

I. Excel 도수분포표 및 기술통계량

II. R 도수분포표 및 기술통계량

1. 도수분포표

- 도수분포표를 만들기 위해서는 두 가지가 필요한데 하나는 도수분포를 수행하게 될 영역의 값(입력범위)이고 다른 하나는 위의 영역 값에 대한 범위가 나열된 계급구간임
- 먼저 계급구간을 오름차순으로 설정한 후 데이터-분석-데이터 분석을 실행하면 통계 데이터분석 대화상자가 나타나고, 히스토그램을 선택하면 히스토그램대화상자가 나타나는데 입력범위와 계급구간 및 몇 가지(차트출력)를 선택한 후 확인을 누르면 도수분포표와 히스토그램을 만들어 줌
- 도수 결과는 자동적으로 갱신이 되지 않으므로 값영역이나 구간값 영역의 데이터를 변경하는 경우 자료/도수분포를 재실행해야 함
- (예) 우리나라 지자체의 지방재정자립도를 나타내는 finance-k.xlsx로 제주도 지방재정자립도의 도수분포표 및 히스토그램을 작성해 보라
- <http://kanggc iptime.org/book/data/finance-k.xlsx>로 파일을 다운로드
- R열은 제주자료를 나타내 주고 있는데 T8셀부터 T12셀에 계급 구간 26, 30,34,38,42를 입력
- 데이터-데이터분석-히스토그램을 선택하고 확인을 누르면 아래 왼쪽 그림과 같은 히스토그램 대화상자가 나타나는데 입력 범위는 제주의 지방재정자립도를 나타내는 데이터이고, 계급 구간은 앞에서 설정하여 입력한 것으로 확인을 누르면 아래 오른쪽 그림과 같은 도수분포표 및 히스토그램을 작성

R	S	T	U	V	W	X	Y	Z	AA	AB
제주										
31.2										
33.6										
36.6										
37.4										
34.7										
39.3										
33.8		26.0								
26.4		30.0								
26.3		34.0								
25.2		38.0								
26.1		42.0								
25.1										
28.5										
30.6										
34.0										
36.4										
38.2										
39.6										

히스토그램

입력 범위(\$R\$2:\$R\$19)

계급 구간(\$T\$8:\$T\$12)

