Korean Income Distribution

coop711

2015년 4월 12일

Data Management

자료 입력

```
income.kor<-read.table("labor_income_kor.txt", header=T, row.names=1)
income.kor
```

```
##          percentage.workers percentage.labour.income
## 0-5                    19.1                       1.7
## 5-10                   12.3                       3.6
## 10-20                  22.8                      12.8
## 20-30                  14.4                      13.6
## 30-40                   9.8                      13.0
## 40-60                  12.0                      22.5
## 60-80                   5.4                      14.1
## 80-100                  2.3                       7.8
## 100-200                 1.6                       7.7
## 200-300                 0.1                       1.2
## 300-500                 0.1                       0.8
## 500-1000                0.0                       0.6
## 1000-                   0.0                       0.6
```

```
str(income.kor)
```

```
## 'data.frame':    13 obs. of  2 variables:
##  $ percentage.workers      : num  19.1 12.3 22.8 14.4 9.8 12 5.4 2.3 1.6 0.1
...
##  $ percentage.labour.income: num  1.7 3.6 12.8 13.6 13 22.5 14.1 7.8 7.7 1.2
...
```

barplot() 을 그리기 위하여 height 를 설정하려면 width를 파악하여야 함. 그러기 위해서 소득 구간을 row.names의 구간으로부터 설정.

```
income.breaks<-c(0,5,10,20,30,40,60,80,100,200,300,500,1000,2000)
```

width에 해당하는 각 소득구간의 폭을 계산

```
income.widths<-diff(income.breaks)
income.widths
```

```
##  [1]    5    5   10   10   10   20   20   20  100  100  200  500 1000
```

각 기둥의 면적이 해당 소득구간의 퍼센티지와 같게 해주려면 각 퍼센티지를 width 로 나눠줘야 함.

```
height.workers<-income.kor[,1]/income.widths
round(height.workers, digits=3)
```

```
##  [1] 3.820 2.460 2.280 1.440 0.980 0.600 0.270 0.115 0.016 0.001 0.000
## [12] 0.000 0.000
```

아무런 argument 도 설정하지 않고 barplot() 을 그리면

```
barplot(height.workers, income.widths)
```

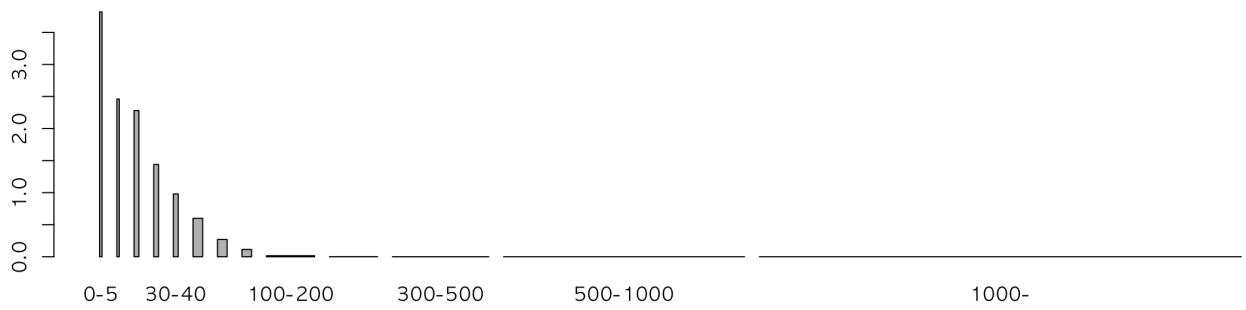
각 bar 의 이름을 row.names에서 가져오면

```
names.bar<-rownames(income.kor)
names.bar
```

```
##  [1] "0-5"      "5-10"     "10-20"    "20-30"    "30-40"    "40-60"
##  [7] "60-80"    "80-100"   "100-200"  "200-300"  "300-500"  "500-1000"
## [13] "1000-"
```

bar 의 이름을 넣어 다시 그리되, bar 사이의 공간을 없애면

```
barplot(height.workers, income.widths, names.arg=names.bar)
```



실제 인원은 거의 없는 것처럼 보이는 5억원 이상의 구간을 합쳐야 할 필요. 자료를 재구성하면,

```
income.kor.2<-income.kor[1:11,]
income.kor.2[11,]<-apply(income.kor[11:13,], 2, sum)
income.kor.2
```

```
##         percentage.workers percentage.labour.income
## 0-5                   19.1                       1.7
## 5-10                  12.3                       3.6
## 10-20                 22.8                      12.8
## 20-30                 14.4                      13.6
## 30-40                  9.8                      13.0
## 40-60                 12.0                      22.5
## 60-80                  5.4                      14.1
## 80-100                 2.3                       7.8
## 100-200                1.6                       7.7
## 200-300                0.1                       1.2
## 300-500                0.1                       2.0
```

```
rownames(income.kor.2)
```

```
##  [1] "0-5"     "5-10"    "10-20"   "20-30"   "30-40"   "40-60"   "60-80"
##  [8] "80-100"  "100-200" "200-300" "300-500"
```

