

- 68페이지 11번 문제 -

R명령어 리스트와 그래프

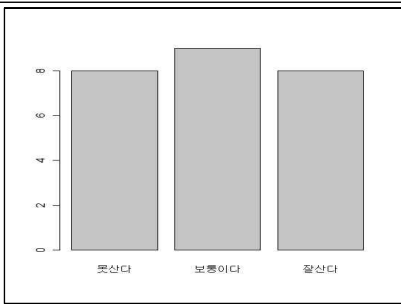
```
income <- read.csv("d:/knou/11.csv")

#학력과 생활수준 변수를 명목형 자료형으로 변환해 준다.
income$학력 <- as.factor(income$학력)
income$생활수준 <- as.factor(income$생활수준)

levels(income$학력) <- c("중졸이하", "고졸", "대졸이상")
levels(income$생활수준) <- c("못산다", "보통이다", "잘산다")

#데이터 출력
income
#   일련번호 생활수준   학력 수입
#1         1   잘산다 대졸이상 5000
#2         2   못산다 중졸이하 2000
#3         3   보통이다   고졸 3000
#4         4   잘산다 대졸이상 4700
#5         5   못산다 대졸이상 2700
#6         6   잘산다 대졸이상 5500
#7         7   못산다 대졸이상 2800
#8         8   보통이다 대졸이상 3875
#9         9   보통이다   고졸 3400
#10        10   잘산다 대졸이상 5000
#11        11   보통이다   고졸 3400
#12        12   못산다 중졸이하 1800
#13        13   잘산다 대졸이상 5000
#14        14   보통이다   고졸 3400
#15        15   보통이다 대졸이상 3875
#16        16   보통이다 대졸이상 3875
#17        17   잘산다 대졸이상 5000
#18        18   잘산다 대졸이상 5000
#19        19   못산다 중졸이하 1800
#20        20   보통이다   고졸 3400
#21        21   못산다 중졸이하 1800
#22        22   못산다   고졸 2275
#23        23   보통이다 대졸이상 3875
#24        24   잘산다 중졸이하 4050
#25        25   못산다   고졸 2275

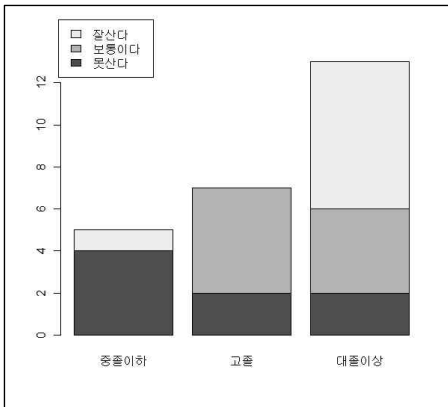
#생활수준 변수에 대한 막대 그래프
barplot(table(income$생활수준))
```



```
tab <- table(income$생활수준, income$학력)
```

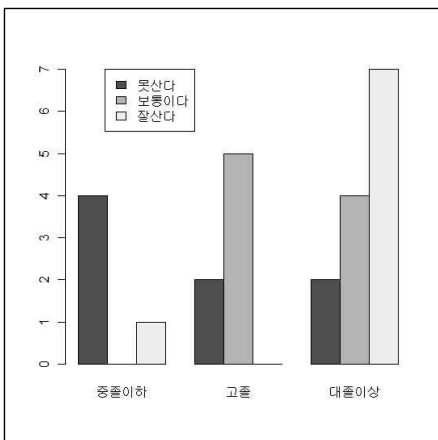
#쌓는형

```
barplot(tab, legend.text=T, args.legend=list(x=1, y=15))
```



