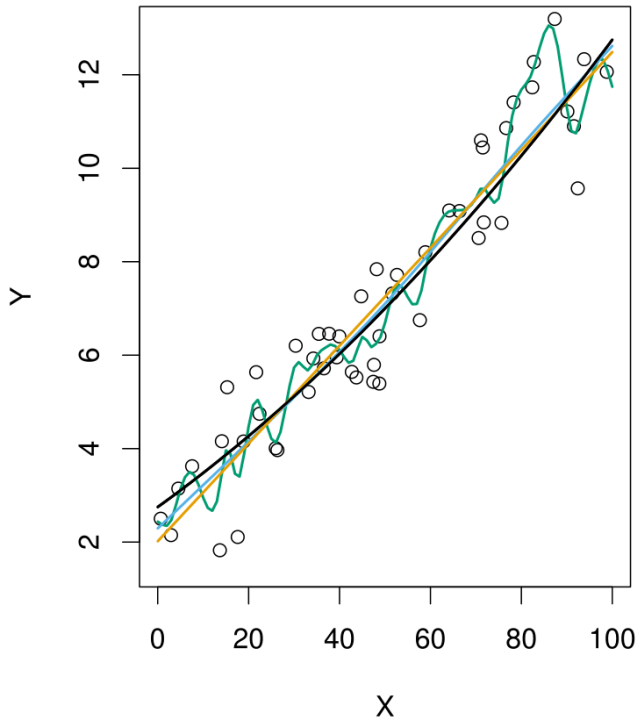
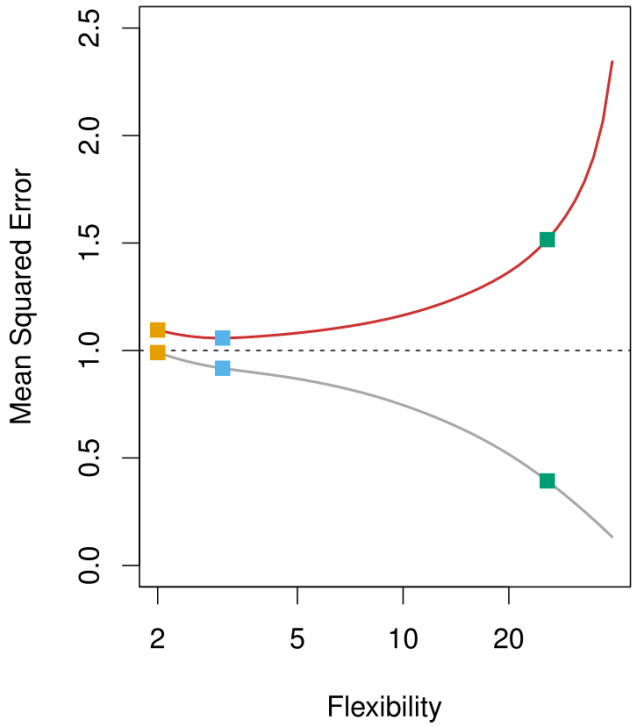
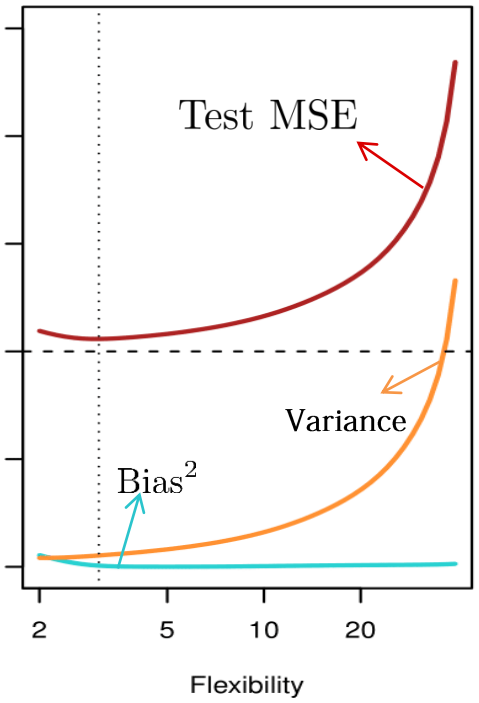
$$E\left(Y - \hat{f}(X)\right)^2 = \mathrm{Var}\left(\hat{f}(X)\right) + \left[E\left(\hat{f}(X)\right) - f(X)\right]^2 + \mathrm{Var}(\epsilon)$$

Test MSE | Variance | Bias^2 | Irreducible error







단순회귀의 Unbiasedness**를**
눈으로 직접 확인해보자!!

