Example.1.1

EXAMPLE 1.1 The Location of Sparrow Nests: A Frequency Table of Nominal Data

The variable is nest site, and there are four recorded categories of this variable. The numbers recorded in these categories constitute the frequency distribution.

Nest Site	*Number of Nests Observed*
A. Vines	56
B. Building eaves	60
C. Low tree branches	46
D. Tree and building cavities	49

결과 해석은 잘 했어요.

R

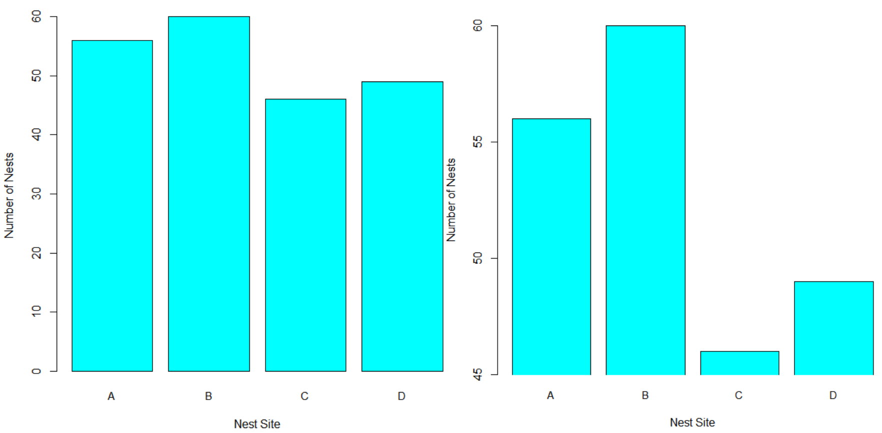
CODE

```
#Ex1.1a
counts=c(56,60,46,49)
barplot(counts, ylab="Number of Nests", xlab="Nest Site",
names.arg = c("A","B","C","D"), cex.names = 0.9, col="cyan")
#Ex1.1b
barplot(counts, ylab="Number of Nests", xlab="Nest Site", names.arg =
c("A","B","C","D"), cex.names = 0.9, col="cyan",
          ylim=c(45,60),xpd=F, yaxp=c(45,60,3))
```

Fair

그래프에 보다 많은 정보를 수치로 포함시켜 보세요.

OUTPUT



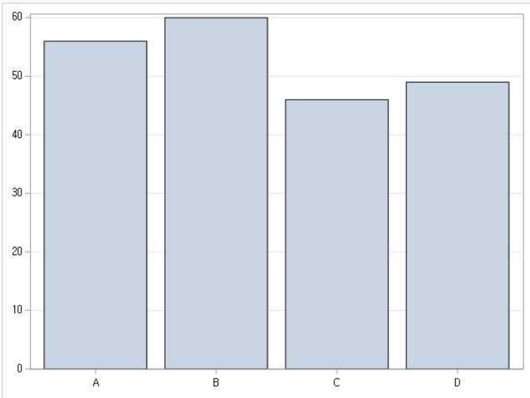
SAS

CODE

```
data ex1_1;
input nests$ num;
cards;
A 56
B 60
C 46
D 49
;
run;
```

```
proc freq data=ex1_1;
weight num;
tables nests/ out= Freqout;
run;
proc print data= Freqout;
run;
proc sgplot data=Freqout;
vbar nests/response= count;
yaxis grid display=(nolabel);
xaxis display=(nolabel);
run;
```

OUTPUT



결과해석

R에서의 첫 번째 그래프를 보면 장소에 상관없이 발견되는 둥지의 수가 큰 차이가 없어 보인다. 그러나 y축을 45부터 시작해서 그린 두 번째 그래프를 보면 C(Low tree branches)에서 발견되는 둥지 수가 현저히 적어보인다. 이를 보고 그래프 축의 단위를 어떻게 하느냐에 따라 결과 해석의 차이가 클 수 있다는 것을 알 수 있다.

Example.1.2

EXAMPLE 1.2 Numbers of Sunfish, Tabulated According to Amount of Black Pigmentation: A Frequency Table of Ordinal Data

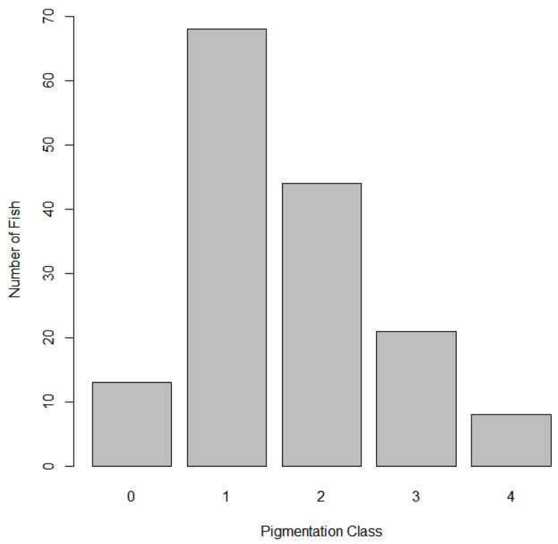
The variable is amount of pigmentation, which is expressed by numerically ordered classes. The numbers recorded for the five pigmentation classes compose the frequency distribution.

Pigmentation Class	*Amount of Pigmentation*	*Number of Fish*
0	No black pigmentation	13
1	Faintly speckled	68
2	Moderately speckled	44
3	Heavily speckled	21
4	Solid black pigmentation	8

R

CODE

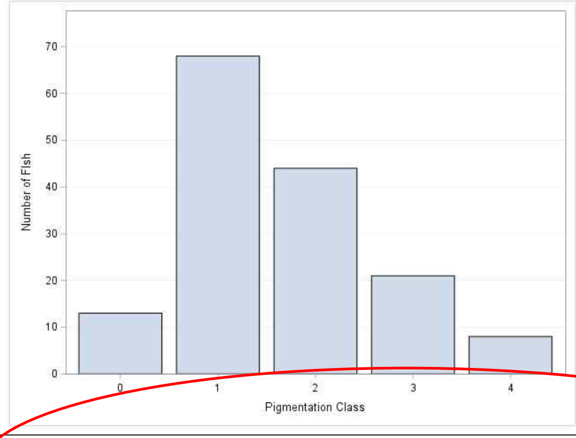
```
#Ex1.2
counts=c(13,68,44,21,8)
barplot(counts, ylab="Number of Fish", xlab="Pigmentation Class",
```

<table>
<tr><td></td><td>names.arg = c("0","1","2","3","4"), cex.names = 1.0)</td></tr>
<tr><td>OUTPUT</td><td>
</td></tr>
<tr><td colspan="2">SAS</td></tr>
<tr><td>CODE</td><td>

```
data ex1_2;
input class$ num;
cards;
0 13
1 68
2 44
3 21
4 8
; run;
proc freq data=ex1_2;
weight num;
tables class/ out= Freqout;
run;
proc sgplot data=Freqout;
vbar class/response= count ;
yaxis grid  values=(0 to 70 by 10) label="Number of FIsh"
offsetmax=0.1;
xaxis label="Pigmentation Class";
run;
```

</td></tr>
</table>

OUTPUT



결과해석

그래프를 보면 관찰한 Sunfish(개복치)들 중 대부분이 희미하더라도 점박이 무늬를 띄고, 반면에 무늬가 없거나, 검정 색상을 띄는 sunfish는 드물게 관찰되었다는 것을 알 수 있다.

Example.1.3

EXAMPLE 1.3 Frequency of Occurrence of Various Litter Sizes in Foxes: A Frequency Table of Discrete, Ratio-Scale Data

The variable is litter size, and the numbers recorded for the five litter sizes make up frequency distribution.