#나란형

```
barplot(tab, legend.text=T, args.legend=list(x=5, y=7), beside=T)
```



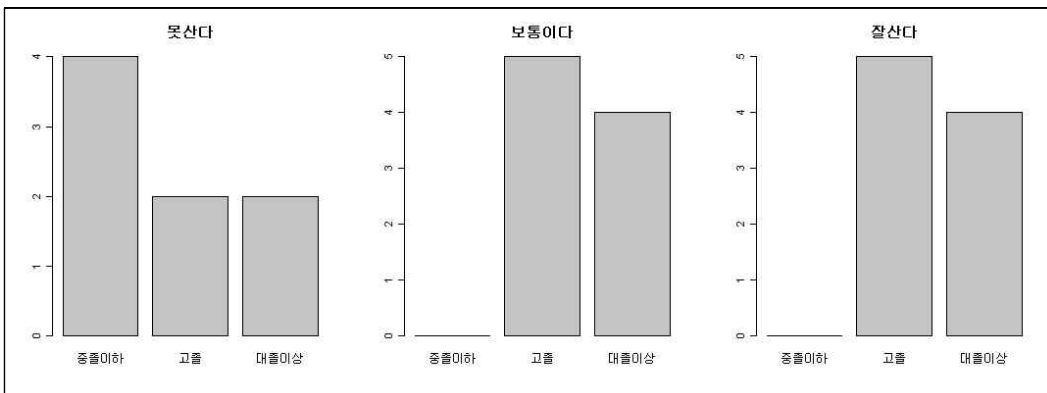
#분리형

```
par(mfrow=c(1,3))
```

```
barplot(tab[1,], main="못산다")
```

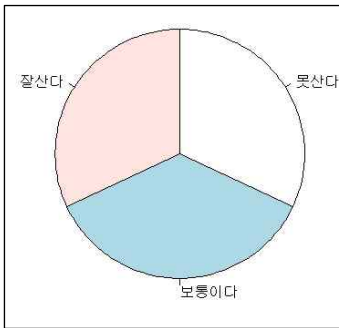
```
barplot(tab[2,], main="보통이다")
```

```
barplot(tab[3,], main="잘산다")
```

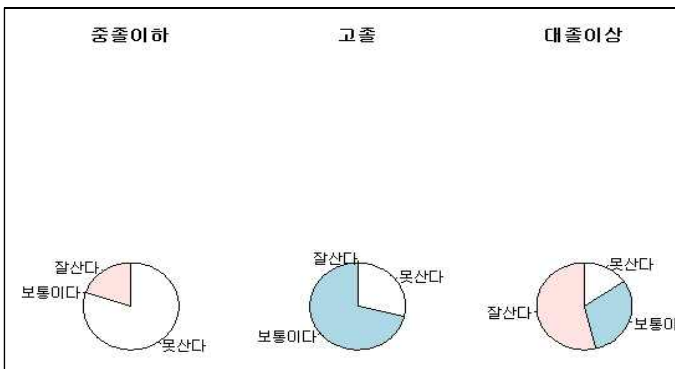


#생활 수준에 따른 원그림

```
pie(table(income$생활수준), clockwise=T )
```



```
par(mfrow = c(1,3))
pie(tab[,1], main=colnames(tab)[1], clockwise=T)
pie(tab[,2], main=colnames(tab)[2], clockwise=T)
pie(tab[,3], main=colnames(tab)[3], clockwise=T)
```



```
dev.off()
```

#수입에 대한 기본 통계량

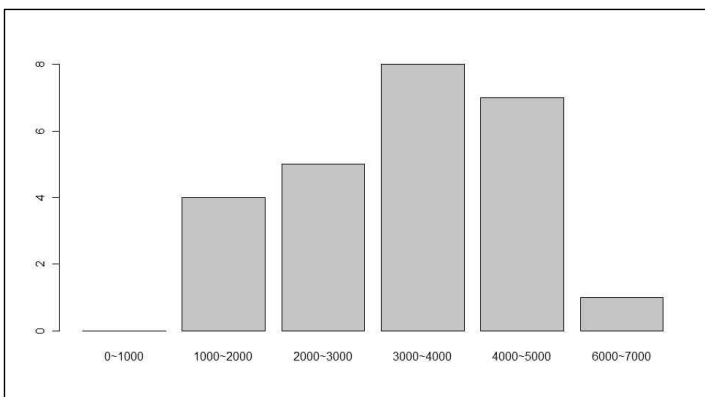
```
summary(income$수입)
```

```
#  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
# 1800  2700   3400   3552  4700   5500
```

#수입에 대한 도수분포표

```
discritizedincome <- cut(income$수입, breaks=c(0, 1000, 2000, 3000, 4000,5000, 6000), labels=c("0~1000",
"1000~2000", "2000~3000" ,"3000~4000", "4000~5000", "6000~7000"))
```

