

0.1. Өгөгдлийг илэрхийлэх

Осолд орсон дугуйчдын насны судалгаа гаргахад дараах хэлбэртэй байжээ.

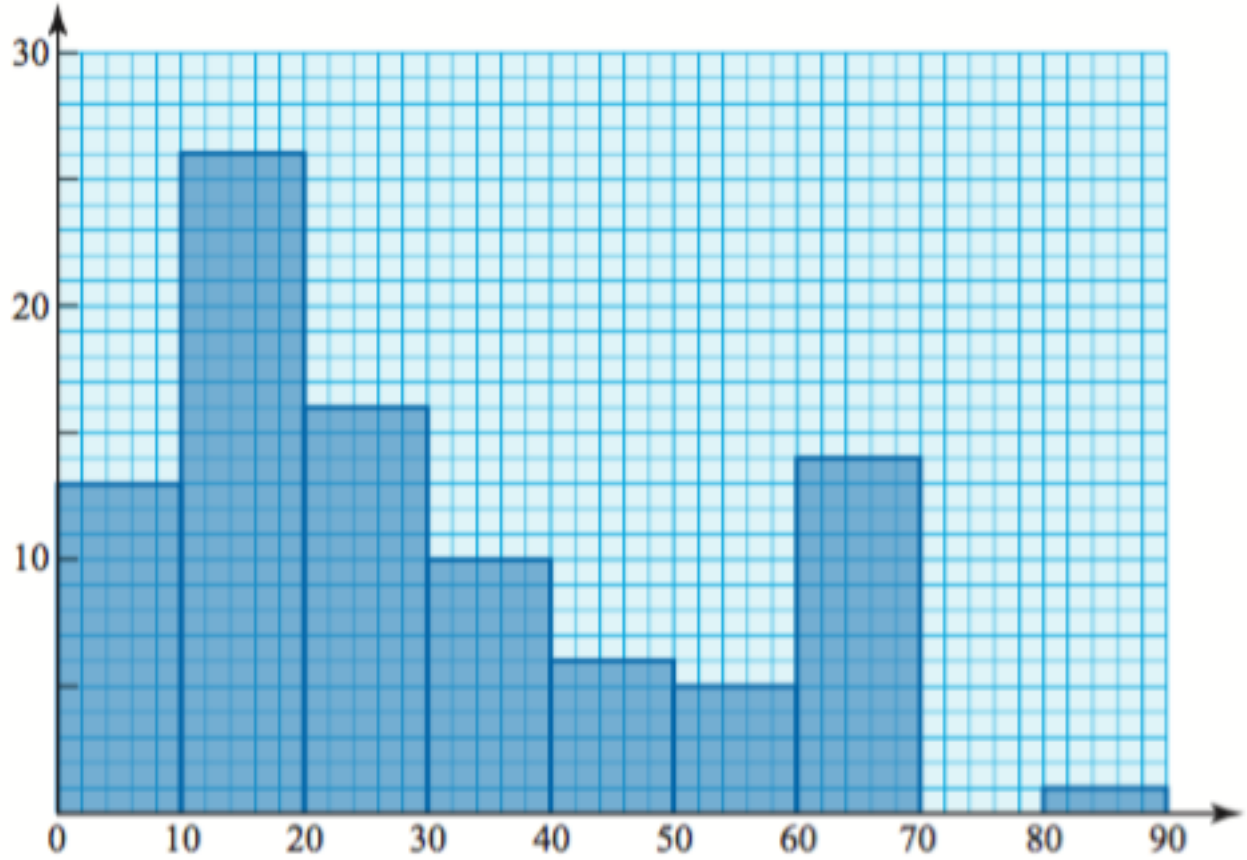
66 6 62 19 20 15 21 8 21 63 44 10 44 34 18
 35 26 61 13 61 28 21 7 10 52 13 52 20 17 26
 64 11 39 22 9 13 9 17 64 32 8 9 31 19 22
 37 18 138 16 67 45 10 55 14 66 67 14 62 28 36
 9 23 12 9 37 7 36 9 88 46 12 59 61 22 49
 18 20 11 25 7 42 29 6 60 60 16 50 16 34 14
 18 15

Ийм хэлбэрийн өгөгдлийг **боловсруулаагүй өгөгдөл** гэж нэрлэнэ.

