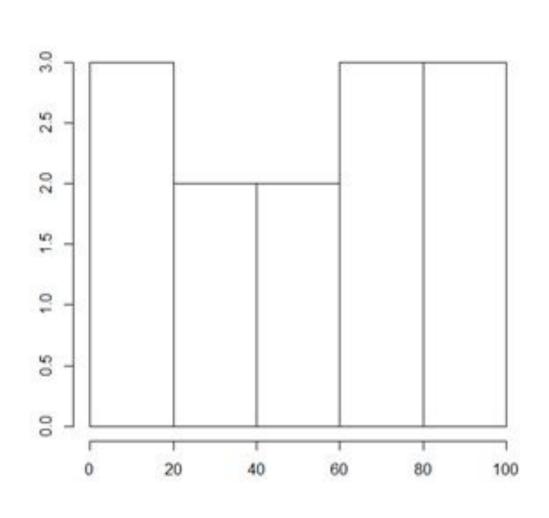
- (i) Mr. Yamamoto is higher.
- (ii) Mr. Takahashi is higher.
- (iii) It is not determined.

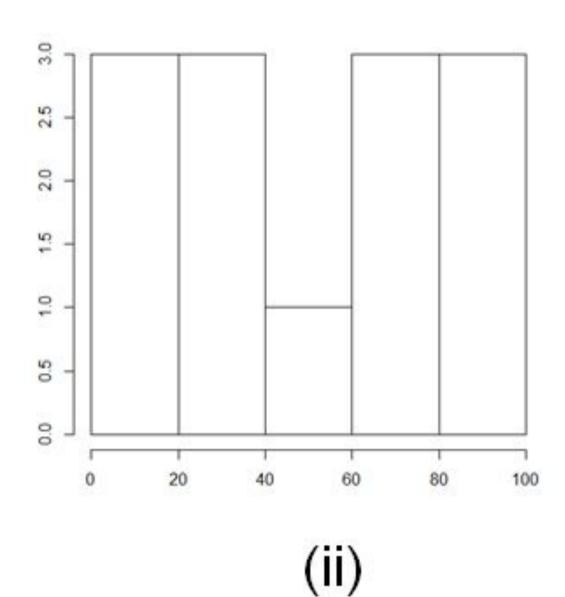


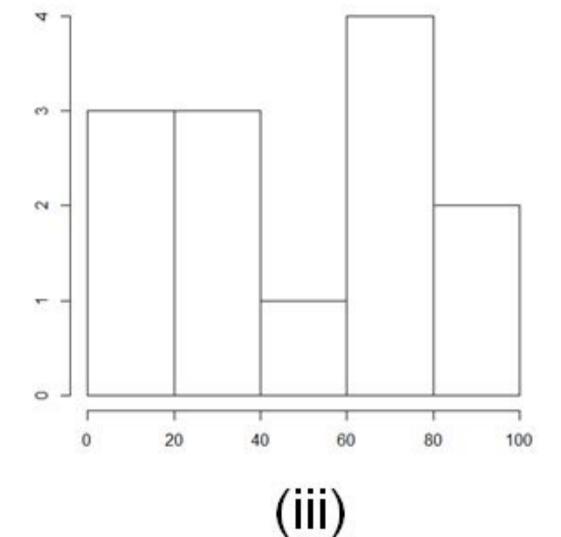
Choose a suitable histogram that matches the following data.