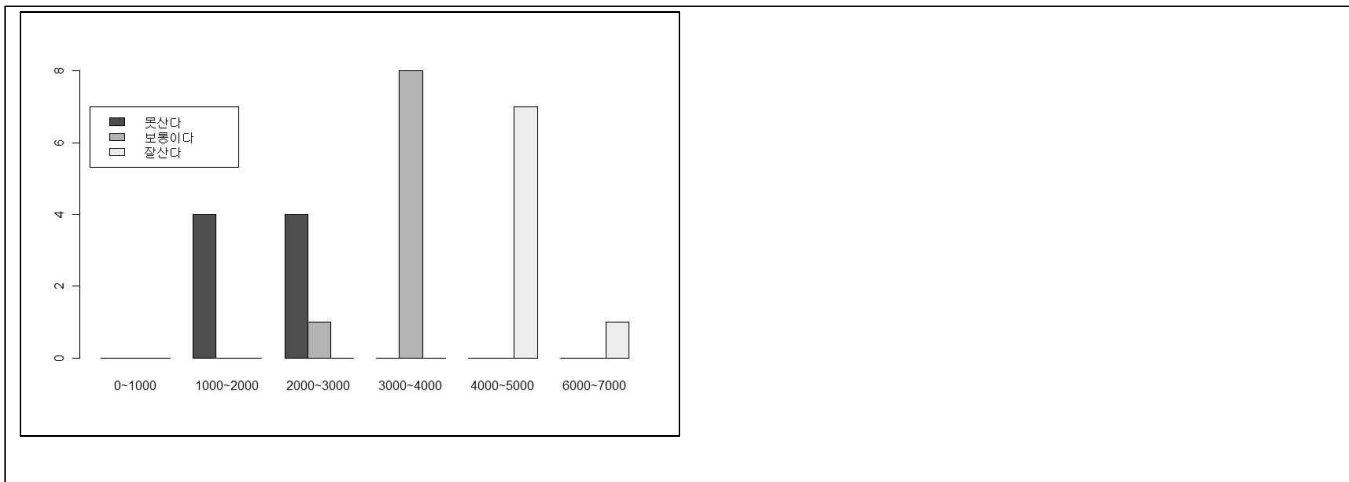
```
barplot(table(discritizedincome ))
```



#생활 수준에 따른 수입의 도수분포표(나란형)

```
tab <- table(income$생활수준, discritizedincome)
```

```
barplot(tab, legend.text=T, args.legend=list(x=7, y=7), beside=T)
```



- 219페이지 22번 문제 -

```
elec <- read.csv("d:/knou/elec.csv")
elec
#  sex score
#1  M   73
#2  F   57
#3  M   96
#4  F   78
#5  M   74
#6  M   42
#7  F   55
#8  M   44
#9  M   91
#10 F   91
#11 F   50
#12 F   65
#13 M   46
#14 F   63
#15 M   82
#16 F   60
#17 M   97
#18 F   79
#19 M   85
#20 F   79
#21 F   92
#22 M   50
#23 M   42
#24 M   46
#25 F   86
#26 F   81
#27 M   81
#28 M   83
#29 F   64
#30 F   76
#31 M   40
#32 M   57
#33 F   78
#34 M   66
```

```
#35 M 84
#36 F 96
#37 M 94
#38 M 70
#39 M 70
#40 F 81
```

#모든 근로자들의 평균적응 점수가 60보다 크다고 주장하는 것에 대해서 유의수준 0.05에서 검정하라.

```
t.test(elec$score, alternative="greater", mu= 60)
```

```
# One Sample t-test
```

```
#
```

```
#data: elec$score
```

```
#t = 4.077, df = 39, p-value = 0.0001087
```

```
#alternative hypothesis: true mean is greater than 60
```

```
#95 percent confidence interval:
```

```
# 66.51275 Inf
```

```
#sample estimates:
```

```
#mean of x
```

```
# 71.1
```

#p-value가 0.0001087로 유의수준0.05보다 작게 나와

#근로자의 평균적응 점수가 60이다라는 귀무가설을 기각하고 60보다 크다는 대립가설을 선택한다.

#남,녀에 따라서 평균적응 점수가 다른지 유의수준 0.05로 검정한다.

