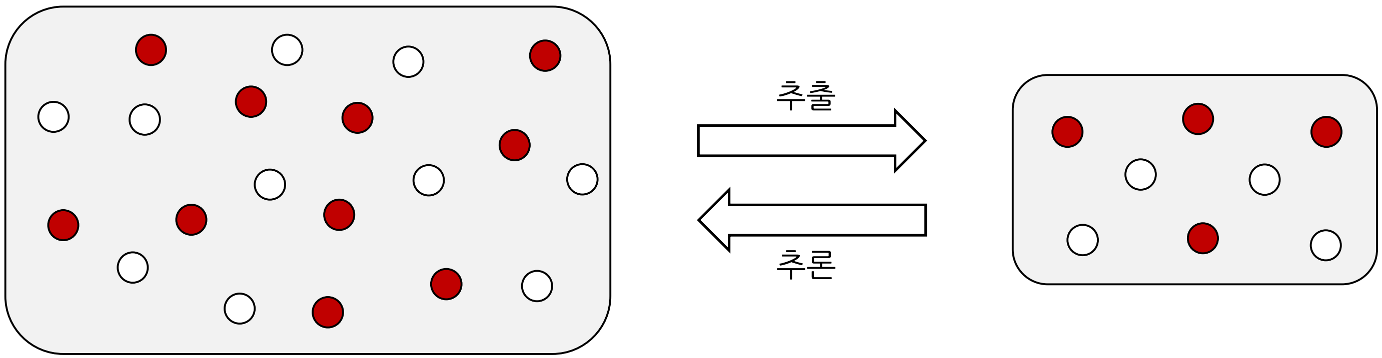
9장 표집분포

통계적 추론(statistical inference)

모집단(population) 표본(sample)



parameters : μ, σ statistics : $\bar{X}, s$

- 모수(parameter) : 수치로 표현되는 모집단의 특성
- 통계량(statistic) : 표본의 관측 값에 의하여 결정되는 수치

통계량의 확률분포

- 표집분포(sampling distribution)

  - 통계량의 값은 표본추출마다 달라진다
  - 통계량은 확률 변수로 자신의 확률분포를 가짐

- 예제 1

  - sample() 함수를 이용하여 1~10 사이의 정수 5개 복원 추출을 10회 반복 총 10개의 표본평균을 구하고 히스토그램을 통해 표본평균의 분포를 확인

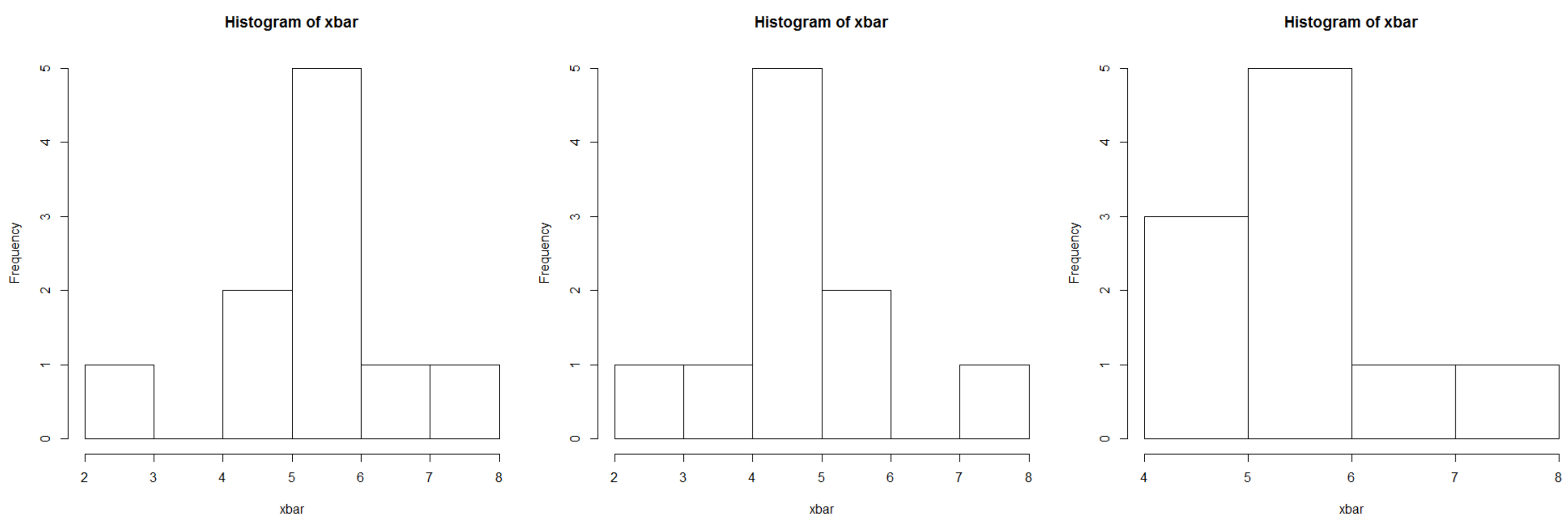
  - hint : for(), 1:10, sample(), replace=T, mean(), hist()