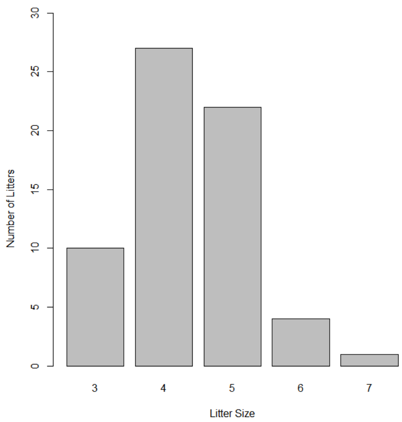
Litter Size	*Frequency*
3	10
4	27
5	22
6	4
7	1

R

CODE

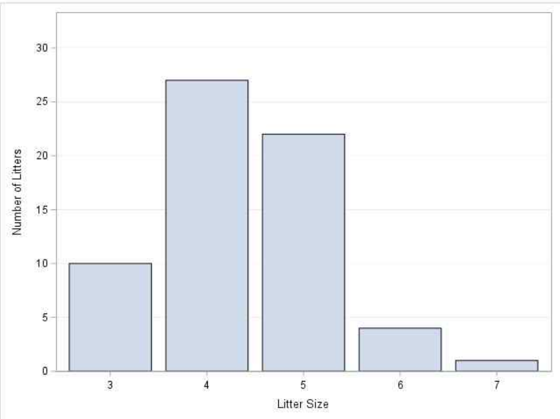
```
#Ex1.3
counts=c(10,27,22,4,1)
barplot(counts, ylab="Number of Litters", xlab="Litter Size", names.arg = c("3","4","5","6","7"), yaxp=c(0,30,6))
```

OUTPUT



<table>
<tr><th colspan="2">SAS</th></tr>
<tr><td>CODE</td><td>

```
data ex1_3;
input size$ num;
cards;
3 10
4 27
5 22
6 4
7 1
;
run;
proc freq data=ex1_3;
weight num;
tables size/ out= Freqout;
run;
proc sgplot data=Freqout;
vbar size/response= count ;
yaxis grid  values=(0 to 30 by 5) label="Number of Litters"
offsetmax=0.1;
xaxis label="Litter Size";
run;
```

</td></tr>
<tr><td>OUTPUT</td><td>


</td></tr>
<tr><td>결과해석</td><td>위의 그래프를 보면 데이터가 왼쪽으로 치우쳐 있다는 것을 확인할 수 있다. 즉 여우의 litter size(한배 산란수)는 보통 4~5마리이고 더 많이 새끼를 낳는 여우는 드물다는 것을 확인할 수 있다.</td></tr>
</table>

Example.1.4a

EXAMPLE 1.4a Number of Aphids Observed per Clover Plant: A Frequency Table of Discrete, Ratio-Scale Data

Number of Aphids on a Plant	*Number of Plants Observed*	*Number of Aphids on a Plant*	*Number of Plants Observed*
0	3	20	17
1	1	21	18
2	1	22	23
3	1	23	17
4	2	24	19
5	3	25	18
6	5	26	19
7	7	27	21
8	8	28	18
9	11	29	13
10	10	30	10
11	11	31	14
12	13	32	9
13	12	33	10
14	16	34	8
15	13	35	5
16	14	36	4
17	16	37	1
18	15	38	2
19	14	39	1
		40	0
		41	1

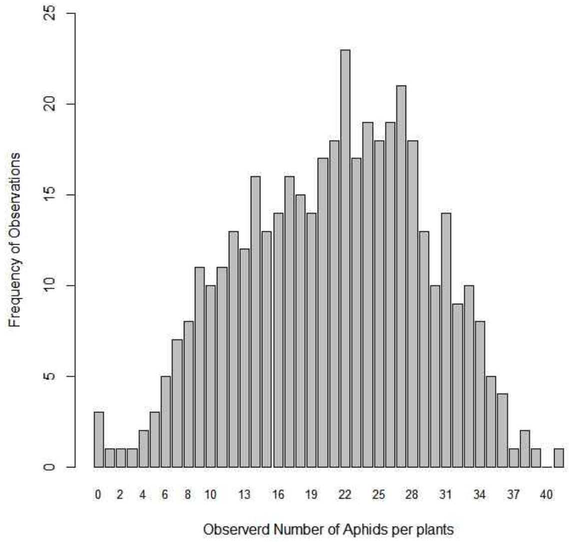
Total number of observations = 424

R

CODE

```
#Ex1.4a
aphids = seq(0, 41)
plants =
c(3,1,1,1,2,3,5,7,8,11,10,11,13,12,16,13,14,16,15,14,17,18,23,17,19,18,19,21,
18,13,10,14,9,10,8,5,4,1,2,1,0,1)
barplot(plants, names.arg = aphids, cex.names = 0.8, ylim = c(0,25),
ylab = "Frequency of Observations", xlab = "Observerd Number of
Aphids per plants")
```

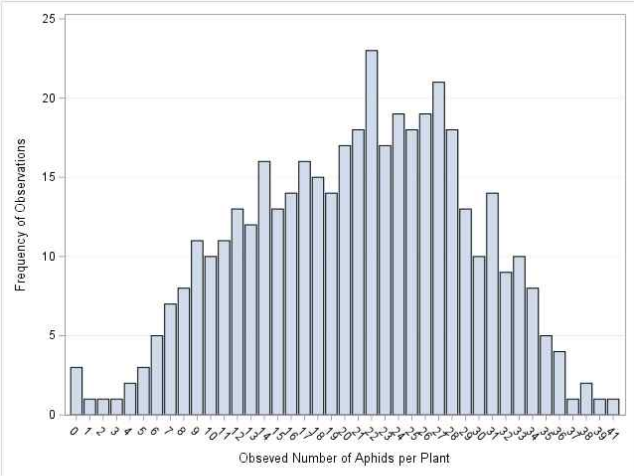
OUTPUT



SAS

CODE

```
data ex1_4a;
input aphids num@@;
cards;
0 3 1 1 2 1 3 1 4 2 5 3 6 5 7 7 8 8 9 11 10 10 11 11 12 13 13 12
14 16 15 13 16 14 17 16 18 15 19 14
20 17 21 18 22 23 23 17 24 19 25 18 26 19 27 21 28 18 29 13 30
10 31 14 32 9 33 10 34 8 35 5 36 4 37 1 38 2 39 1 40 0 41 1
;
run;
proc freq data=ex1_4a;
weight num;
tables aphids/ out= Freqout;
run;
proc sgplot data=Freqout;
vbar aphids/response= count ;
yaxis grid  values=(0 to 25 by 5) label="Frequency of
Observations" ;
xaxis label="Obseved Number of Aphids per Plant";
run;
```

OUTPUT	
결과해석	Aphids(진딧물)의 수에 따른 관찰된 Cover plant(피복식물)의 수를 나타낸 것으로 평균적으로 20~23마리의 진딧물이 대부분의 Cover plant에서 관측되는 것이라 생각할 수 있다.

Example.1.4b

EXAMPLE 1.4b Number of Aphids Observed per Clover Plant: A Frequency Table Grouping the Discrete, Ratio-Scale Data of Example 1.4a

Number of Aphids on a Plant	*Number of Plants Observed*
0–3	6
4–7	17
8–11	40
12–15	54
16–19	59
20–23	75
24–27	77
28–31	55
32–35	32
36–39	8
40–43	1

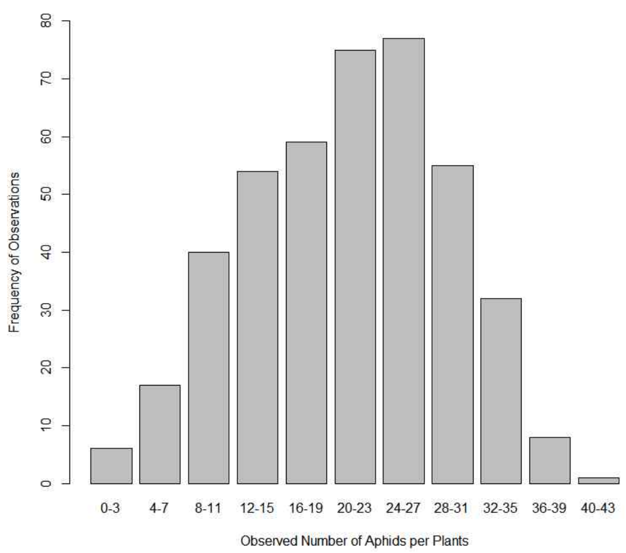
Total number of observations = 424

R

CODE

```
#Ex1.4b
plants=c(6,17,40,54,59,75,77,55,32,8,1)
barplot(plants, names.arg= c("0-3","4-7","8-11","12-15","16-19",
"20-23","24-27","28-31","32-35","36-39","40-43"),
ylab="Frequency of Observations", xlab="Observed Number of Aphids
per Plants", ylim=c(0,80), yaxp=c(0,80,8))
```

