

1. 다음은 Carseats dataset에 대한 summary() 결과이다.

아래 결과물에 대한 해석으로 가장 적절하지 않은 것은?

summary(Carseats)					
ComPrice		Income		Advertising	
Min.	: 77	Min.	: 21.00	Min.	: 0.000
1st Qu.	:115	1st Qu.	: 42.75	1st Qu.	: 0.000
Median	:125	Median	: 69.00	Median	: 5.000
Mean	:125	Mean	: 68.66	Mean	: 6.635
3rd Qu.	:135	3rd Qu.	: 91.00	3rd Qu.	:12.000
Max.	:175	Max.	:120.00	Max.	:29.000
Price		ShelveLoc		US	
Min.	: 24.0	Bad	: 96	No	:142
1st Qu.	:100.0	Good	: 85	Yes	:258
Median	:117.0	Medium	:219		
Mean	:115.8				
3rd Qu.	:131.0				
Max.	:191.0				

- ① US 변수는 구간척도에 해당된다.
- ② ShelveLoc 변수는 서열척도에 해당된다.
- ③ Price 변수에는 결측값이 없다.
- ④ Income에 평균은 68.66이다.

2. summary(Wage)에 대한 설명 중 부적절한 것은?

summary(Wage)									
year		age		health_ins		logwage		wage	
Min.	:2003	Min.	:18.00	1. Yes	:2083	Min.	:3.000	Min.	: 20.09
1st Qu.	:2004	1st Qu.	:33.75	2. No	: 917	1st Qu.	:4.447	1st Qu.	: 85.38
Median	:2006	Median	:42.00			Median	:4.653	Median	:104.92
Mean	:2006	Mean	:42.41			Mean	:4.654	Mean	:111.70
3rd Qu.	:2008	3rd Qu.	:51.00			3rd Qu.	:4.857	3rd Qu.	:128.68
Max.	:2009	Max.	:80.00			Max.	:5.763	Max.	:318.34

- ① age 변수의 중위값은 42.00이다.
- ② logwage 변수를 벡터를 추출하기 위해서는 Wage\$logwage를 실행한다.
- ③ wage 변수는 범주형 데이터 타입이다.
- ④ year 변수의 최댓값은 2009이다.

3. 다음 중 상자 그림에 대한 설명 중 올바른 것은?

```
> df<-boxplot(airquality$Ozone)
> df$stats

      [,1]
[1,]  1.0
[2,] 18.0   #1사분위
[3,] 31.5
[4,] 63.5   #3사분위
[5,] 122.0
```

- ① $Q1-1.5IQR$ 또는 $Q3+1.5IQR$ 을 벗어난 데이터를 이상치로 판단한다.
- ② 평균은 31.5이다.
- ③ 데이터의 IQR는 45이다.
- ④ Q3보다 큰 데이터는 전체 데이터의 50% 이상이다.

답) 1, 3, 1