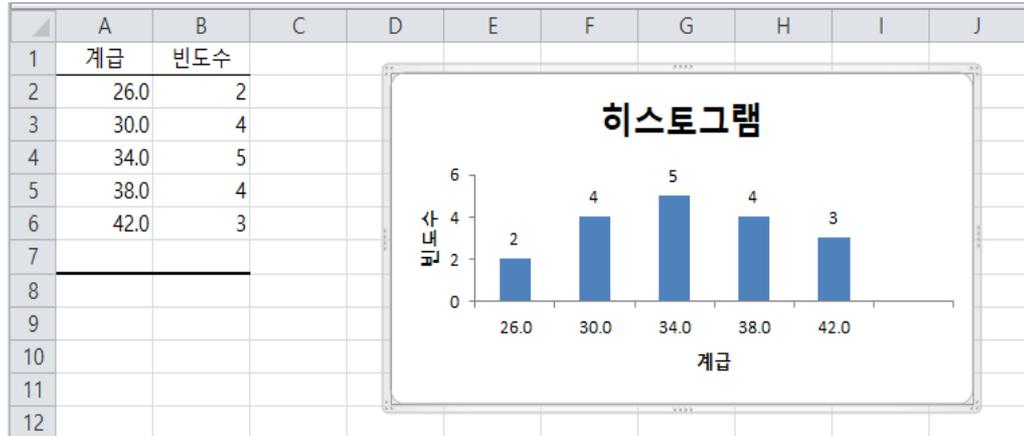
출력 옵션

새로운 워크시트

파레토: 순차적 히스토그램

누적 백분율

차트 출력



2. 기술통계량

- 기술통계량이란 연속형 데이터의 집중화 경향, 분산도, 분포 등에 대한 특성을 파악하기 위하여 의미 있는 수치로 요약된 것
 - 집중화 경향 : 평균, 중위수, 최빈값 등
 - 흩어짐의 정도 : 최댓값, 최솟값, 범위, 분산, 표준편차, 변동계수 등
 - 분포의 형태 및 대칭성의 정도 : 첨도, 왜도 등
- (예) 우리나라 주요 거시경제변수를 나타내는 describe.xlsx로 기술통계량 계산해 보라
 - <http://kanggc.ipetime.org/book/data/describe.xlsx>로 파일을 다운로드
 - 데이터-분석-데이터 분석을 실행하면 나타나는 통계 데이터분석대화상자에서 기술통계법을 선택하면 나타나는 아래 왼쪽 그림과 같은 기술통계법 대화상자에서 다음을 선택
 - 데이터 입력 범위를 선정하는데 이때 계열이름이 있는 셀과 자료가 입력되어 있는 셀인 B1셀부터 E16까지 모두 선택
 - 첫째 행 이름표 사용 및 요약통계량에 체크한 후 확인을 클릭하면 아래 오른쪽 그림과 같이 선택된 자료에 관한 각종 요약통계량을 계산

기술 통계법 ? X

입력
 입력 범위(I):

데이터 방향: 열(C) 행(R)

첫째 행 이름표 사용(L)

출력 옵션
 출력 범위(O):
 새로운 워크시트(P):
 새로운 통합 문서(W)

요약 통계량(S)
 평균에 대한 신뢰 수준(N): %
 K번째 큰 값(A):
 K번째 작은 값(M):

	A	B	C	D	E	F	G	H
1	소비자물가상승률		주택담보대출금리		GDP성장률		GDP디플레이터상승률	
2								
3	평균	2.553333333	평균	5.103333333	평균	3.84	평균	2.053333
4	표준 오차	0.294693815	표준 오차	0.332921651	표준 오차	0.45269563	표준 오차	0.300138
5	중앙값	2.8	중앙값	5.39	중앙값	3.3	중앙값	2.4
6	최빈값	2.8	최빈값	#N/A	최빈값	2.8	최빈값	3
7	표준 편차	1.141344237	표준 편차	1.289400008	표준 편차	1.753282636	표준 편차	1.162428
8	분산	1.302666667	분산	1.662552381	분산	3.074	분산	1.351238
9	첨도	-0.516421948	첨도	-0.861333643	첨도	0.07950617	첨도	-1.213
10	왜도	0.047769618	왜도	-0.408992108	왜도	0.495655071	왜도	-0.37418
11	범위	4	범위	4.09	범위	6.7	범위	3.6
12	최소값	0.7	최소값	2.91	최소값	0.7	최소값	-0.1
13	최대값	4.7	최대값	7	최대값	7.4	최대값	3.5
14	합	38.3	합	76.55	합	57.6	합	30.8
15	관측수	15	관측수	15	관측수	15	관측수	15

1. 도수분포표

```

b3-ch4-4.R

library(openxlsx)

df<-read.xlsx("http://kanggc.ipetime.org/book/data/finance-k.xlsx"
)
df_dat<-data.matrix(df)

year<-df_dat[,1]
jj<-df_dat[,18]
jj

bins<-c(22,26,30,34,38,42)
bins

class<-cut(jj,breaks=bins)
class

table(class)

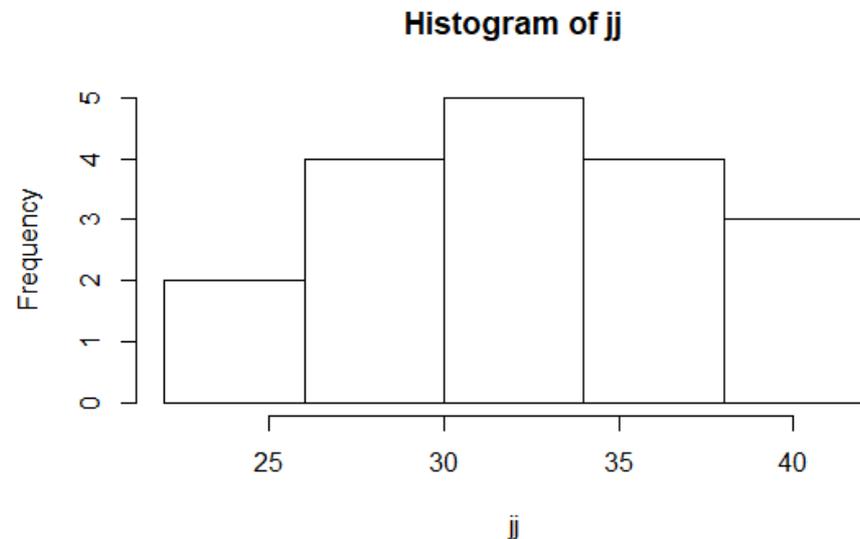
transform(table(class),Rel_Freq=prop.table(Freq))

hist(jj, breaks=bins, xlim=c(22,42))
    
```

```

b3-ch4-4.R(명령어 설명)

# 5개의 계급으로 계급 구간을 나눔
# 데이터가 어느 계급 구간에 속하는지 식별하고 코드화
# 각 계급 구간에 속하는 데이터 수를 계산
# 계급 구간 및 도수로 도수분포표를 생성
    
```



