

lsfit 함수

📖 lsfit 함수의 사용법

📖 두 변수 사이의 회귀식 사용

Loren ipsum dolor sit amet, ius an molestie facilisi erroribus, mutat natorum delectus ei vis. Has ornatus conclusionemque id, an videri maiestatis sit. In etqui praesent sit. An vel agan porro comprehensan, ad ludus constituto nea, et ius utroque scaevola assumaverit.

Via cu nodus nulla feugait, oratio facilisi ex usu, eilit vitae sea te. Ea fabulas accusamus dissentias sea, facete tacinates definitiones et per. Nihil dicant mediocrem pro eu, no mei nostro sensibus platonem. Qui id sunno perpetua neglegantur. Vel ipsum novum copiosae ut. Quo et liber detracto probatus. Nam augue scribentur an. Sea oporteat percipitur incidereat ab. Qui viris nemore an.

회귀분석은?

보기: 키와 몸무게 사이의 관계식 찾기

- + 이 관계식을 찾기 위해 100명의 키와 몸무게 조사한 결과
 - 첫 번째 사람의 키와 몸무게는 (x_1, y_1) , 두 번째 사람의 키와 몸무게는 (x_2, y_2) , ..., 100번째 사람의 키와 몸무게는 (x_{100}, y_{100}) 으로 자료가 짝으로 얻어짐
- + i 번째 사람의 자료는 (x_i, y_i) 로 표시
- + n 명을 조사한다면 n 번째 자료는 (x_n, y_n) 으로 표현
- ? n 개의 자료가 있을 때 두 변수의 관계식을 찾는 회귀분석 문제를 R-언어에서 해결하는 방법을 알아보자.



lsfit 함수의 사용법



두 변수에 대한 자료 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 가 있을 경우



이 자료들 사이의 관계를 직선의 방정식 $y = b_0 + b_1x$ 로 표현하고자 b_0, b_1 을 얻을 때 사용할 할 수 있는 함수로 lsfit 함수가 있음



사용함수

```
lsfit(x, y, intercept = TRUE, ...)
```



매개변수

- x : 직선 $y = b_0 + b_1x$ 에서 x 에 사용할 변수(설명변수; 독립변수)
- y : 직선 $y = b_0 + b_1x$ 에서 y 에 사용할 변수(반응변수; 종속변수)
- `intercept` : 직선 $y = b_0 + b_1x$ 에서 강제로 y -절편 b_0 를 0으로 할지 설정함
 - TRUE가 기본값으로 0으로 설정하지 않으며, F이면 b_0 가 0으로 설정됨



두 변수 사이의 회귀식 사용



예시



자료

BMI 자료를 사용하여 키와 몸무게의 관계식(회귀식)을 찾아보자.

사용할 자료

```
> BMI <- read.table(url("http://jupiter.hallym.ac.kr/ftpdata/data/bmi.txt"),  
  col.names=c("height", "weight", "year", "religion", "gender", "marriage"))
```



내용 : 2000년, 177명에 대한 조사 결과

- 키, 몸무게, 출생년도
- 종교(Bu=불교, C1=개신교, C2=가톨릭, No=없음)
- 성별(F=여자, M=남자)
- 결혼여부(N=미혼, Y=기혼)



두 변수 사이의 회귀식 사용



예시

+ 사용함수

```
> lsfit(BMI$weight, BMI$height)
```

```
$coefficients
```

```
Intercept      X
```

```
135.6921380 0.4938178
```

