

[프라임칼리지]통계학개론 - 2차 과제물(총 100점 만점)

정시제출기한: 24.06.02(일) 23:59

추가연장제출: 24.06.03(월) 23:59

- [과제] 메뉴

이 름: 김민서

□ 모든 문항에 대해 답만 쓰지 말고 코드 및 결과에 대한 자신의 해석을 반드시 포함해야 합니다.

모든 문항에 대해 “R 또는 R studio 또는 Google Colab(R 언어)”를 사용해 계산하여 1)작성한 '코드'와 2)코드를 실행한 '프로그래밍 결과(캡처이미지)'를 각각 첨부하고 3)결과에 대한 자신의 해석을 기술해주세요.

(자신의 해석은 R코드 동작에 대한 해석이 아닌 실행한 결과에 대한 해석을 의미합니다.)

- R 실습을 다루는 과제이므로 엑셀이나 공학용 계산기, 파이썬 등 다른 프로그램이나 손으로 직접 계산한 답안은 평가가 불가능합니다.

문제1. [7강. 표본분포 - 중심극한정리]

모집단의 분포가 정규분포, 감마분포, 균등분포, 베타분포이고 표본 수를 각각 2, 5, 10, 30으로 할 때, 1)표본평균의 분포를 그리고 2)특징을 서술하시오. (25점)

```
install.packages("TeachingDemos")
library(TeachingDemos)

# 표본 수가 2일 때, 표본평균의 분포
clt.examp(2)

# 표본 수가 5일 때, 표본평균의 분포
clt.examp(5)

# 표본 수가 10일 때, 표본평균의 분포
clt.examp(10)

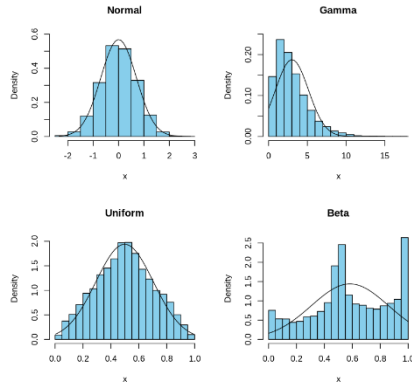
# 표본 수가 30일 때, 표본평균의 분포
clt.examp(30)
```

Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

1) 표본평균의 분포

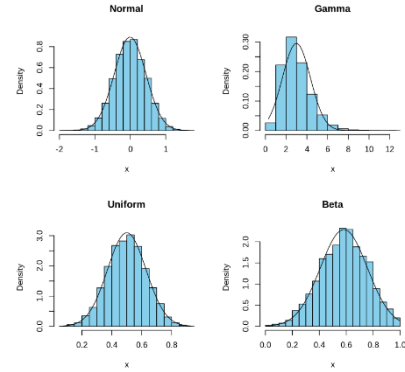
표본수가 2일 때, 표본평균의 분포

sample size = 2



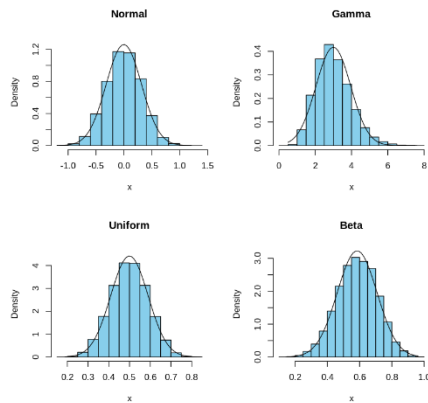
표본수가 5일 때 표본평균의 분포

sample size = 5



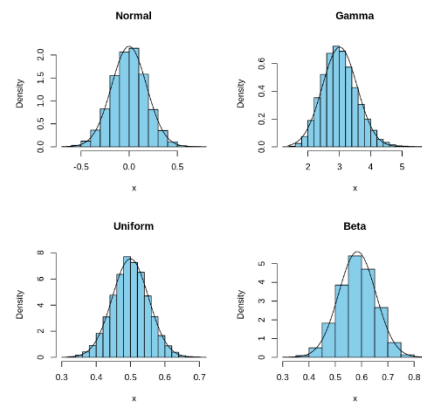
표본수가 10일 때, 표본평균의 분포

sample size = 10



표본수가 30일 때 표본평균의 분포

sample size = 30



2) 특징

표본의 수 N 이 작으면 모 집단의 분포가 표본평균의 분포에 영향을 미치지만

중심극한정리에 따라 N 이 커질수록 표본평균의 분포가 정규분포에 가까워지는 것을 볼 수 있으며 30일 때 모든 분포가 정규분포와 유사해짐을 볼 수 있다.