#일단 var.test로 두 집단의 분산이 같은지 검정한다.

```
var.test(score ~ sex, data=elec)
```

```
# F test to compare two variances
```

```
#data: elec$score[elec$sex == "M"] and elec$score[elec$sex == "F"]
```

```
#F = 2.0951, num df = 21, denom df = 17, p-value = 0.1262
```

```
#alternative hypothesis: true ratio of variances is not equal to 1
```

```
#95 percent confidence interval:
```

```
# 0.805833 5.202849
```

```
#sample estimates:
```

```
#ratio of variances
```

```
# 2.095149
```

#p-value가 0.1262로 0.05보다 크며 따라서 두 집단의 분산이 같다고 하는 귀무가설을 지지하게 된다.

```
t.test(score ~ sex, data=elec, var.equal=T)
```

```
# Two Sample t-test
```

```
#data: score by sex
```

```
#t = 0.9437, df = 38, p-value = 0.3513
```

```
#alternative hypothesis: true difference in means is not equal to 0
#95 percent confidence interval:
# -5.922636 16.266071
#sample estimates:
#mean in group F mean in group M
#      73.94444      68.77273

#p-value가 0.3513으로 유의수준 0.05보다 크므로 남녀에 따른 근로자 평균적응 점수가 같다는 귀무가설을 지지하게 된다.
```

- 221페이지 27번 문제

```
accident <- read.csv("d:/knou/27.csv")
accident

# 위치 전의사고수 후의사고수
#1 A 8 5
#2 B 12 3
#3 C 5 2
#4 D 4 1
#5 E 6 4
#6 F 3 2
#7 G 4 2
#8 H 3 4
#9 I 2 3
#10 J 6 5
#11 K 6 4
#12 L 9 3

t.test(accident$후의사고수, accident$전의사고수, alternative="less", mu=0, paired=T)

# Paired t-test

#data: accident$후의사고수 and accident$전의사고수
#t = -3.1154, df = 11, p-value = 0.004915
#alternative hypothesis: true difference in means is less than 0
#95 percent confidence interval:
# -Inf -1.058877
#sample estimates:
#mean of the differences
# -2.5

# p-value가 0.004915로 유의수준 0.05 보다 작은 값을 가지므로 가로등을 설치 한 이후의 사고감소가 유의하다는 것을 알 수 있다.
```

256페이지 21번 문제

```
#그룹에 따른 만족 평균 만족도가 다르다고 할 수 있는가를 유의수준 0.05로 검정하라!

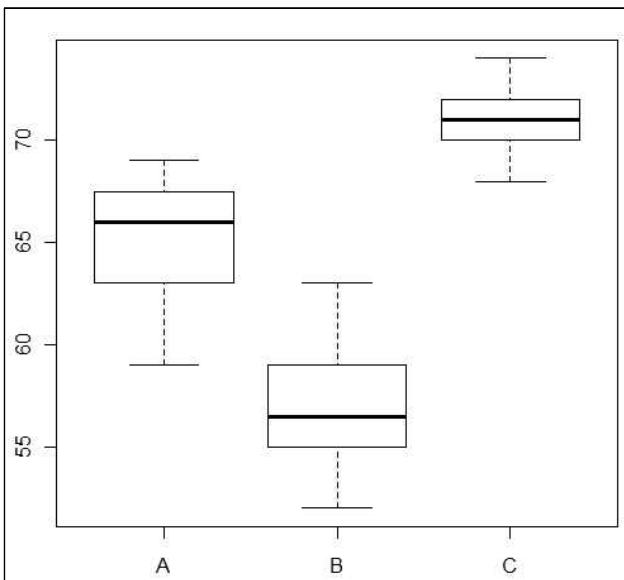
fullfill <- read.csv("d:/knou/21.csv")
```

fullfill

```
# 그룹 만족도
#1 A 69
#2 A 67
#3 A 65
#4 A 59
#5 A 68
#6 A 61
#7 A 66
#8 B 56
#9 B 63
#10 B 55
#11 B 59
#12 B 52
#13 B 57
#14 C 71
#15 C 72
#16 C 70
#17 C 68
#18 C 74
```

#그룹에 따른 박스플롯을 그렸을 때 어느정도 집단내의 데이터가 정규분포 형태를 띠고  
#있다는 것을 알 수 있으며 그룹간의 만족도가 차이가 있다는 것을 어느정도 짐작할 수 있다.