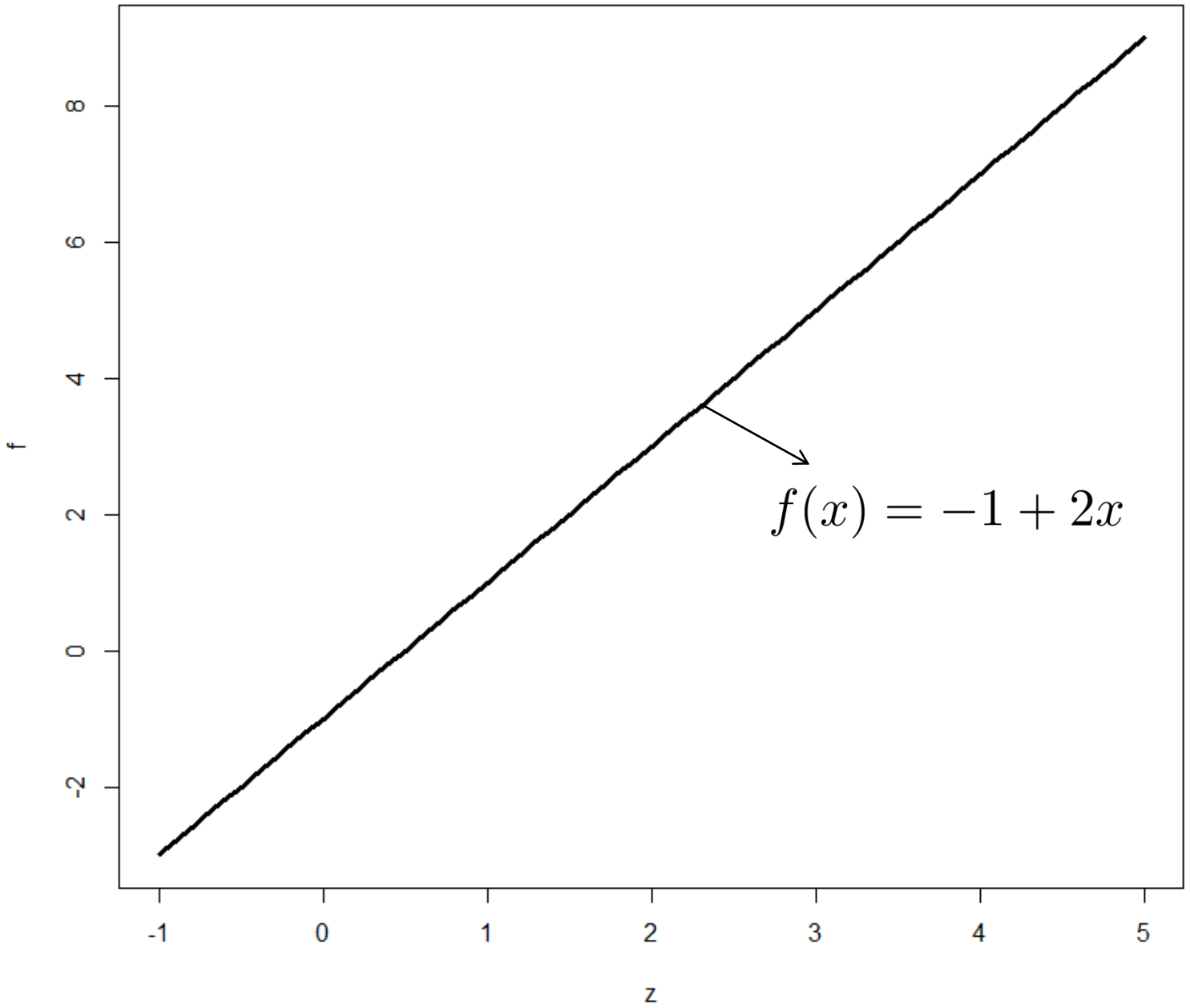
회귀분석의 기본 가정

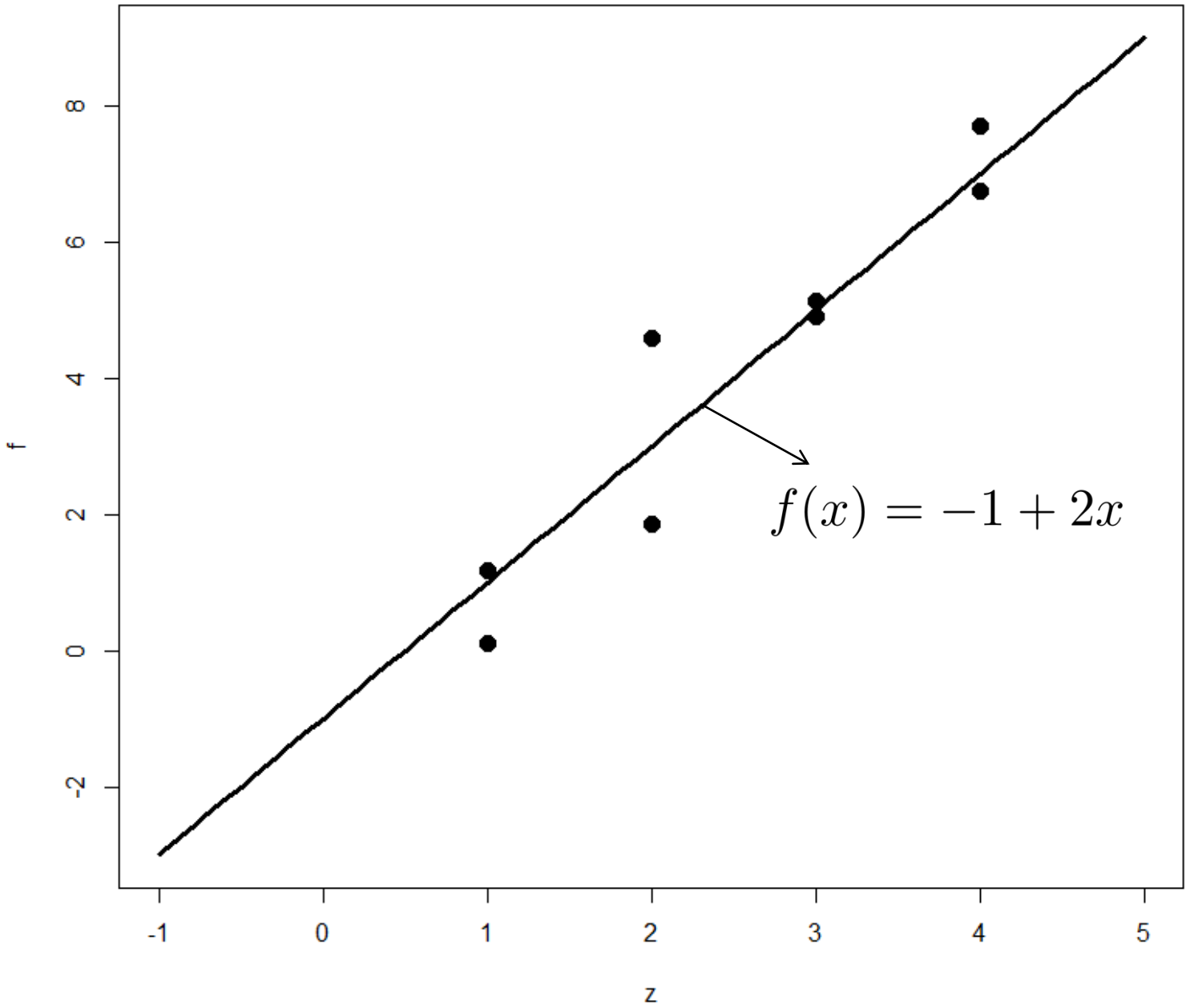
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, \cdots, n$$

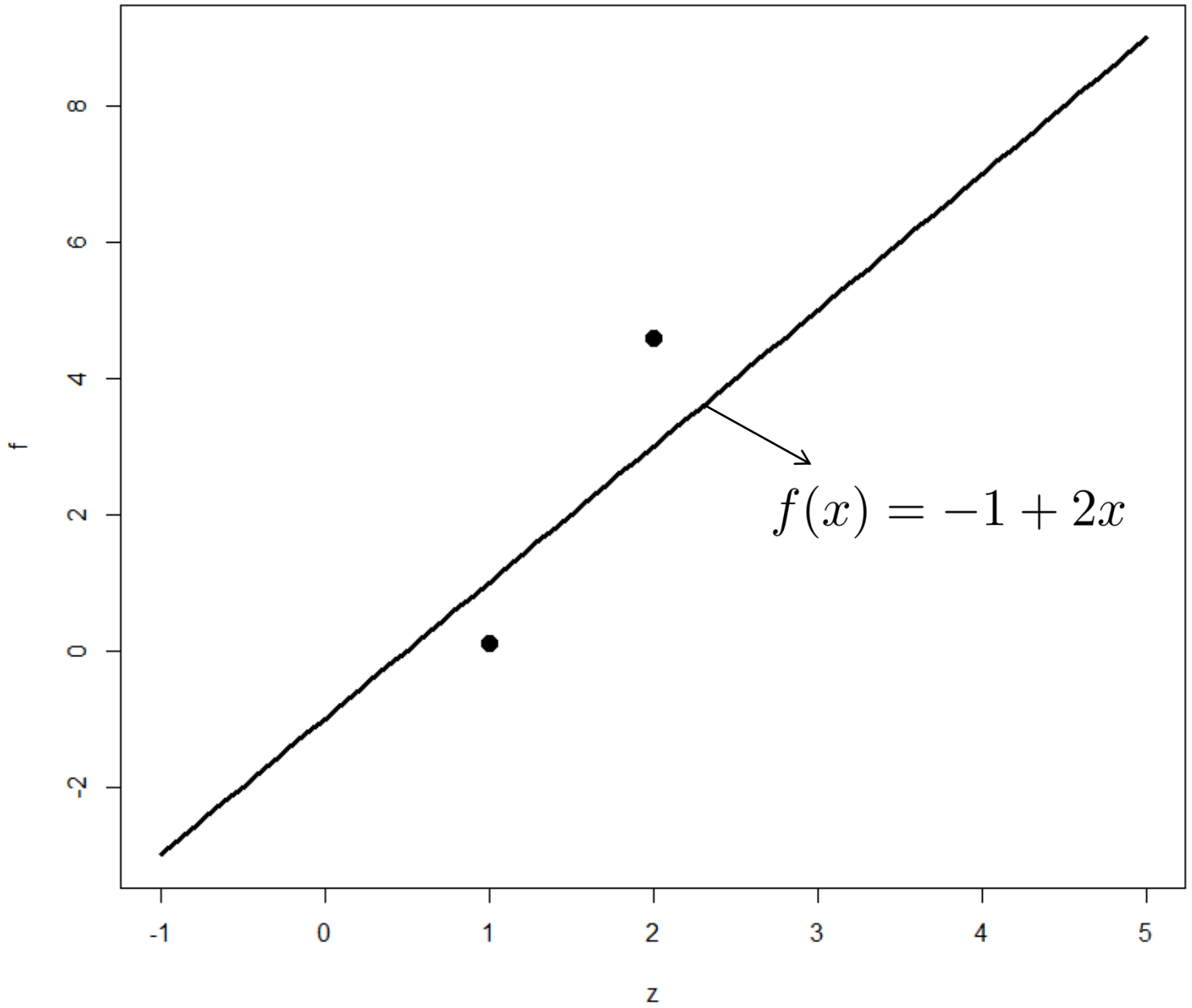
1. 오차항의 평균은 0이고 분산은 σ^2이다.
   즉, $E(\epsilon_i) = 0, \quad Var(\epsilon_i) = \sigma^2$

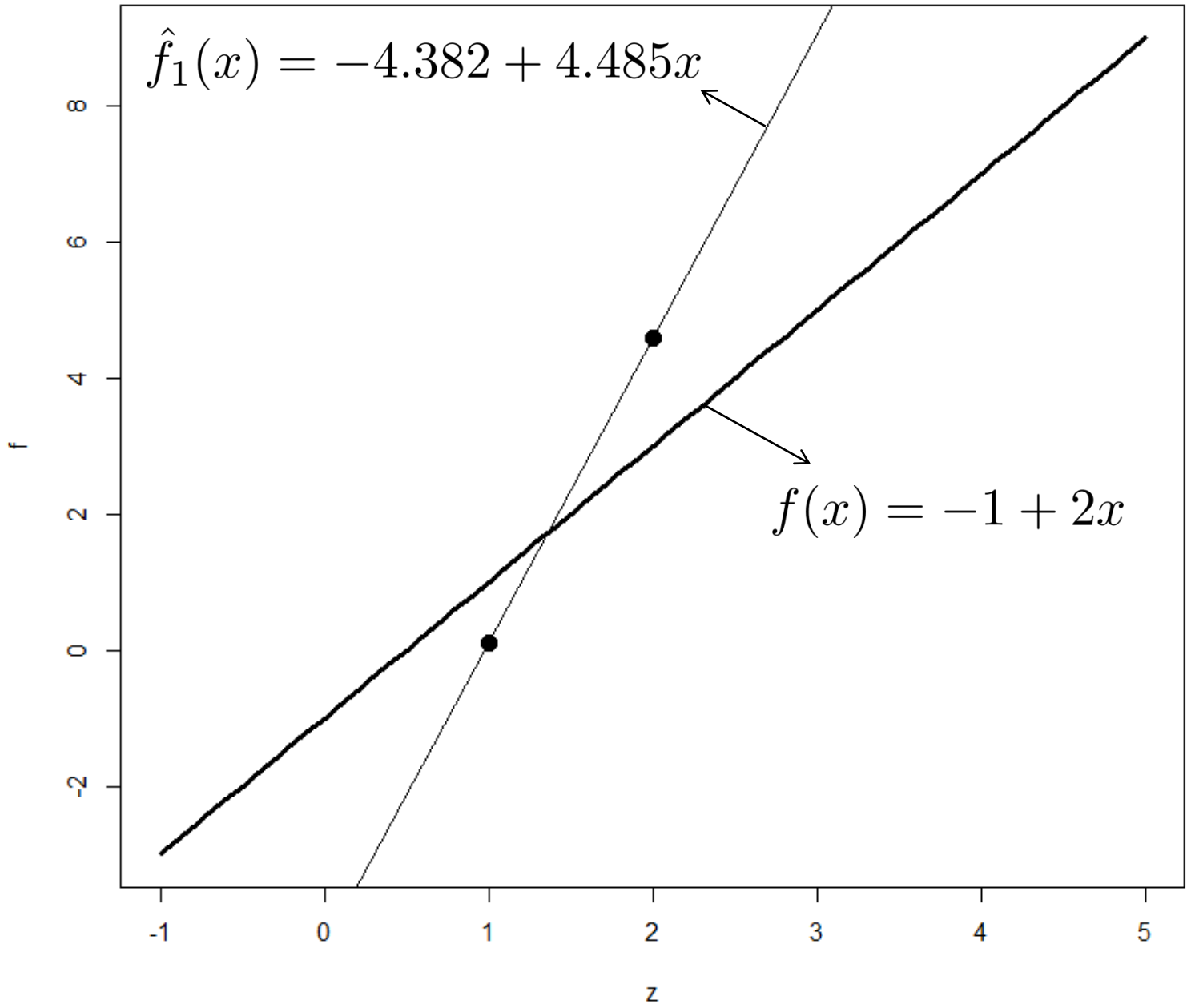
2. 오차항들은 서로 독립이다.

