DFMBA 경영통계 – 3rd assignment

학번 : [redacted]

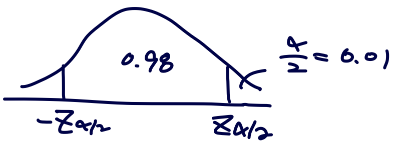
이름 :

1. 어느 신용카드 회사는 월말 미결제 잔금 때문에 이자를 지불해야 하는 신용카드 보유자의 비율을 추정하고자 한다. 특정 달에 고객 중 120명을 무작위로 선택하여 조사한 결과 이 중 월말 미결제 잔금이 남아있는 경우는 78명이었다고 한다. 이를 이용하여 구간추정을 한다면, 신뢰수준 98%에서 요구되는 오차한계가 얼마인지 구하여라. (10점)

$X_1 X_2 \cdots X_{120} \overset{iid}{\sim} Ber(p)$ & $np \ge 5$, $n(1-p) \ge 5$ 만족하므로

$\hat{p} \dot\sim N\left(p, \frac{p(1-p)}{n}\right)$ by CLT, p에 관한 CI $1-\alpha = 0.98$, $\hat{p} = \frac{78}{120} = 0.65$

$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ 신뢰구간 형성시,




```
> # 1번 R코드
> p <- 78/120
> n <- 120
> n*p
[1] 78
> n*(1-p)
[1] 42
> z <- -qnorm(0.01)
> z*sqrt(p*(1-p)/n)
[1] 0.1012919
```

오차한계 $= z_{0.01} \times \sqrt{\frac{0.65 \times 0.35}{120}} = 0.1012919$

2. 코스피200지수에 편입되어 있는 상장사 중 7개의 회사를 임의로 추출하였다. 추출된 7개 회사의 순이익의 평균은 17%, 표준편차는 8.5%였다고 하자. 코스피200지수에 편입된 상장사 전체에 대해 순이익은 정규분포를 따른다고 할 때, 전체 상장사의 순이익의 평균에 대한 95%의 신뢰구간을 구하고 이를 해석하여라. (10점)

$X_1 X_2 \cdots X_7 \overset{iid}{\sim} N(\mu, \sigma^2)$

$\bar{x} = 17$, $s = 8.5$, 모집단은 정규분포이며, 모분산은 알지못할때,

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

$\bar{x} \pm t_{\alpha/2, n-1} \times \frac{s}{\sqrt{n}}$

$= 17 \pm t(0.025, 6) \times \frac{8.5}{\sqrt{7}}$

```
> # 2번 R코드
> xbar <- 17
> sd <- 8.5
> n <- 7
> t <- -qt(0.025, df= n-1)
> xbar + c(-1,1) * t * sd/sqrt(n)
[1]  9.138811 24.861189
```

$(9.138811, 24.861189)$

⇒ 전체 상장사의 순이익의 평균(μ)에 대한 95%의 신뢰구간은 (9.138811, 24.861189)이며, 이는 임의 추출된 7개의 표본을 100번 반복적으로 뽑은뒤 신뢰구간을 구성시 100개 중 95개 정도는 모평균(μ)을 포함할 것으로 예상된다

```
> # 3번 R코드
> xbar <- 29.5
> sd <- 6.39
> n <- 49
> z <- -qnorm(0.005)
> xbar + c(-1,1) * z * sd/sqrt(n)
[1] 27.14864 31.85136
```

3. 현재 주식시장에서 거래되고 있는 소형주에 대한 3월만기 옵션가격의 평균을 알아보고자 한다. 이를 위해 49개의 옵션을 무작위로 선택하여 가격을 조사한 결과 그 평균은 29.5, 표준편차는 6.39였다고 한다. 소형주 전체에 대한 평균 옵션가격의 99% 신뢰구간을 구하고 이를 해석하여라. (10점)