{7,8,15,21,29,41,53,64,64,69,88,92,95}



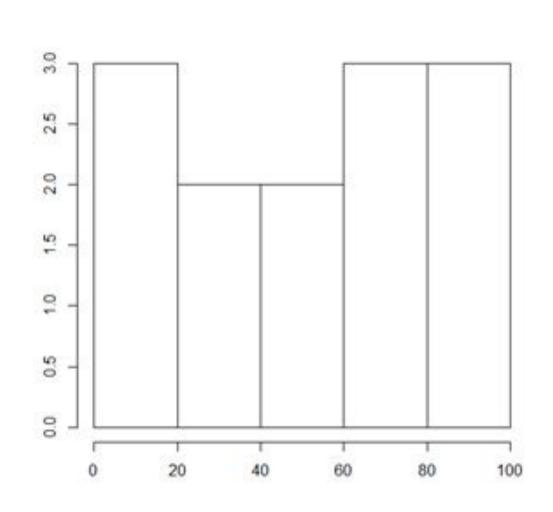


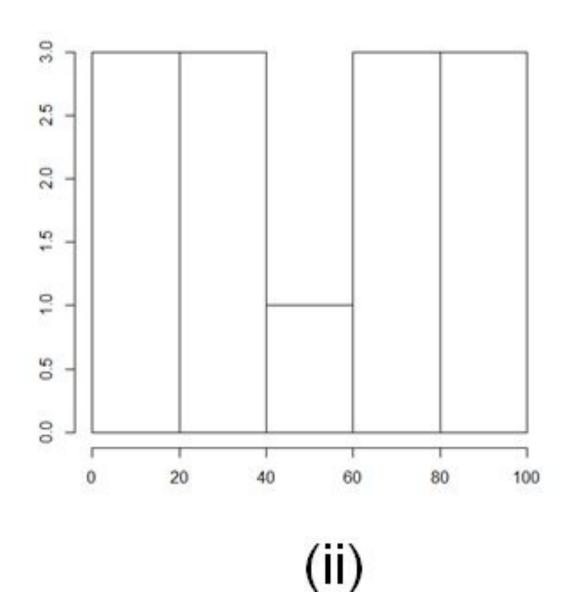


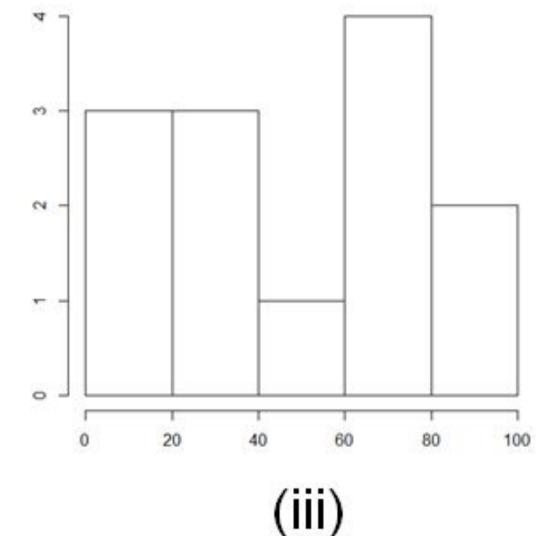


問16[Answer]

Choose a suitable histogram that matches the following data. {7,8,15,21,29,41,53,64,64,69,88,92,95}







The table below shows the monthly earnings of part-time job of 8 students. Answer the suitable representative value in this case.

- (i) Arithmetic mean.
- (ii) Geometric mean.
- (iii) Median.
- (iv)Range.

Students	A	В	C	D	E	F	G	Н
Earnings	30	31	29	90	30	28	32	31
[one thousand								
Yen]								



Q.17 Ans.

The table below shows the monthly earnings of part-time job of 8 students. Answer the suitable representative value in this case.

- (i) Arithmetic mean.
- (ii) Geometric mean.
- (iii) Median.
- (iv)Range.

Students	A	В	C	D	E	F	G	Н
Earnings	30	31	29	90	30	28	32	31
[one thousand								
Yen]								



In a certain store, they record the number of visits of 100 customers during a month. It is summarized in the table below.

Choose the most suitable sentence.



Quartile	1 st quartile	2 nd quartile	3 rd quartile
Number of visits	3	8	16

- (i) The majority of customers visit the store less than 8 times in the month.
- (ii) No customer visits the store more than 100 times.
- (iii) If you arrange the customers in the ascending order, the 25th person visits the store just 3 times.
- (iv) If you arrange the customers in the ascending order, the upper 20 % of them visit the store 16 times or more.
- (v) You cannot say (i)-(iv) from the table.