통계량의 확률분포

- 예제 1 코드

```
> xbar <- c()
> for(i in 1:10){ xbar[i] <- mean(sample(1:10,5,replace=T))}
> hist(xbar)
```

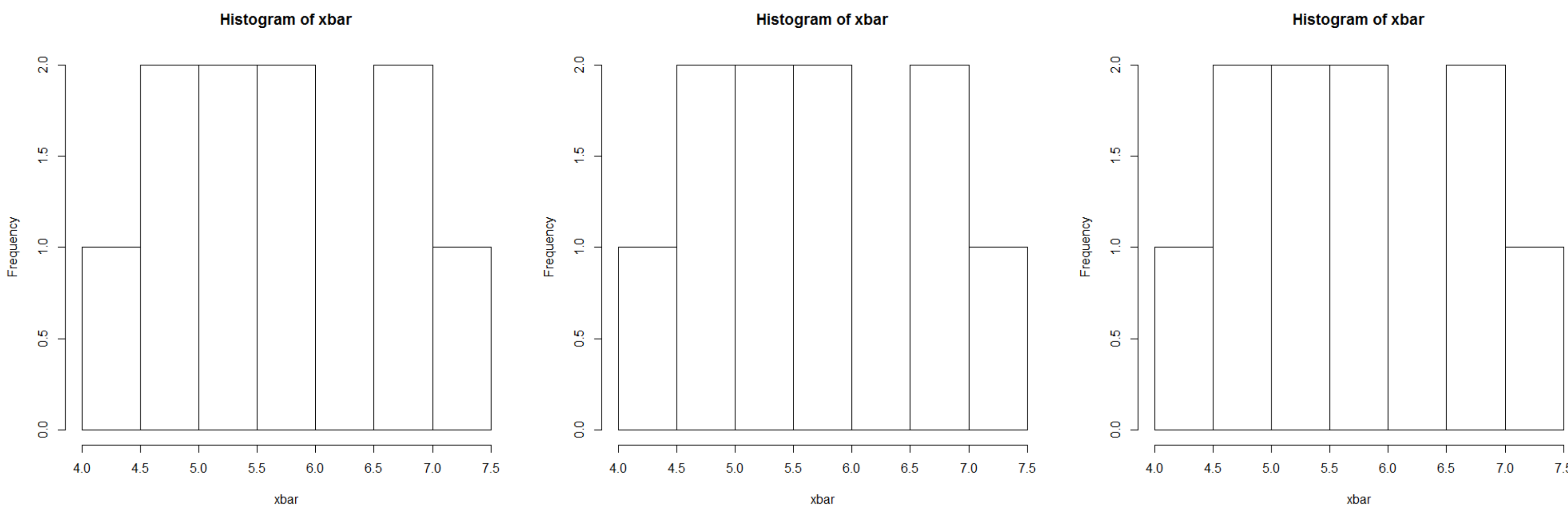
- 반복할 때마다 값이 바뀌는 것을 확인



■ 예제 1 참고

```
> set.seed(1) # set.seed(n)을 입력하여 표본추출 값을 고정
> xbar <- c()
> for(i in 1:10){xbar[i] <- mean(sample(1:10,5,replace=T))}
> hist(xbar)
```

- set.seed(n) : 동일한 n을 지정하면 반복시 동일한 표본 추출값을 제공



표본평균의 분포와 중심극한정리

- Central Limit Theorem(C.L.T.)