+ 관계식

$$y = 135.7 + 0.49x$$



두 변수 사이의 회귀식 사용



예시

+ 회귀식의 적용 1 : 몸무게 50(kg)인 경우 키(cm) 예측

$$> 135.7 + 50 * 0.49$$

[1] 160.2

+ 회귀식의 적용 2 : 몸무게 60(kg), 70(kg)인 경우 키(cm) 예측

$$> 135.7 + 60 * 0.49$$

[1] 165.1

$$> 135.7 + 70 * 0.49$$

[1] 170





두 변수 사이의 회귀식 사용



예시

+ 회귀식의 적용 3 : 여러 개의 값에 대해서 한 번에 알고 싶은 경우

```
> reg.coef <- lsfit(BMI$weight, BMI$height)$coef  
> x <- seq(45, 80, by=5)  
> y <- reg.coef[1] + reg.coef[2] * x  
> yhat <- data.frame(x,y)  
> yhat
```

	x	y
1	45	157.9139
2	50	160.3830
3	55	162.8521
4	60	165.3212
5	65	167.7903
6	70	170.2594
7	75	172.7285
8	80	175.1976

* 몸무게 45부터 80까지 5 단위에 대해서 키의 예측값을 얻을 수 있음

Im 함수

📖 Im 함수의 사용법

📖 일반적인 선형모형에 대한 식 사용

Lorem ipsum dolor sit amet, ius an molestie facilisi erroribus, mutat natorum delectus ei vis. Has ornatus conclusionemque id, an videri molestatis sit. In etqui praesent sit. An vel agan porro comprehensan, ad ludus constituto nea, et ius utroque scaevola assuaverit.

Vis cu nodus nulla feugait, oratio facilisi in usu, eili vitae sea te. Ea fabulas accusamus dissonantia sea, facete tacinates definitiones et per. Nihil dicant mediocrem pro eu, no mei nostro sensibus platonem. Qui id sunno perpetua neglegentur. Vel ipsum novum copiosae ut. Quo et liber detracto probatus. Nam augue scribentur an. Sea oporteat percipitur incidere et. Qui viris nemore an.



lm 함수의 사용법



 회귀분석에 상용할 수 있는 또 다른 R 함수로 lm 함수가 있음(linear model=lm)

 사용함수


```
lm(formula, data, ...)
```

 매개변수

➤ formula : $y \sim x$ 형태로 설정하며 $y = b_0 + b_1x$ 의 기울기와 절편을 얻음

- ~의 왼쪽엔 y, 오른쪽에 x를 설정함

➤ data : x, y가 데이터 프레임의 변수 이름일 때 데이터 프레임의 이름을 설정함

 **주의** : lm 함수의 formula 설정은 y가 먼저 나오지만 lsfit 함수는 lsfit(x,y)로 x가 먼저 나오므로 착오가 없어야 함



일반적인 선형모형에 대한 식 사용



예시 1



BMI 자료에서 키와 몸무게의 관계

lm 함수를 사용하여 lsfit과 같은 결과를 얻는다.

```
> lm(height ~ weight, data=BMI)
```

Call:

```
lm(formula = height ~ weight, data = BMI)
```

Coefficients:

(Intercept)	weight
135.6921	0.4938



일반적인 선형모형에 대한 식 사용



예시 2



산점도 회귀직선 그리기

가상자료로 lm 함수를 사용하여 산점도를 설명하는 직선의 식을 얻어보자.

```
x <- seq(1,10)
y <- jitter(2*x+1)
lm(y ~ x)
```

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept) x
0.7396 2.0313

*** $y = 0.793 + 2.0313x$ 로 산점도를 설명하는 직선 얻음**

- 이 결과는 jitter 함수를 사용하기 전의 자료 $y = 1 + 2x$ 와 근접할 결과임
- jitter 함수는 난수를 발생하므로 매번 다른 결과를 얻음에 유의해야 함

plot 함수와 abline 함수

 abline 함수


 plot과 abline에 대한 적용

Lorem ipsum dolor sit amet, ius an molestie facilisi erroribus, mutat nalerum delectus ei vis. Has ornatus conclusionemque id, an vide molestatis sit. In etqui praesent sit. An vel agan porro comprehensan, ad ludus constituto nea, et ius utroque scaevola assuaverit.