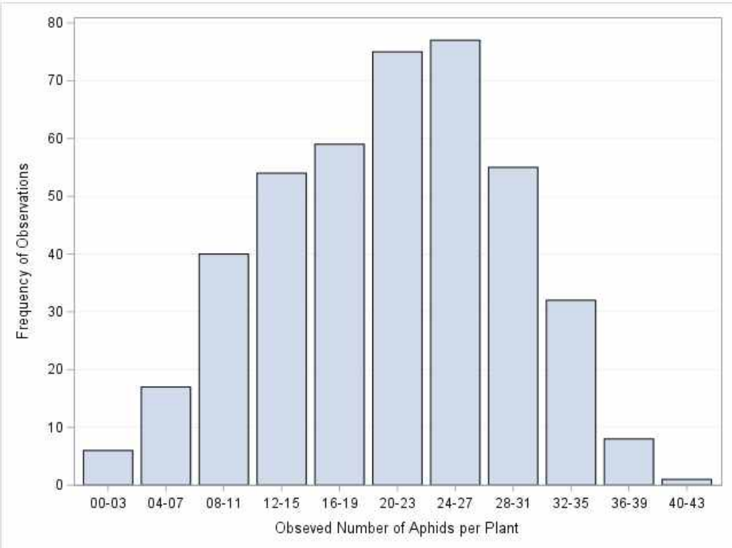
OUTPUT



SAS

CODE

```
data ex1_4b;
input aphids$ num @@;
cards;
00-03 6 04-07 17 08-11 40 12-15 54 16-19 59 20-23 75 24-27 77
```

	`28-31 55 32-35 32 36-39 8 40-43 1` `;` `run;` `proc freq data=ex1_4b;` `weight num;` `tables aphids/ out= Freqout ;` `run;` `proc sgplot data=Freqout;` `vbar aphids/response= count ;` `yaxis grid  values=(0 to 80 by 10) label="Frequency of Observations" ;` `xaxis label="Obseved Number of Aphids per Plant";` `run;`
OUTPUT	
결과해석	1.4a와 동일한 데이터의 그래프지만 x축을 4단위씩 묶어서 합쳐 그린 그래프이다. 위 그래프를 보면 1.4a의 그래프보다 조금 더 대칭적으로 그래프가 그려졌다는 것을 확인 할 수 있다. 이렇게 x축의 구간설정에 따라 그래프의 형태가 완전히 바뀌어 정보를 다르게 해석할 수 있으므로 구간설정을 잘 하는 것이 중요하다.

Example.1.5

EXAMPLE 1.5 Determinations of the Amount of Phosphorus in Leaves: A Frequency Table of Continuous Data

Phosphorus (mg/g of leaf)	Frequency (i.e., number of determinations)	Cumulative frequency Starting with Low Values	Cumulative frequency Starting with High Values
8.15–8.25	2	2	130
8.25–8.35	6	8	128
8.35–8.45	8	16	122
8.45–8.55	11	27	114
8.55–8.65	17	44	103
8.65–8.75	17	61	86
8.75–8.85	24	85	69
8.85–8.95	18	103	45
8.95–9.05	13	116	27
9.05–9.15	10	126	14
9.15–9.25	4	130	4

Total frequency = 130 = n

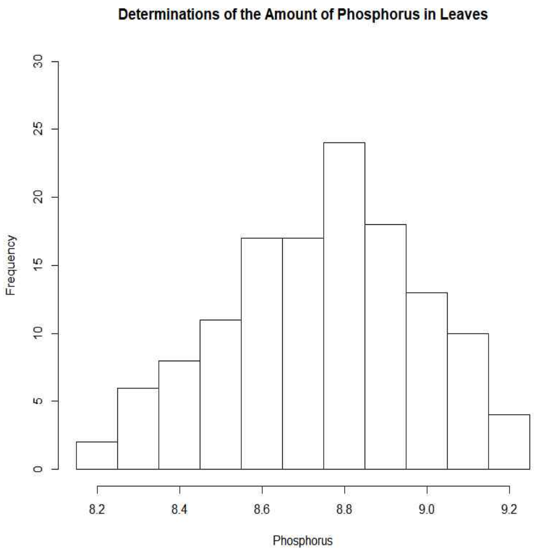
R

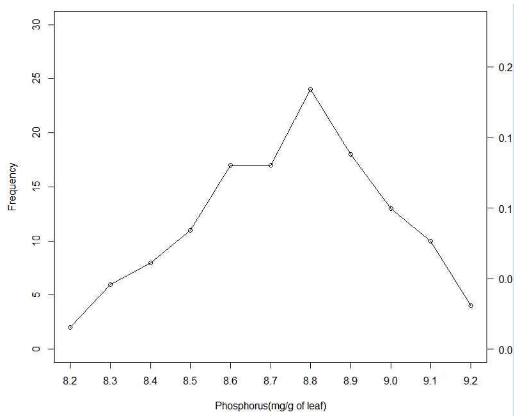
CODE

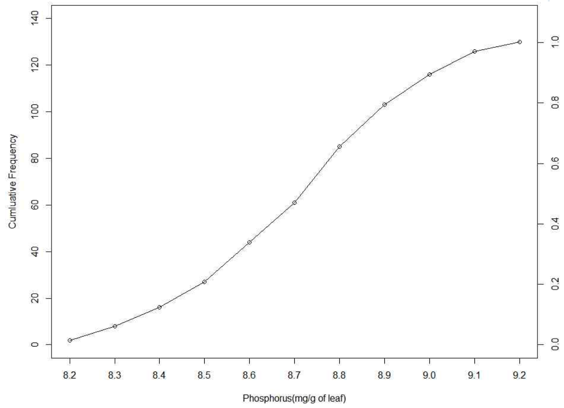
```
#Ex1.5a
df= as.data.frame(cbind(x= seq(8.15,9.25,length=11),
Freq= c(2,6,8,11,17,17,24,18,13,10,4)))
df.freq= as.vector(rep(df$x, df$Freq))
hist(df.freq, breaks=seq(8.15,9.25,length=12),
main="Determinations of the Amount of Phosphorus in Leaves",
ylab="Frequency", xlab="Phosphorus", ylim=c(0,30),yaxp=c(0,30,6))
lines(df$x,df$Freq)

#Ex1.5b
x=seq(8.2, 9.2, 0.1)
frq=c(2,6,8,11,17,17,24,18,13,10,4)
y=data.frame(x,frq)
plot(y, type='o', ylab= "Frequency", xlab= "Phosphorus(mg/g of leaf)",
ylim=c(0,30), xaxp=c(8.2,9.2,10))
rfrq=frq/sum(frq)
par(new=T)
plot(rfrq, type="n",ylim=c(0,0.23), ylab="", xlab="", axes=F)
axis(side=4, at=NULL, labels = T, las=1)

#EX1.5c
cumfrq=cumsum(frq)
z=data.frame(x,cumfrq)
rcumfrq= cumfrq/sum(frq)
```

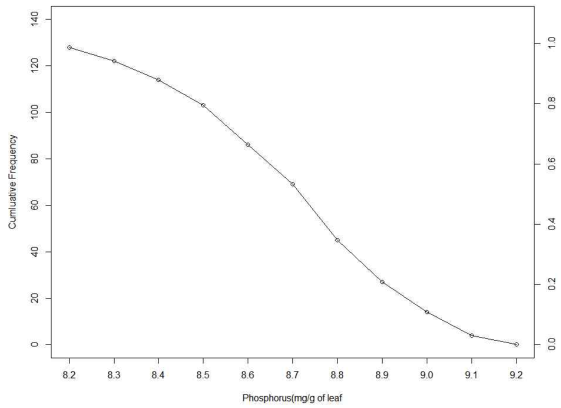
```
plot(z, type="o", ylab="Cumluative Frequency",
xlab="Phosphorus(mg/g of leaf",ylim=c(0,140),xaxp=c(8.2,9.2,10))
par(new=T)
zr=data.frame(x,rcumfrq)
plot(zr, type="n",ylim=c(0,1.08), ylab="", xlab="",xaxp=c(8.2,9.2,10), axes =
F)
axis(side=4, at=NULL, labels = T)

#Ex1.5d
w=rep(130,11)
v=w-cumfrq
rv=v/sum(frq)
inv=data.frame(x,v)
plot(inv,type="o", ylab="Cumluative Frequency",
xlab="Phosphorus(mg/g of leaf",ylim=c(0,140),xaxp=c(8.2,9.2,10))
par(new=T)
rinv=data.frame(x,rv)
plot(rinv, type="n",ylim=c(0,1.08), ylab="", xlab="",xaxp=c(8.2,9.2,10), axes
= F)
axis(side=4, at=NULL, labels = T)
```