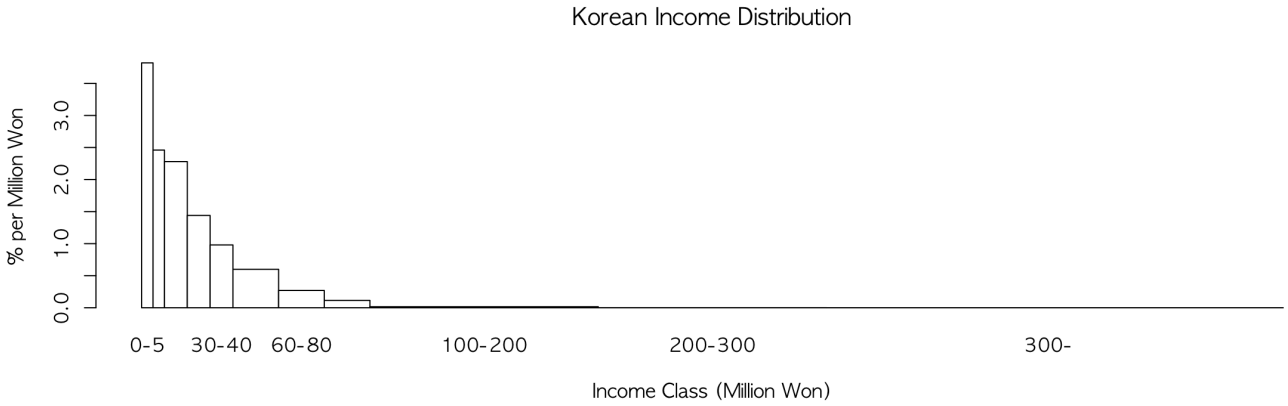
```
rownames(income.kor.2)[11]<-"300-  "
income.kor.2
```

```
##         percentage.workers percentage.labour.income
## 0-5                   19.1                       1.7
## 5-10                  12.3                       3.6
## 10-20                 22.8                      12.8
## 20-30                 14.4                      13.6
## 30-40                  9.8                      13.0
## 40-60                 12.0                      22.5
## 60-80                  5.4                      14.1
## 80-100                 2.3                       7.8
## 100-200                1.6                       7.7
## 200-300                0.1                       1.2
## 300-                   0.1                       2.0
```

```
income.breaks.2<-c(0,5,10,20,30,40,60,80,100,200,300,500)
income.widths.2<-diff(income.breaks.2)
height.workers.2<-income.kor.2[,1]/income.widths.2
names.bar.2<-rownames(income.kor.2)
```

다시 barplot()을 작동시키되 회색 대신 흰색을 넣고, bar 사이의 공간을 없애면

```
barplot(height.workers.2, income.widths.2, names.arg=names.bar.2, space=0, co
l="white")
title(main="Korean Income Distribution", xlab="Income Class (Million Won)", yla
b="% per Million Won")
```

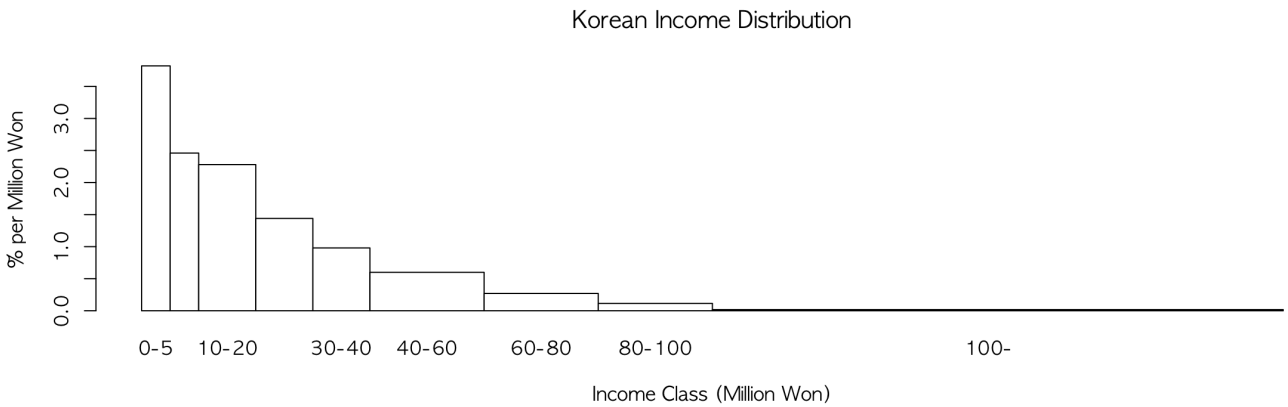


1억 이상의 구간을 합치기 위하여 자료를 다시 손보면,

```
income.kor.3<-income.kor.2[1:9,]
income.kor.3[9,]<-apply(income.kor.2[9:11,], 2, sum)
rownames(income.kor.3)[9]<-"100-   "
income.breaks.3<-c(0,5,10,20,30,40,60,80,100,200)
income.widths.3<-diff(income.breaks.3)
height.workers.3<-income.kor.3[,1]/income.widths.3
names.bar.3<-rownames(income.kor.3)
```