Via cu nodus nulla feugait, oratio facilisi ex usu, eilit vitae sea te. Ea fabulas accusamus dissentias sea, facete tacinates definitiones et per. Nihil dicant mediocram pro eu, no mei nostro sensibus platonem. Qui id sunno perpetua neglegentur. Vel ipsun novum copiosae ut. Quo et liber detracto probatus. Nam augue scriben- tur an. Sea oporteat percipitur incidereat ab- Qui viris nemore an.

abline 함수



 y-절편이 a이고 기울기가 b인 직선을 기존 그래프에 추가하는 함수

 사용함수

```
abline(a = NULL, b = NULL, h = NULL, v = NULL, coef = NULL, ...)
```

매개변수

- a, b : 그릴 직선의 y 절편 및 기울기
- h : 수평선을 그릴 때 y값만 설정하며 이 때 y값을 h에 설정함
- v : 수직선을 그릴 때 x 축의 값
- coef : 위의 a, b값을 벡터로 설정할 때 사용함

- coef=c(a, b)





plot과 abline에 대한 적용



예시 1



자료

```
x <- seq(1,10)  
y <- jitter(2*x+1)
```

에 대한 산점도에 겹쳐 그린 회귀직선 그림은 다음 두 명령의 결과이다.

```
plot(x,y)  
abline(lm(y ~ x)$coef)
```

