2024년 1학기 컴퓨터 응용통계 답안지

학번 .....C389008..... 이름...김동혁..... 서명 ...김동혁.....

문항번호	답	문항번호	답
1	1	11	3
2	2	12	3
3	1	13	2
4	2	14	4
5	2	15	3
6	1	16	2
7	4	17	5
8	4	18	5
9	2	19	3
10	5	20	4

여기에 번호순으로 손풀이, R코드와 결과 붙여서, PDF로 저장하기

1 ~ 2

```
> #1
> 198-(1.96*16/sqrt(25))
[1] 191.728
> 198+(1.96*16/sqrt(25)) # 1
[1] 204.272
>
> -2/(16/5) #2 검정통계량
[1] -0.625
> 2 * (1 - pnorm(0.625)) # 2 p -값
[1] 0.5319711
>
```

3. p-값이 유의수준 0.05 보다 크므로 귀무가설을 기각하지 않음. 평균이 200이라고 말할 수 있다.

4~6

```
#4>
에러: 예기치 않은 '>'입니다 in ">"
> library(MASS)
> data(Cars93)
>
> x <- subset(Cars93, Type=="Midsize")$MPG.highway
> x
 [1] 25 26 30 31 27 25 29 27 30 27 22 24 23 26 25 26 24 26 31 27 29 28
> mean(x)
[1] 26.72727
> sd(x) # 틀림 2번
[1] 2.510584
> max(x)
[1] 31
> t.test(x)

        One Sample t-test
```

```
data:  x
t = 49.933, df = 21, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 25.61414 27.84040
sample estimates:
mean of x
 26.72727

>   # 95% 신뢰 (25.61414 27.84040)
> # 5
>
>   # 95% 신뢰 (25.61414 27.84040)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.96504, p-value = 0.5972
P-값이 0.05 보다 작으므로 귀무가설을 기각한다
```

7.

8~9

```
#8>
에러: 예기치 않은 '>'입니다 in ">"
> data(painters)
> painters$School
 [1] A A A A A A A A A A B B B B B B C C C C C C D D D D D D D D D D E E E E E E E
[40] F F F F G G G G G G G H H H H
Levels: A B C D E F G H
> a <- subset(painters, School=="A")$Composition
> b <- subset(painters, School=="B")$Composition
>
> var.test(a, b) #8

        F test to compare two variances
```

```
data:  a and b
F = 7.9764, num df = 9, denom df = 5, p-value = 0.03415
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.193884 35.769464
sample estimates:
ratio of variances
            7.976401

> t.test(a, b) #9

        Welch Two Sample t-test

data:  a and b
t = -0.92698, df = 12.196, p-value = 0.3719
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.911758  2.378425
sample estimates:
mean of x mean of y
 10.40000  12.16667

> t.test(a, b, var.equal=T)

        Two Sample t-test

data:  a and b
t = -0.75268, df = 14, p-value = 0.4641
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.800845  3.267512
sample estimates:
mean of x mean of y
 10.40000  12.16667
```

10.

11~14

```
#11>
+ data(UScereal)
[1] FALSE
>
>
> x <- UScereal$protein
> y<- UScereal$calories
>
> plot(x, y)
> fit <- lm(y~x)
> fit

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)                x
      87.99            16.67

> abline(fit)
> anova(fit)
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq F value       Pr(>F)
x           1 124262  124262  62.611 5.071e-11 ***
Residuals 63 125034    1985
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(fit)
```

```
Call:
lm(formula = y ~ x)

Residuals:
      Min        1Q    Median        3Q       Max
-121.379  -21.379    0.883   16.458  151.925

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   87.986      9.528   9.234 2.55e-13 ***
x             16.674      2.107   7.913 5.07e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44.55 on 63 degrees of freedom
Multiple R-squared:  0.4985,    Adjusted R-squared:  0.4905
F-statistic: 62.61 on 1 and 63 DF,  p-value: 5.071e-11

>
> abline(fit)
> par(mfrow=c(2,2))
> plot(fit)
> shapiro.test(fit$resid)

        Shapiro-Wilk normality test

data:  fit$resid
W = 0.94939, p-value = 0.009889

> #12번 b,c
> 124262/249296
[1] 0.4984516
> sqrt(1985)
[1] 44.55334
```

15~16

```
> #15
```

```
> install.packages("agricolae")
경고: 패키지 'agricolae'가 사용중이므로 설치되지 않을 것입니다
> library(agricolae)
> fit <- lm(weight~feed, data = chickwts)
> duncan.test(fit, "feed", alpha=0.05, console=TRUE)

Study: fit ~ "feed"

Duncan's new multiple range test
for weight

Mean Square Error:  3008.554

feed,  means

                weight       std  r       se Min Max     Q25    Q50    Q75
casein      323.5833 64.43384 12 15.83391 216 404 277.25 342.0 370.75
horsebean 160.2000 38.62584 10 17.34518 108 227 137.00 151.5 176.25
linseed     218.7500 52.23570 12 15.83391 141 309 178.00 221.0 257.75
meatmeal   276.9091 64.90062 11 16.53798 153 380 249.50 263.0 320.00
soybean     246.4286 54.12907 14 14.65936 158 329 206.75 248.0 270.00
sunflower 328.9167 48.83638 12 15.83391 226 423 312.75 328.0 340.25

Groups according to probability of means differences and alpha level( 0.05 )

Means with the same letter are not significantly different.

                weight groups
sunflower 328.9167      a
casein      323.5833      a
meatmeal   276.9091       b
soybean     246.4286      bc
linseed     218.7500      c
horsebean 160.2000       d
> anova(fit)
Analysis of Variance Table

Response: weight
```

```
             Df Sum Sq Mean Sq F value       Pr(>F)
feed          5 231129   46226  15.365 5.936e-10 ***
Residuals 65 195556    3009
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> par(mfrow=c(1,1))
> boxplot(weight~feed, data = chickwts)
```

17.

```
> #17
> data(HairEyeColor)
> HairEyeColor[1:3, 1:3, Sex = "Female"]
          Eye
Hair      Brown Blue Hazel
  Black      36    9     5
  Brown      66   34    29
  Red        16    7     7
> my <- HairEyeColor[1:3, 1:3, Sex = "Female"]
>
```

18.

1. 표본 크기가 작아지면 신뢰구간은 넓어진다.
2. 유의수준이 작아지면 신뢰구간이 짧아진다
3. 95% 신뢰구간은 동일분포에서 동일한 크기의 표본 100개로 100개의 신뢰구간을 만들 때, 이들 중 약 5개가 실제 모평균을 포함하지 못한다는 의미이다.
4. p -값이 유의수준보다 작으면 귀무가설을 기각한다

19.

20.