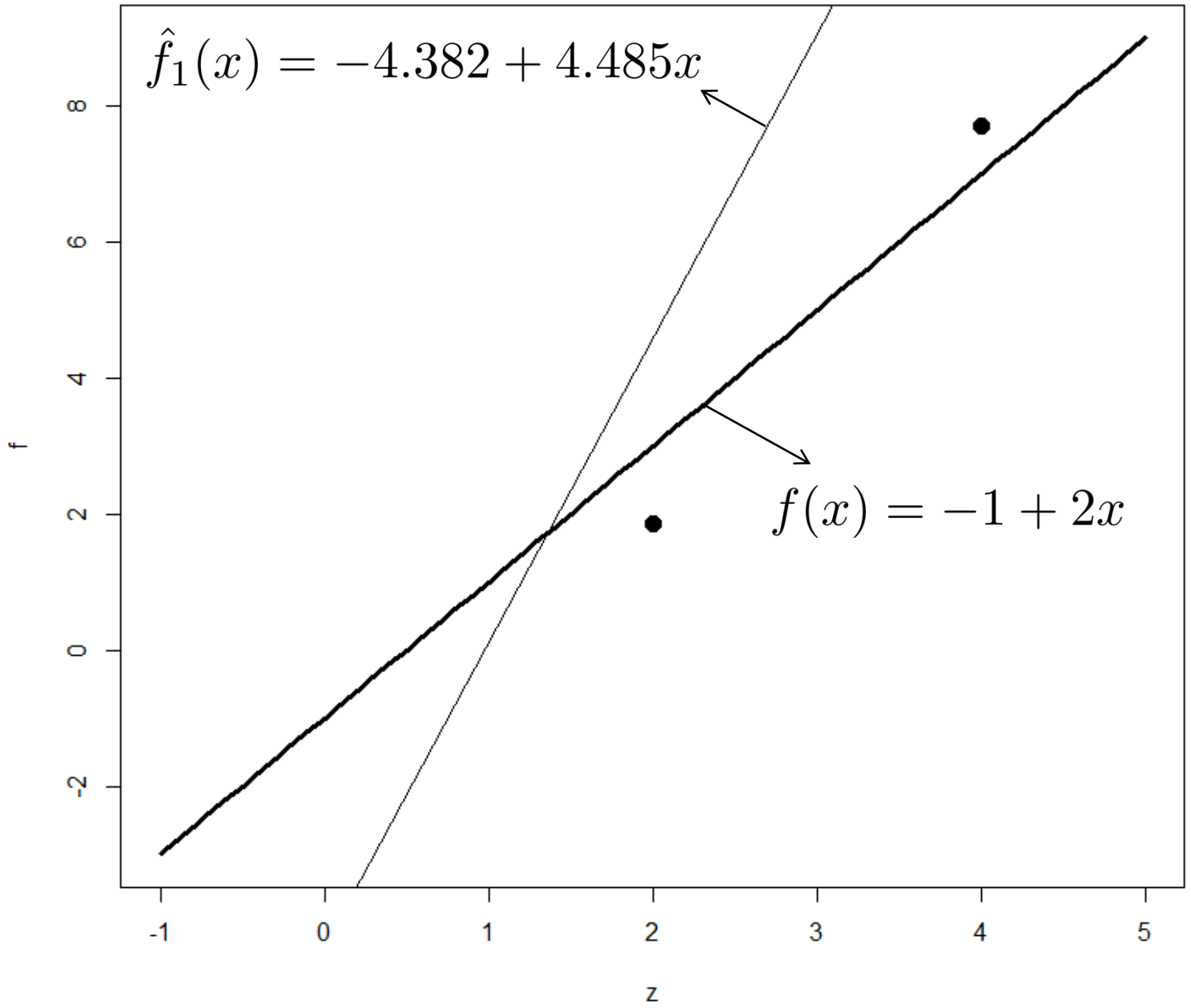
3. 오차항은 정규분포를 따른다: $\epsilon_i \sim N(0, \sigma^2)$

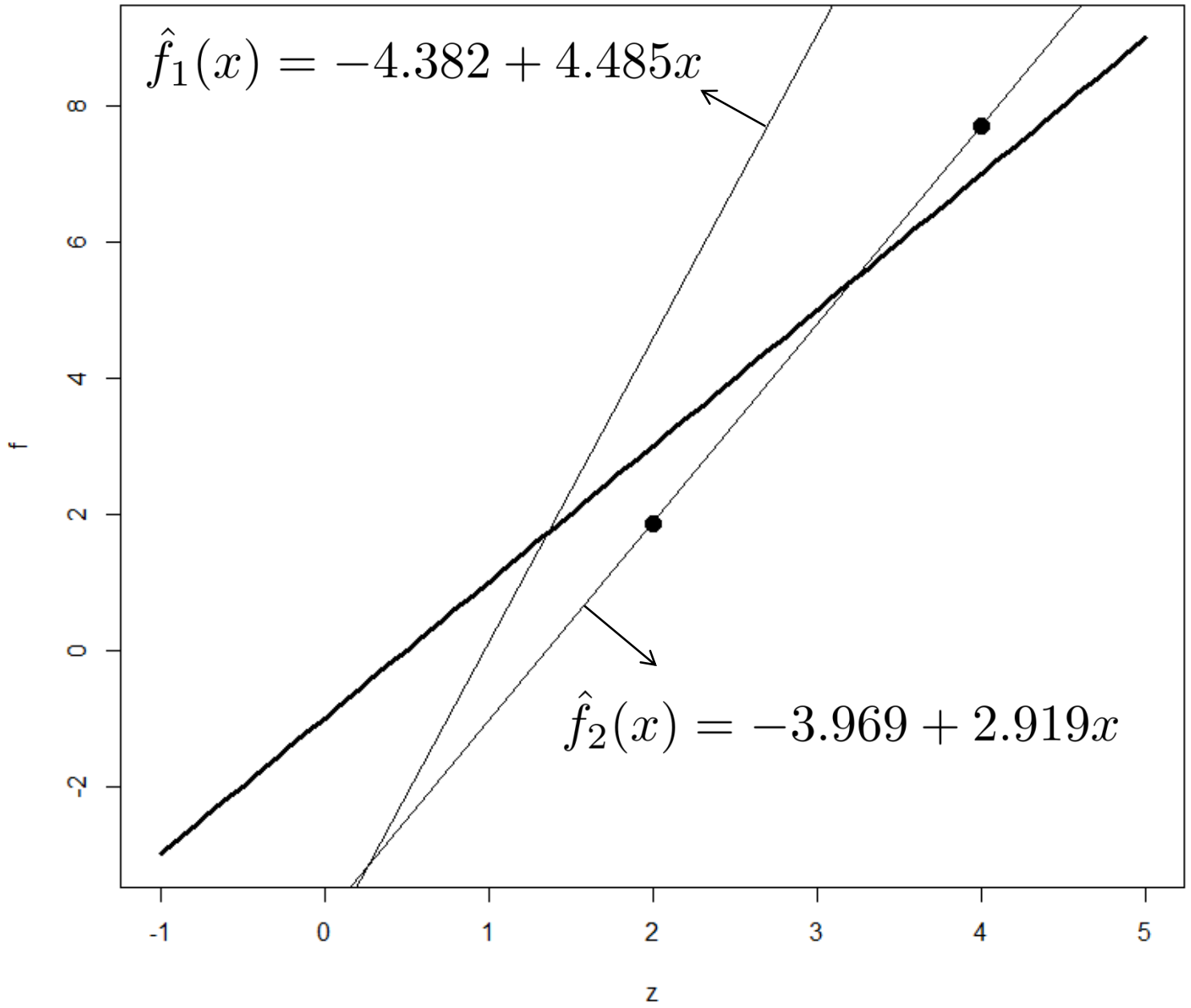

f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$f(x) = -1 + 2x$


f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$f(x) = -1 + 2x$


f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$f(x) = -1 + 2x$


$\hat{f}_1(x) = -4.382 + 4.485x$
$f(x) = -1 + 2x$
f
z
8
6
4
2
0
-2
-1
0
1
2
3
4
5

