

```

> ##### test -simulation
> N <- 100
> set.seed(1234)
> tX1 <- rnorm(N,0,1)
> tX2 <- rnorm(N,3,1)
> tX3 <- rgamma(N,1,3)
> tX4 <- sample(c(0,1),N,replace=T)
> tX5 <- sample(1:3,N,replace=T,prob=c(1,2,3)/6)
> tX6 <- rbinom(N,4,0.3)
>
> tX <- cbind(rep(1,length(tX1)),tX1,tX2,tX3,tX4,tX5,tX6)
> tbeta <- (0:6)/5
> ty <- as.vector(tX%*%tbeta)+rnorm(N,0,2)
>
> dat <- data.frame(y=ty, age=tX1, height=tX2, weight=tX3, smoking=tX4,
therapy=tX5, surgery=tX6)
>
> res <- MyReg(dat)

```

여기에 주어진 방식으로 자료를 생성하도록 한다. 따라서 결과는 동일하게 나와야 한다. 이와는 별도로 실제 자료를 이용하여 그 결과도 함께 리포트에 작성하도록 한다.

#### == ANALYSIS OF VARIANCE ==

Source	df	SS	MS	F	P-value
Regression	6	195.3	32.55	8.72	0 ***
Error	93	347.11	3.73		
Total	99	542.41			

Estimated error variance : 3.7323  
R-squares : 0.3601

필요한 통계량들을 자동으로 계산하도록 작성. 공식은 문제파일에 수록하였으니 참고할 것.

#### == PARAMETER ESTIMATES ==

	Estimate	Std.Error	t value	Pr(> t )
(Intercept)	-0.7866	1.182	-0.6655	0.5058
age	0.0198	0.1951	0.1012	0.9194
height	0.5986	0.1953	3.0653	0.0022 **
weight	-0.0803	0.6073	-0.1322	0.8948
smoking	0.2927	0.4105	0.713	0.4758
therapy	1.2801	0.2794	4.5809	0 ***
surgery	1.0944	0.2018	5.4219	0 ***

```

>
> str(res)
List of 10
 $ beta      : num [1:7] -0.7866 0.0198 0.5986 -0.0803 0.2927 ...
 $ sig2      : num 3.73
 $ pred      : num [1:100] 7.1
 $ residuals : num [1:100] 7.1
 $ stdresid  : num [1:100] 7.1
 $ SSR       : num 195
 $ SSE       : num 347
 $ F         : num 8.72