Боловсруулаагүй өгөгдлийг хэрхэн тоолох вэ? өгөгдлийг багцлан тоолох нь өгөгдлийг тохирсон интервал дотор түргэн тоолох боломж олгоно. Ихэвчлэн 5-р багцалж тоолдог. Жишээ нь дээрх өгөгдлийг багцлан тоолж дараах хүснэгтийг гаргаж болно. Хүснэгтэнд насыг 10 насны интервалд тооцон, багц мөн давтамжыг харуулав.

Stated age (years)	Tally	Frequency
0-9		13
10-19		26
20-29		16
30-39		10
40-49		6
50-59		5
60-69		14
70-79		0
80-89		1
:		
130-139		1
Total		92

Эндээс бид өгөгдлийн график дүрслэлийг харуулбал



болно.

Энэ өгөгдлийг ашиглан иш-навчны диаграм байгуулбал

0	6 6 7 7 7 8 8 9 9 9 9 9 9
1	0 0 0 1 1 2 2 3 3 3 4 4 4 5 5 6 6 6 7 7 8 8 8 8 9 9
2	0 0 0 1 1 1 2 2 2 3 5 6 6 8 8 9
3	1 2 4 4 5 6 6 7 7 9
4	2 4 4 5 6 9
5	0 2 2 5 9
6	0 0 1 1 1 2 2 3 4 4 6 6 7 7
7	
8	8

болно.

Олон оронтой тооны хувьд иш-навчны диаграмыг хэрхэн байгуулах вэ?

Тухайлбал гурван оронтой туунуудыг илэрхийлэхдээ эхний хоёр оронг иш болгон үлдсэн нэг оронг навч болгож авч болно. Үүнээс гадна эхний оронг иш

болгон авч болох жишээг доор үзүүлэв.

			9	5	1	7														
			2	6	0	2	3	5	8											
		5	3	0	7	1	2	5	6	6	7									
9	7	5	1	1	8	3	5													
		8	6	2	1	9	2													

Ийм хэлбэрийн диаграмаар хоёр төрлийн өгөгдлийг илэрхийлж бас болно. Жишээ нь дээрх диаграм нь тухайн ангийн охид, хөвгүүдийн шалгалтын оноо байж болно. Зүүн гар талын багана нь охидын авсан оноо, баруун гар талын багана нь хөвгүүдийн оноо байж болно.

Тоон болон чанарын өгөгдөл

Хувьсагч. Шоог орхих үед 1, 2, 3, 4, 5, 6 тоотой нүдний аль нэгээр бууна. Энэ тохиолдолд үүнийг хийсвэрлэн X гэх мэт том үсгээр тэмдэглэдэг. Энэ нь шоог орхих үед буух нүдэн дээрх тоог илэрхийлнэ.

Үүнтэй ижлээр, хэрэв ангийн температурыг хэмжин T үсгээр тэмдэглэж болно. Эдгээр нь хувьсагчид юм.

Хувьсагчдын авах утгыг харгалзах жижиг үсгээр тэмдэглэдэг. Тухайлбал, X хувьсагчийн хувьд x гэж тэмдэглэдэг ба $x = 1, 2, 3, 4, 5, 6$ болно. T хувьсагчийн хувьд $t = 18, 21, 20, 19, 23, \dots$ гэх мэт утга авах боломжтой.

Дискрет ба тасралтгүй хувьсагчид

Шоог орхиход буух нүдэн дээрх тоо нь 1, 2, 3, 4, 5, 6 байна. Мөн зоосыг орхиход унах боломж нь сүлд эсвэл тоо байна. Эдгээр нь дискрет хувьсагчийн жишээ бөгөөд авч болох утгыг бүгдийг жагсаах боломжтой хувьсагчийг **дискрет** хувьсагч гэж нэрлэдэг.