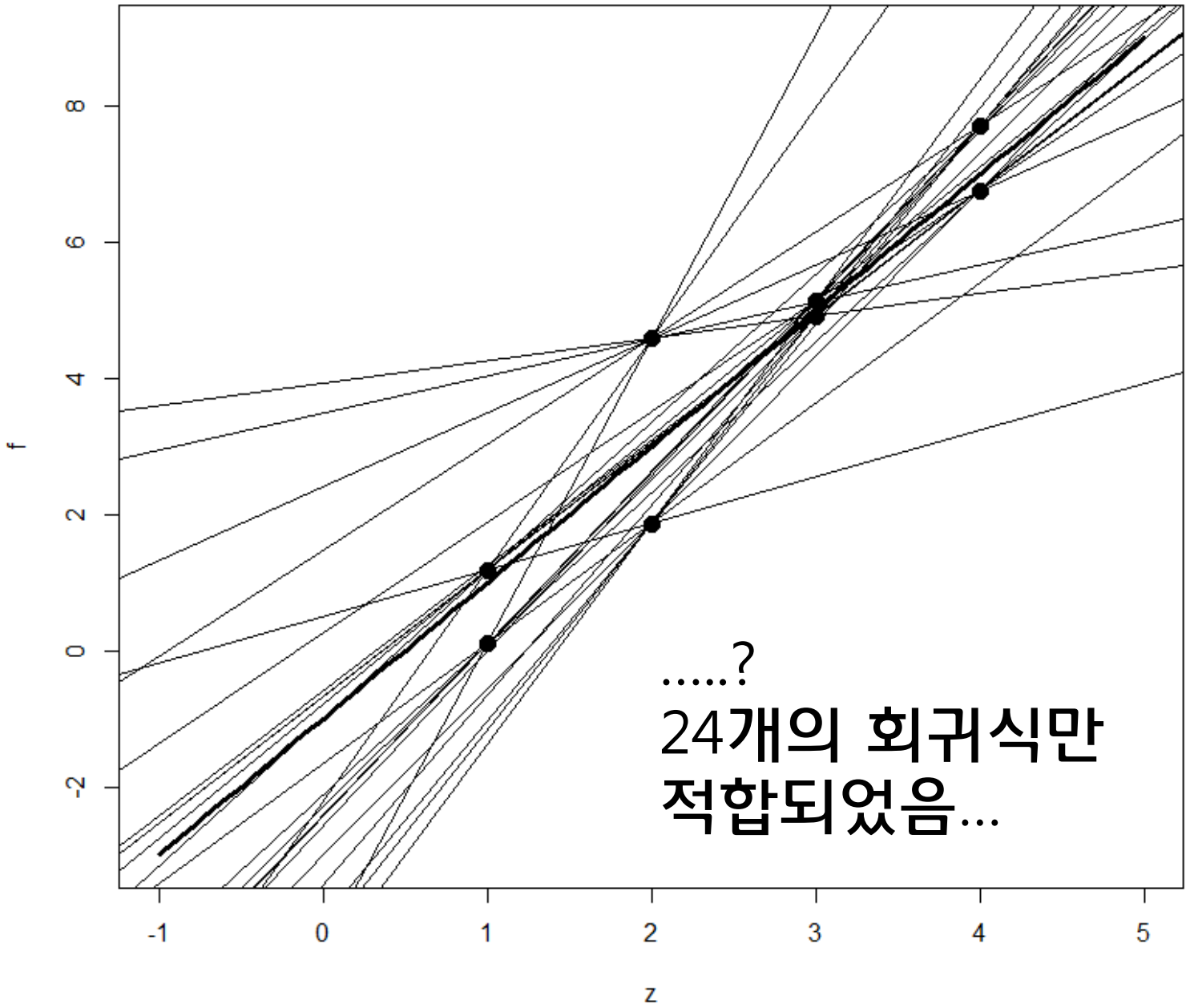

$\hat{f}_1(x) = -4.382 + 4.485x$
$f(x) = -1 + 2x$
f
z
8
6
4
2
0
-2
-1
0
1
2
3
4
5


$\hat{f}_1(x) = -4.382 + 4.485x$
$\hat{f}_2(x) = -3.969 + 2.919x$
f
z
8
6
4
2
0
-2
-1
0
1
2
3
4
5

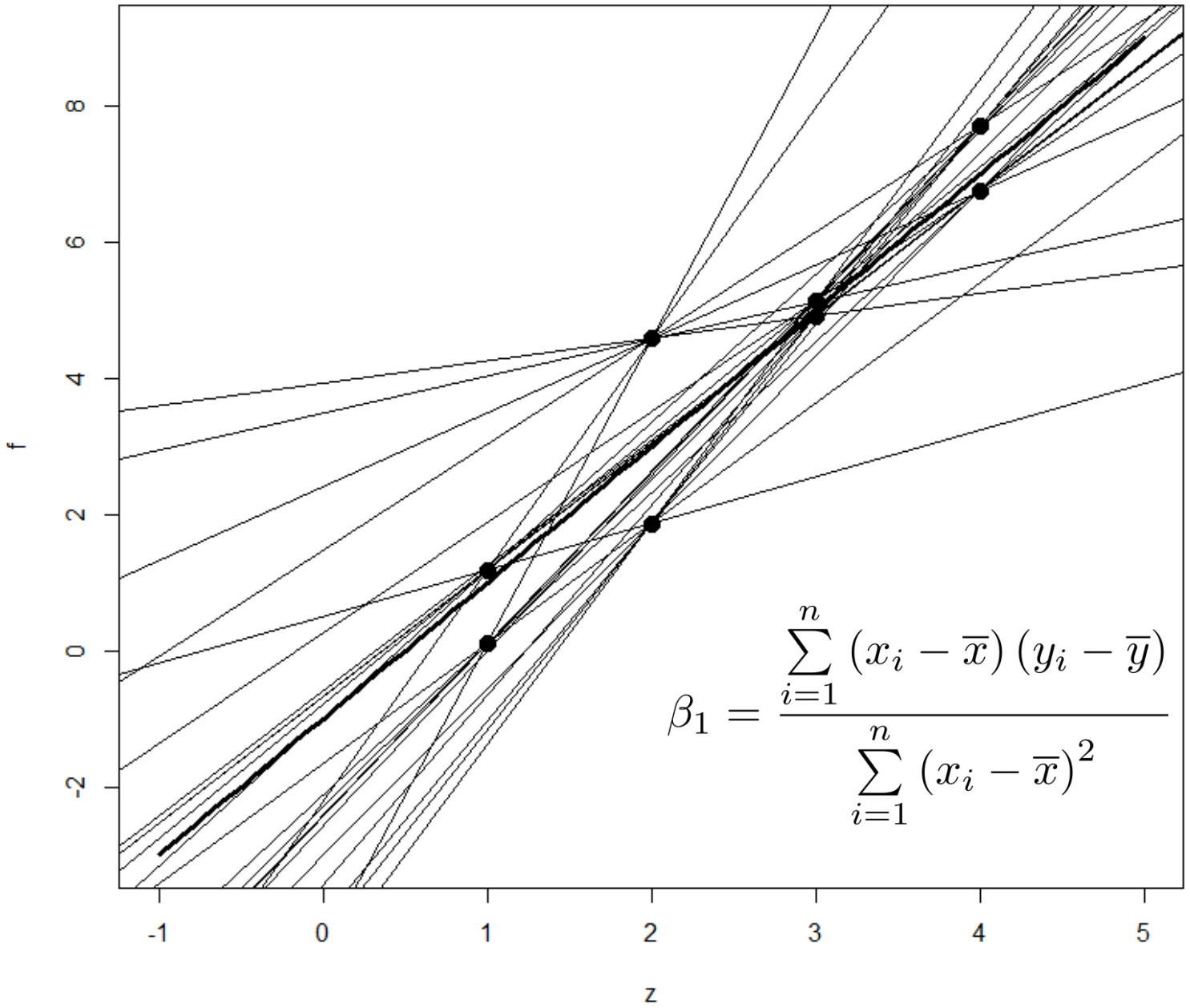
8C2 = 28개의 모든 가능한 회귀직선을 적합하여 평균내면 되겠구나!

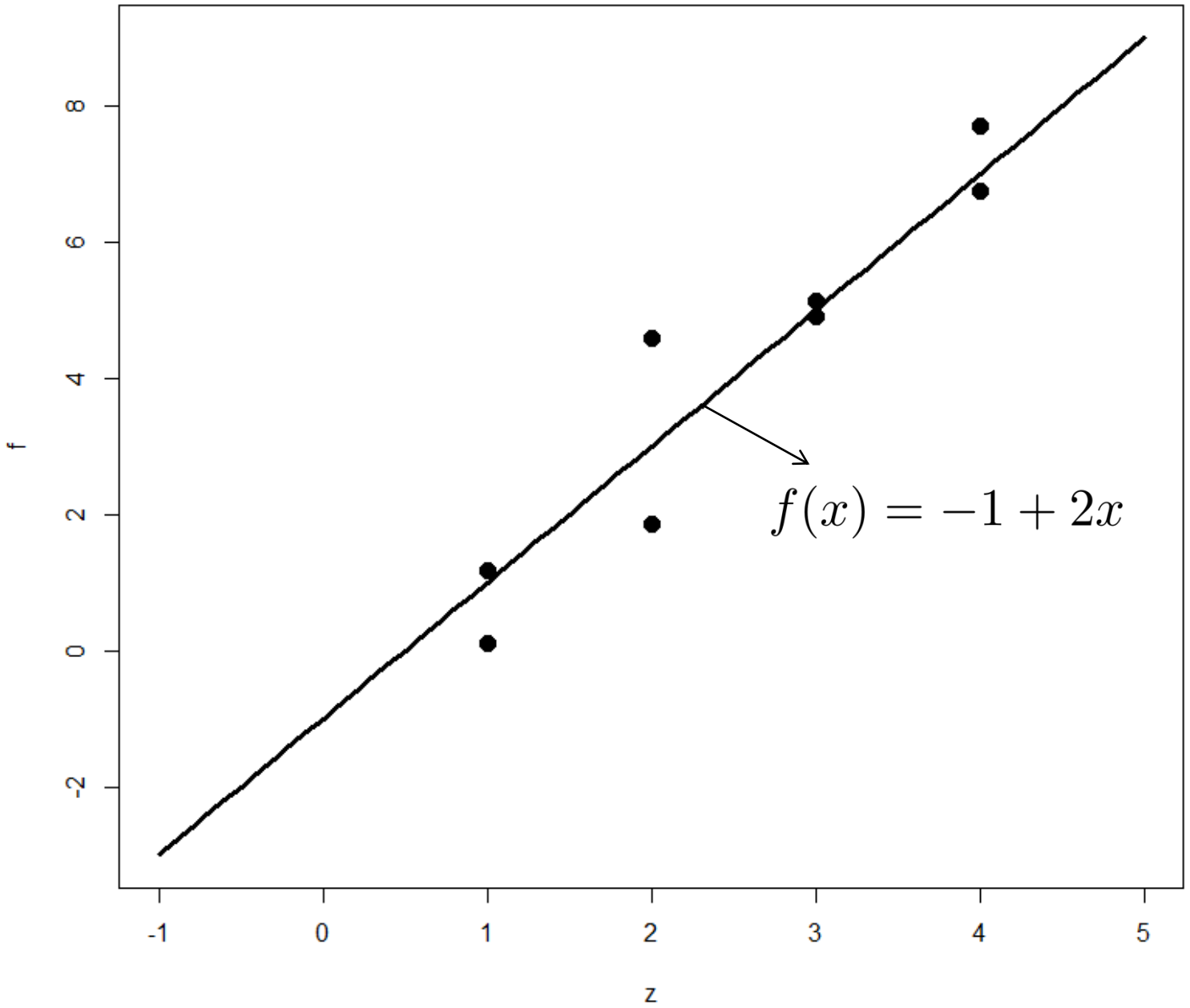
$$\frac{\hat{f}_1(x) + \hat{f}_2(x) + \cdots + \hat{f}_{28}(x)}{28} = -1 + 2x$$