```

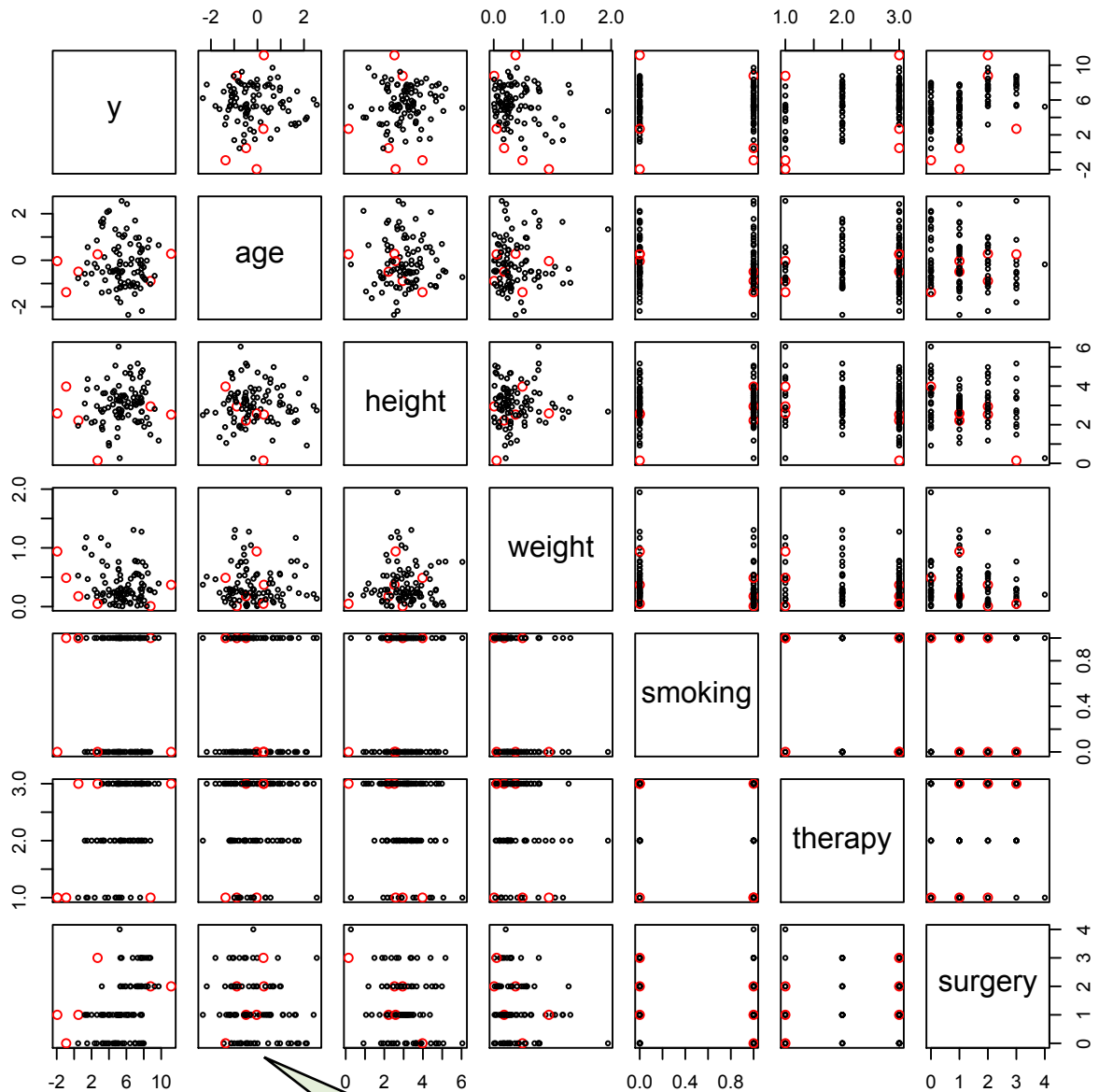
함수의 결과값(return)은 리스트 형식임을 알 수 있으며 리스트의 원소는 beta(회귀계수 추정량), sig2(분산추정량), pred(예측값), residuals(잔차), stdresid(표준화잔차), SSR(회귀제곱합), SSE(잔차제곱합), F(F값), P.value(회귀식에 대한 유의확률), dat(원자료)로 구성되어야 한다.

```

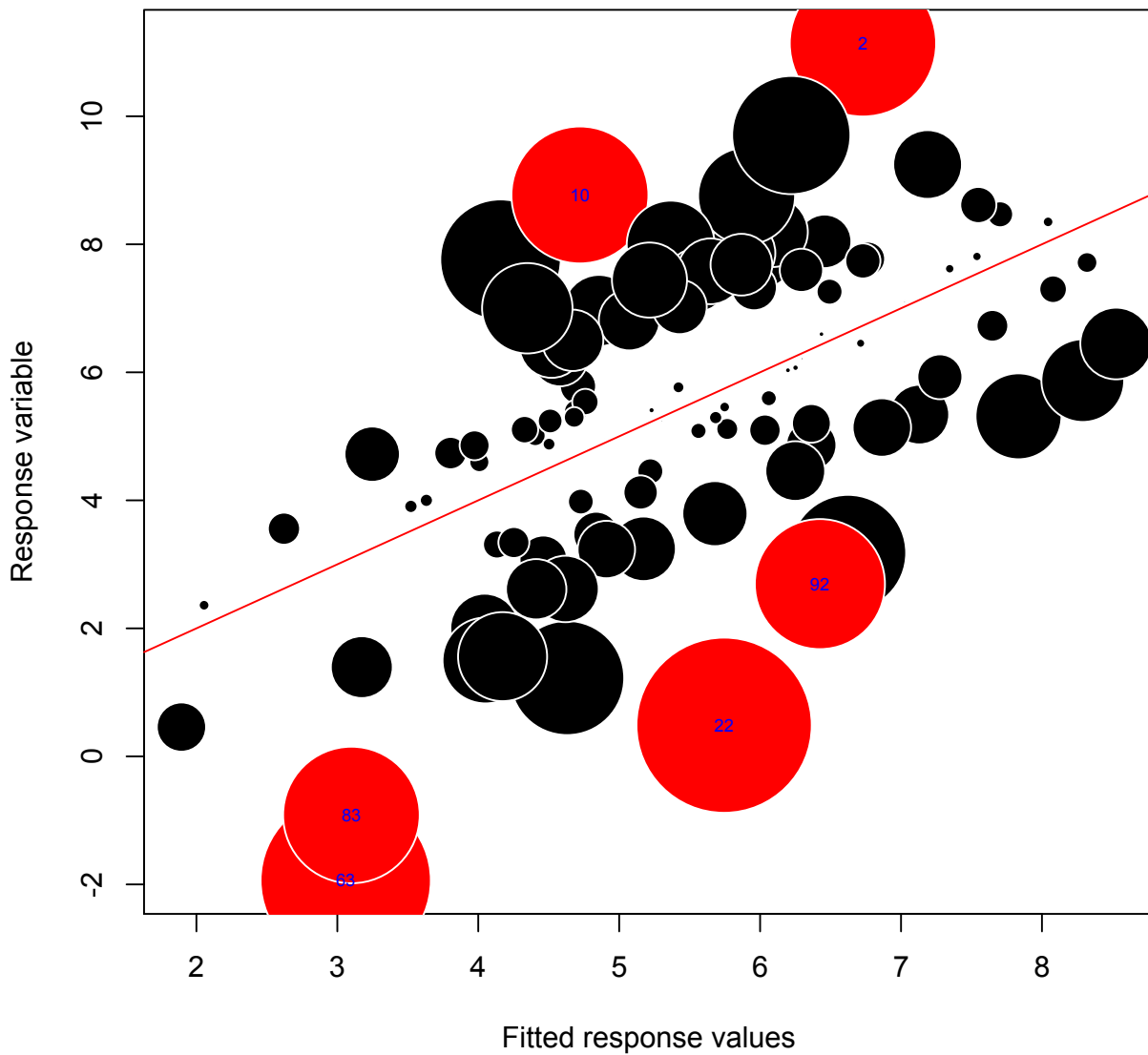
$ P.value : num 1.56e-07
$ dat     :'data.frame':  100 obs. of  7 variables:
..$ y      : num [1:100] 7.62 11.14 5.78 6.21 7.76 ...
..$ age    : num [1:100] -1.207 0.277 1.084 -2.346 0.429 ...
..$ height : num [1:100] 3.41 2.53 3.07 2.5 2.17 ...
..$ weight : num [1:100] 0.301 0.371 0.213 0.373 0.248 ...
..$ smoking: num [1:100] 1 0 0 1 0 0 0 1 0 1 ...
..$ therapy: int [1:100] 2 3 2 2 2 3 2 2 3 1 ...
..$ surgery: num [1:100] 3 2 1 1 1 1 1 1 3 2 ...