1억 이상의 구간을 합쳐 barplot을 그리면,

```
barplot(height.workers.3, income.widths.3, names.arg=names.bar.3, space=0, co
l="white")
title(main="Korean Income Distribution", xlab="Income Class (Million Won)", yla
b="% per Million Won")
```

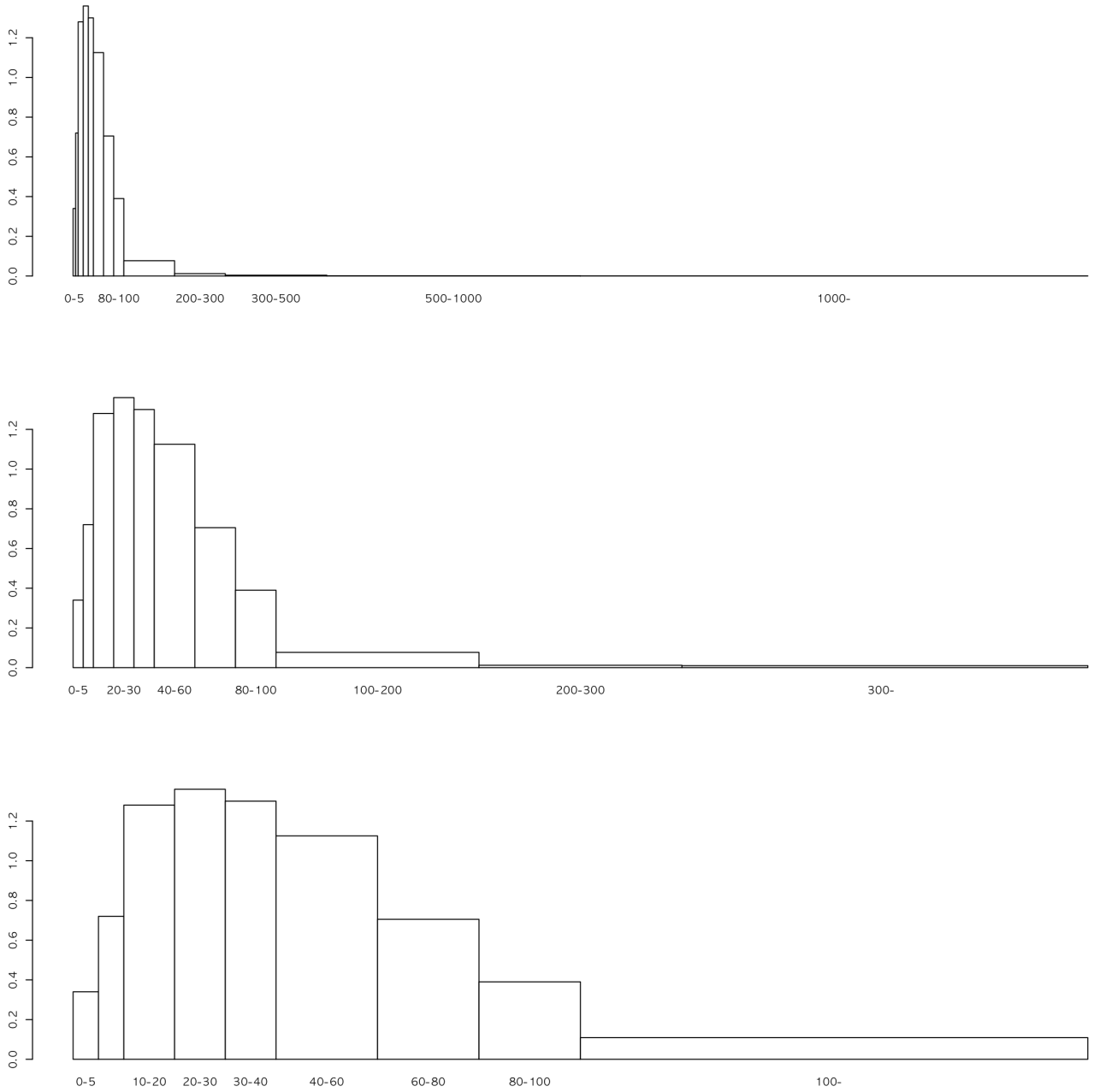


같은 방법으로 소득규모에 대하여 세 개의 barplot을 그리려면, 우선 자료를 정리하고.

```
height.income<-income.kor[,2]/income.widths
height.income.2<-income.kor.2[,2]/income.widths.2
height.income.3<-income.kor.3[,2]/income.widths.3
```