  - 모집단의 평균이 μ이고 분산이 σ^2일 때, 임의 추출된 표본의 표본평균 $\bar{X}$는 표본의 크기 n이 큰 경우(보통 30 이상) 근사적으로 정규분포를 따르며, 그 평균은 μ이고 표준편차는 $\sigma/\sqrt{n}$이 된다

  - 모집단이 정규분포를 따르지 않더라도 성립하는 정리이다.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

표본평균의 분포와 중심극한정리

각 분포에서의 난수 추출 함수

- 1) 정규분포 : rnorm(표본 수, 평균, 표준편차)

```
> rnorm(5, 0, 1)
[1] 0.07434132 -0.58952095 -0.56866873 -0.13517862  1.17808700
```

- 2) 이항분포 : rbinom(표본 수, 전체 시행 수, 성공 확률)

```
> rbinom(5, 2, 0.5)
[1] 0 2 1 2 1
```

- 3) 포아송분포 : rpois(표본 수, 평균)

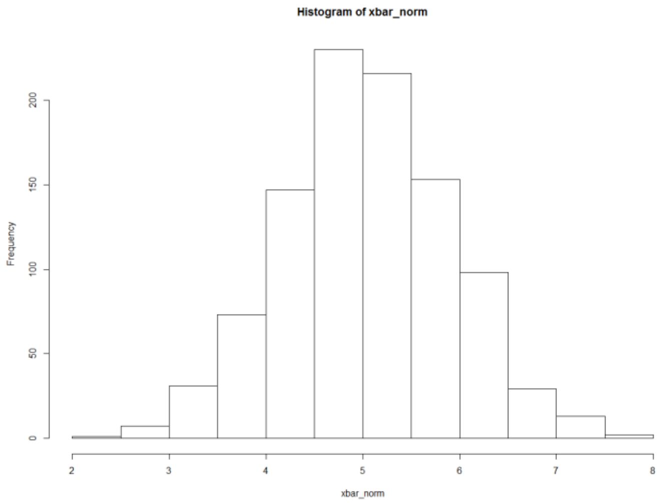
```
> rpois(5, 5)
[1] 2 4 6 4 6
```

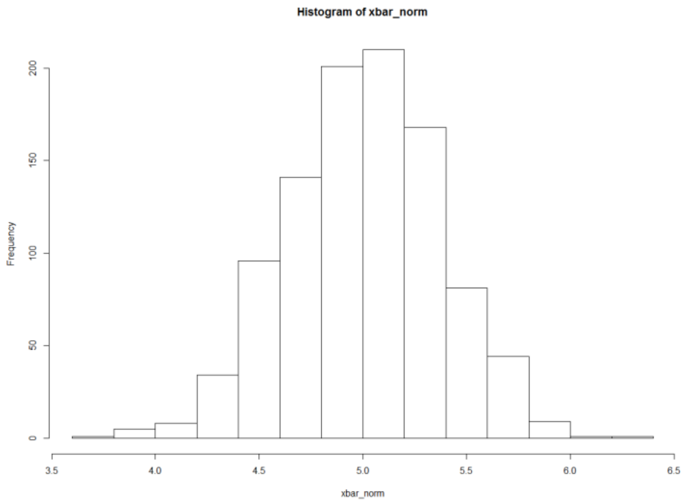
- 4) 카이제곱분포 : rchisq(표본 수, 자유도)

```
> rchisq(5, 1)
[1] 0.26355478 0.04959776 0.02030710 5.40862711 0.03048265
```