```

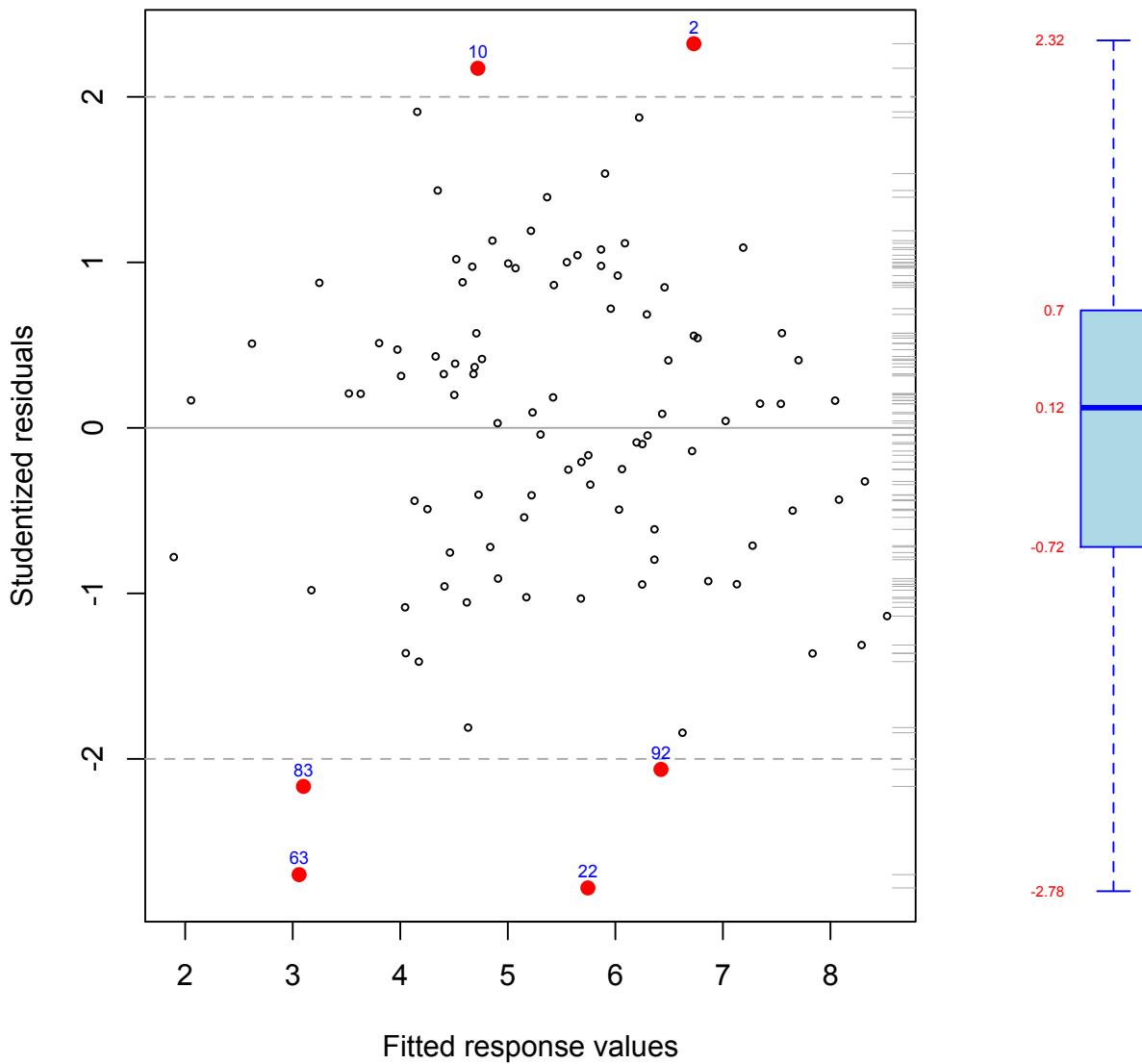
>



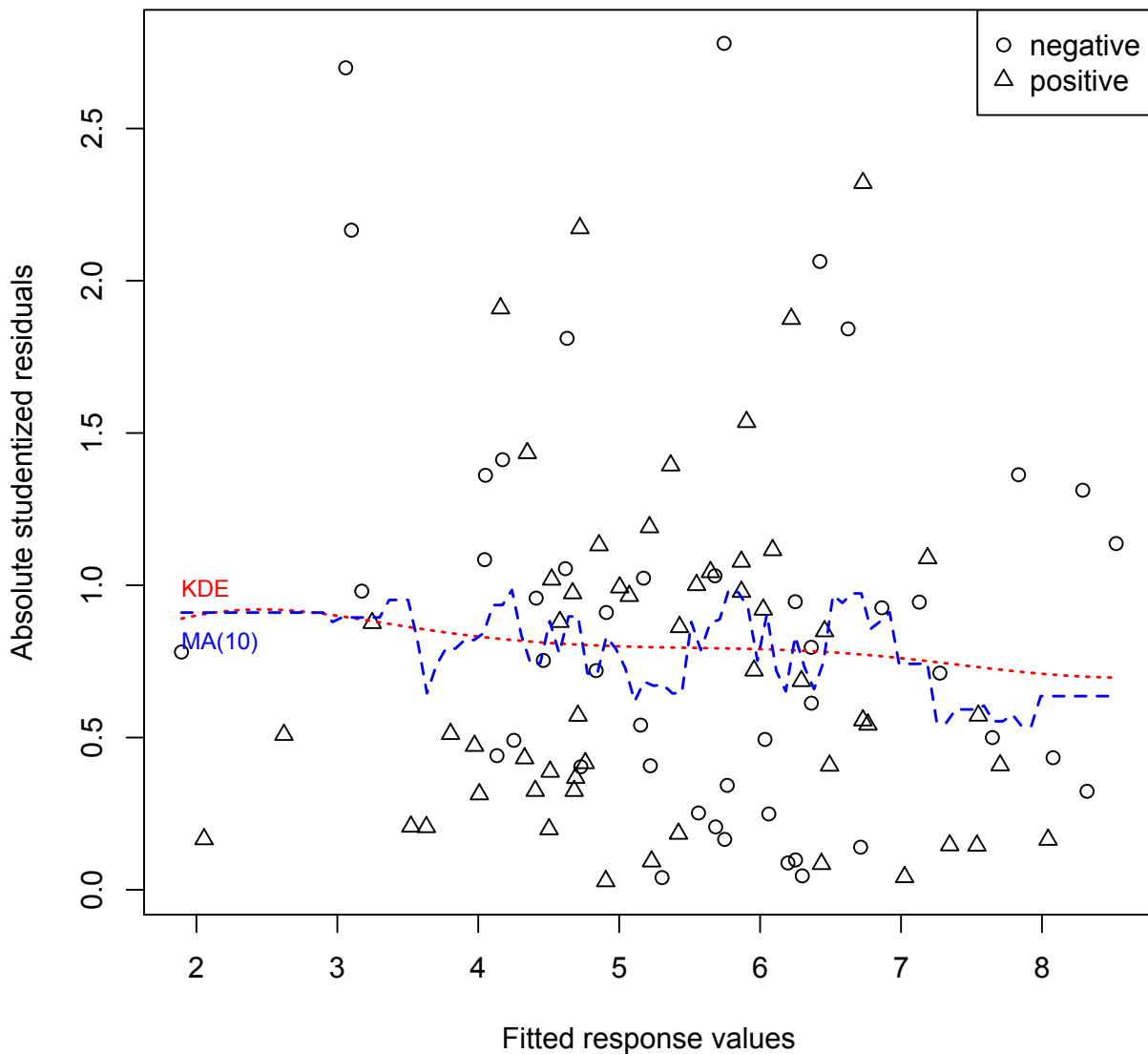
모든 자료들에 대한 pairwise scatterplot을 그리되, 회귀분석 결과에서 이상점(outlier)로 판단된 자료를 구분할 수 있도록 색상 및 모양, 또는 크기를 적절히 정하여 보기 좋도록 한다.



예측값(fitted response)에 대한 반응변수(y)의 산점도를 그린다. 각각의 점은 표준화잔차의 절대값의 크기에 비례하도록 하며, 이상점에 대해서는 색상을 다르게 하고 몇번째 관측치 인지를 원안에 표시하도록 한다. 가운데 직선은 절편 0, 기울기 1인 직선이다.



예측값(fitted response)에 대한 표준화잔차의 산점도로서 잔차그림을 그린다. 참조선이 -2, 0, 2에 그려져 있으며 표준화잔차의 절대값이 2이상인 경우에 이상점으로 판단하여 다른 모양으로 표시하고 해당 관측치의 번호를 표시한다. 표준화잔차의 분포를 보기 위해 우측에 표준화잔차에 대한 상자그림을 그리고 산점도의 우측 좌표축에 rug함수를 이용하여 표시를 한다.



표준화잔차의 분산의 변화를 보기 위해, 예측값 (fitted response)에 대한 표준화잔차절대값의 산점도로서 잔차그림을 그린다. 잔차의 부호(양, 음)에 따라 점의 모양을 다르게 그리도록 하며 범례(legend)를 두어 이해를 돕도록 한다. 전체적인 경향을 표시하기 위해 kernel density estimation(KDE) 및 moving average(MA)를 통하여 참조곡선을 그려 넣도록 한다. 자세한 방법은 문제파일을 참조하도록 한다.