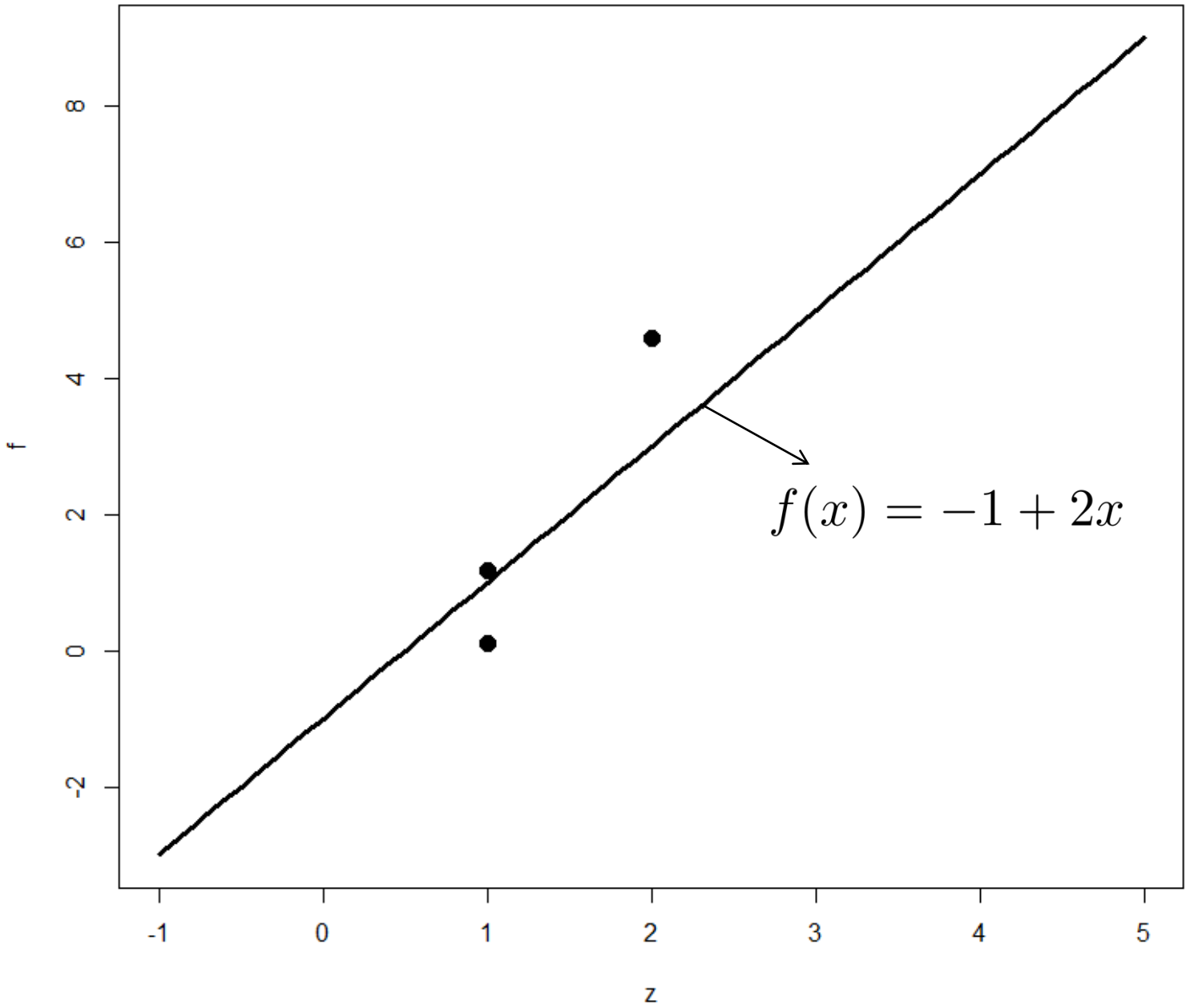
$$= f(x)$$

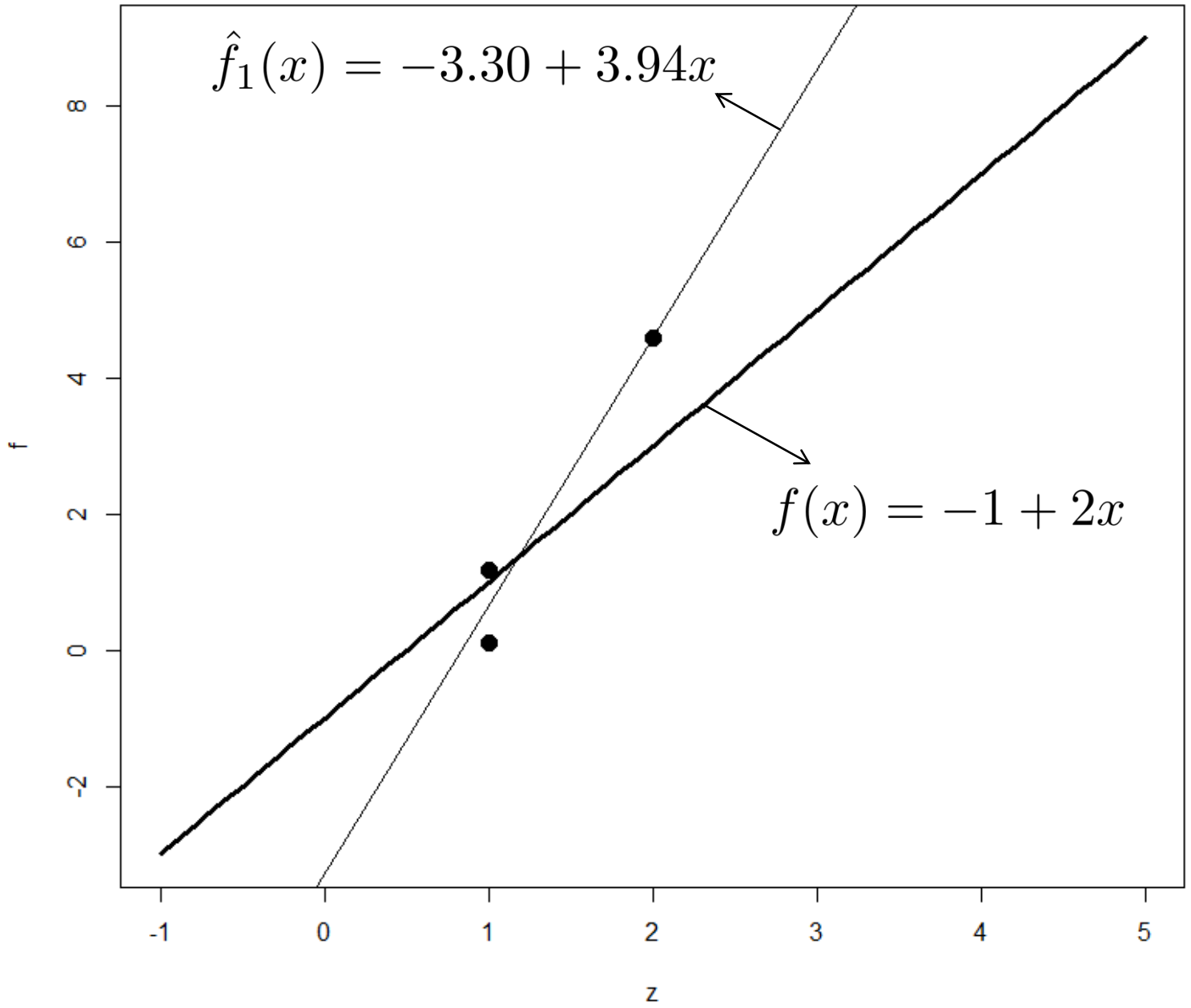

f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
.....?
24개의 회귀식만
적합되었음…

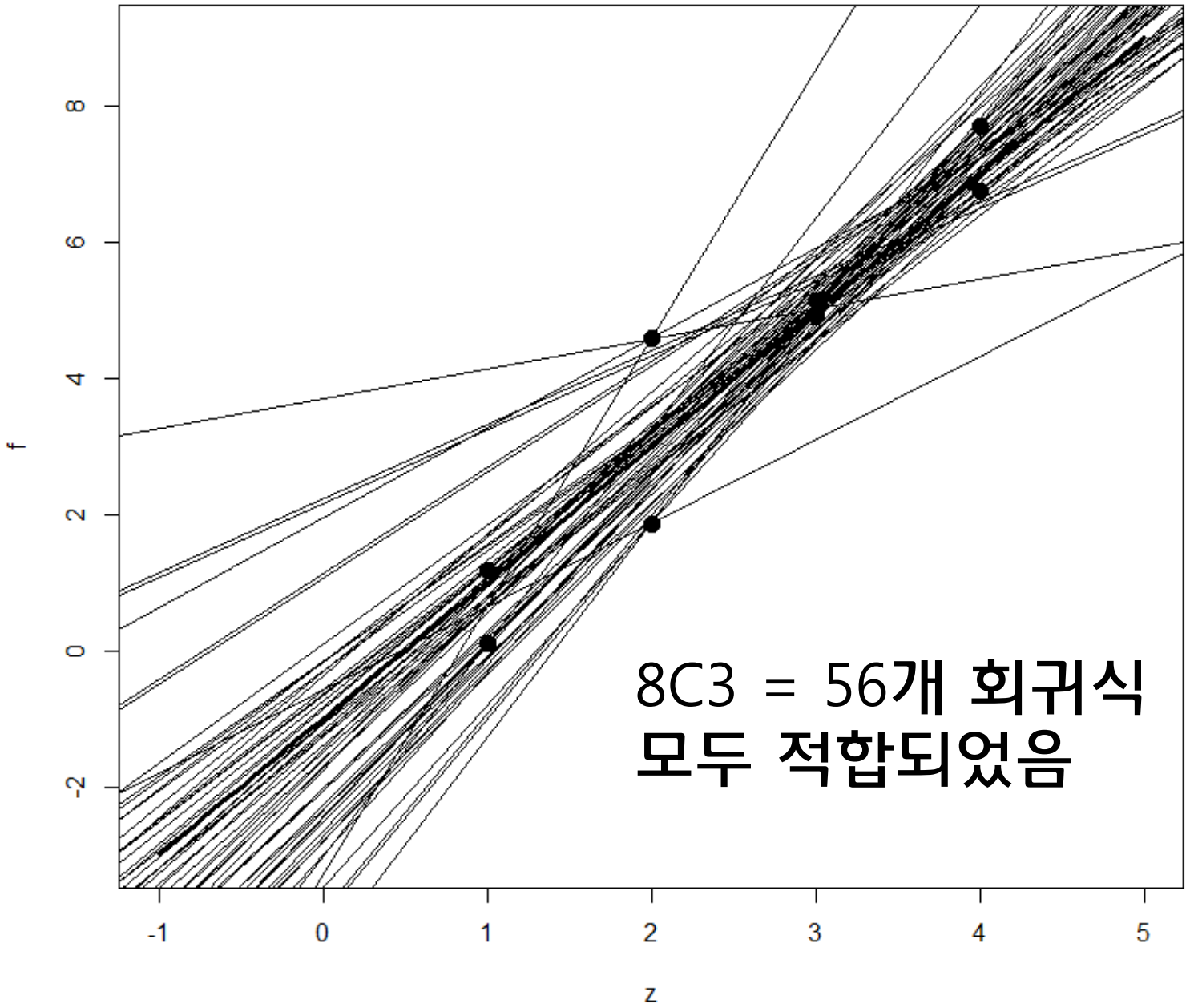
Q1. 왜 24개의 회귀직선만 적합 되었을까...?
나머지 4개는 왜 적합이 되지 않지...?