문제2. [8강. 통계적 추정 I - 모평균의 구간 추정]

금년도 대출자의 초임(모집단)을 알아보기 위하여 100명의 표본을 단순임의확률추출하여 조사하니 표본평균이 75만원이었다. 전체 대출자의 초임이 정규분포를 따르고 표본표준편차가 5만원으로 추정되었다고 할 때 다음 물음에 답하시오. (25점)

(1) 전체 대출자 초임의 평균을 95%의 신뢰도로 구간추정하라. t 분포를 이용하는 것과 정규분포를 이용하는 것에 차이가 있는지 여부와 그렇게 판단한 이유를 서술하시오.

```
# 95% 신뢰구간
# t분포 이용
c(75 + qt(0.025,99)*5/sqrt(100), 75 + qt(0.975,99)*5/sqrt(100))
# 정규분포 이용
c(75 + qnorm(0.025)*5/sqrt(100), 75 + qnorm(0.975)*5/sqrt(100))
```

```
74.0078915242068 · 75.9921084757932
74.02001800773 · 75.97998199227
```

표본의 크기가 $n=100$ 으로 충분히 크기 때문에 t분포를 사용한 신뢰구간과 정규분포를 사용한 신뢰구간의 차이는 각각

- t분포 : 74.00만원 $\leq \mu \leq$ 75.99만원
- 정규분포 : 74.02만원 $\leq \mu \leq$ 75.98만원

으로 차이가 거의 없다.

(2) 전체 대출자 초임의 평균을 99%의 신뢰도로 구간추정하라. 이 구간의 너비가 문제 (1)의 95% 신뢰구간과 비교하여 어떠한지 서술하시오.

```
# 99% 신뢰구간
# t분포 이용, 표본 수 100
c(75 + qt(0.005,99)*5/sqrt(100), 75 + qt(0.995,99)*5/sqrt(100))
```

```
↔ 73.6867972713596 · 76.3132027286404
```

```
[6] # 95% 신뢰구간
# t분포 이용, 표본 수 100
c(75 + qt(0.025,99)*5/sqrt(100), 75 + qt(0.975,99)*5/sqrt(100))
```

```
↔ 74.0078915242068 · 75.9921084757932
```

- 99% 신뢰구간을 t분포를 사용시 $73.68 \leq \mu \leq 76.31$ 나오므로

구간의 너비를 계산시

95%의 신뢰구간의 너비 : $75.99 - 74.00 = 1.99$ 만원

99%의 신뢰구간의 너비 : $76.31 - 73.68 = 2.63$ 만원

즉, 99%의 신뢰구간의 너비가 95%보다 더 크다. 따라서, 신뢰수준이 높아질수록 모집단 평균을 포함할 가능성이 더 높아지기 때문에 구간의 범위가 넓어지는 것을 의미한다.

(3) 표본의 크기를 400명으로 늘렸을 때 전체 대출자 초임의 평균에 대한 95% 신뢰구간을 구하라. 이 구간의 너비가 문제 (1)의 95% 신뢰구간과 비교하여 어떠한지 서술하시오.

```
# 95% 신뢰구간
# t 분포 이용, 표본 수 400
c(75 + qt(0.025, 399)*5/sqrt(400), 75 + qt(0.975, 399)*5/sqrt(400))
```

- 신뢰구간 95%에서 표본 크기 100명과 400명 비교시

```
[4] # 표본수 400명에 대한 95% 신뢰구간
c(75 + qt(0.025, 399)*5/sqrt(400), 75 + qt(0.975, 399)*5/sqrt(400))
```