Q.18 Ans.

In a certain store, they record the number of visits of 100 customers during a month. It is summarized in the table below.

Choose the most suitable sentence.



Quartile	1 st quartile	2 nd quartile	3 rd quartile
Number of visits	3	8	16

- (i) The majority of customers visit the store less than 8 times in the month.
- (ii) No customer visits the store more than 100 times.
- (iii) If you arrange the customers in the ascending order, the 25th person visits the store just 3 times.
- (iv) If you arrange the customers in the ascending order, the upper 20 % of them visit the store 16 times or more.
- (v) You cannot say (i)-(iv) from the table.

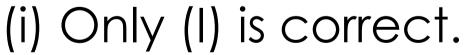


The following boxplots show the result of the handball throwing in classes A and B. The sentences (I) and (II) below concern this result.

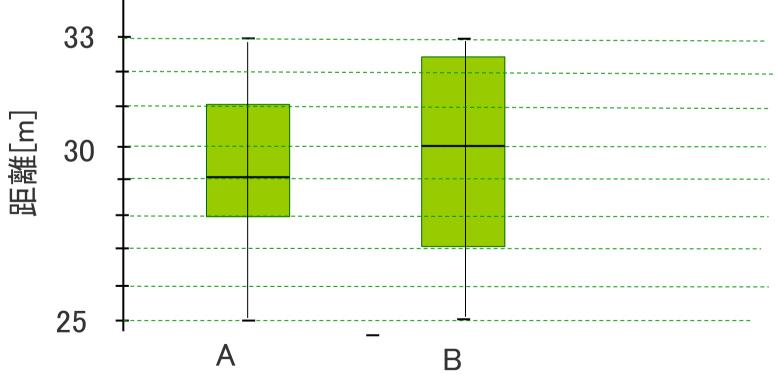
Then, answer the correct choice from (i)-(iv) below.

(I): The range of class-B is larger than that of class-A.

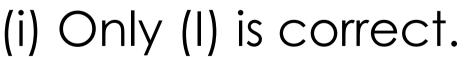
(II): The IQR of class-B is larger than that of class-A.



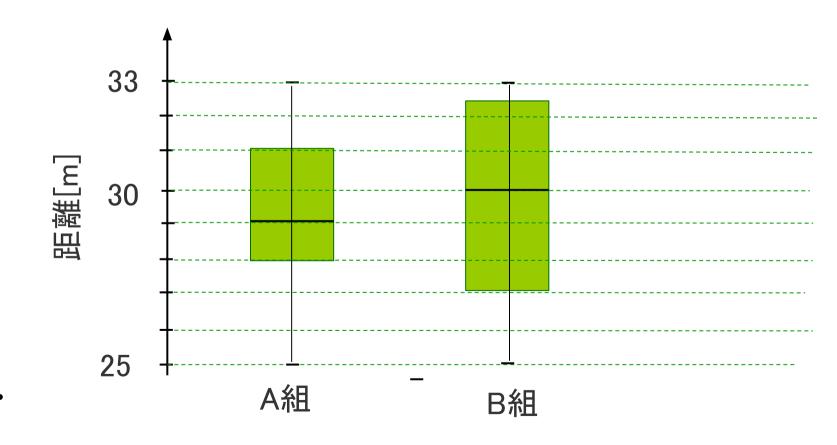
- (ii) Only (II) is correct.
- (iii) Both (I) and (II) are incorrect.
- (iv) Both (I) and (II) are correct.