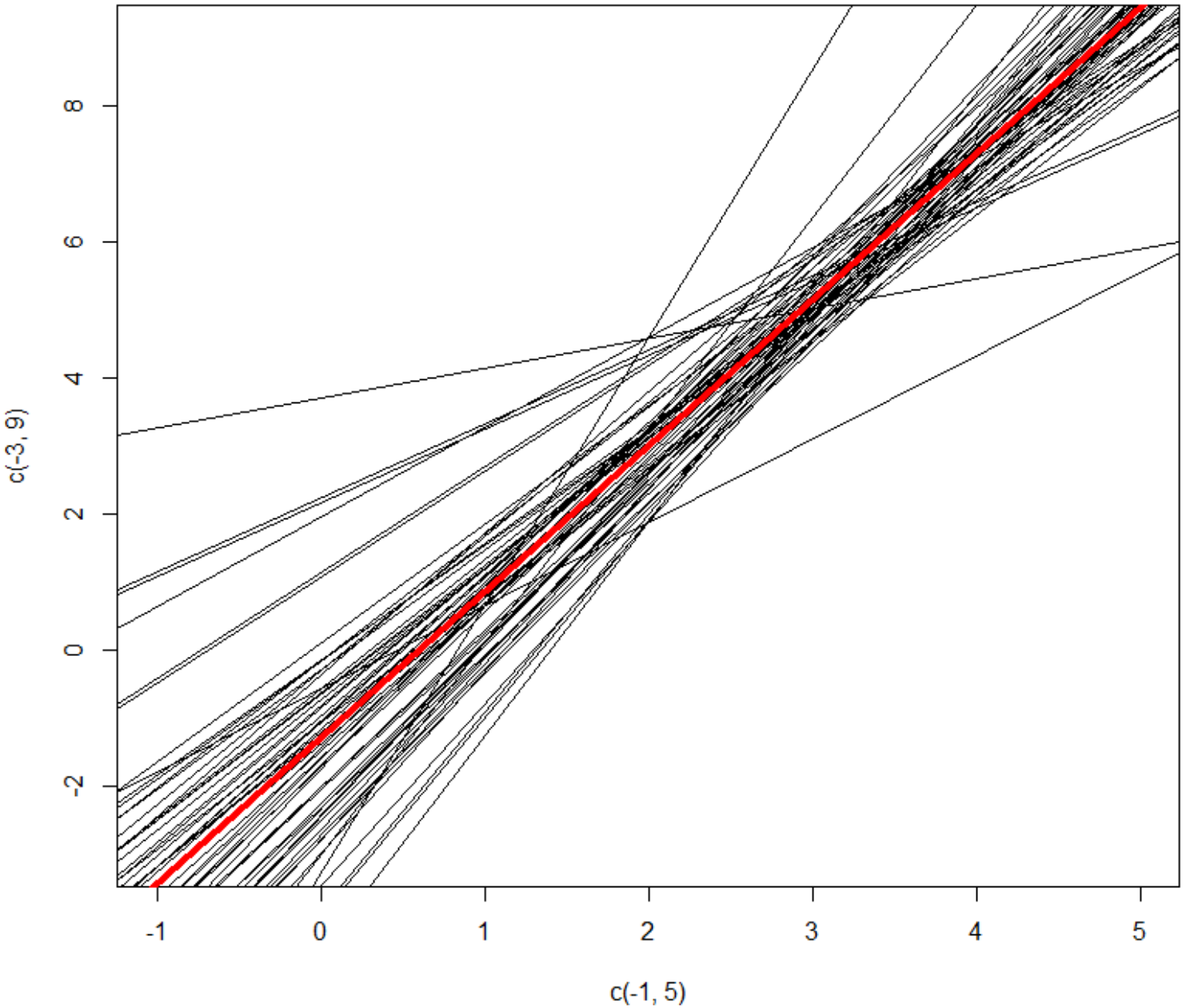

f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$

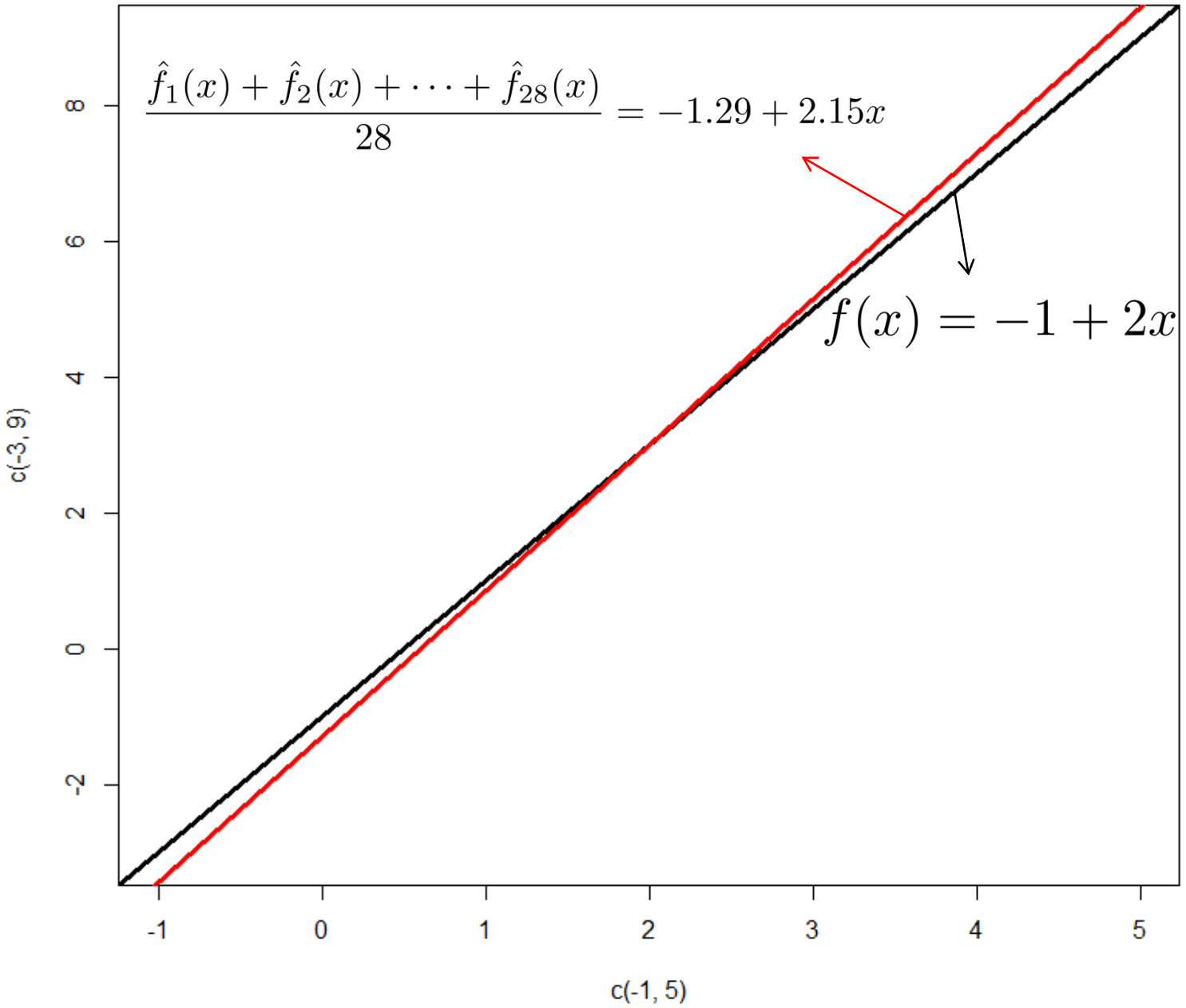

f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$f(x) = -1 + 2x$


f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$f(x) = -1 + 2x$

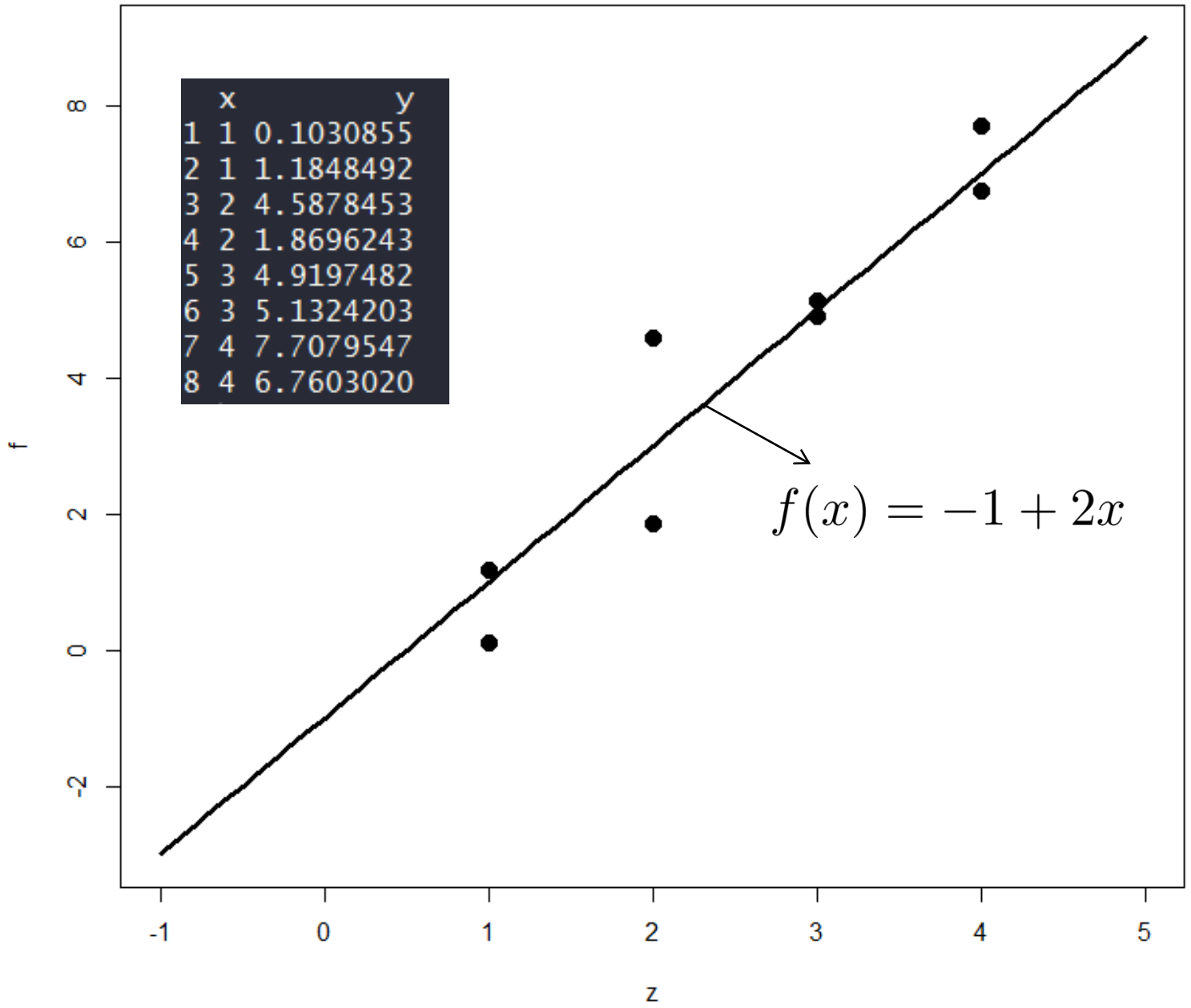