OUTPUT









SAS

CODE

```
data ex1_5;
input pho frq @@;
cards;
8.2 2
8.3 6
8.4 8
8.5 11
8.6 17
```

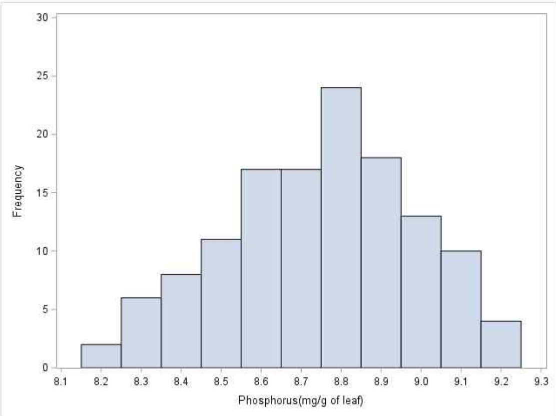
```
8.7 17
8.8 24
8.9 18
9.0 13
9.1 10
9.2 4
;
run;
proc freq data=ex1_5;
weight frq;
tables pho/ out= Freqout;
run;
proc print data=Freqout;
run;
proc sgplot data= Freqout;
histogram pho/ scale=count freq= count binstart=8.2 nbins=11
;
xaxis values=(8.1 to 9.3 by 0.1) label="Phosphorus(mg/g of
leaf)";
yaxis values=(0 to 30 by 5) label="Frequency";
run;
data ex1_5r;
set ex1_5;
rel=frq/130;
run;
proc sgplot data=ex1_5r;
series x= pho y= frq /markers;
series x= pho y=rel/ y2axis transparency=1;
xaxis values=(8.1 to 9.3 by 0.1) label="Phosphorus(mg/g of
leaf)";
yaxis values=(0 to 30 by 5) label="Frequency";
y2axis values=(0 to 0.2 by 0.05) label="Relative Frequency";
run;
data ex1_5c;
set ex1_5;
retain cum 0;
cum+frq ;
rcum=cum/130;
run;
proc print data=ex1_5c;run;
```

정확하지 않아요. 두번째 series에서도 markers를 이용하여 2개의 선분이 겹치도록 만드세요.

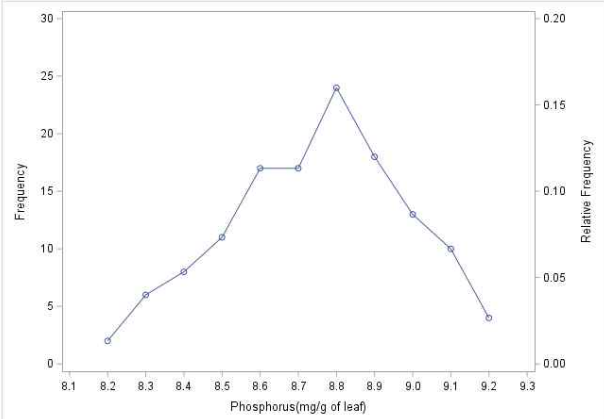
<table>
<tr><td></td><td>

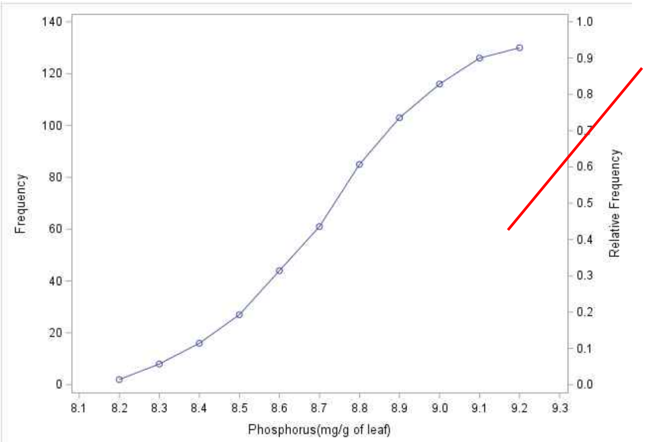
```
proc sgplot data=ex1_5c;
series x=pho y=cum / markers;
series x=pho y=rcum/y2axis transparency=1;
xaxis values=(8.1 to 9.3 by 0.1) label="Phosphorus(mg/g of
leaf)";
yaxis values=(0 to 140 by 20) label="Frequency";
y2axis values=(0 to 1 by 0.10) label="Relative Frequency";
run;
data ex1_5ic;
set ex1_5;
retain cum 130;
cum + (-frq);
rcum=cum/130;
run;
proc sgplot data=ex1_5ic;
series x=pho y=cum / markers;
series x=pho y=rcum/y2axis transparency=1;
xaxis values=(8.1 to 9.3 by 0.1) label="Phosphorus(mg/g of
leaf)";
yaxis values=(0 to 140 by 20) label="Frequency";
y2axis values=(0 to 1 by 0.10) label="Relative Frequency";
run;
```

</td></tr>
<tr><td>OUTPUT</td><td>

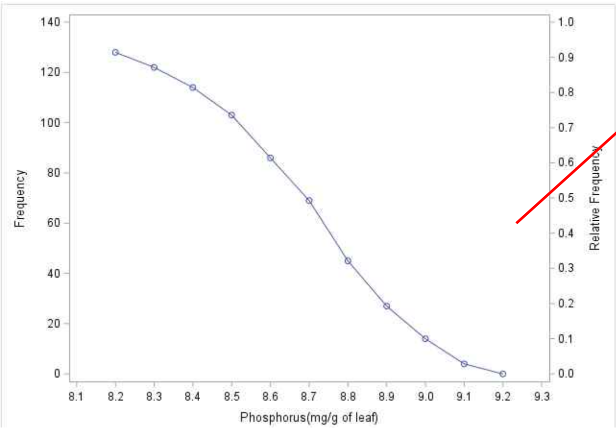


</td></tr>
</table>





표본크기와 오른쪽 눈금 1이 대응되어야 합니다.



결과해석	위의 자료는 잎에 들어있는 phosphorus(인)의 함유량을 측정한 것으로 예를 들어 8.15~8.25 정도의 함유량을 가진 잎이 2번 측정되었다는 것이다. 위의 자료의 분포가 살짝 오른쪽으로 치우쳐 보이지만 어느 정도 8.7정도를 중심으로 대칭적인 분포가 나타나고 있는 것을 히스토그램을 보고 알 수 있다. 그리고 누적분포도수다각형에서 확인할 수 있듯이 자료의 중심부근에서 기울기가 증가하고 자료의 끝부분에서의 기울기는 완만하므로 여기서도 자료가 자료의 중심으로부터 대칭적 형태를 나타내고 있다는 것을 확인할 수 있다. 즉, 잎에 들어있는 phosphorus의 함유량은 대체적으로 8.7~8.8 정도라고 생각할 수 있다.

Example.3.1

EXAMPLE 3.1 A Sample of 24 from a Population of Butterfly Wing Lengths

X_i (in centimeters): 3.3, 3.5, 3.6, 3.6, 3.7, 3.8, 3.8, 3.8, 3.9, 3.9, 3.9, 4.0, 4.0, 4.0, 4.0, 4.1, 4.1, 4.1, 4.2, 4.2, 4.3, 4.3, 4.4, 4.5.

$$\sum X_i = 95.0 \text{ cm}$$
$$n = 24$$
$$\bar{X} = \frac{\sum X_i}{n} = \frac{95.0 \text{ cm}}{24} = 3.96 \text{ cm}$$

R

CODE

```
#Ex3.1
X=c(3.3,3.5,3.6,3.6,3.7,3.8,3.8,3.8,3.9,3.9,3.9,4.0,4.0,4.0,4.0,4.1,4.1,4.1,4.2,4.2,4.3,4.3,4.4,4.5)
Xbar=mean(X)
round(Xbar,2)
```

OUTPUT

```
> #Ex3.1
> X=c(3.3,3.5,3.6,3.6,3.7,3.8,3.8,3.8,3.9,3.9,3.9,4.0,4.0,4.0,4.0,4.1,4.1,4.1,4.2,4.2,4.3,4.3,4.4,4.5)
> Xbar=mean(X)
> round(Xbar,2)
[1] 3.96
```

SAS

CODE

```
data ex3_1;
input x @@;
cards;
3.3 3.5 3.6 3.6 3.7 3.8 3.8 3.8 3.9 3.9 3.9
4.0 4.0 4.0 4.0 4.1 4.1 4.1 4.2 4.2 4.3 4.3 4.4 4.5
;
run;
proc means data=ex3_1 mean;
var x;
run;
```

OUTPUT

Analysis Variable : x
Mean
3.9583333

결과해석

나비의 날개 길이를 측정하기 위해 24마리의 나비 표본을 추출해 날개 길이를 측정해본 결과 평균 3.96cm라는 값을 얻었다.

