

I. 기술통계량 계산

II. 평균의 계산

1. 자료의 중심

- 평균(산술평균) : mean 함수를 이용하여 계산
- 중위수 : median 함수를 이용하여 계산
- 최빈값 : base로 주어진 함수가 없어 programming을 하거나 prettyR 패키지의 Mode 함수를 이용

```
> ch2_1
[1] 65 74 65 36 81 60 43 21 83 64 12 91 60 24 54 69 89 96 86 85 95 85 51 81 47 62 85 46 49 76 44 72 33 46 49 74 78 48 62 97 31 96
[43] 97 88 61 54 89 77 72 35
```

```
b1-ch3-9.rev.R

library(prettyR)
ch2_1<-scan("http://kanggc.iptime.org/book/data/ch2_1.txt")
ch2_1
max(ch2_1)
min(ch2_1)
mean(ch2_1)
median(ch2_1)
summary(ch2_1)
table(ch2_1)
max(table(ch2_1))
table(ch2_1)[table(ch2_1)==max(table(ch2_1))]
Mode(ch2_1) )
```

```
> max(ch2_1)
[1] 97
> min(ch2_1)
[1] 12
> mean(ch2_1)
[1] 64.76
> median(ch2_1)
[1] 65
> summary(ch2_1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 12.00  48.25   65.00   64.76  84.50   97.00
```

```
> max(table(ch2_1))
[1] 3
```

```
> table(ch2_1)[table(ch2_1)==max(table(ch2_1))]
85
3
```

```
> Mode(ch2_1)
[1] "85"
```

```
> table(ch2_1)
ch2_1
12 21 24 31 33 35 36 43 44 46 47 48 49 51 54 60 61 62 64 65 69 72 74 76 77 78 81 83 85 86 88 89 91 95 96 97
 1  1  1  1  1  1  1  1  1  1  2  1  1  2  1  2  2  1  2  1  2  1  2  2  1  1  1  2  1  3  1  1  2  1  1  2  2
```

2. 자료의 변화량

- 범위 : range 및 diff 함수를 이용하여 계산
- 분산 : var 함수를 이용하여 계산 또는 편차 및 변동을 먼저 구하고 변동을 자유도(=자료 -1)로 나눔
- 표준편차 : sd 함수를 이용하여 계산

b1-ch3-9.rev.R

앞에서 계속

```
diff(range(ch2_1))
```

```
(dev<-ch2_1-mean(ch2_1))
```

```
(devsq<-dev^2)
```

```
(variation<-sum(devsq))
```

```
(variance<-variation/(length(ch2_1)-1))
```

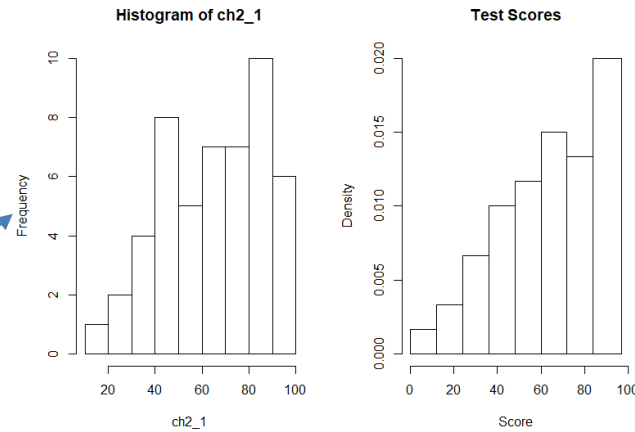
```
var(ch2_1)
```

```
sd(ch2_1)
```

```
> diff(range(ch2_1))
[1] 85
> (dev<-ch2_1-mean(ch2_1))
[1] 0.24 9.24 0.24 -28.76 16.24 -4.76 -21.76 -43.76 18.24 -0.76 -52.76 26.24 -4.76 -40.76 -10.76 4.24 24.24 31.24
[19] 21.24 20.24 30.24 20.24 -13.76 16.24 -17.76 -2.76 20.24 -18.76 -15.76 11.24 -20.76 7.24 -31.76 -18.76 -15.76 9.24
[37] 13.24 -16.76 -2.76 32.24 -33.76 31.24 32.24 23.24 -3.76 -10.76 24.24 12.24 7.24 -29.76
> (devsq<-dev^2)
[1] 0.0576 85.3776 0.0576 827.1376 263.7376 22.6576 473.4976 1914.9376 332.6976 0.5776 2783.6176 688.5376
[13] 22.6576 1661.3776 115.7776 17.9776 587.5776 975.9376 451.1376 409.6576 914.4576 409.6576 189.3376 263.7376
[25] 315.4176 7.6176 409.6576 351.9376 248.3776 126.3376 430.9776 52.4176 1008.6976 351.9376 248.3776 85.3776
[37] 175.2976 280.8976 7.6176 1039.4176 1139.7376 975.9376 1039.4176 540.0976 14.1376 115.7776 587.5776 149.8176
[49] 52.4176 885.6576
> (variation<-sum(devsq))
[1] 24053.12
> (variance<-variation/(length(ch2_1)-1))
[1] 490.88
> var(ch2_1)
[1] 490.88
> sd(ch2_1)
[1] 22.15581
```