세 개의 barplot을 연속으로 그리기 위하여 par(mfrow=c(3,1)) 설정

```
par(mfrow=c(3,1))
barplot(height.income, income.widths, names.arg=names.bar, space=0, col="whit
e")
barplot(height.income.2, income.widths.2, names.arg=names.bar.2, space=0, co
l="white")
barplot(height.income.3, income.widths.3, names.arg=names.bar.3, space=0, co
l="white")
```




```
par(mfrow=c(1,1))
```

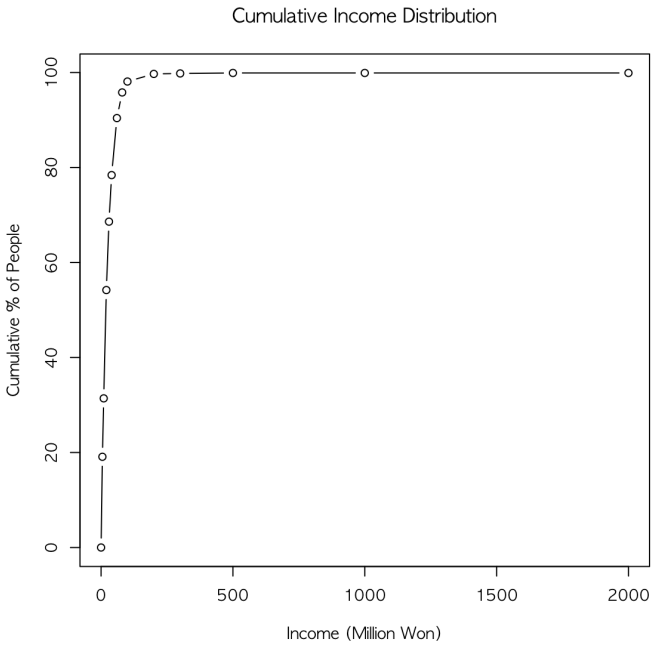
barplot 보다 누적도표가 분포의 윤곽을 살피는 데 더 낫다는 점을 상기하면, 누적분포를 구하는 일부터 시작하여야 함. 자료로부터 이미 아는 사실이지만, cumsum()함수의 활용겸 확인차 계산해보면

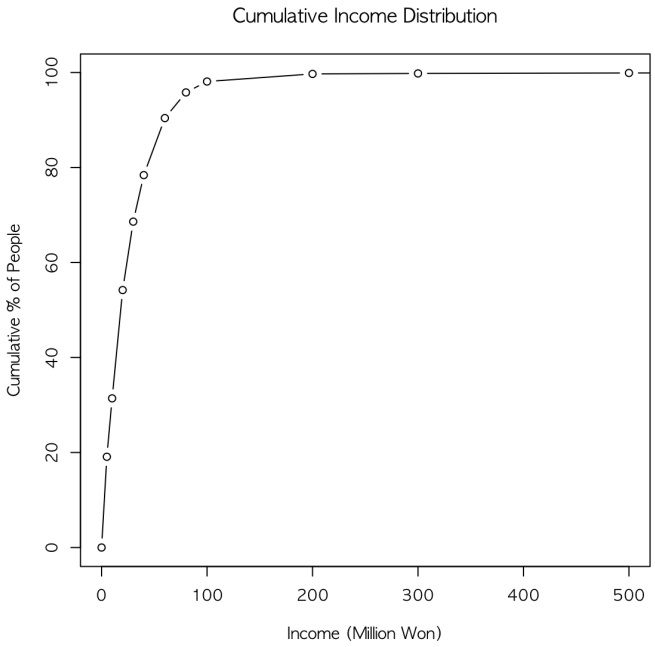
```
income.kor.cum<-apply(income.kor, 2, cumsum)
income.kor.cum
```