boxplot(만족도 ~ 그룹, data=fullfill)



```
result <- aov(만족도 ~ 그룹, data=fullfill)
```

```
summary(result)
```

```
#           Df Sum Sq Mean Sq F value    Pr(>F)
#그룹         2     546  273.000  23.808 2.216e-05 ***
#Residuals   15     172   11.467
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0
```

```
#세 그룹에 따른 평균 만족도가 같을 거라는 귀무가설은
#p-value가 2.216e-05로 0.05보다 작게 나와 기각되며 세그룹에 따른 평균 만족도가 다르다는
#대립가설을 지지하게 된다.
```

256 페이지 22번 문제

```
strong <- read.csv("d:/knou/22.csv")
```

```
strong
```

```
# 역청탄 타르피치 내압강도
```

#	A	B	Value
#1	A1	B1	79
#2	A2	B1	72
#3	A3	B1	51
#4	A4	B1	58
#5	A5	B1	68
#6	A1	B2	75
#7	A2	B2	66
#8	A3	B2	48
#9	A4	B2	56
#10	A5	B2	65
#11	A1	B3	69
#12	A2	B3	64
#13	A3	B3	44
#14	A4	B3	51
#15	A5	B3	61
#16	A1	B4	65
#17	A2	B4	62
#18	A3	B4	41
#19	A4	B4	45
#20	A5	B4	58

```
#그룹별 통계량
```

```
attach(strong)
```

```
tapply(내압강도, 역청탄, summary)
```

```
#$A1
```

#	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
#	65	68	72	72	76	79

```
#$A2
```

#	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
#	62.0	63.5	65.0	66.0	67.5	72.0

```
#$A3
```

#	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
#	41.00	43.25	46.00	46.00	48.75	51.00

```
#$A4
```

#	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
---	------	---------	--------	------	---------	------



```
# 45.0 49.5 53.5 52.5 56.5 58.0
```

```
#$A5
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
# 58.00 60.25 63.00 63.00 65.75 68.00
```

```
tapply(내압강도, 타르피치, summary)
```

```
#$B1
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
# 51.0 58.0 68.0 65.6 72.0 79.0
```

```
#$B2
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
# 48 56 65 62 66 75
```

```
#$B3
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
# 44.0 51.0 61.0 57.8 64.0 69.0
```

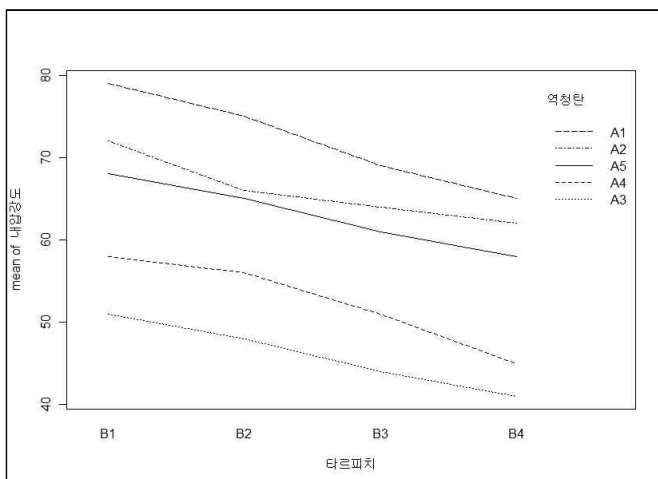
```
#$B4
```

```
# Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
# 41.0 45.0 58.0 54.2 62.0 65.0
```

#타르피치와 역청탄 각각의 명복값이 내압 강도 평균 값에 서로 영향을 주는지 확인해 본다.

```
interaction.plot(타르피치, 역청탄, 내압강도)
```



#그래프가 서로 겹치는 곳이 없는 것을 봐서 타르피치가 역청탄과 교호작용같은 것은 일으키지 않는 것을 알 수 있다.

```
result <- aov(내압강도 ~ 타르피치 + 역청탄)
```

```
summary(result)
```

```
#           Df Sum Sq Mean Sq F value    Pr(>F)
#타르피치    3  369.0  123.00   73.80 5.296e-08 ***
#역청탄      4 1764.8  441.20  264.72 1.373e-11 ***
#Residuals  12   20.0    1.67
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0
```

#타르피치, 역청탄에 모두 p-value가 0.05보다 적은 값을 가지므로 평균 내압강도는 이들의 변화에 따라 달라진다는 대립가설을 지지하게 된다.  
#그리고 평균 내압강도에 역청탄의 변화가 타르피치의 변화보다 더 많은 영향을 끼친다는 것을 p-value의 비교에 따라 알 수 있다.