3. 도수분포표 및 히스토그램

- 도수분포표 작성 방법
 - 구간의 개수는 자료 개수의 제곱근
 - 구간의 너비는 범위/구간의 수



```
b1-ch3-9.rev.R

앞에서 계속

bins<-c(0, 12, 24, 36, 48, 60, 72, 84, 97)
class<-cut(ch2_1, breaks=bins)
table(class)
table(class)/length(ch2_1)
transform(table(class), Rel_Freq=prop.table(Freq))
par(mfrow=c(1,2))
hist(ch2_1)
hist(ch2_1, breaks=bins, main="Test Scores", xlab="Score")
```

```
> table(class)
class
(0,12] (12,24] (24,36] (36,48] (48,60] (60,72] (72,84] (84,97]
      1      2      4      6      7      9      8     13

> table(class)/length(ch2_1)
class
(0,12] (12,24] (24,36] (36,48] (48,60] (60,72] (72,84] (84,97]
 0.02  0.04  0.08  0.12  0.14  0.18  0.16  0.26

> transform(table(class), Rel_Freq=prop.table(Freq))
      class Freq Rel_Freq
1 (0,12]      1    0.02
2 (12,24]      2    0.04
3 (24,36]      4    0.08
4 (36,48]      6    0.12
5 (48,60]      7    0.14
6 (60,72]      9    0.18
7 (72,84]      8    0.16
8 (84,97]     13    0.26
```

```
> ch2_1
[1] 65 74 65 36 81 60 43 21 83 64 12 91 60 24 54 69 89 96 86 85 95 85 51 81 47 62 85 46 49 76 44 72 33 46 49 74 78 48 62 97 31 96
[43] 97 88 61 54 89 77 72 35

> class<-cut(ch2_1, breaks=bins)
> class
[1] (60,72] (72,84] (60,72] (24,36] (72,84] (48,60] (36,48] (12,24] (72,84] (60,72] (0,12] (84,97] (48,60] (12,24] (48,60]
[16] (60,72] (84,97] (84,97] (84,97] (84,97] (84,97] (84,97] (48,60] (72,84] (36,48] (60,72] (84,97] (36,48] (48,60] (72,84]
[31] (36,48] (60,72] (24,36] (36,48] (48,60] (72,84] (72,84] (36,48] (60,72] (84,97] (24,36] (84,97] (84,97] (84,97] (60,72]
[46] (48,60] (84,97] (72,84] (60,72] (24,36]
Levels: (0,12] (12,24] (24,36] (36,48] (48,60] (60,72] (72,84] (84,97]
```

1. 산술평균

- 산술평균 : mean 함수를 이용하여 계산

2. 조화평균(harmonic mean : H)

$$\frac{1}{H} = \frac{1}{n} \left(\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n} \right) = \frac{\sum_{i=1}^n \frac{1}{X_i}}{n}$$

- 일반적으로 역수의 형태로 된 변수를 평균할 경우 사용
- n개의 양수인 변수의 조화평균은 그 변수의 역수를 산술평균한 것의 역수
- (예) 자동차가 처음 10km를 시속 30km로 달리고, 다음 10km를 시속 60km로 달렸을 경우 평균시속은?



- 산술평균 : $\frac{1}{2}(30 + 60) = 45km/h$
- 조화평균 : $\frac{2}{\frac{1}{30} + \frac{1}{60}} = 40km/h$
- 20km를 가는데 30분이 소요되는데 평균시속이 45km/h이면 20km를 가는데 30분이 소요되지 않고, 40km/h이면 30분이 소요됨

3. 기하평균(geometric mean : G)

$$G = \sqrt[n]{X_1 X_2 \dots X_n}$$

- 여러 개의 수를 연속으로 곱해 그 개수의 거듭제곱근으로 구한 수
- 비율을 평균하는 경우에 적합하며, 연평균 인구증가율이나 연평균 경제성장률을 구할 때 활용
- 최초년도와 최종년도의 데이터만 있으면 연평균 증가율 계산이 가능

연도	인구	전년대비비율(%)	$\bar{X} \times$ 전년도인구	$G \times$ 전년도인구
2014	5000	-	-	-
2015	6000	120	6133	6128
2016	7800	130	7523	7510
2017	9204	118	9228	9204
-	-	-	$\bar{X}=122.67$	$G=122.56$

$$G = \left(\left(\frac{\text{최종년도의 값}}{\text{최초년도의 값}} \right)^{(1/n)} - 1 \right) \times 100, \text{ n은 경과년도}$$

b1-ch3-10.R

```
a<-c(10, 2, 19, 24, 6, 23, 47, 24, 54, 77)
n<-length(a)
```

```
mean(a)
1/mean(1/a)
prod(a)^(1/n)
```

```
> mean(a)
[1] 28.6
> 1/mean(1/a) # compute the harmonic mean
[1] 10.01109
> prod(a)^(1/n) # compute the geometric mean
[1] 18.92809
```

```
b<-c(30,60)
mean(b)
1/mean(1/b)
```

```
> b<-c(30,60)
> mean(b)
[1] 45
> 1/mean(1/b)
[1] 40
```

```
c<-c(120,130,118)
m<-length(c)
round(mean(c),digits=2)
round(prod(c)^(1/m),digits=2)
```

```
> c<-c(120,130,118)
> m<-length(c)
> round(mean(c),digits=2)
[1] 122.67
> round(prod(c)^(1/m),digits=2)
[1] 122.56
```

```
g<-((9204/5000)^(1/3))*100
round(g, digits=2)
g<-((9204/5000)^(1/3)-1)*100
round(g, digits=2)
```

```
> g<-((9204/5000)^(1/3))*100
> round(g, digits=2)
[1] 122.56
> g<-((9204/5000)^(1/3)-1)*100
> round(g, digits=2)
[1] 22.56
```