

2. 범주형 자료의 요약

범주형 자료 요약

다수의 범주가
반복해서 관측

관측값의 크기보다
포함되는 범주에 관심

범주형 자료
요약 필요

각 범주에 속하는
관측값의 개수를 측정

전체에서 차지하는
각 범주의 비율 파악

효율적으로 범주 간의
차이점 비교 가능

도수분포표

도수
(Frequency)

각 범주에 속하는 관측값의 **개수**

```
value_counts()
```

상대도수
(Relative Frequency)

도수를 자료의 **전체 개수**로 나눈 **비율**

```
value_counts(normalize=True)
```

도수분포표
(Frequency Table)

범주형 자료에서 **범주와 그 범주에 대응**하는 도수, 상대도수를 나열해 표로 만든 것

도수분포표를 구하는 파이썬 코드

```
# 한 가지 범주의 도수분포표  
pd.crosstab(index = 범주 ,  
             columns = "count")  
  
# index로 설정한 범주에 해당하는  
도수를 계산하여 도수분포표를 제작
```

```
# 두 가지 범주의 도수분포표  
pd.crosstab(index = 범주 ,  
            columns = 또 다른 범주)  
  
# index로 설정한 범주와 cloums로  
설정한 범주를 모두 만족하는 도수를  
계산하여 도수분포표를 제작
```

몇 개의 범주를 기준으로 둘 것인지에 따라
다양한 도수분포표를 만들 수 있습니다.

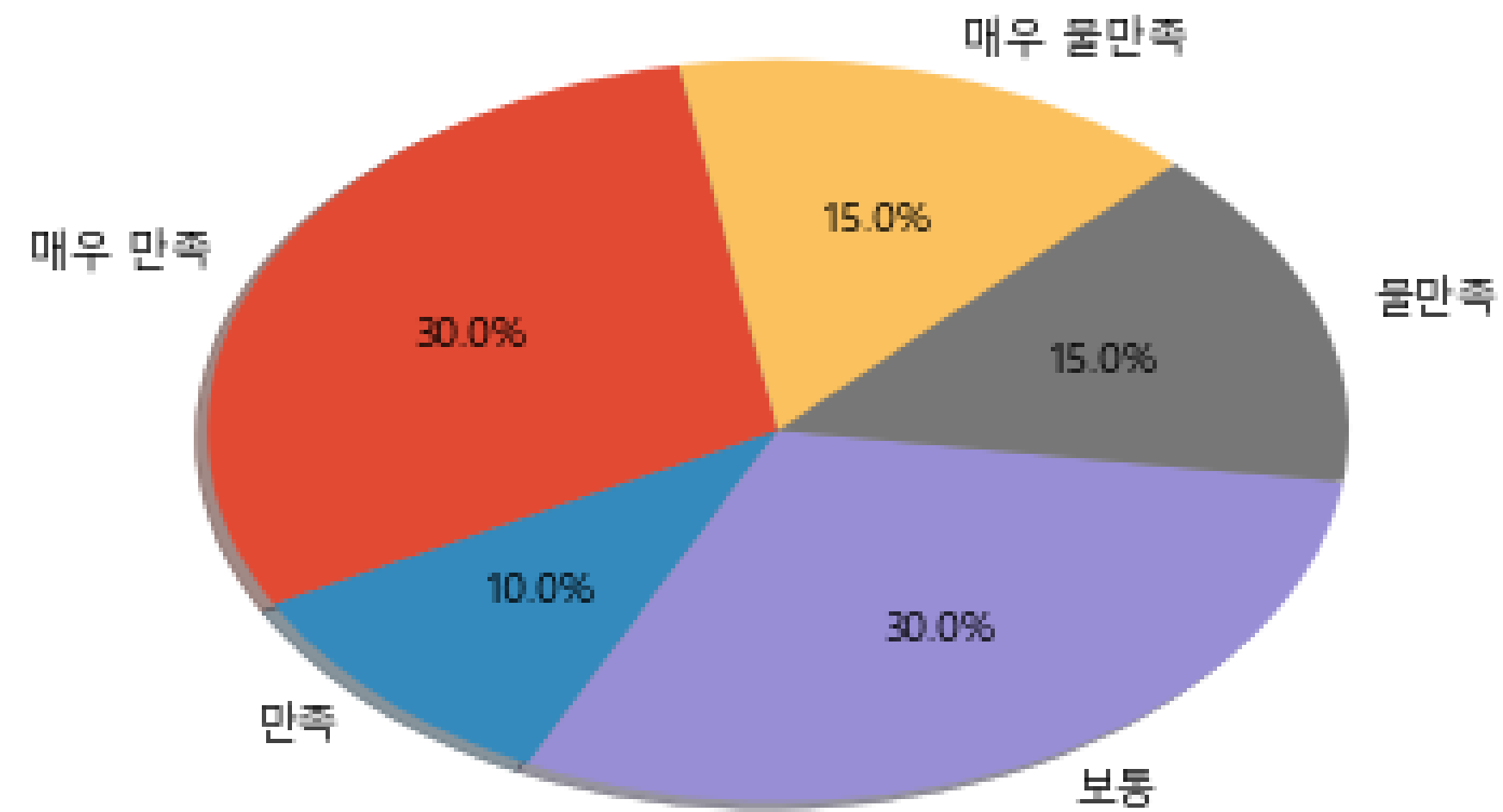
도수분포표 예시

강의 만족도 설문 (100명 조사)

