

Assignment 1

2023-12675 박지호

2.1

2.1-(1)

- 히스토그램에서는 2.0~2.2시간과 2.4~2.6시간에 우승자 기록이 다수 분포되어 있다는 사실을 확인할 수 있으나, 상자 그림에서는 이를 확인할 수 없다.
- 상자 그림에서는 우승자 기록의 중앙값과 산포를 쉽게 확인할 수 있으며, 시간이 긴 쪽에서 이상치가 나타난다는 사실을 확인할 수 있으나, 히스토그램에서는 이를 확인하기 어렵다.

2.1-(2)

- 해당 데이터는 남성 우승자와 여성 우승자의 기록을 전부 포함하기 때문에, 각 봉우리에 남성과 여성 우승자의 기록이 주로 분포되어 있을 것으로 추정된다

2.1-(3)

- 기록의 중앙값의 경우 남성의 우승 기록이 여성의 우승 기록보다 그 시간이 짧다.
- 남성의 우승 기록이 여성의 우승 기록보다 더 조밀하게 나타난다.
- 남성의 우승 기록과 여성의 우승 기록 모두 시간이 짧은 쪽에 치우친 형태로 나타나며, 시간이 긴 쪽에서 이상치가 여럿 나타난다.

2.2

2.2-(1)

- 모자이크 플롯을 확인하였을 때, 심장이식 수술을 받은 환자들이 심장이식 수술을 받지 못한 환자보다 생존 비율이 유의미한 정도로 높은 것을 확인할 수 있다. 따라서 심장이식이 환자의 생존과 연관이 있다고 추정할 수 있다.

2.2-(2)

- 오른쪽 상자 그림을 확인하였을 때, 대조 그룹의 경우 생존 기간이 0 부근에 조밀하게 배치되어 있는 것을 확인할 수 있다. 반면 심장이식 수술을 받은 처리 그룹의 경우 생존 기간이 대체적으로 대조 그룹보다 긴 쪽에 분포한다는 것을 확인할 수 있다. 따라서 심장이식 수술이 환자의 생존 기간 연장에 유의미한 효과가 있다고 추정할 수 있다.

2.2-(3)

- 처리 그룹: $\frac{45}{69}$
- 대조 그룹: $\frac{30}{34}$

3.1

3.1-(1)

- $$\frac{0.99 \times 0.006}{0.99 \times 0.006 + 0.02 \times 0.994} \approx 0.23$$

- 검사 결과 양성 판정을 받은 사람이 실제 감염자일 확률은 23%이다

3.1-(2)

- $$\frac{0.99 \times 0.99 \times 0.006}{0.99 \times 0.99 \times 0.006 + 0.02 \times 0.02 \times 0.994} \approx 0.94$$

- 2회 검사 결과 모두 양성 판정을 받은 사람이 실제 감염자일 확률은 94%로, (1)의 결과와 비교하였을 때 훨씬 높은 확률이다

3.2

3.2-(1)

	X = 0	X = 1	X = 2	Total
Y = 0	0.22222	0.22222	0.02222	0.46667
Y = 1	0.33333	0.13333	0	0.46667
Y = 2	0.06667	0	0	0.06667
Total	0.62222	0.35556	0.02222	1

3.2-(2)

- 아니오, $P(X = x \wedge Y = y) = P(X = x)P(Y = y)$ 가 모든 x 와 y 에 대해 성립하지는 않기 때문이다. 기어 변속기에만 문제가 생긴 차량, 엔진에만 문제가 생긴 차량, 모두 문제가 없는 차량 중 두 대를 뽑는 것이기 때문에, $X = 2, Y = 2$ 와 같은 경우는 불가능해 이러한 결과가 나타난다.

3.2-(3)

- $$P(X + Y = k) = \begin{cases} \frac{2}{9} & \text{if } (k = 0) \\ \frac{5}{9} & \text{if } (k = 1) \\ \frac{2}{9} & \text{if } (k = 2) \end{cases}$$

- 평균: 1

- 분산: $\frac{4}{9}$