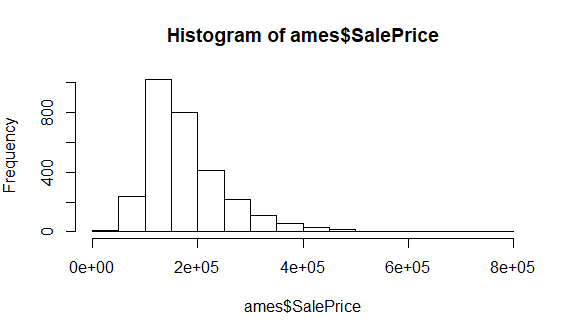
통계학 실험 과제 #2

2017-17497 김현규

1.

|  |  |  |
| --- | --- | --- |
| 1  2  3 | ames=read.csv("E:\\Rstudio\_dataset\\dataset\\ch04\\ames.csv", header=T)  hist(ames$SalePrice)  summary(ames$SalePrice)  [*Colored by Color Scripter*](http://colorscripter.com/info#e) | [cs](http://colorscripter.com/info#e) |



첫번째 줄은 read.csv함수를 통해 주어진 자료를 불러온다. 두번째 줄에서는 ames데이터의 saleprice 변수에 대해 히스토그램을 그려주고 세 번재 줄에서는 이 변수의 수치적 요약값을 구해준다. 히스토그램과 summary에서 알 수 있듯이 right-skewed 되어 있음을 볼 수 있다.

2.

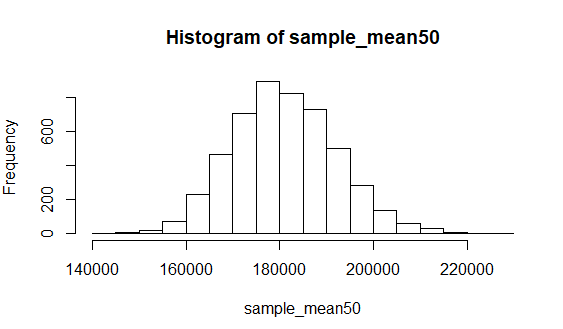
|  |  |  |
| --- | --- | --- |
| 1  2 | x<-sample(ames$SalePrice, size = 50)  mean(x) | [cs](http://colorscripter.com/info#e) |



첫번째 줄은 변수 x에 50개의 랜덤 표본을 저장하는 것이고 두번째 줄은 x의 평균값을 출력하는 프로그램이다. 실제 모평균은 180796인데 반해 50개의 랜덤 표본을 선택하여 얻은 모평균의 추정값은 176727.9으로 나타났다.

3.

|  |  |  |
| --- | --- | --- |
| 1  2  3  4  5  6 | sample\_mean50<-c()  for(i in 1:5000){    x<-sample(ames$SalePrice, size = 50)    sample\_mean50[i]<-mean(x)  }  hist(sample\_mean50) | [cs](http://colorscripter.com/info#e) |



첫 줄에서는 5000번 반복하며 표본 평균을 저장할 변수를 선언한다. 두번째 줄은 i를 1부터 5000까지 증가시키며 for문 내부를 반복하는 것을 의미한다. x에는 2번 문제와 같이 50개의 표본을 랜덤 추출하고 sample\_mean50[i]에는 추출한 표본들의 평균을 대입한다. 표본 평균의 분포를 히스토그램으로 그린 결과 위와 같이 정규분포를 따르고 있음을 알 수 있다.

4.

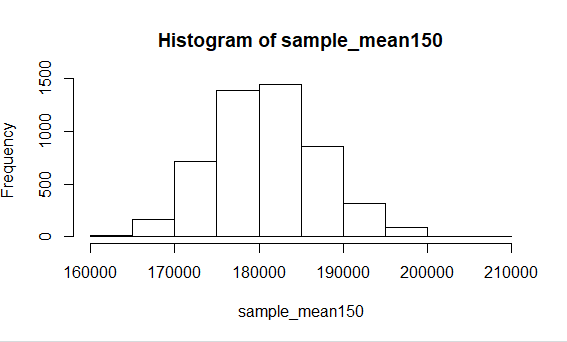
|  |  |  |
| --- | --- | --- |
| 1  2 | mean(sample\_mean50)  var(sample\_mean50) | [cs](http://colorscripter.com/info#e) |



Sample\_mean의 평균과 분산을 계산한 결과는 위와 같다. 이론적으로 모분산의 분산값과 표본 평균의 분산값은 같으며, 이 경우에도 실제 모평균과 표본 평균이 0.6%의 작은 차이를 보이고 있다. 표본 평균의 분산은 모분산을 표본의 크기로 나눈 값이 된다. 모분산의 값이 6381883616이므로 모분산의 값을 50으로 나눈 값인 127637672가 되고, 직접 구한 표본 분산과 유사함을 알 수 있다.

5.

|  |  |  |
| --- | --- | --- |
| 1  2  3  4  5  6  7  8 | sample\_mean150<-c()  for(i in 1:5000){    x<-sample(ames$SalePrice, size = 150)    sample\_mean150[i]<-mean(x)  }  hist(sample\_mean150)  mean(sample\_mean150)  var(sample\_mean150) | [cs](http://colorscripter.com/info#e) |





표본의 크기가 150인 경우에도 마찬가지인 코드를 작성하였다. 히스토그램을 보면 sample\_mean50인 경우에 비해서 더 중앙으로 밀집되었음을 확인할 수 있고 실제로 표본 분산도 더 작게 구해졌다. 표본의 크기가 커질수록 표본 평균이 모평균 근처에 분포하게 된다.