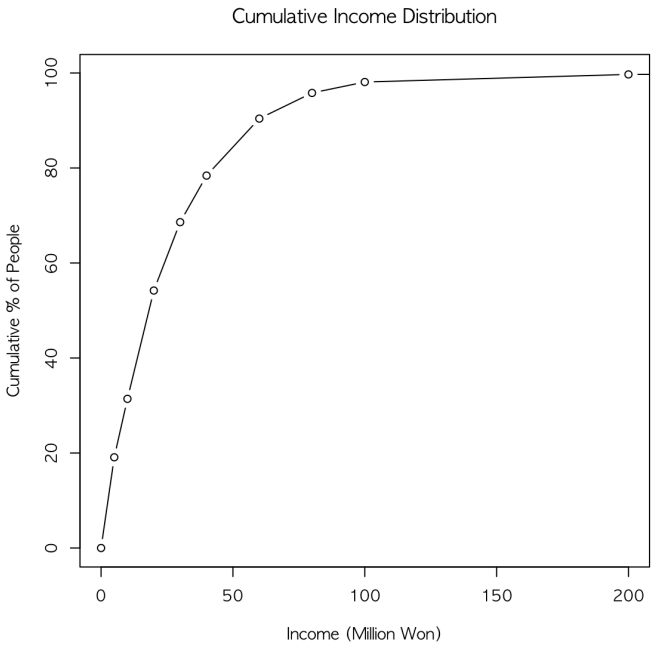
```
##          percentage.workers percentage.labour.income
## 0-5                    19.1                       1.7
## 5-10                   31.4                       5.3
## 10-20                  54.2                      18.1
## 20-30                  68.6                      31.7
## 30-40                  78.4                      44.7
## 40-60                  90.4                      67.2
## 60-80                  95.8                      81.3
## 80-100                 98.1                      89.1
## 100-200                99.7                      96.8
## 200-300                99.8                      98.0
## 300-500                99.9                      98.8
## 500-1000               99.9                      99.4
## 1000-                  99.9                     100.0
```

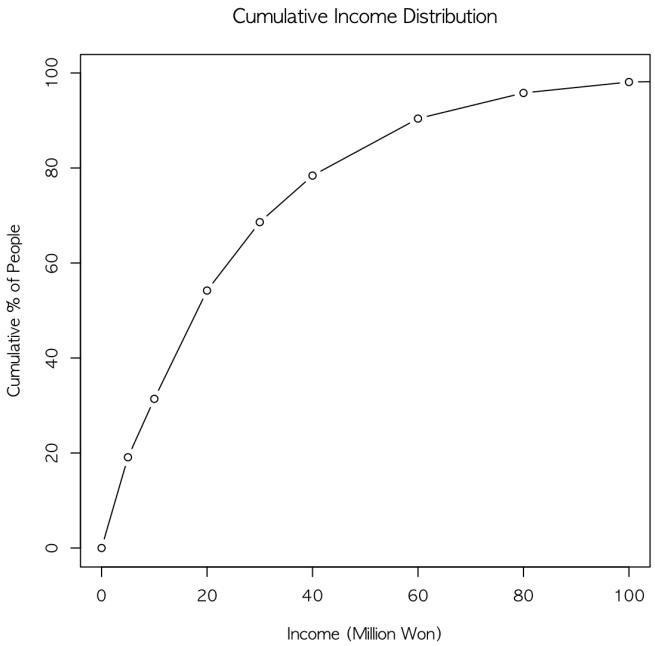
누적도표를 그리려면 첫 시작은 (0, 0)이어야 함에 유의. 마침 income.breaks 와 맞춰보면 income.kor.cum 의 첫 행을 0으로만 추가해 주면 되는 일임. xlim 을 좁혀가면서 분포 윤곽 파악.

```
par(mfrow=c(2,2))
plot(x=income.breaks, y=c(0,income.kor.cum[,1]), type="b", ann=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of People")
plot(x=income.breaks, y=c(0,income.kor.cum[,1]), type="b", xlim=c(0,500), an
n=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of People")
plot(x=income.breaks, y=c(0,income.kor.cum[,1]), type="b", xlim=c(0,200), an
n=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of People")
plot(x=income.breaks, y=c(0,income.kor.cum[,1]), type="b", xlim=c(0,100), an
n=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of People")
```