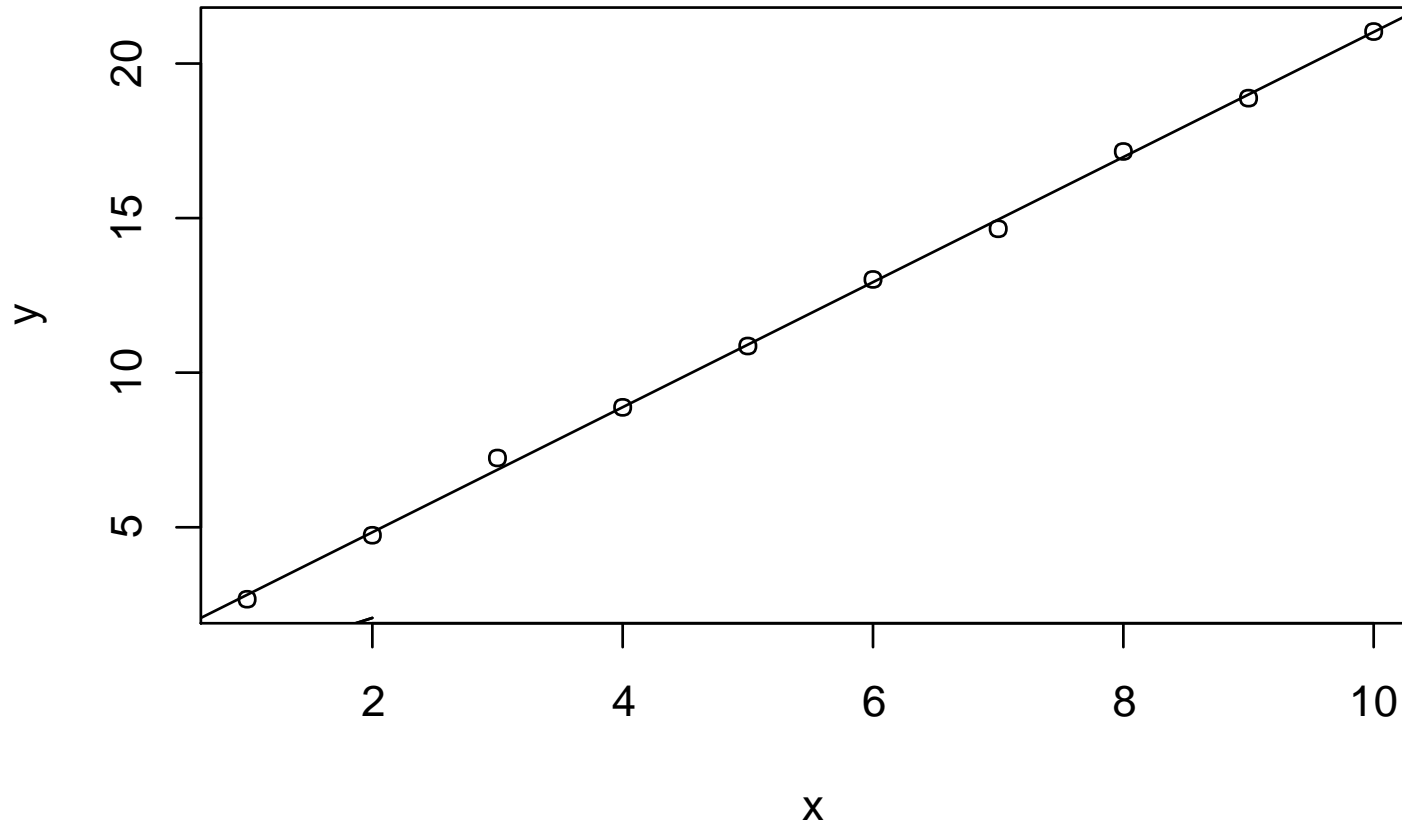


일반적인 선형모형에 대한 식 사용



예시 1

 자료 x, y 의 산점도와 산점도 위에 겹쳐 그린 회귀직선 $y = 0.793 + 2.0313x$





plot과 abline에 대한 적용



예시 2



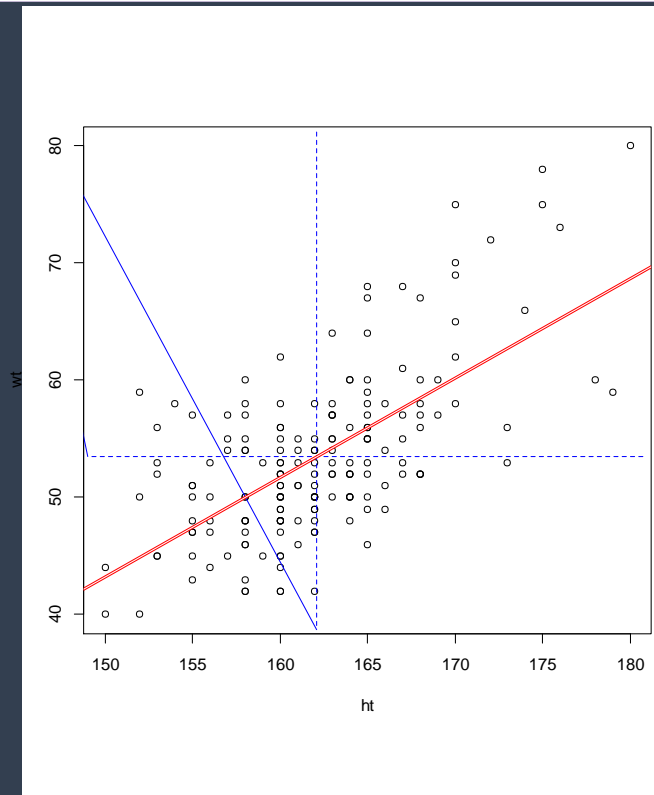
자료

abline 함수로 직선을 추가하자.

```

abline.test <- function() {
  ht <- BMI$height
  wt <- BMI$weight
  plot(ht, wt)
  abline(v=mean(ht), lty=2,
        col="blue")
  abline(h=mean(wt), lty=2,
        col="blue")
  abline(lsfht, wt), lwd=3,
        col="red")
  abline(lsfht, wt)$coef,
        lw=1, col="white")
} # end function

```



- 먼저 `plot(ht, wt)`로 산점도를 그림
- `abline(v=mean(ht))`로 ht의 평균을 지나는 수직선을 그림
- `abline(h=mean(wt))`로 wt의 평균을 지나는 수평선을 그림
- `lsfit(ht, wt)$coef`로 ht와 wt의 관계를 잘 설명하는 직선의 방정식을 얻어 이를 `abline(lsfht, wt)$coef`로 빨간 색의 선으로 그림