범주	도수	상대도수	누적 상대도수
매우 만족	30	0.3	0.3
만족	10	0.1	0.4
보통	30	0.3	0.7
불만족	15	0.15	0.85
매우 불만족	15	0.15	1.00

원그래프(Pie Chart)

```
plt.pie(수치, labels = 라벨)
```



원형그래프(Pie Chart)

숫자의 나열보다 전체적인 분포를 이해하기 쉬운 그래프

원을 각 범주가 차지하는 비율로 중심 각을 나누어
피자처럼 조각을 나눈 형태의 그림

장점

전체에서 범주가 차지하는
비율을 파악하기 쉬움

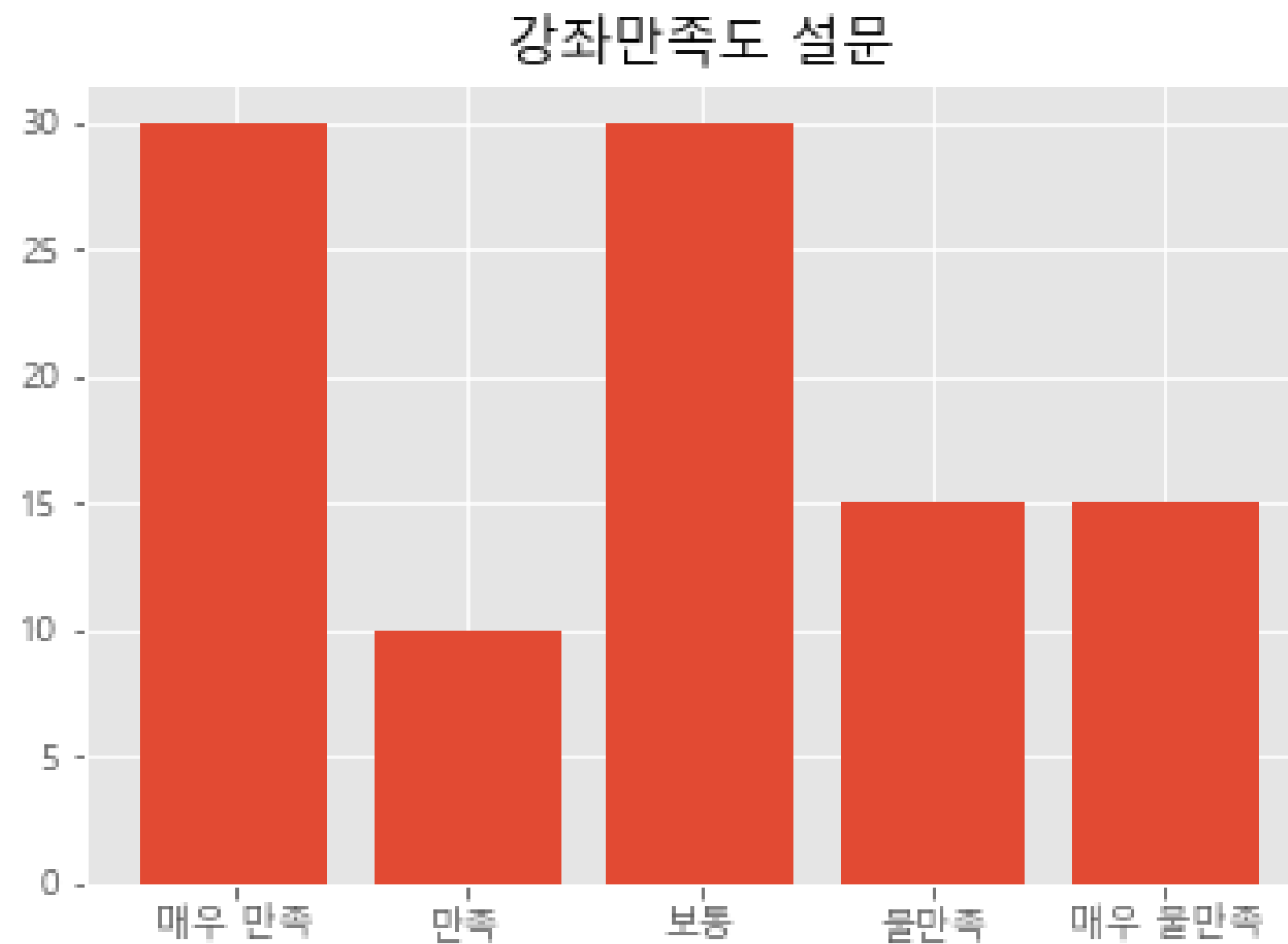
단점

범주 간 도수 비교 및 도수
크기 차이 파악이 어려움

막대그래프 (Bar Chart)

```
plt.bar(수치, labels = 라벨)
```

y축 =
도수의 크기



막대그래프 (Bar Chart)

각 범주에서 도수의 크기를 막대로 그림

그래프의 y 축 : 도수에 대한 눈금

그래프의 x 축 : 범주를 나열

장점

각 범주가 가지는 도수의
크기 차이를 비교하기 쉬움

단점

각 범주가 차지하는 비율의
비교는 어려움