
통계학실험 Exercise

컴퓨터공학부 16 김종범

March 11, 2016



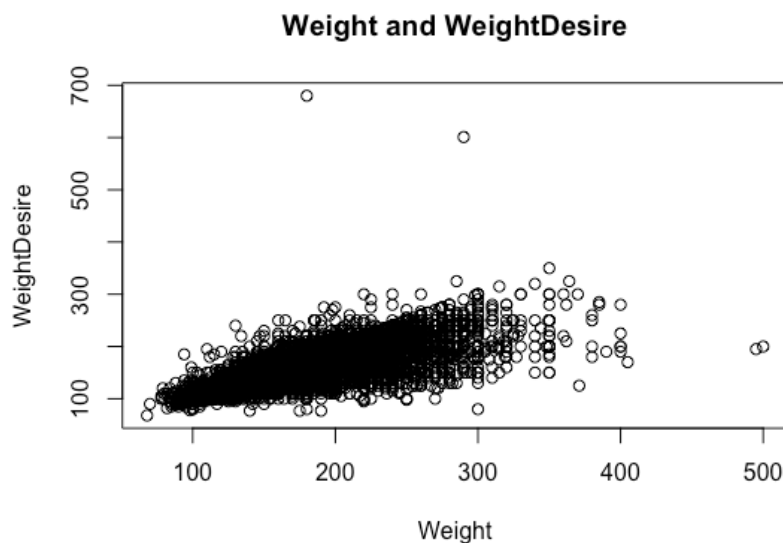
예제 1. genhlth 변수에 대해 적절한 방법을 이용하여 요약해보자. 범주형 자료의 경우에는 어떠한 요약 방법을 사용할 수 있는가?

genhlth 변수는 범주형 자료이다. 범주형 자료의 요약은 분할표를 이용할 수 있다.

excellent	fair	good	poor	very good
4657	2019	5676	677	6972

예제 2. weight 변수에 대한 수치적 요약 값을 구해보자. 전체 응답자의 평균 몸무게는 얼마인가? 평균 몸무게는 169.683(pound) 이다.

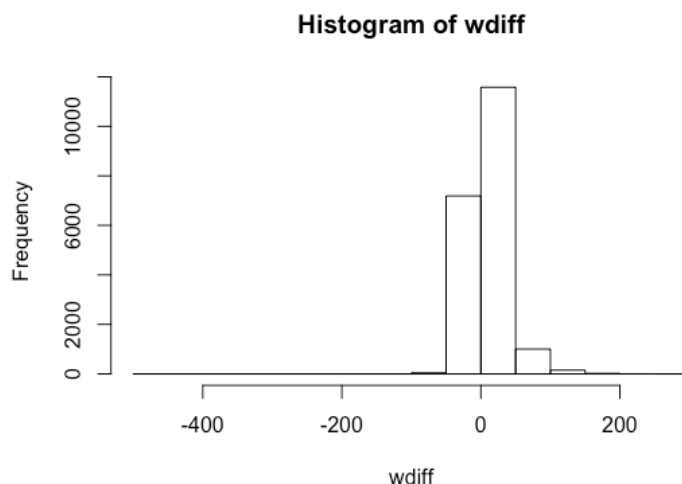
예제 3. weight 변수와 wtdesired 변수의 산점도를 그려보자. 두 변수 사이에는 어떠한 관계가 존재한다고 보여지는가? 두 변수의 상관계수는 무엇을 나타내고 있는가?



두 변수의 산점도를 그려보면, 두 변수는 선형적인 관계가 존재한다고 볼 수 있다.

두 변수의 상관계수는 약 0.8이며, 이는 강한 양적 선형관계로 해석할 수 있다.

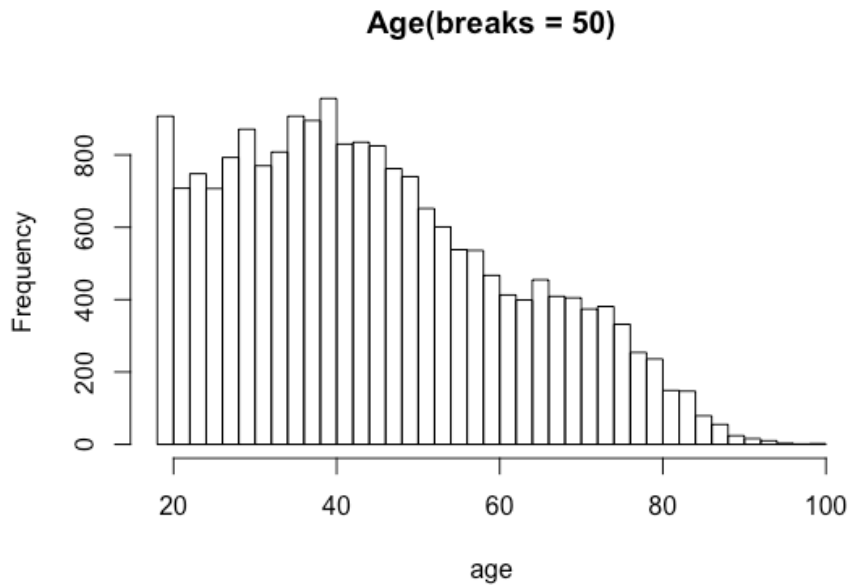
예제 4. wtdesired 변수와 weight 변수의 차를 계산하여 새로운 변수 wdiff 를 만들어보자. wdiff의 분포는 어떠한가? 수치적 요약과 그래프 요약을 통해 살펴보자. 이것이 의미하는 바는 무엇인가?



수치적 요약으로는 평균이 14.59, 표준편차가 약 24이다. 최소값은 -500, 최댓값은 300이며, 중간값은 10이다.

그래프 요약으로는 왼쪽 그림과 같다. 주로 -50부터 50 사이에 대부분의 사람들이 몰려있는 형태이다.

예제 5. age 변수를 이용하여 히스토그램을 그려보자. 그리고 구간의 수를 50, 100으로 바꿔가며 동일한 히스토그램을 그린 후 비교해보자.



위 그림은 구간의 수를 50으로 하였을 때 히스토그램이고, 아래 그림은 구간의 수를 100으로 하였을 때 히스토그램이다.

전체적으로 비슷한 모양을 띄나 구간의 수가 100인 경우가 더 조밀한 모양을 보여준다.