Example.3.2

EXAMPLE 3.2 The Data from Example 3.1 Recorded as a Frequency Table

X_i (cm)	f_i	f_iX_i (cm)
3.3	1	3.3
3.4	0	0
3.5	1	3.5
3.6	2	7.2
3.7	1	3.7
3.8	3	11.4
3.9	3	11.7
4.0	4	16.0
4.1	3	12.3
4.2	2	8.4
4.3	2	8.6
4.4	1	4.4
4.5	1	4.5
	$\sum f_i = 24$	$\sum f_iX_i = 95.0$ cm

$$k = 13$$

$$\sum_{i=1}^{k} f_i = n = 24$$

$$\overline{X} = \frac{\sum_{i=1}^{k} f_iX_i}{n} = \frac{95.0\text{ cm}}{24} = 3.96\text{ cm}$$

$$\text{median} = 3.95\text{ cm} + \left(\frac{1}{4}\right)(0.1\text{ cm})$$

$$= 3.95\text{ cm} + 0.025\text{ cm}$$

$$= 3.975\text{ cm}$$

R

CODE

```
#Ex3.2
install.packages("Hmisc")
library(Hmisc)
xi=seq(3.3,4.5,0.1)
fi=c(1,0,1,2,1,3,3,4,3,2,2,1,1)
wtd.mean(xi, weights=fi)
x.median=wtd.quantile(xi, weights=fi, probs = c(.5), type=c("(i-1)/(n-1)"))
x.median
```

OUTPUT

```
> xi=seq(3.3,4.5,0.1)
> fi=c(1,0,1,2,1,3,3,4,3,2,2,1,1)
> wtd.mean(xi, weights=fi)
[1] 3.958333
> x.median=wtd.quantile(xi, weights=fi, probs = c(.5), type=c("(i-1)/(n-1)
> x.median
   50%
3.9375
```

SAS

CODE

```
data ex3_2;
input x w @@;
cards;
3.3 1 3.4 0 3.5 1 3.6 2 3.7 1 3.8 3 3.9 3 4.0 4 4.1 3 4.2 2 4.3
2 4.4 1 4.5 1
;
run;
proc means data=ex3_2 mean median;
var x;
```

<table>
<tr><td></td><td><code>weight w; run;</code></td></tr>
<tr><td>OUTPUT</td><td><table><tr><th colspan="2">Analysis Variable : x</th></tr><tr><th>Mean</th><th>Median</th></tr><tr><td>3.9583333</td><td>4.0000000</td></tr></table></td></tr>
<tr><td>결과해석</td><td>도수분포표로 나타나 있는 경우에는 변수값에 빈도수만큼의 가중치를 곱해서 전체 수로 나눠서 구할 수 있다. 평균값은 Ex3.1과 동일한 결과를 얻는다.</td></tr>
</table>

Example.3.3

EXAMPLE 3.3 Life Span for Two Species of Birds in Captivity

The data for each species are arranged in order of magnitude

Species A X_i (mo)	*Species B* X_i (mo)
16	34
32	36
37	38
39	45
40	50
41	54
42	56
50	59
82	69
	91
$n = 9$	$n = 10$
median $= X_{(n+1)/2} = X_{(9+1)/2} = X_5 = 40$ mo	median $= X_{(n+1)/2} = X_{(10+1)/2} = X_{5.5} = 52$ mo
$\overline{X} = 42.11$ mo	$\overline{X} = 53.20$ mo

R

CODE

```
#Ex3.3
A=c(16,32,37,39,40,41,42,50,82)
B=c(34,36,38,45,50,54,56,59,69,91)
xa=c(mean(A),median(A))
xb=c(mean(B),median(B))
xa
xb
```

OUTPUT

```
> #Ex3.3
> A=c(16,32,37,39,40,41,42,50,82)
> B=c(34,36,38,45,50,54,56,59,69,91)
> xa=c(mean(A),median(A))
> xb=c(mean(B),median(B))
> xa
[1] 42.11111 40.00000
> xb
[1] 53.2 52.0
```

SAS

CODE

```
data ex3_3;
input a b @@;
cards;
16 34
32 36
37 38
39 45
```

```
40 50
41 54
42 56
50 59
82 69
. 91
;
run;
proc means data=ex3_3 mean median maxdec=2;
var a;
var b;
run;
```

OUTPUT

Variable	Mean	Median
a	42.11	40.00
b	53.20	52.00

결과해석

철장에 갇힌 A종 새의 수명의 중위수는 40개월이고 평균은 42.11개월이고 B종 새의 수명의 중위수는 52개월이고 평균은 53개월이므로 두 종 모두 평균이 중위수보다 크므로 데이터가 살짝 왼쪽으로 치우칠 가능성이 있다고 판단할 수 있다.

Example.3.4

EXAMPLE 3.4 The Geometric Mean of Ratios of Change

Decade	*Population Size*	*Ratio of Change* X_i
0	10,000	
1	10,500	$\frac{10{,}500}{10{,}000} = 1.05$
2	11,550	$\frac{11{,}550}{10{,}500} = 1.10$
3	13,860	$\frac{13{,}860}{11{,}550} = 1.20$
4	18,156	$\frac{18{,}156}{13{,}860} = 1.31$

$$\overline{X} = \frac{1.05 + 1.10 + 1.20 + 1.31}{4} = \frac{4.66}{4} = 1.1650$$

and $(10{,}000)(0.1650)(1.650)(1.650)(1.650) = 18{,}421$

But,

$$\overline{X}_G = \sqrt[4]{(1.05)(1.10)(1.20)(1.31)} = \sqrt[4]{1.8157} = 1.1608$$

or

$$\overline{X}_G = \text{antilog}\left[\frac{\log(1.05) + \log(1.10) + \log(1.20) + \log(1.31)}{4}\right]$$

$$= \frac{\text{antilog}(0.0212 + 0.0414 + 0.0792 + 0.1173)}{4} = \frac{\text{antilog}(0.2591)}{4}$$

$$= \text{antilog } 0.0648 = 1.1608$$

and $(10{,}000)(1.1608)(1.1608)(1.1608)(1.1608) = 18{,}156$

R

CODE

```
#Ex3.4
install.packages("psych")
library(psych)
RX=c(1.05,1.10,1.20,1.31)
XRX=c(mean(RX),geometric.mean(RX))
XRX
```

OUTPUT

```
> library(psych)
> RX=c(1.05,1.10,1.20,1.31)
> XRX=c(mean(RX),geometric.mean(RX))
> XRX
[1] 1.165000 1.160803
```

SAS

CODE

```
data ex3_4;
input x @@;
cards;
```

	1.05 1.10 1.20 1.31 ; run; proc means data=ex3_4 mean; var x; run; proc ttest data=ex3_4 dist=lognormal; var x; run;
OUTPUT	

Analysis Variable : x
Mean
1.1650000

N	Geometric Mean
4	1.1608

결과해석	한 세기마다 인구 규모의 변화의 평균비율을 산술평균으로 구한 값을 4번 곱하였더니 4세기의 최종 인구 수와 다른 값이 나왔다. 이렇게 변화의 비율에 대한 평균값을 계산할 때에는 기하평균을 이용하는 것이 좀 더 정확한 결과를 얻을 수 있다. 위 결과를 보면 기하평균을 0세기부터 4세기동안 곱해보면 최종 인구 수와 같은 값이 나오는 것을 확인할 수 있다.