2. 기술통계량

b3-ch4-5.R

```
library(openxlsx)
library(fBasics)
df<-read.xlsx("http://kanggc.iptime.org/book/data/describe-e.xls
x")
df_dat<-data.matrix(df)
cpi<-df_dat[,2]
interest<-df_dat[,3]
gdp<-df_dat[,4]
deflator<-df_dat[,5]
df_new<-cbind(cpi, interest, gdp, deflator)
df_new
summary(df_new)
(var<-var(df_new))
(sd1<-sd(cpi))
(sd2<-sd(interest))
(sd3<-sd(gdp))
(sd4<-sd(deflator))
basicStats(df_new)
```

```
> (sd1<-sd(cpi))
[1] 1.141344

> (sd2<-sd(interest))
[1] 1.2894

> (sd3<-sd(gdp))
[1] 1.753283

> (sd4<-sd(deflator))
[1] 1.162428
```

b3-ch4-5.R(명령어 설명)

```
# R 객체를 열 또는 행으로 결합
# 분산-공분산을 계산
# fBasics 패키지를 이용하여 기술통계량을 계산

> summary(df_new)
      cpi      interest      gdp      deflator
Min.   :0.700   Min.   :2.910   Min.   :0.70   Min.   : -0.100
1st Qu.:1.750   1st Qu.:4.245   1st Qu.:2.80   1st Qu.: 1.000
Median :2.800   Median :5.390   Median :3.30   Median : 2.400
Mean   :2.553   Mean   :5.103   Mean   :3.84   Mean   : 2.053
3rd Qu.:3.200   3rd Qu.:6.035   3rd Qu.:5.05   3rd Qu.: 3.050
Max.   :4.700   Max.   :7.000   Max.   :7.40   Max.   : 3.500
```

```
> (var<-var(df_new))
      cpi      interest      gdp      deflator
cpi      1.3026667  1.2029524  0.2998571  0.6090952
interest 1.2029524  1.6625524  0.8371429  0.6597381
gdp      0.2998571  0.8371429  3.0740000  0.1448571
deflator 0.6090952  0.6597381  0.1448571  1.3512381
```

```
> basicStats(df_new)
      cpi      interest      gdp      deflator
nobs      15.000000  15.000000  15.000000  15.000000
NAS        0.000000   0.000000   0.000000   0.000000
Minimum    0.700000   2.910000   0.700000  -0.100000
Maximum    4.700000   7.000000   7.400000   3.500000
1. Quartile 1.750000   4.245000   2.800000   1.000000
3. Quartile 3.200000   6.035000   5.050000   3.050000
Mean       2.553333   5.103333   3.840000   2.053333
Median     2.800000   5.390000   3.300000   2.400000
Sum        38.300000  76.550000  57.600000  30.800000
SE Mean    0.294694   0.332922   0.452696   0.300138
LCL Mean   1.921278   4.389287   2.869064   1.409602
UCL Mean   3.185389   5.817379   4.810936   2.697064
Variance   1.302667   1.662552   3.074000   1.351238
Stdev     1.141344   1.289400   1.753283   1.162428
Skewness   0.038640  -0.330829   0.400930  -0.302666
Kurtosis  -1.026629  -1.235876  -0.665100  -1.449217
```