↔ 74.5085181760198 · 75.4914818239802

```
# 표본수 100명에 대한 95% 신뢰구간
c(75 + qt(0.025, 99)*5/sqrt(100), 75 + qt(0.975, 99)*5/sqrt(100))
```

↔ 74.0078915242068 · 75.9921084757932

- 표본크기 400명의 경우 95% 신뢰구간 $74.50 \leq \mu \leq 75.49$

구간의 너비 : $75.49 - 74.50 = 0.99$ 만원

- 표본크기 100명의 경우 95% 신뢰구간 $74.00 \leq \mu \leq 75.99$

구간의 너비 : $75.99 - 74.00 = 1.99$ 만원

표본의 크기를 100에서 400으로 늘릴 경우 표본크기가 증가함에 따라 표준오차가 감소하기 때문에 신뢰구간이 작아진다. 즉, 표본 크기가 커지면 모집단 평균에 대한 정확도가 높아지며 신뢰구간의 폭이 줄어들게 된다.

※ 문제2의 (2), (3)번 소문항의 경우(각각 99% 신뢰구간 / 표본크기 400명), 표본크기 100명에 대한 95% 신뢰구간과 비교하는 내용에 대한 서술을 반드시 할 것!

문제3. [10강. 통계적 가설검정 I - 모평균에 대한 가설검정]

어느 아파트 한 동에서 10가구에 대해 한 달간 은행방문횟수를 조사하여 우리나라 전체 가구(모집단)의 평균 은행방문횟수인 12회보다 큰지를 검정해보고자 한다. 표본평균은 12.2, 표본표준편차는 0.2로 추정되었고 모집단은 정규분포를 따른다고 가정한다. (25점)

(1) 주어진 문제 상황에 대해 귀무가설과 대립가설을 각각 기술하시오.

- 귀무가설(H_0): 은행 방문횟수가 12회다
- 대립가설(H_1): 은행 방문횟수가 12회보다 크다

(2) 검정통계량 T를 구하시오.

```
# 모평균에 대한 검정
mu = 12; barX = 12.2
S = 0.2; n = 10; alpha = 0.05

# t통계량 값
T = (12.2-12)/(0.2/sqrt(10))

print(T)
```

[1] 3.162278

답 : 약 3.162

(3) 유의수준 0.05에서 기각여부와 가설검정에 대한 결론을 서술하시오.

```
# 유의확률 : 단측검정
PV = 1 - pt(T,9)
PV
if(PV < alpha) print("5% 유의수준에서 귀무가설 기각")
if(PV >= alpha) print("5% 유의수준에서 귀무가설 기각하지 못함")

0.00575399258297193
[1] "5% 유의수준에서 귀무가설 기각"
```

p-value가 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 채택한다.

문제4. [12강. 통계적 비교 I – 대응 표본에 대한 두 모집단 평균의 비교]

한 사람의 연주 실력에 대해 10회의 연주동안 두 명의 평가자가 평가하였는데 차례대로 다음과 같았다.

평가자1: 90, 80, 75, 60, 65, 70, 95, 90, 88, 50

평가자2: 88, 84, 72, 65, 68, 76, 96, 93, 90, 60

두 평가자 간 평가에 실제로 차이가 있는지 5% 유의수준에서 검정하시오. 단, 평가자 간 평가의 차이는 정규분포를 따른다고 가정한다. (25점)

```
# 두 평가자의 데이터 수집
x = c(90, 80, 75, 60, 65, 70, 95, 90, 88, 50)
y = c(88, 84, 72, 65, 68, 76, 96, 93, 90, 60)

# t 검정, 짝진 두 모집단 평균 검정
it = t.test(x,y, paired = TRUE)
it
```

Paired t-test

```
data: x and y
t = -2.4232, df = 9, p-value = 0.03841
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -5.6072476 -0.1927524
sample estimates:
mean difference
      -2.9
```

```
t = -2.4232, df = 9, p-value = 0.03841
```

p-value가 유의수준 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 채택한다.

따라서, 두 평가자간 평가가 차이가 있다.