Example.3.5

EXAMPLE 3.5 The Harmonic Mean of Rates

$$X_1 = 40 \text{ km/hr}, \ X_2 = 20 \text{ km/hr}$$

$$\overline{X} = \frac{40 \text{ km/hr} + 20 \text{ km/hr}}{2} = \frac{60 \text{ km/hr}}{2} = 30 \text{ km/hr}$$

But

$$\overline{X}_H = \frac{2}{\dfrac{1}{40 \text{ km/hr}} + \dfrac{1}{20 \text{ km/hr}}} = \frac{2}{0.0250 \text{ hr/km} + 0.0500 \text{ hr/km}}$$

$$= \frac{2}{0.075 \text{ hr/km}} = 26.67 \text{ km/hr}$$

R

CODE

```
#Ex3.5
v=c(40,20)
xv=c(mean(v),harmonic.mean(v))
xv
```

OUTPUT

```
> #Ex3.5
> v=c(40,20)
> xv=c(mean(v),harmonic.mean(v))
> xv
[1] 30.00000 26.66667
```

SAS

CODE

```
proc iml;
x={40, 20} ;
mean = mean(x);
harm_mean= 2 / (sum(1/x));
print mean harm_mean ;
run; quit;
```

OUTPUT

mean	harm_mean
30	26.666667

결과해석

속도의 변화같이 변화율의 평균을 구하고 싶을 때에는 조화평균을 사용하는 것이 바람직하다.

Example.3.6

EXAMPLE 3.6 Coding Data to Facilitate Calculations

Sample 1 (Coding by Subtraction: $A = -840$ g)		Sample 2 (Coding by Division: $M = 0.001$ liters/ml)	
X_i (g)	coded $X_i = X_i - 840$ g	X_i (ml)	coded $X_i = (X_i)(0.001 \text{ liters/ml}) = X_i$ liters
842	2	8,000	8.000
844	4	9,000	9.000
846	6	9,500	9.500
846	6	11,000	11.000
847	7	12,500	12.500
848	8	13,000	13.000
849	9		

Sample 1:

$\sum X_i = 5922$ g coded $\sum X_i = 42$ g

$\bar{X} = \frac{5922 \text{ g}}{7} = 846$ g coded $\bar{X} = \frac{42 \text{ g}}{7} = 6$ g

$\bar{X} = \text{coded } \bar{X} - A = 6 \text{ g} - (-840 \text{ g}) = 846$ g

Sample 2:

$\sum X_i = 63{,}000$ ml coded $\sum X_i = 63.000$ liters

$\bar{X} = 10{,}500$ ml coded $\bar{X} = 10.500$ liters

$\bar{X} = \text{coded } \frac{\bar{X}}{M} = \frac{10.500 \text{ liters}}{0.001 \text{ liters/ml}} = 10{,}500$ ml

R

CODE

```
#Ex3.6
x1=c(842,844,846,846,847,848,849)
cx1=x1-840
mean(x1)
mean(cx1)
```

<table>
<tr><td></td><td>mean(cx1)-(-840)
x2=c(8000,9000,9500,11000,12500,13000)
cx2=x2/1000
mean(x2)
mean(cx2)
mean(cx2)/0.001</td></tr>
<tr><td>OUTPUT</td><td>

```
> #Ex3.6
> x1=c(842,844,846,846,847,848,849)
> cx1=x1-840
> mean(x1)
[1] 846
> mean(cx1)
[1] 6
> mean(cx1)-(-840)
[1] 846
> x2=c(8000,9000,9500,11000,12500,13000)
> cx2=x2/1000
> mean(x2)
[1] 10500
> mean(cx2)
[1] 10.5
> mean(cx2)/0.001
[1] 10500
```

</td></tr>
<tr><td colspan="2">SAS</td></tr>
<tr><td>CODE</td><td>

```
proc iml;
x1={842,844,846,846,847,848,849};
x2=x1-840;
mean1=mean(x1);
mean2=mean(x2);
cv=mean2 + 840;
print mean1 mean2 cv;
run; quit;
proc iml;
x1={8000,9000,9500,11000,12500,13000};
x2=x1/1000;
mean1=mean(x1);
mean2=mean(x2);
cv=mean2 * 1000;
print mean1 mean2 cv;
run; quit;
```

</td></tr>
<tr><td>OUTPUT</td><td>

mean1	mean2	cv
846	6	846

mean1	mean2	cv
10500	10.5	10500

</td></tr>
</table>

결과해석	데이터에 일정 숫자를 빼거나 나눠주고 평균을 구한 다음에 빼거나 나눴던 수만큼 다시 더하거나 곱해주면 원래 데이터의 평균과 같은 값이 된다는 것을 확인할 수 있다.

Example.4.1

EXAMPLE 4.1 **Calculation of Measures of Dispersion for Two Hypothetical Samples of 7 Insect Body Weights**

Sample 1

X_i (g)	$X_i - \overline{X}$ (g)	$\lvert X_i - \overline{X}\rvert$ (g)	$(X_i - \overline{X})^2$ (g^2)
1.2	−0.6	0.6	0.36
1.4	−0.4	0.4	0.16
1.6	−0.2	0.2	0.04
1.8	0.0	0.0	0.00
2.0	0.2	0.2	0.04
2.2	0.4	0.4	0.16
2.4	0.6	0.6	0.36
$\sum X_i$ = 12.6 g	$\sum(X_i - \overline{X})$ = 0.0 g	$\sum\lvert X_i - \overline{X}\rvert$ = 2.4 g	$\sum(X_i - \overline{X})^2$ = 1.12 g^2

= sum of squared deviations from the mean
= "sum of squares"

$$n = 7;\ \overline{X} = \frac{\sum X_i}{n} = \frac{12.6\text{ g}}{7} = 1.8\text{ g}$$

$$\text{range} = X_7 - X_1 = 2.4\text{ g} - 1.2\text{ g} = 1.2\text{ g}$$

$$\text{interquartile range} = Q_3 - Q_1 = 2.2\text{ g} - 1.4\text{ g} = 0.8\text{ g}$$

$$\text{mean deviation} = \frac{\sum\lvert X_i - \overline{X}\rvert}{n} = \frac{2.4\text{ g}}{7} = 0.34\text{ g}$$

$$\text{variance} = s^2 = \frac{\sum(X_i - \overline{X})^2}{n-1} = \frac{1.12\text{ g}^2}{6} = 0.1867\text{ g}^2$$

$$\text{standard deviation} = s = \sqrt{0.1867\text{ g}^2} = 0.43\text{ g}$$

Sample 2

X_i (g)	$X_i - \overline{X}$ (g)	$\lvert X_i - \overline{X}\rvert$ (g)	$(X_i - \overline{X})^2$ (g^2)
1.2	−0.6	0.6	0.36
1.6	−0.2	0.2	0.04
1.7	−0.1	0.1	0.01
1.8	0.0	0.0	0.00
1.9	0.1	0.1	0.01
2.0	0.2	0.2	0.04
2.4	0.6	0.6	0.36
$\sum X_i$ = 12.6 g	$\sum(X_i - \overline{X})$ = 0.0 g	$\sum\lvert X_i - \overline{X}\rvert$ = 1.8 g	$\sum(X_i - \overline{X})^2$ = 0.82 g^2

= sum of squared deviations from the mean
= "sum of squares"

$$n = 7;\ \overline{X} = \frac{\sum X_i}{n} = \frac{12.6\text{ g}}{7} = 1.8\text{ g}$$

$$\text{range} = X_7 - X_1 = 2.4\text{ g} - 1.2\text{ g} = 1.2\text{ g}$$

$$\text{interquartile range} = Q_3 - Q_1 = 2.0\ \text{g} - 1.6\ \text{g} = 0.4\ \text{g}$$

$$\text{mean deviation} = \frac{\sum |X_i - \overline{X}|}{n} = \frac{1.8\ \text{g}}{7} = 0.26\ \text{g}$$

$$\text{variance} = s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1} = \frac{0.82\ \text{g}^2}{6} = 0.1367\ \text{g}^2$$

$$\text{standard deviation} = s = \sqrt{0.1367\ \text{g}^2} = 0.37\ \text{g}$$

R

CODE

```
#Ex4.1
xi=c(1.2,1.4,1.6,1.8,2.0,2.2,2.4)
round(sum(xi-mean(xi)),2)
md=sum(abs(xi-mean(xi)))
sum((xi-mean(xi))^2)
mean(xi)
rangex=max(xi)-min(xi)
library(stats)
IQR(xi, type=1)
md/length(xi)
var(xi)
sd(xi)
xi2=c(1.2,1.6,1.7,1.8,1.9,2.0,2.4)
round(sum(xi2-mean(xi2)),2)
md2=sum(abs(xi2-mean(xi2)))
sum((xi2-mean(xi2))^2)
mean(xi2)
rangex2=max(xi2)-min(xi2)
IQR(xi2, type=1)
md2/length(xi2)
var(xi2)
sd(xi2)
```

data가 바뀔 때마다 동일한 계산을 하므로 함수를 만들어 이용해보세요.