$X_1 X_2 \cdots X_{49} \overset{iid}{\sim} ?$

μ에 대한 0.99 CI, $n=49$, $\bar{X}=29.5$, $S=6.39$일때,

정규가 아닌 모집단으로 표본 $n \geq 30$으로 충분히 크므로 중심극한정리에 의해 정규가정

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0,1)$ by CLT

$\bar{X} \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} = 29.5 \pm z_{0.005} \times \frac{6.39}{\sqrt{49}}$ (27.14864, 31.85136)

⇒ 소형주 전체에 대한 평균 옵션가격(μ)의 신뢰구간은 (27.14864, 31.85136) 이며,
이는 임의추출된 49개의 표본을 반복적으로 뽑은뒤 신뢰구간 구성시, 100개중
99개 정도는 모평균(μ)을 포함할 것으로 기대된다.

4. 'assign3data.csv' 파일에는 무작위로 추출한 40개 회사채에 대한 만기년수와 수익률이 기록되어 있다. 만기년수와 수익률의 모분포가 정규분포라고 할 때 다음 물음에 답하여라.

(1) 만기년수의 모평균에 대한 95% 신뢰구간을 구하여라. (10점)

$X_1 X_2 \cdots X_{40} \overset{iid}{\sim} N(\mu, \sigma^2)$. 정규분포이나 모분산 알지못하므로 t분포 사용

μ, 0.95 CI 이고, $\bar{X}=9.70625$, $S=7.980523$, $n=40$

$\bar{X} \pm t_{\alpha/2, n-1} \times \frac{S}{\sqrt{n}}$

⇒ 모평균에 대한 신뢰구간은

(7.153955, 12.258545)

```
> # 4-1번 R코드
> yr_xbar <- mean(assign3data$year)
> yr_xbar
[1] 9.70625
> yr_sd <- sd(assign3data$year)
> yr_sd
[1] 7.980523
> n <- 40
> clevel <- 0.95
> yr_coeff <- qt( 1 - (1-clevel)/2, df=(n-1))
> yr_xbar + c(-1,1) * yr_coeff * yr_sd /sqrt(n)
[1]  7.153955 12.258545
```

(2) 회사채 수익률의 모표준편차에 대한 99% 신뢰구간을 구하여라. (10점)

$X_1 X_2 \cdots X_{40} \overset{iid}{\sim} N(\mu, \sigma^2)$

σ^2 0.99 CI 이고, $\bar{X}=3.88535$, $S=1.619403$, $n=40$ 이므로

$$\left(\frac{\overset{39}{(n-1)} \cdot S^2}{\chi^2_{0.005,39}} , \frac{\overset{39}{(n-1)} \cdot S^2}{\chi^2_{0.995,39}} \right)$$

⇒ 모표준편차에 대한 신뢰구간은

(1.249820, 2.261606)

```
> # 4-2번 R코드
> return_xbar <- mean(assign3data$return)
> return_xbar
[1] 3.88535
> return_sd <- sd(assign3data$return)
> return_sd
[1] 1.619403
> n <- 40
> sqrt( c( (n-1) * return_sd^2 / qchisq(0.995, df=n-1 ), (n-1) *
 return_sd^2 / qchisq( 0.005, df=n-1) ) )
[1] 1.249820 2.261606
```

5. $X_1, X_2, \ldots, X_n$은 모평균이 μ 이고 모분산이 σ^2인 어느 모집단으로부터의 확률 표본이라고 하자. 다음 중 중심극한의 정리(Central Limit Theorem)가 적용된 사실은 무엇인가? (3) (5점)

① 표본 수 n 이 많아질 수록, 표본 $X_1, X_2, \ldots, X_n$는 정규 분포에 근사한다.

② 표본 수 n 이 많아질 수록, 표본 평균 $\bar{X}$의 분산은 작아진다.

③ 표본 수 n 이 많아질 수록, 표본 평균 $\bar{X}$는 정규 분포에 근사한다.

④ 표본 수 n 에 관계없이, 표본 평균 $\bar{X}$의 기대값은 모평균 μ가 된다.

