
통계학실험 제 4장

모집단과 표본



과목명	통계학실험
담당교수명	정상아
제출일	2016.04.08
학과	공과대학 컴퓨터공학부
학번, 이름	2016-17101, 김종범

예제 1. 현재 주어진 자료는 일정 기간동안 지역 내의 모든 부동산 거래를 기록한 자료이므로 일종의 모집단이라고 생각할 수 있다. SalePrice 변수에 대해 히스토그램을 그려보고 수치적 요약값을 구해보자. 모집단의 분포는 어떠한가?

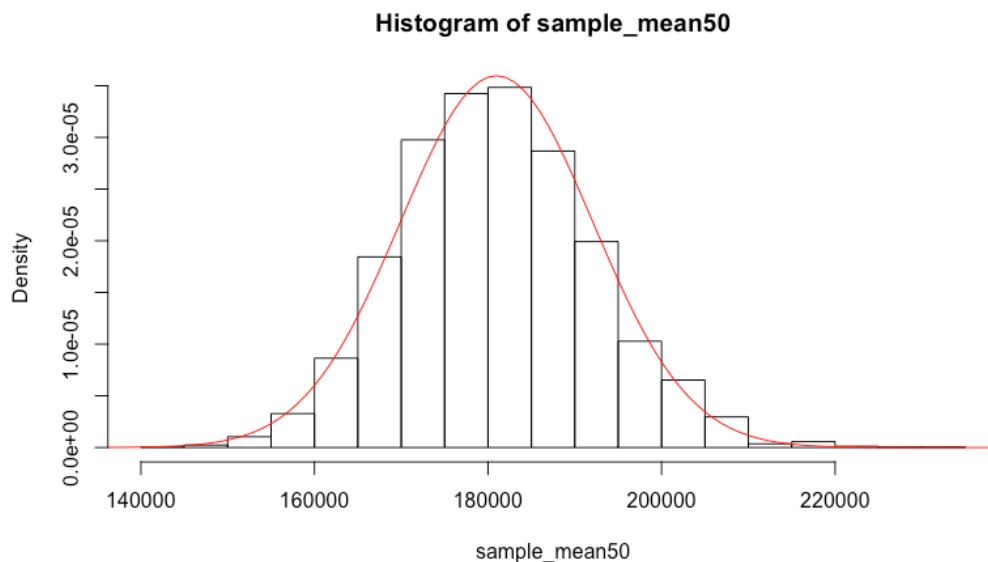
수치적 요약값들은 다음과 같다.

```
> mean      =      mean(salePrice);      mean;
[1] 180796.1
> sd        =      sd(salePrice);      sd;
[1] 79886.69
> var       =      var(salePrice);      var;
[1] 6381883616
```

예제 2. 이 지역에서 발생한 전체 부동산 거래 가격의 평균값을 추정해보려고 한다. 지 금처럼 모집단 전체를 알게 되는 경우는 매우 드물기 때문에 대부분의 경우에는 모집단의 부분집합인 표본을 선택하여 모수를 추정하게 된다. SalePrice에서 50개의 랜덤 표본을 선택해보자. 이 때, 모평균의 추정값은 무엇인가?

```
> pick = sample(salePrice, 50)
> mean(pick)
[1] 186989.1
```

예제 3. 예제 2의 과정을 5000번 반복하여 표본 평균의 표본 분포를 구해보자. 즉, 크기가 50인 랜덤 표본을 선택하여 표본평균을 구하는 과정을 5000번 반복하고 이 결과를 sample_mean50이라는 이름의 벡터에 저장을 한다. sample_mean50을 이용하여 히스토그램을 sample_mean50 그려보자. 표본 평균의 분포는 어떠한가?



표본 평균의 분포는 정규분포꼴이 됨을 볼 수 있다.

예제 4. 예제 3의 sample_mean50의 평균과 분산을 계산해보자. sample_mean50의 평균값은 모집단의 평균과 어떠한 관계가 있는가? sample_mean50의 분산값은 모분산과 어떠한 관계가 있는가?

```
> mean(sample_mean50)
```

```
[1] 180972.1
```

```
> var(sample_mean50)
```

```
[1] 126537995
```

표본 평균의 평균값은 모집단의 평균과 매우 비슷하며, 표본 평균의 분산값은 모집단의 분산의 $1/50$ 배임을 확인할 수 있다.

예제 5. 예제 3의 과정을 표본의 크기를 150으로 증가시켜 반복해보자. 이 결과는 sample_mean150에 저장한다. 표본의 크기에 따른 표본 평균의 분포는 어떠한가?

표본의 크기가 커질수록 표본평균의 분산이 작아지게 되어, 표준편차가 작아지게 된다. 또한 표본의 크기가 크면 클수록 정규분포에 근접한다.