표본평균의 분포와 중심극한정리

- 예제 2 : 정규 모집단에서의 $\bar{X}$의 분포
  - 아래와 같은 결과를 출력하라
  - 평균이 5, 표준편차가 2인 정규분포에서 표본 수 5일 때 $\bar{X}$의 분포 (반복 1000회)
  - 평균이 5, 표준편차가 2인 정규분포에서 표본 수 30일 때 $\bar{X}$의 분포 (반복 1000회)





표본평균의 분포와 중심극한정리

- 예제 2 : 정규 모집단에서의 $\bar{X}$의 분포
  - 평균이 5, 표준편차가 2인 정규분포에서 표본 수 5일 때 $\bar{X}$의 분포 (반복 1000회)

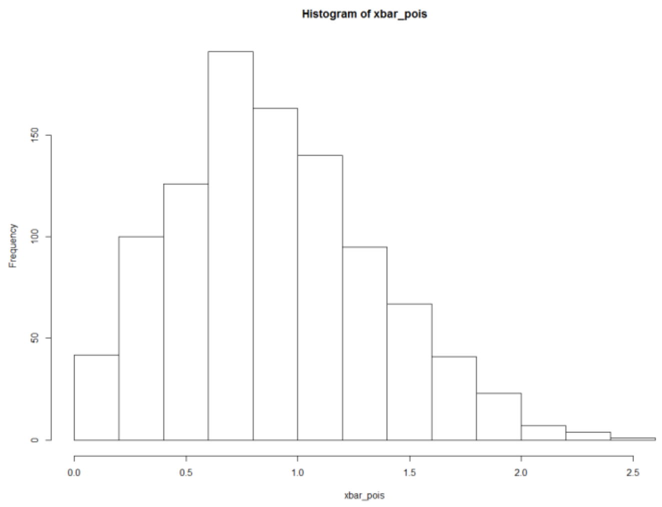
```
> xbar_norm <- c()
> for(i in 1:1000) xbar_norm[i] <- mean(rnorm(5,5,2))
> hist(xbar_norm)
```

  - 평균이 5, 표준편차가 2인 정규분포에서 표본 수 30일 때 $\bar{X}$의 분포 (반복 1000회)

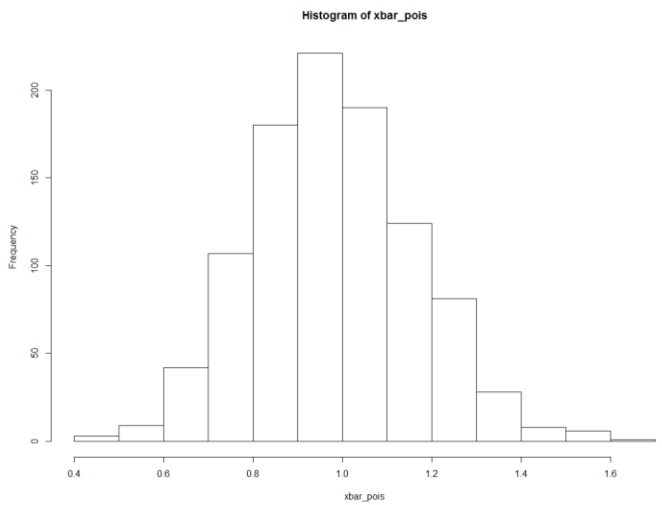
```
> xbar_norm <- c()
> for(i in 1:1000) xbar_norm[i] <- mean(rnorm(30,5,2))
> hist(xbar_norm)
```

표본평균의 분포와 중심극한정리

- 예제 3 : 비대칭형 모집단에서의 $\bar{X}$의 분포
  - 아래와 같은 결과를 출력하라
  - 평균이 1인 포아송분포에서 표본 수 5일 때 $\bar{X}$의 분포 (반복 1000회)



  - 평균이 1인 포아송분포에서 표본 수 30일 때 $\bar{X}$의 분포 (반복 1000회)



표본평균의 분포와 중심극한정리

- 예제 3 : 비대칭형 모집단에서의 $\bar{X}$의 분포
  - 평균이 1인 포아송분포에서 표본 수 5일 때 $\bar{X}$의 분포 (반복 1000회)