OUTPUT

	`` > #Ex4.1 > xi=c(1.2,1.4,1.6,1.8,2.0,2.2,2.4) > round(sum(xi-mean(xi)),2) [1] 0 > md=sum(abs(xi-mean(xi))) > sum((xi-mean(xi))^2) [1] 1.12 > mean(xi) [1] 1.8 > rangex=max(xi)-min(xi) > library(stats) > IQR(xi, type=1) [1] 0.8 > md/length(xi) [1] 0.3428571 > var(xi) [1] 0.1866667 > sd(xi) [1] 0.4320494 > xi2=c(1.2,1.6,1.7,1.8,1.9,2.0,2.4) > round(sum(xi2-mean(xi2)),2) [1] 0 > md2=sum(abs(xi2-mean(xi2))) > sum((xi2-mean(xi2))^2) [1] 0.82 > mean(xi2) [1] 1.8 > rangex2=max(xi2)-min(xi2) > IQR(xi2, type=1) [1] 0.4 > md2/length(xi2) [1] 0.2571429 > var(xi2) [1] 0.1366667 > sd(xi2) [1] 0.3696846 ```
SAS	
CODE	data ex4_1; input s1 s2 @@; cards; 1.2 1.2 1.4 1.6 1.6 1.7 1.8 1.8 2.0 1.9 2.2 2.0 2.4 2.4 ; run;

```
proc univariate data=ex4_1;
var s1;
var s2;
run;
```

OUTPUT

기본 통계 측도			
위치측도		변이측도	
평균	1.800000	표준 편차	0.43205
중위수	1.800000	분산	0.18667
최빈값	.	범위	1.20000
		사분위수 범위	0.80000

기본 통계 측도			
위치측도		변이측도	
평균	1.800000	표준 편차	0.36968
중위수	1.800000	분산	0.13667
최빈값	.	범위	1.20000
		사분위수 범위	0.40000

결과해석

7마리의 곤충의 무게에 관한 산포들을 구한 것으로 순수한 값들로 봤을 때는 편차가 적다고 생각할 수도 있으나 곤충 자체가 작으므로 상대적으로 산포가 크다고 판단할 수도 있다. 이렇게 산포는 단위에 따라 상대적으로 결정되기 때문에 절대적인 판단 기준이 모호하다.

Example.4.2

EXAMPLE 4.2 "Machine Formula" Calculation of Variance, Standard Deviation, and Coefficient of Variation (These are the data of Example 4.1)

Sample 1

X_i (g)	X_i^2 (g^2)
1.2	1.44
1.4	1.96
1.6	2.56
1.8	3.24
2.0	4.00
2.2	4.84
2.4	5.76

$\sum X_i = 12.6 \text{ g} \quad \sum X_i^2 = 23.80 \text{ g}^2$

$n = 7$

$$\bar{X} = \frac{12.6 \text{ g}}{7} = 1.8 \text{ g}$$

$$SS = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$= 23.80 \text{ g}^2 - \frac{(12.6 \text{ g})^2}{7}$$

$$= 23.80 \text{ g}^2 - 22.68 \text{ g}^2$$

$$= 1.12 \text{ g}^2$$

$$s^2 = \frac{SS}{n-1}$$

$$= \frac{1.12 \text{ g}^2}{6} = 0.1867 \text{ g}^2$$

$$s = \sqrt{0.1867 \text{ g}^2} = 0.43 \text{ g}$$

$$V = \frac{s}{\bar{X}} = \frac{0.43 \text{ g}}{1.8 \text{ g}} = 0.24 = 24\%$$

Sample 2

X_i (g)	X_i^2 (g^2)
1.2	1.44
1.6	2.56
1.7	2.89
1.8	3.24
1.9	3.61
2.0	4.00
2.4	5.76

$\sum X_i = 12.6 \text{ g} \quad \sum X_i^2 = 23.50 \text{ g}^2$

$n = 7$

$$\bar{X} = \frac{12.6 \text{ g}}{7} = 1.8 \text{ g}$$

$$SS = 23.50 \text{ g}^2 - \frac{(12.6 \text{ g})^2}{7}$$

$$= 0.82 \text{ g}^2$$

$$s^2 = \frac{0.82 \text{ g}^2}{6} = 0.1367 \text{ g}^2$$

$$s = \sqrt{0.1367 \text{ g}^2} = 0.37 \text{ g}$$

$$V = \frac{0.37 \text{ g}}{1.8 \text{ g}} = 0.21 = 21\%$$

R	
CODE	`#Ex4.2` `x1=c(1.2,1.4,1.6,1.8,2.0,2.2,2.4)` `mean(x1)` `ss1=sum(x1^2)-sum(x1)^2/length(x1)` `sv1=ss1/(length(x1)-1)` `sqrt(sv1)` `cv1=sqrt(sv1)/mean(x1)` `cv1` `x2=c(1.2,1.6,1.7,1.8,1.9,2.0,2.4)` `mean(x2)` `ss2=sum(x2^2)-sum(x2)^2/length(x2)` `sv2=ss2/(length(x2)-1)` `sqrt(sv2)` `cv2=sqrt(sv2)/mean(x2)` `cv2`
OUTPUT	`> #Ex4.2` `> x1=c(1.2,1.4,1.6,1.8,2.0,2.2,2.4)` `> mean(x1)` `[1] 1.8` `> ss1=sum(x1^2)-sum(x1)^2/length(x1)` `> sv1=ss1/(length(x1)-1)` `> sqrt(sv1)` `[1] 0.4320494` `> cv1=sqrt(sv1)/mean(x1)` `> cv1` `[1] 0.2400274` `> x2=c(1.2,1.6,1.7,1.8,1.9,2.0,2.4)` `> mean(x2)` `[1] 1.8` `> ss2=sum(x2^2)-sum(x2)^2/length(x2)` `> sv2=ss2/(length(x2)-1)` `> sqrt(sv2)` `[1] 0.3696846` `> cv2=sqrt(sv2)/mean(x2)` `> cv2` `[1] 0.2053803`
SAS	
CODE	`data ex4_2;` `input s1 s2 @@;` `cards;` `1.2 1.2` `1.4 1.6` `1.6 1.7`

```
1.8 1.8
2.0 1.9
2.2 2.0
2.4 2.4
;
run;
proc univariate data=ex4_2;
var s1;
var s2;
run;
```

OUTPUT

N	7	**가중합**	7
평균	1.8	**관측값 합**	12.6
표준 편차	0.43204938	**분산**	0.18666667

N	7	**가중합**	7
평균	1.8	**관측값 합**	12.6
표준 편차	0.36968455	**분산**	0.13666667

결과해석

두 표본을 뽑아서 산포를 구한 것으로 값이 어느정도 비슷하게 나왔다고 볼 수 있다. 만약 정밀한 결과를 요구한다면 산포가 크다고 판단할 수도 있다. 여기서는 같은 모집단에서 추출한 표본들이므로 단위가 같아 변동계수는 큰 의미가 없다.

Example.4.3

EXAMPLE 4.3 Indices of Diversity for Nominal Scale Data: The Nesting Sites of Sparrows

Category (i)	Observed Frequencies (f_i)
	Sample 1
Vines	5
Eaves	5
Branches	5
Cavities	5

$$H' = \frac{n \log n - \sum f_i \log f_i}{n} = [20 \log 20 - (5 \log 5 + 5 \log 5 + 5 \log 5 + 5 \log 5)]/20$$
$$= [26.0206 - (3.4949 + 3.4949 + 3.4949 + 3.4949)]/20$$
$$= 12.0410/20 = 0.602$$
$$H'_{\max} = \log 4 = 0.602$$
$$J' = \frac{0.602}{0.602} = 1.00$$

	Sample 2
Vines	1
Eaves	1
Branches	1
Cavities	17

$$H' = \frac{n \log n - \sum f_i \log f_i}{n} = [20 \log 20 - (1 \log 1 + 1 \log 1 + 1 \log 1 + 17 \log 17)]/20$$
$$= [26.0206 - (0 + 0 + 0 + 20.9176)]/20$$
$$= 5.1030/20 = 0.255$$
$$H'_{\max} = \log 4 = 0.602$$
$$J' = \frac{0.255}{0.602} = 0.42$$

	Sample 3
Vines	2
Eaves	2
Branches	2
Cavities	34

$$H' = \frac{n \log n - \sum f_i \log f_i}{n} = [40 \log 40 - (2 \log 2 + 2 \log 2 + 2 \log 2 + 34 \log 34)]/40$$
$$= [64.0824 - (0.6021 + 0.6021 + 0.6021 + 52.0703)]/40$$
$$= 10.2058/40 = 0.255$$
$$H'_{\max} = \log 4 = 0.602$$
$$J' = \frac{0.255}{0.602} = 0.42$$