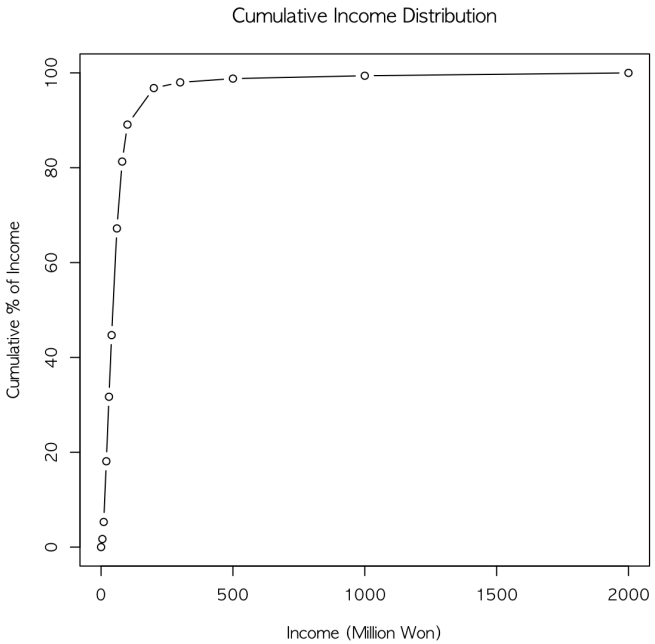


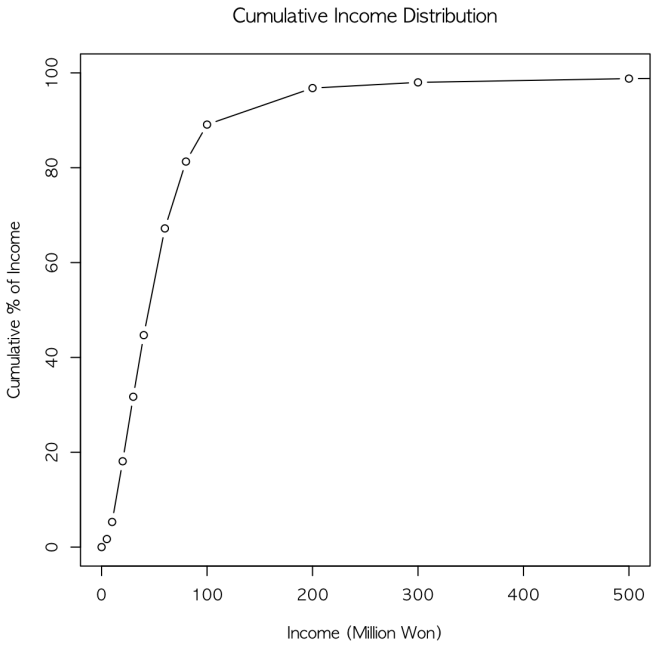


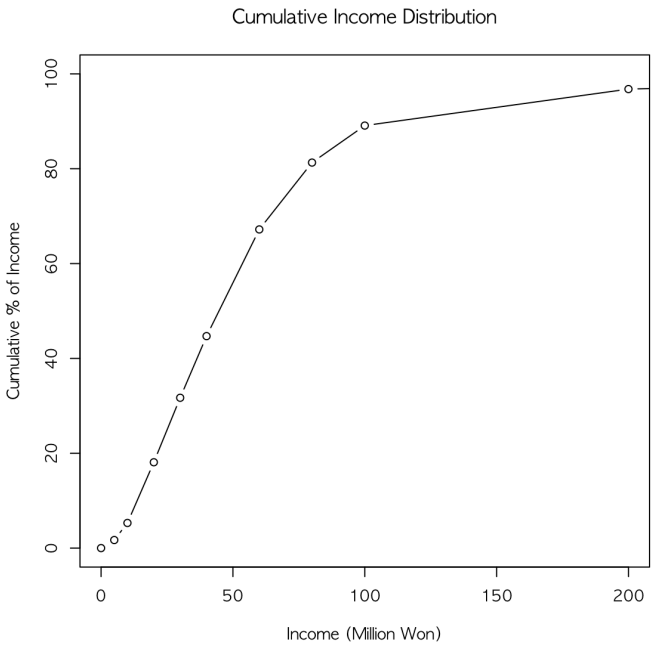

```
par(mfrow=c(1,1))
```

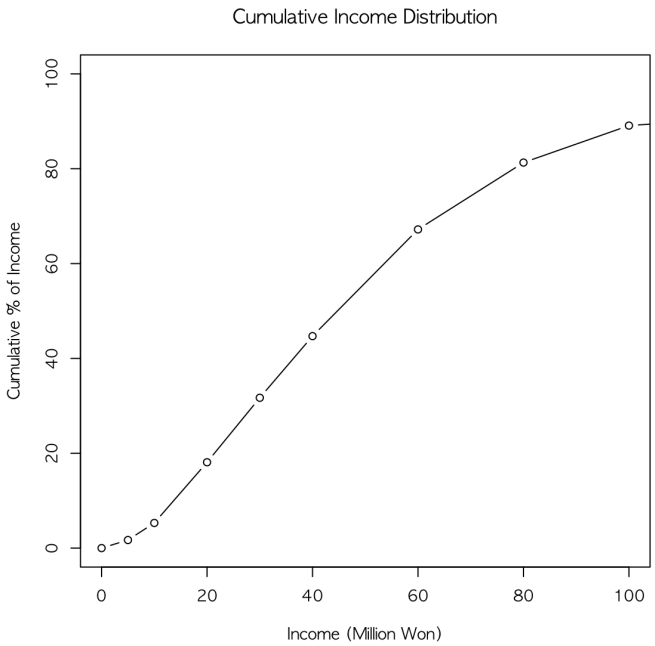
소득 자체의 누적분포에 대해서도 같은 방법으로 그려보면

```
par(mfrow=c(2,2))
plot(x=income.breaks, y=c(0,income.kor.cum[,2]), type="b", ann=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of Income")
plot(x=income.breaks, y=c(0,income.kor.cum[,2]), type="b", xlim=c(0,500), an
n=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of Income")
plot(x=income.breaks, y=c(0,income.kor.cum[,2]), type="b", xlim=c(0,200), an
n=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of Income")
plot(x=income.breaks, y=c(0,income.kor.cum[,2]), type="b", xlim=c(0,100), an
n=F)
title(main="Cumulative Income Distribution", xlab="Income (Million Won)", yla
b="Cumulative % of Income")
```