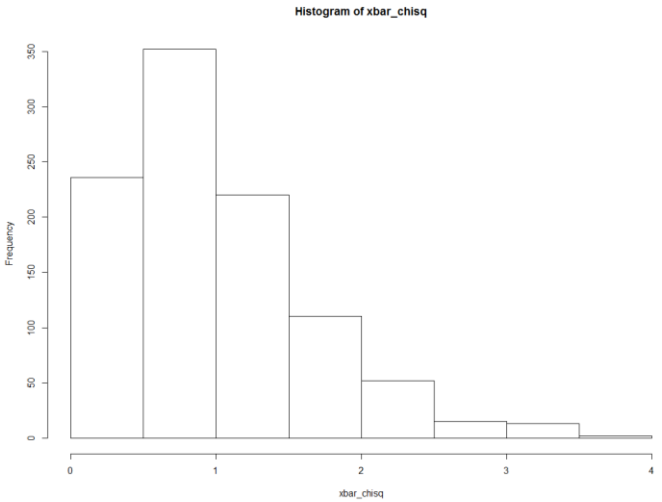
```
> xbar_pois <- c()
> for(i in 1:1000) xbar_pois[i] <- mean(rpois(5,1))
> hist(xbar_pois)
```

  - 평균이 1인 포아송분포에서 표본 수 30일 때 $\bar{X}$의 분포 (반복 1000회)

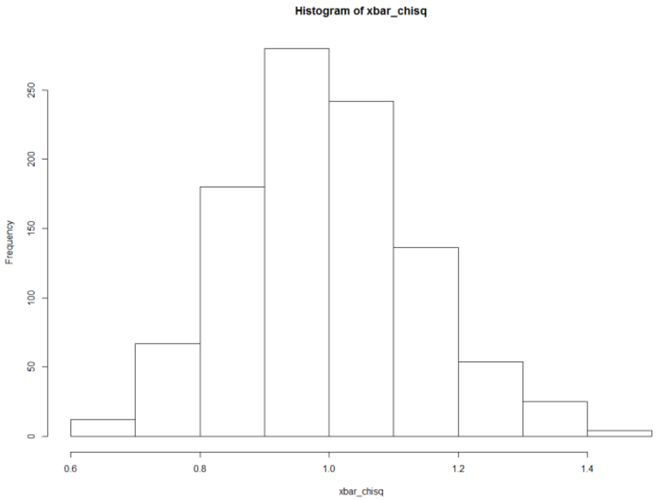
```
> xbar_pois <- c()
> for(i in 1:1000) xbar_pois[i] <- mean(rpois(30,1))
> hist(xbar_pois)
```

표본평균의 분포와 중심극한정리

- 예제 4 : 비대칭형 모집단에서의 $\bar{X}$의 분포
  - 아래와 같은 결과를 출력하라
  - 자유도가 1인 카이제곱분포에서 표본 수 5일 때 $\bar{X}$의 분포 (반복 1000회)



  - 자유도가 1인 카이제곱분포에서 표본 수 100일 때 $\bar{X}$의 분포 (반복 1000회)



표본평균의 분포와 중심극한정리

- 예제 4 : 비대칭형 모집단에서의 $\bar{X}$의 분포
  - 자유도가 1인 카이제곱분포에서 표본 수 5일 때 $\bar{X}$의 분포 (반복 1000회)

```
> xbar_chisq <- c()
> for(i in 1:1000) xbar_chisq[i] <- mean(rchisq(5,1))
> hist(xbar_chisq)
```

  - 자유도가 1인 카이제곱분포에서 표본 수 100일 때 $\bar{X}$의 분포 (반복 1000회)

```
> xbar_chisq <- c()
> for(i in 1:1000) xbar_chisq[i] <- mean(rchisq(100,1))
> hist(xbar_chisq)
```

표본평균의 분포와 중심극한정리

- 결론

  - 모집단이 정규분포를 따르지 않더라도 n이 충분히 클 때, 표본평균이 정규분포를 따른다. (CLT)

  - CLT는 카이제곱분포 등 비대칭형 분포에서도 동일하게 적용된다.