$\hat{f}_1(x) = -3.30 + 3.94x$
$f(x) = -1 + 2x$
f
z
8
6
4
2
0
-2
-1
0
1
2
3
4
5

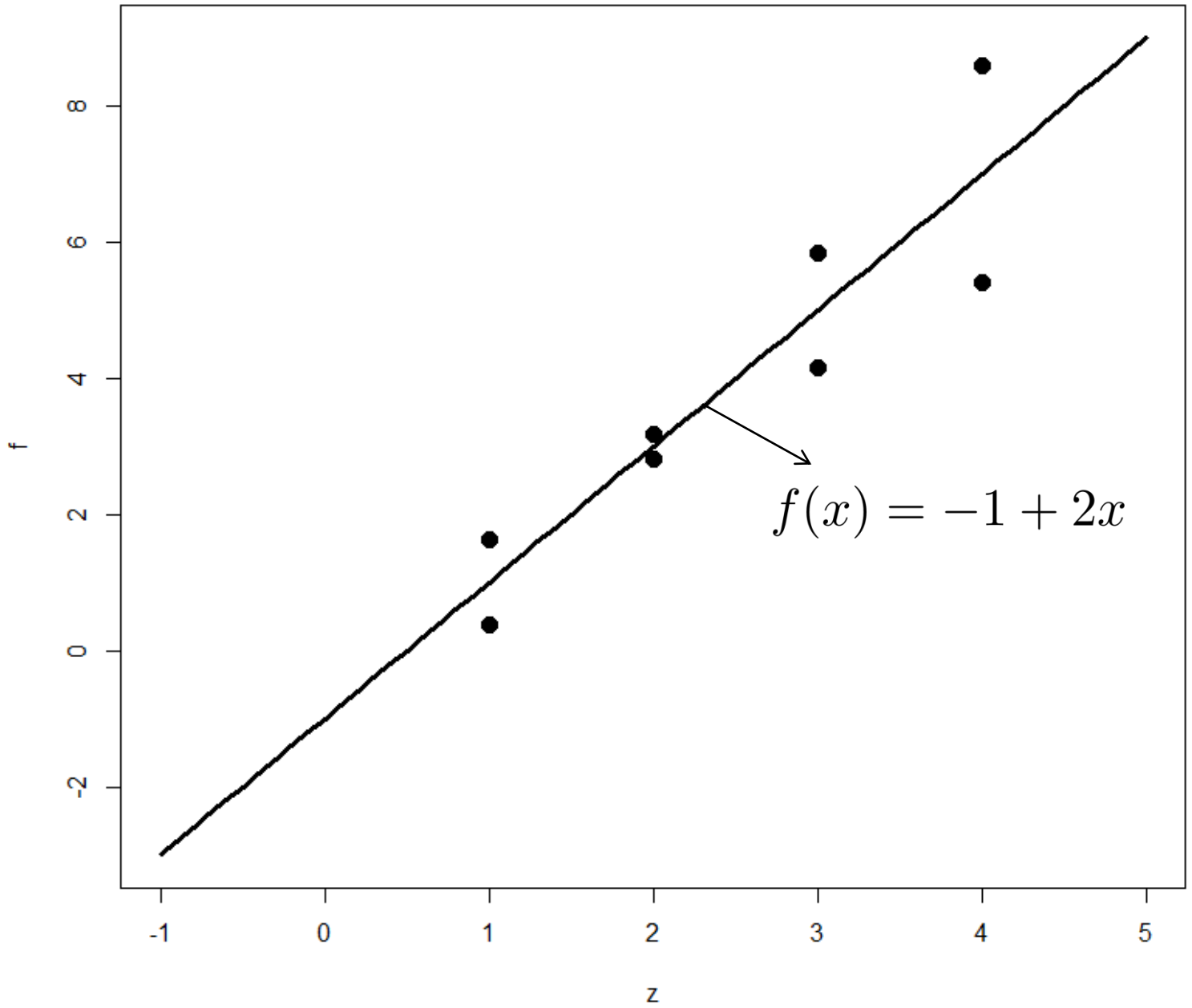

f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
8C3 = 56개 회귀식
모두 적합되었음

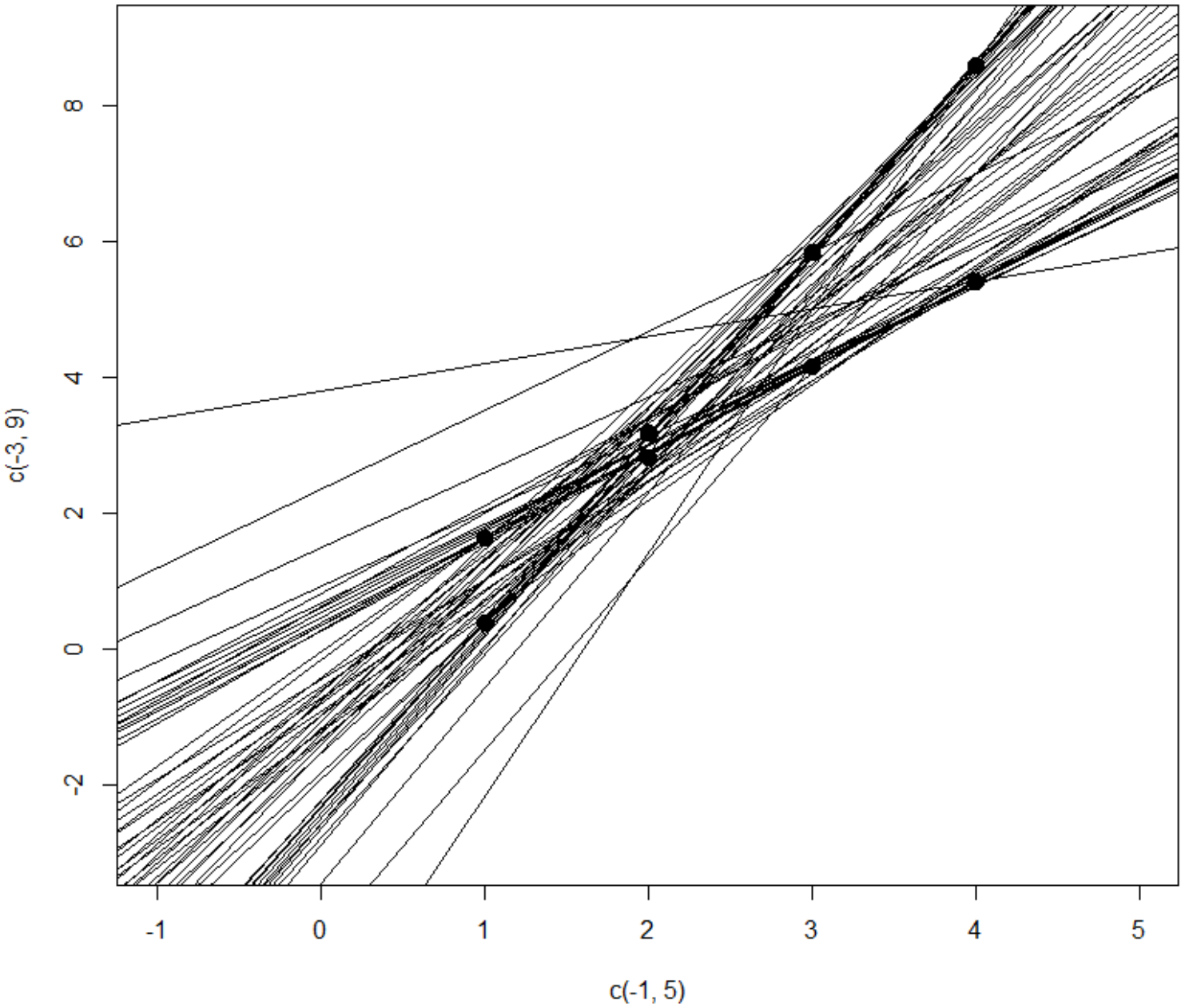

c(-3, 9)
8
6
4
2
0
-2
-1
0
1
2
3
4
5
c(-1, 5)