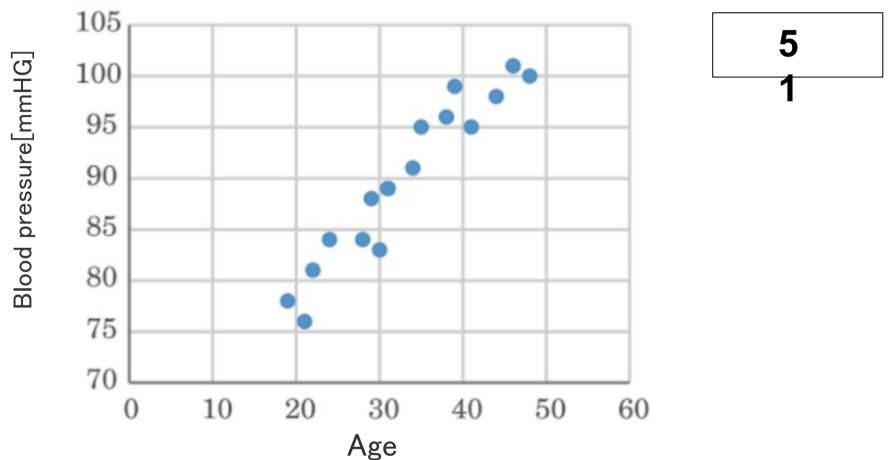




- (ii) Only (II) is correct.
- (iii) Both (I) and (II) are incorrect.
- (iv) Both (I) and (II) are correct.



The figure below shows the relationship between age and blood pressure. Then, choose the appropriate sentence from below.

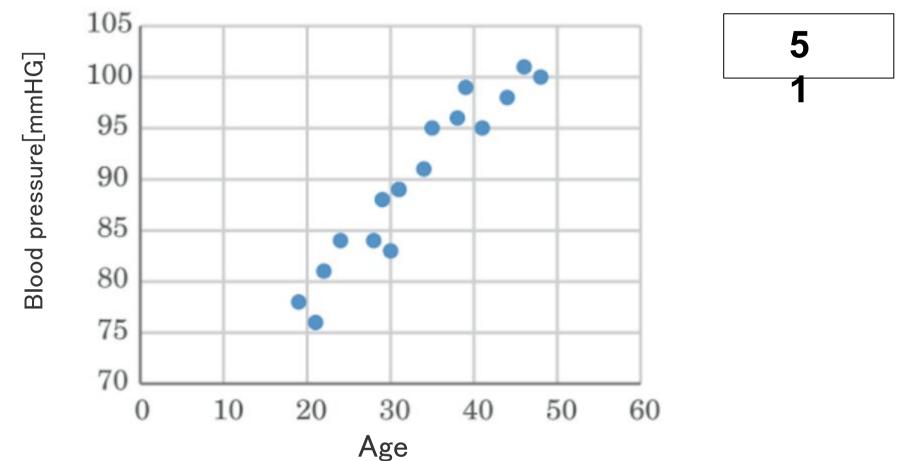


- (i) It seems that there is a positive correlation between age and blood pressure.
- (ii) It seems that there is a negative correlation between age and blood pressure.
- (iii) It seems that there is no correlation between age and blood pressure.
- (iv) As age increases by 10, then the blood pressure increases by about 30.



Q.20 Ans.

The figure below shows the relationship between age and blood pressure. Then, choose the appropriate sentence from below.



- (i) It seems that there is a positive correlation between age and blood pressure.
- (ii) It seems that there is a negative correlation between age and blood pressure.
- (iii) It seems that there is no correlation between age and blood pressure.
- (iv) As age increases by 10, then the blood pressure increases by about 30.



Fill in the blanks with correct words.

- If N samples are observed, the quotient of dividing the sum of data by N.
- $\frac{5}{\text{used}^3}$ in is the N-th root of the product of data. For instance, it is used to find the of a co $\frac{5}{4}$
- an is the reciprocal of the arithmetic mean of reciprocals of samples. For instance, it is used to find the of a $\frac{5}{6}$ regular route.
- (i) arithmetic (ii) geometric (iii) harmonic (iv) growth rate (v) mean velocity

Fill in the blanks with correct words.

- If N samples are observed, the quotient of dividing the sum of data by N.
- Geometri an is the N-th root of the product of data. For instance, it is used to find the of a contract of data.
- Harmon an is the reciprocal of the arithmetic mean of reciprocals of samples. For instance, it is used to find the regular route.
- (i) arithmetic (ii) geometric (iii) harmonic (iv) growth rate (v) mean velocity