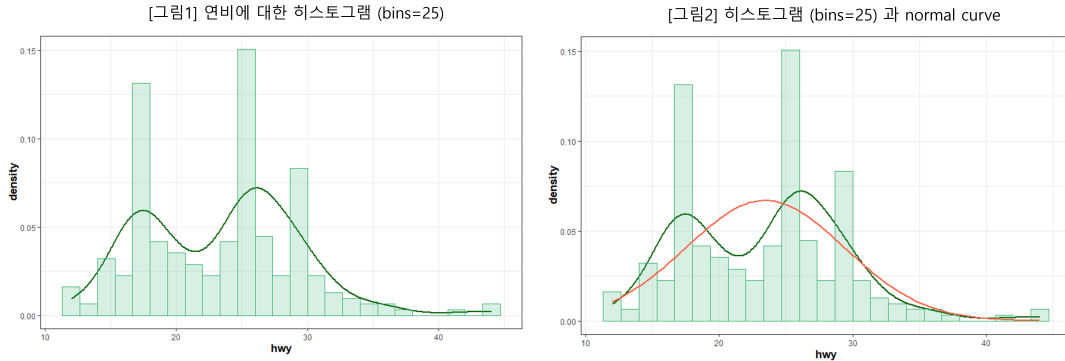


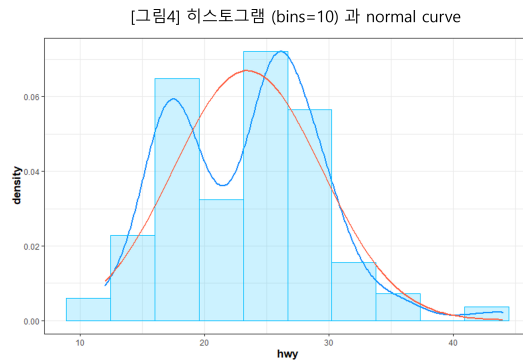
GGPLOT 활용하기

연비는 정규 분포를 따른다고 할 수 없다

[1] 고속도로 연비(hwy)에 대한 히스토그램



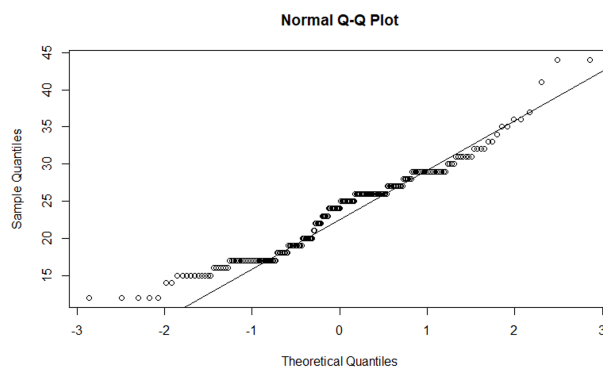
[2] 고속도로 연비(hwy)에 대한 히스토그램



[3] 고속도로 연비는 정규분포와 유사하다고 할 수 있을까?

[그림2] bins=25인 경우와 [그림4] bins=10인 경우를 비교해서 보면 [그림4]로 인해 마치 연비가 정규분포와 유사하다고 생각할 수 있겠지만 연비는 정규분포라고 할 수 없다. 단지 bins를 작은 값으로 설정했다고 정규분포로 보는 것은 타당하지 않다. Q-Q Plot은 그래프 상의 점들이 직선 모양을 보이면 해당 데이터가 정규 분포를 따른다고 할 수 있다. 실제로 정규성을 검정할 수 있는 'Q-Q Plot'을 그려보면 연비는 정규분포가 아님을 알 수 있다. [그림5]는 연비에 대해 Q-Q Plot을 그린 그래프이다. 그래프 상의 점이 직선을 유지한다고 할 수 없으므로 연비는 정규분포를 따른다고 할 수 없다.

[그림5] 연비에 대한 Q-Q plot



"Written by Lee, Da-In"