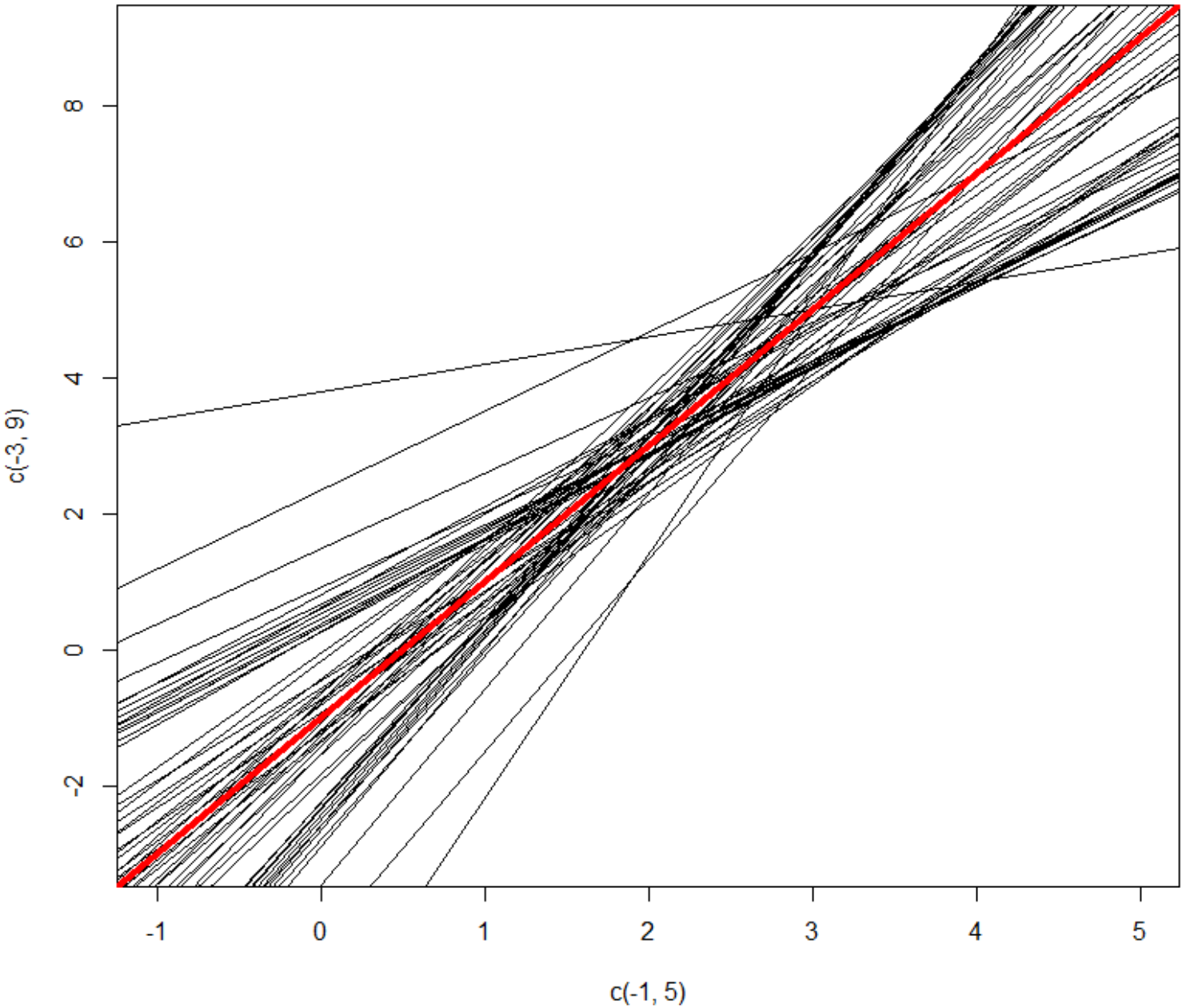

$\frac{\hat{f}_1(x) + \hat{f}_2(x) + \cdots + \hat{f}_{28}(x)}{28} = -1.29 + 2.15x$
$f(x) = -1 + 2x$
c(-3, 9)
8
6
4
2
0
-2
-1
0
1
2
3
4
5
c(-1, 5)

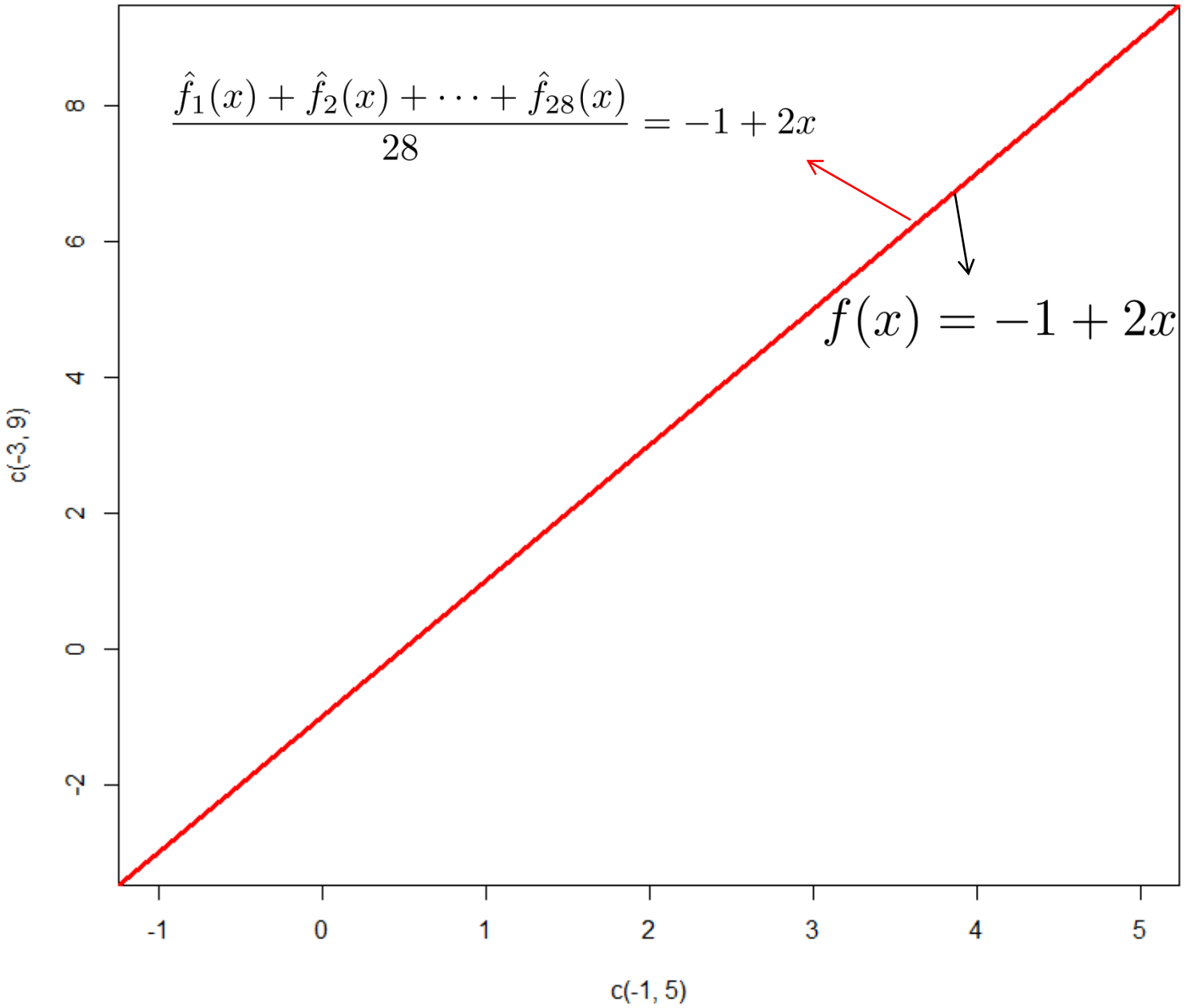
Q2. 왜 Unbiasedness가 아니지????


x y
1 1 0.1030855
2 1 1.1848492
3 2 4.5878453
4 2 1.8696243
5 3 4.9197482
6 3 5.1324203
7 4 7.7079547
8 4 6.7603020
$f(x) = -1 + 2x$
f
z
-2
0
2
4
6
8
-1
0
1
2
3
4
5


f
8
6
4
2
0
-2
-1
0
1
2
3
4
5
z
$f(x) = -1 + 2x$


c(-3, 9)
8
6
4
2
0
-2
-1
0
1
2
3
4
5
c(-1, 5)


c(-3, 9)
8
6
4
2
0
-2
-1
0
1
2
3
4
5
c(-1, 5)


$\frac{\hat{f}_1(x) + \hat{f}_2(x) + \cdots + \hat{f}_{28}(x)}{28} = -1 + 2x$
$f(x) = -1 + 2x$
c(-3, 9)
8
6
4
2
0
-2
-1
0
1
2
3
4
5
c(-1, 5)