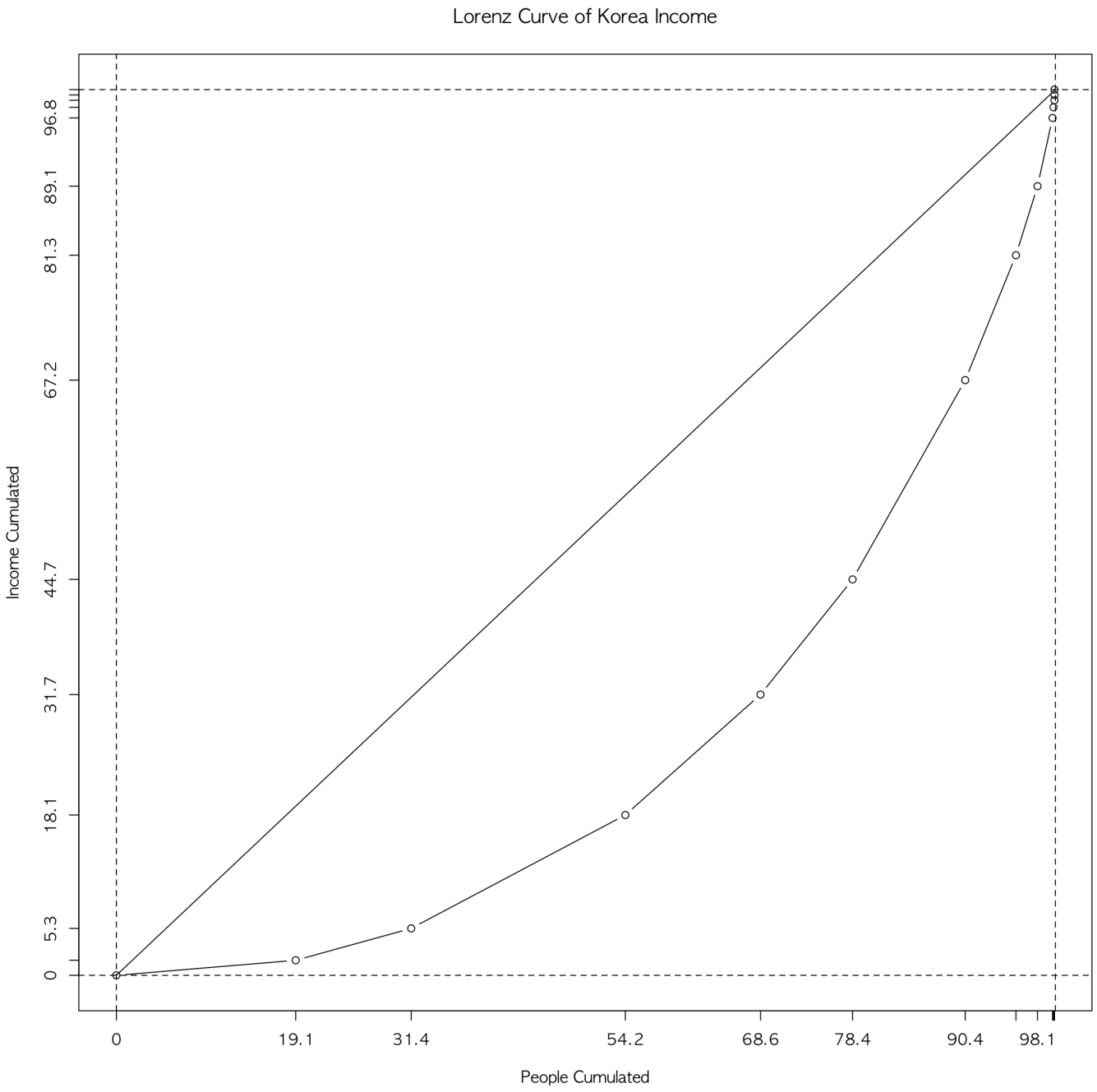




```
par(mfrow=c(1,1))
```

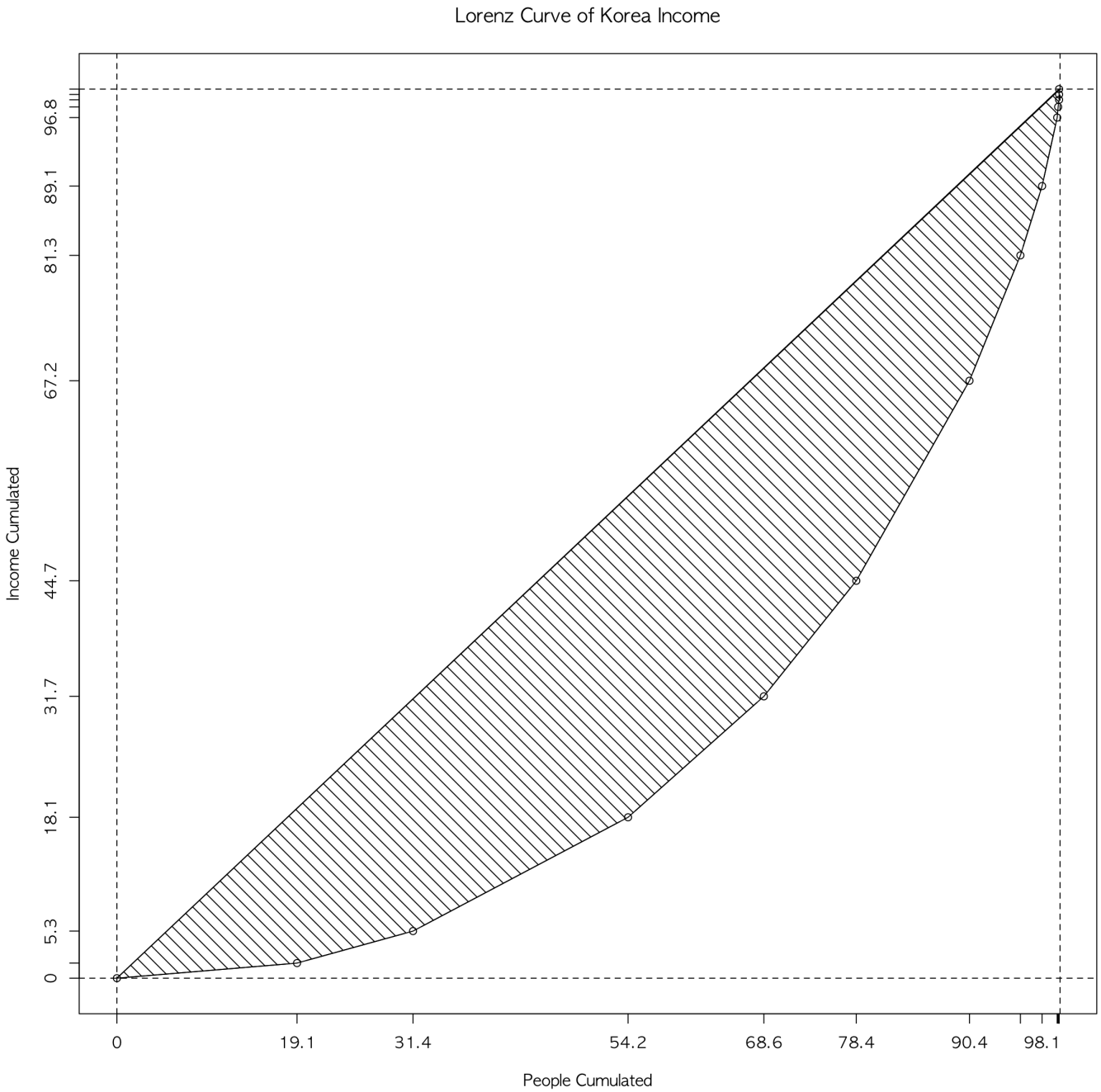
이제 두 누적분포를 한 장에 살피는 방법을 생각해보자. x 축을 사람, y 축을 소득으로 하여 두 점을 이어주면 어떤 결과가 나오는 지 살펴 보자.

```
people<-c(0, income.kor.cum[,1])
income<-c(0, income.kor.cum[,2])
plot(x=people, y=income, type="b", ann=F, xaxt="n", yaxt="n")
lines(x=0:100, y=0:100, type="l")
axis(side=1, at=people, labels=people)
axis(side=2, at=income, labels=income)
abline(h=c(0,100), lty=2)
abline(v=c(0,100), lty=2)
title(main="Lorenz Curve of Korea Income", xlab="People Cumulated", ylab="Incom
e Cumulated")
```



초승달 부분에 빗금을 치고 싶다면,

```
plot(x=people, y=income, type="b", ann=F, xaxt="n", yaxt="n")
lines(x=0:100, y=0:100, type="l")
axis(side=1, at=people, labels=people)
axis(side=2, at=income, labels=income)
abline(h=c(0,100), lty=2)
abline(v=c(0,100), lty=2)
title(main="Lorenz Curve of Korea Income", xlab="People Cumulated", ylab="Incom
e Cumulated")
polygon(x=c(people, 0), y=c(income,0), density=10, angle=135)
```



이 곡선의 이름은?