R	
CODE	#Ex4.3

	install.packages("vegan") library(vegan) h1=diversity(c(5,5,5,5),base = 10) hmax=log10(length(c(5,5,5,5))) h1/hmax h2=diversity(c(1,1,1,17), base = 10) h2/hmax h3=diversity(c(2,2,2,34), base = 10) h3/hmax
OUTPUT	> h1=diversity(c(5,5,5,5),base = 10) > hmax=log10(length(c(5,5,5,5))) > h1/hmax [1] 1 > h2=diversity(c(1,1,1,17), base = 10) > h2/hmax [1] 0.4237923 > h3=diversity(c(2,2,2,34), base = 10) > h3/hmax [1] 0.4237923
SAS	
CODE	proc iml; s1={5,5,5,5}; h1=(sum(s1)*log10(sum(s1))-t(s1)*log10(s1))/sum(s1); hmax=log10(nrow(s1)); j1=h1/hmax; s2={1,1,1,17}; h2=(sum(s2)*log10(sum(s2))-t(s2)*log10(s2))/sum(s2); j2=h2/hmax; s3={2,2,2,34}; h3=(sum(s3)*log10(sum(s3))-t(s3)*log10(s3))/sum(s3); j3=h3/hmax; print hmax h1 j1 h2 j2 h3 j3 ; run; quit;
OUTPUT	hmax: 0.60206, h1: 0.60206, j1: 1, h2: 0.2551484, j2: 0.4237923, h3: 0.2551484, j3: 0.4237923
결과해석	첫 번째 표본은 종부유도, 종풍부도 모두 균등하고 종균등도 척도가 1로 다양한 종이 고르고 분포되어 있다고 볼 수 있다. 두 번째 표본은 종부유도는 첫 번째 표본과 같지만 한 종의 나무가 거의 대부분의 비율을 차지하고 종균등도 척도도 0.4238로 다양하지 못하게 분포되어 있다고 볼 수 있다. 세 번째 표본은 두 번째 표본의 각 개체 수의 정확히 두배로 분포되어 있는데 종균등도 척도가 두 번째 표본과 같은 값인 0.4238이 나오는 것을 확인할 수 있다.

SAS OUTPUT:

hmax	h1	j1	h2	j2	h3	j3
0.60206	0.60206	1	0.2551484	0.4237923	0.2551484	0.4237923

Example.4.4

EXAMPLE 4.4 Coding Data to Facilitate the Calculation of Measures of Dispersion

Sample 1 (Coding by Subtraction: $A = -840$ g)

Without Coding X_i		*Using Coding* $[X_i]$	
X_i (g)	X_i^2 (g^2)	$[X_i]$ (g)	$[X_i]^2$ (g^2)
842	708,964	2	4
843	710,649	3	9
844	712,336	4	16
846	715,716	6	36
846	715,716	6	36
847	717,409	7	49
848	719,104	8	64
849	720,801	9	81

$\sum X_i = 6765$ g $\quad \sum X_i^2 = 5{,}720{,}695$ g^2 $\qquad \sum[X_i] = 45$ g $\quad \sum[X_i]^2 = 295$ g^2

$$s^2 = \frac{5720695\ g^2 - \frac{(6765\ g)^2}{8}}{7}$$
$$= 5.98\ g^2$$
$$s = 2.45\ g$$
$$\bar{X} = 845.6\ g$$
$$V = \frac{s}{\bar{X}} = \frac{2.45\ g}{845.6\ g}$$
$$= 0.0029 = 0.29\%$$

$$[s^2] = \frac{295\ g^2 - \frac{(45\ g)^2}{8}}{7}$$
$$= 5.98\ g^2$$
$$[s] = 2.44\ g$$
$$[\bar{X}] = 5.6\ g$$

Sample 2 (Coding by Division: $M = 0.01$)

Without Coding X_i		*Using Coding* $[X_i]$	
X_i (sec)	X_i^2 (sec^2)	$[X_i]$ (sec)	$[X_i]^2$ (sec^2)
800	640,000	8.00	64.00
900	810,000	9.00	81.00
950	902,500	9.50	90.25
1100	1,210,000	11.00	121.00
1250	1,562,500	12.50	156.25
1300	1,690,000	13.00	169.00

$\sum X_i = 6300$ sec $\sum X_i^2 = 6{,}815{,}000$ sec^2 $\qquad \sum[X_i] = 63.00$ sec $\sum[X_i]^2 = 681.50$ sec^2

$$s^2 = \frac{6815000\ sec^2 - \frac{(6300\ sec)^2}{6}}{5}$$
$$= 40{,}000\ sec^2$$
$$s = 200\ sec$$
$$\bar{X} = 1050\ sec$$
$$V = 0.19 = 19\%$$

$$[s^2] = \frac{681.50\ sec^2 - \frac{(63.00\ sec)^2}{6}}{5}$$
$$= 4\ sec^2$$
$$[s] = 2.00\ sec$$
$$[\bar{X}] = 10.50\ sec$$
$$[V] = 0.19 = 19\%$$

R

CODE

```
#Ex4.4
x41=c(842,843,844,846,846,847,848,849)
var(x41)
sd(x41)
mean(x41)
```

	sd(x41)/mean(x41) cx41=c(2,3,4,6,6,7,8,9) var(cx41) sd(x41) mean(cx41) sd(cx41)/mean(x41) x42=c(800,900,950,1100,1250,1300) var(x42) sd(x42) mean(x42) sd(x42)/mean(x42) cx42=c(8,9,9.5,11,12.5,13) var(cx42) sd(x42) mean(cx42) sd(cx42)/mean(x42)
OUTPUT	

	`` > #Ex4.4 > x41=c(842,843,844,846,846,847,848,849) > var(x41) [1] 5.982143 > sd(x41) [1] 2.445842 > mean(x41) [1] 845.625 > sd(x41)/mean(x41) [1] 0.002892348 > cx41=c(2,3,4,6,6,7,8,9) > var(cx41) [1] 5.982143 > sd(x41) [1] 2.445842 > mean(cx41) [1] 5.625 > sd(cx41)/mean(x41) [1] 0.002892348 > > x42=c(800,900,950,1100,1250,1300) > var(x42) [1] 40000 > sd(x42) [1] 200 > mean(x42) [1] 1050 > sd(x42)/mean(x42) [1] 0.1904762 > cx42=c(8,9,9.5,11,12.5,13) > var(cx42) [1] 4 > sd(x42) [1] 200 > mean(cx42) [1] 10.5 > sd(cx42)/mean(x42) [1] 0.001904762 ```
SAS	
CODE	data ex4_4; input ucs1 cs1 ucs2 cs2 @@; cards; 842 2 800 8 843 3 900 9 844 4 950 9.5 846 6 1100 11 846 6 1250 12.5 847 7 1300 13 848 8 . . 849 9 . .

	`` ; run; proc univariate data=ex4_4; var ucs1; var cs1; var ucs2; var cs2; run; ```
OUTPUT	(see tables below)

변수: ucs1

적률			
N	8	가중합	8
평균	845.625	관측값 합	6765
표준 편차	2.44584195	분산	5.98214286

변수: cs1

적률			
N	8	가중합	8
평균	5.625	관측값 합	45
표준 편차	2.44584195	분산	5.98214286

변수: ucs2

적률			
N	6	가중합	6
평균	1050	관측값 합	6300
표준 편차	200	분산	40000

변수: cs2

적률			
N	6	가중합	6
평균	10.5	관측값 합	63
표준 편차	2	분산	4

결과해석	첫 번째 표본을 보면 기존 자료가 숫자가 커서 각각에 840을 빼고 구해봤더니 기존의 분산, 표준편차와 똑같은 값을 얻었다. 이를 통해 분산, 표준편차는 원자료에 일정한 값을 빼줘도 같은 값을 얻는다는 것을 알 수 있다. 두 번째 표본은 원자료에 100씩 나눠서 분산을 구한 것으로, 원자료의 분산보다 100의 제곱, 즉 1만 배만큼 적은 값을 얻을 수 있다. 즉 원자료에 일정한 값을 나눈 자료의 분산은 기존 자료의 분산에 나눠준 일정한 값의 제곱만큼 나눈 값이다