Хувьсагч болгоны авах боломжтой утгыг жагсаах боломжгүй тохиолдол байж болно. Жишээ нь зай, хурд, масс, температур зэрэг нь ийм хувьсагчдын жишээ ба эдгээрийг **тасралтгүй** хувьсагчид гэж нэрлэнэ.

Бодит байдал дээр тасралтгүй хувьсагчдын өгөгдлийг ойролцоолон тасралттай хэлбэрээр илэрхийлэх тохиолдол олон байдаг. Жишээ нь өндөр нь 168 см байх хүн нь ихэнхдээ 167.5-с 168.5 см-ийн хооронд өндөртэй хүмүүсийг төлөөлдөг. Температурын хувьд мөн 21.8° гэдэг нь 21.75° - 21.85° хоорондох температурыг төлөөлж байх боломжтой.

Өгөгдлийг бүлэглэх

Өгөгдлийг анги, бүлгүүдэд хуваах шаардлага гардаг. Тухайн ангид орж буй өгөгдлийн тоог уг ангийн давтамж гэнэ.

Тоон өгөгдлийг бүлэглэсэн үед өгөгдөл болгон тухайн нэг ангид харгалзана. Ангиудыг хилийн утгаар тусгаарлана.

Тухайн ангийн хилийн утгуудыг нэг утгатай тодорхойлох шаардлагатай. Учир нь элемент болгон аль ангид орох нь тодорхой байх ёстой. Жишээ нь

судалгаанд оролцогчдыг 10 насны ангиллаар бүлэглэсэн гэж үзвэл насны бүлгүүдийг 0-9, 10-19, 20-29, гэх мэтээр тодорхойлох нь зүйтэй. Зарим тохиолдолд 0-10, 10-20, 20-30 гэх мэтээр тодорхойлсон байдаг ба энэ тохиолдолд 10,20 насны хүн аль бүлэгт хамаарах нь тодорхойгүй болно. Мөн насыг асууж байгаа асуултыг зөв асуух шаардлагатай. Дээрх 0-9, 10-19, 20-29 бүлгийн хувьд $9\frac{1}{2}$ настай хүн аль ангид орох нь тодорхойгүй. Энэ тохиолдолд

$$0 \leq A < 10, 10 \leq A < 20, 20 \leq A < 30$$

гэж бичвэл оновчтой болно.

Бүлэглэсэн өгөгдөлтэй ажиллах. Боловсруулаагүй өгөгдлийн бүлэглэх хэд хэдэн шалтгаан байдаг. Үүнд:

- (1) өгөгдөл хэт олон
- (2) өгөгдлийн далайц их, маш их тархсан
- (3) маш олон ялгаатай өгөгдлүүд байгаа зэрэг юм.

Бүлэглэсэн өгөгдлүүдэд анализ хийж тэдгээрийг хүснэгт, диаграм хэлбэрээр илэрхийлэхэд хялбар байдаг. Зарим дискрет өгөгдлийн хувьд бүлэглэх шаардлагагүй байж болно. Жишээ нь замаар өнгөрч бул машинд сууж буй зорчигчдын тоо нь 0-4 хүртлэх утгууд авна гэвэл бүлэглэх шаардлагагүй. Гэхдээ өгөгдлийг бүлэглэх нь давуу талтай байх тохиолдлууд байдаг.

0.1.1. Дискрет өгөгдөл. Ажиглах цэгээр дайран өнгөрөн машины тоог долоо хоногийн янз бүрийн хугацаанд тэмдэглэжээ. Өгөгдөл бүр нь тухайн ажиглалтын үед 5 минутын турш ажиглах цэгээр дайрч өнгөрсөн машины тоог илэрхийлнэ. Нийт 100 удаагийн ийм ажиглалт хийсний дүнг дараах хүснэгтэнд харуулав.

Машины тоо x	Давтамж f
0-9	5
10 – 19	8
20 – 29	13
30 – 39	20
40 – 49	22
50 – 59	21
60 – 70	11
Нийт	100

Дунджийг тооцоолох

Өгөгдлийг бүлэглэсэн тохиолдолд элементүүдийн утга алдагдана. Энэ тохиолдолд дунджийг хангалттай нарийвчлалтай олж болно.