6. $N[\mu, \sigma^2]$으로부터의 i.i.d. 표본 $X_1, X_2, \ldots, X_n$의 표본평균이 $\bar{X}$, 표본분산이 S^2이라고 하자. 표본의 개수 n 이 12인 경우에 대하여, 아래 물음에 답하여라. (15점)

(1) $P\left(\frac{\bar{X}-\mu}{\sqrt{S^2/12}} < c\right) = 0.99$를 만족하는 c 값을 구하여라.

$X_1 X_2 \cdots X_{12} \sim N(\mu, \sigma^2)$, 정규분포이나 모평균 알지 못하면.

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ 일때,

$P\left(\frac{\bar{X}-\mu}{S/\sqrt{12}} < c\right) = 0.99$

$\therefore c = t_{0.01,\,11} = 2.718079$

```
> # 6-1번 R코드
> qt(0.99,df=11)
[1] 2.718079
```

(2) $P\left(\frac{(\bar{X}-\mu)^2}{\sigma^2/12} < c\right) = 0.99$를 만족하는 c 값을 구하여라.

```
> # 6-2번 R코드
> qnorm(0.995)^2
[1] 6.634897
```

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

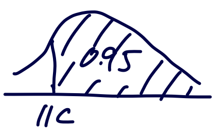
$P\left(\frac{(\bar{X}-\mu)^2}{\sigma^2/12} < c\right) = P\left(-\sqrt{c} < \frac{\bar{X}-\mu}{\sigma/\sqrt{12}} < \sqrt{c}\right) = 0.99$

$\therefore c = {z_{0.005}}^2 = 6.634897$

(3) $P\left(\frac{S^2}{\sigma^2} > c\right) = 0.95$를 만족하는 c 값을 구하여라.

$u = \frac{(n-1)\cdot S^2}{\sigma^2} \sim \chi^2(n-1)$ 이므로

$P\left(\frac{S^2}{\sigma^2} > c\right) = P\left(\frac{(n-1)\cdot S^2}{\sigma^2} > (n-1)\cdot c\right)$



$= P\left(\frac{11\cdot S^2}{\sigma^2} > 11c\right) = 0.95$

$\therefore c = \chi^2_{0.95,\,11}/11 = 0.4158921$

```
> # 6-3번 R코드
> qchisq(0.05,df=11)/11
[1] 0.4158921
```

7. 어느 은행 고객 중 62명을 랜덤하게 뽑아 예금액을 조사한 뒤 이를 이용하여 전체 예금액에 대한 모평균에 대한 신뢰구간을 신뢰수준 95%를 적용하여 구하였더니 (112.62 129.38) 였다고 하자. 다음 중 올바른 설명은 T, 잘못된 설명은 F로 답하여라. (20점)

(1) 모평균이 (112.62 129.38)의 범위에 속하는지 여부는 정확히 알 수 없다. T

(2) 신뢰수준 95%의 의미는 62개의 표본자료 중 약 95%가 (112.62 129.38)의 범위에 속함을 말한다. F

(3) 동일한 자료를 이용하여 신뢰수준 99%로 신뢰구간을 구해 보면 구간의 크기는 더 커질 것이다. T

(4) 다시 62명의 고객을 랜덤으로 뽑았을 때, 새로 계산된 표본 평균은 (112.62 129.38)의 범위에 속할 것이다. F

8. 모집단의 분포 $Normal[\mu, \sigma^2]$에서 추출된 확률표본 $X_1, \ldots, X_n$의 평균 $\bar{X}$을 이용하여 μ에 관한 95%의 신뢰구간을 $(\bar{X} - a,\ \bar{X} + a)$의 형태로 표현하였다. 다음 중 올바른 설명은 T, 잘못된 설명은 F로 답하여라. 단, 모분산 σ^2은 알려지지 않았음. (10점)

(1) 다른 조건은 모두 동일할 때 표본의 수 n이 증가하면 신뢰구간의 길이는 짧아진다. T

(2) 표본을 n개 추출해서 신뢰구간 추정치를 구하는 과정을 반복하는 경우, $\bar{X}$는 랜덤하게 바뀌지만 a는 변하지 않는다. F