298페이지 22번 문제

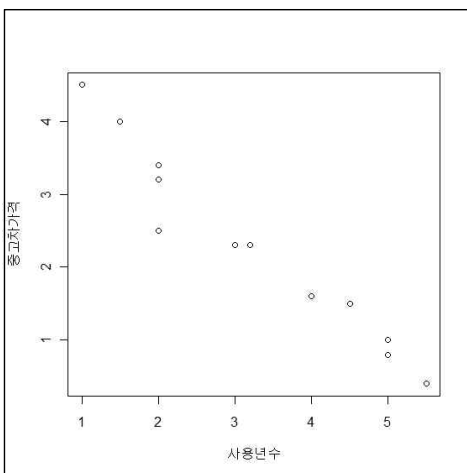
```
car <- read.csv("d:/knou/22-1.csv")
```

car

#	사용연수	중고차가격
#1	1.0	4.5
#2	1.5	4.0
#3	2.0	3.2
#4	2.0	3.4
#5	2.0	2.5
#6	3.0	2.3
#7	3.0	2.3
#8	3.2	2.3
#9	4.0	1.6
#10	4.5	1.5
#11	5.0	1.0
#12	5.0	0.8
#13	5.5	0.4

```
attach(car)
```

```
plot(중고차가격 ~ 사용연수)
```



#산점도 결과 사용연수와 중고차가격은 강한 음의 상관관계가 있음을 예상할 수 있다.

```
cor.test(중고차가격, 사용연수)
```

```
# Pearson's product-moment correlation
```

```
#data: 중고차가격 and 사용연수
```

```
#t = -13.5554, df = 11, p-value = 3.293e-08
```

```
#alternative hypothesis: true correlation is not equal to 0
```

```
#95 percent confidence interval:
```

```
# -0.9916199 -0.9043921
#sample estimates:
#      cor
#-0.9713481

#persrson correlation test결과 p-value는 3.293e-08로 0.01보다 작은 유의한 값이 나왔으며 게다가 상관계
수도 -0.9713481로 강한 음의 상관관계의 값이 도출되어 중고차가격과 사용년수사이에는 강한 음의 상관관
계가 있다고 할 수 있다.

out <- lm(사용년수 ~ 중고차가격)
summary(out)

#Call:
#lm(formula = 사용년수 ~ 중고차가격)

#Residuals:
#      Min        1Q    Median        3Q       Max
#-0.96728 -0.15698  0.06487  0.26906  0.37516

#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept)  5.86117     0.22063   26.57 2.50e-11 ***
#중고차가격  -1.15756     0.08539  -13.55 3.29e-08 ***
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Residual standard error: 0.367 on 11 degrees of freedom
#Multiple R-squared: 0.9435, Adjusted R-squared: 0.9384
#F-statistic: 183.7 on 1 and 11 DF, p-value: 3.293e-08

#회귀분석 결과 분석
# y절편   : 5.86117
# 기울기   : -1.15756 p-value : 3.29e-08
# Multiple R-squared: 0.9435
# Adjusted R-squared: 0.9384
# F-statistic: 183.7 (1 and 11 DF), p-value: 3.293e-08

#추정된 회귀식은 "사용년수 = -1.15*중고차가격 + 5.86" 이며 기울기의 p-value가 3.29e-08로 0.05및 0.01
보다 작어 유의한 수임을 알 수 있다.
#결정계수 R^2은 0.94로 회귀식이 관측값들을 잘 설명하고 있다는 것을 알 수 있다.
#변수가 하나인 관계로 F검정의 p-value는 기울기의 p-value와 같으며 또한 유의한 값을 알 수 있고 x변
수가 y변수를 충분히 잘 설명하고 있다는 것을 알 수 있다.
```

#다시 산점도와 회귀식을 그려 보았다.

```
plot(중고차가격 ~ 사용연수)
```

```
abline(out)
```

