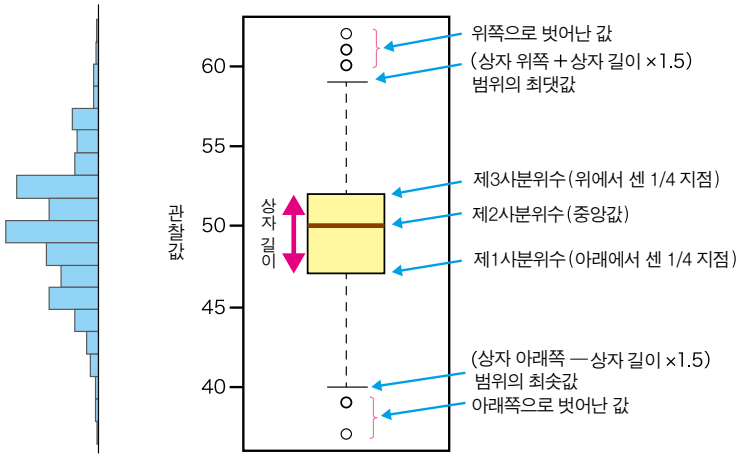


p. 66 그림 3.3.7 제1사분위수 <-> 제3사분위수 자리 바뀔  
위에서 센 1/4 지점 = 제3사분위수 | 아래에서 센 1/4 지점 = 제1사분위수

◆ 그림 3.3.7 상자수염 그림의 정의

히스토그램 → 상자 수염 그림



상자 수염 그림과 비교하고자, 히스토그램은 90도 회전시켜 표시했습니다. 상자 수염 그림은 중앙값이나 사분위수라는 통계량을 나타냄으로써 데이터가 어떤 분포인지를 눈으로 볼 수 있도록 합니다.

부르기도 합니다. 제1사분위수부터 제3사분위수까지의 범위가 바로 상자입니다.

수염은 상자 길이(제1사분위수와 제3사분위수의 차이)의 1.5배 길이를 상자로부터 늘인 범위 안에서, 최댓값 또는 최솟값을 가리킵니다. 이 범위에 포함되지 않은 값은 이상값(그림 3.3.7의 동그라미 부분)으로 정의됩니다.

상자 수염 그림은 중앙값이나 사분위수, 최댓값, 최솟값 등의 통계량은 나타내는 반면, 히스토그램에서 볼 수 있는 상세한 분포 형태 정보는 포함하지 않습니다.

• 분포를 시각화하는 다양한 방법

데이터 분포를 눈으로 보는 방법에는 그 밖에도 몇 가지가 있습니다. 그림

p. 239 그림 8.3.7 감기약을 복용하지 않았을 때 3일 내에 나올 확률  $p \Rightarrow q$

◆ 그림 8.3.7 오즈비

**오즈 (odds)**

확률  $p$ 로 일어나는 사건에서의  $\frac{p}{1-p}$  값

예  $p = 0.5$ 면 오즈 1  
 $p < 0.5$ 면 오즈  $< 1$   
 $p > 0.5$ 면 오즈  $> 1$

**오즈비 (odds ratio, OR)**

확률  $p, q$ 에 대해

$$OR = \frac{\left(\frac{p}{1-p}\right)}{\left(\frac{q}{1-q}\right)}$$

$p$ 와  $q$ 가 어느 정도 다른지 정량화

예  $p = q$ 라면 OR = 1  
 $p > q$ 라면 OR  $> 1$   
 $p < q$ 라면 OR  $< 1$

**예시**

감기약을 복용했을 때 3일 내에 나올 확률  $p = 0.8$   
 감기약을 복용하지 않았을 때 3일 내에 나올 확률  $q = 0.4$

$$OR = \frac{\frac{0.8}{0.2}}{\frac{0.4}{0.6}} = 6$$

의료 통계에서 특히 자주 쓰임(95% 신뢰구간 CI도 함께 기술할 때가 많음)

그런데 사례-대조 연구에서 위험비는 폐암 위험의 적절한 비교 지표가 아닙니다. 왜냐하면, 조사 대상인 폐암에 걸린 사람 수와 걸리지 않은 사람 수의 비율이 달라지면 결과도 달라지기 때문입니다. 반면 오즈비는 조사 대상 수의 비율에 관계없으므로, 적절한 비교 지표입니다.\*



**푸아송 회귀**

데이터가 음수가 되지 않는 정수일 때, 특히 반응변수가 개수인 경우 고려해볼 수 있는 일반화선형모형이 바로 **푸아송 회귀(Poisson regression)**입니다. 예를 들어 반응변수가 한 그루의 나무에서 피는 꽃의 수라면, 음수가 될 수 없으며 항상 정수로 나타나기 때문에 푸아송 회귀를 적용할 수 있습니다.

\*  $M$ (폐암인 사람 수) $=a$ (흡연) $+b$ (비흡연),  $N$ (폐암이 아닌 사람 수) $=c$ (흡연) $+d$ (비흡연)로 위험비와 오즈비를 계산해 보면 알 수 있습니다.