Дунджийг тооцоолоход утгууд нь интервал дээр дундаж цэгээсээ ижил зайтай тархана гэж үзнэ. Иймд дундаж цэгийн утгыг тухайн интервалын төлөөлөл болгон авдаг.

Дээрх жишээний хувьд 0~9 интервалын дундаж цэг нь $x_1 = \frac{0+9}{2} = 4.5.2$

Үүнтэй ижлээр 10~19 интервалын дундаж цэг $x_2 = \frac{10+19}{2} = 14.5$ гэх мэтээр интервал бүрийн дундаж цэгийг тооцоолж болно.

Одоо $x_i \times f_i$ утгыг тооцоолох боломжтой ба дээрх 3 өгөгдлийг ашиглан дундаж олох хүснэгт зохиоё.

Машины тоо x	Давтамж f	$x \times f$
4.5	5	22.5
14.5	8	116
24.5	13	318.5
34.5	20	690
44.5	22	979
54.5	21	1144.5
65	11	715
Нийт	100	3985.5

Эндээс дундаж утгыг олбол

$$\bar{x} = \frac{\sum_{i=1}^7 x_i f_i}{\sum_{i=1}^7 f_i} = \frac{3985.5}{100} = 39.855$$

болно. Энд дунджаас гадна иш навчны диаграм байгуулж болох байсан.

Унадаг дугуйн осолтой холбоотой өмнөх жишээний хувьд осолд орсон хүн бүр гэрээсээ хэр хол зайд осолд орсонг судлан дараах хүснэгтийг гаргажээ.

3000	75	1200	300	50	10	150	1500	250	25
200	4500	35	60	120	400	2400	140	45	5
1250	3500	30	75	250	1200	250	50	250	450
15	400.								

Энд өгөгдлүүдийн тархалт нилээн далайц ихтэй байгаа учир бүлэглэх шаардлагатай болно. Бүлэглэхдээ

- (а) Гэртээ маш ойр, 100 метр дотор,
- (б) Гэртээ ойр, 100-500 метр дотор,
- (в) Хол, 500-1500 метр дотор,
- (г) Нилээн хол, 1500-5000 метр дотор

осолд орсон гэж ангилан бүлэглэв.

Тэгвэл бид дараах хүснэгтийг зохиож болно.

Ангилал	Зай	Интервалын дундаж утга	Давтамж	$x \times f$
Гэртээ маш ойр	$0 \leq x < 100$	50	12	600
Гэртээ ойр	$100 \leq x < 500$	300	11	3300
Хол	$500 \leq x < 1500$	1000	3	3000
Нилээн хол	$1500 \leq x < 5000$	3250	6	19500
Нийт			32	26400

Иймд дунджийг олбол

$$\bar{x} = \frac{26400}{32} = 825 \text{ м}$$

болно.

Анхны өгөгдөл болон бүлгэлсэн өгөгдлийг харьцуулан харуулбал,

	Боловсруулаагүй өгөгдөл	Бүлэглэсэн өгөгдөл
Дундаж	$25785/32 = 806 \text{ м}$	825 м
Моод	250 м	Моодын бүлэг = $0 \leq x < 100$
Медиан	$(200 + 250)/2 = 225 \text{ м}$	

Стандарт хазайлт

x_1, \dots, x_n өгөгдлийн дундаж нь \bar{x} ба x_i элемент бүрийн давтамж f_i байг. Бид эдгээр өгөгдлийн стандарт хазайлтыг $\bar{\sigma}$ гэж тэмдэглэх ба дараах томъёогоор олно.

$$\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}}$$

болно. $(x_i - \bar{x})^2$ илэрхийллийг ашиглах нь түвэгтэй байдаг учир дунджийн тодорхойлолт ашиглан дээрх илэрхийллийг хялбарчилбал

$$\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^n x_i^2 f_i}{n} - \bar{x}^2} = \sqrt{\frac{\sum_{i=1}^n x_i^2 f_i}{\sum_{i=1}^n f_i} - \bar{x}^2}$$

болно.