Minimum Error Entropy

Lesson

- Our measure should not depend upon outlier value
- In general, data values should not be used

No more X axis

- We'll work with Y axis
- i.e. not with data values rather with frequency of data values

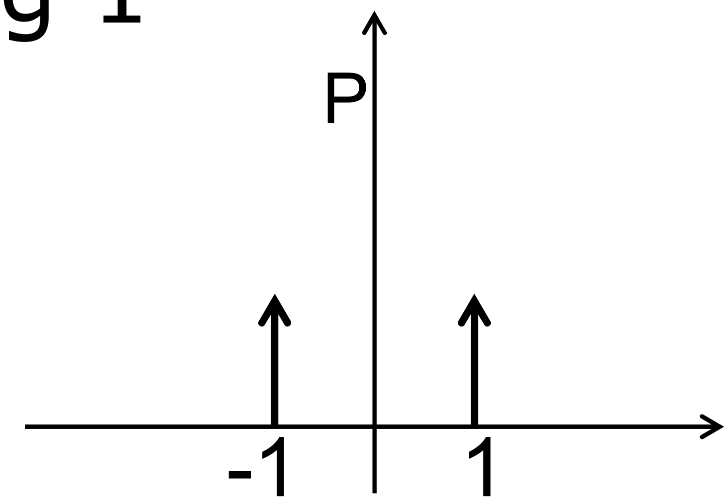
A metric does not depend upon data values

- Entropy
- Doesn't depend on X-axis

$$\sum_{i=1}^{r} p_i \log_2 \frac{1}{p_i} = -\sum_{i=1}^{r} p_i \log_2 p_i$$

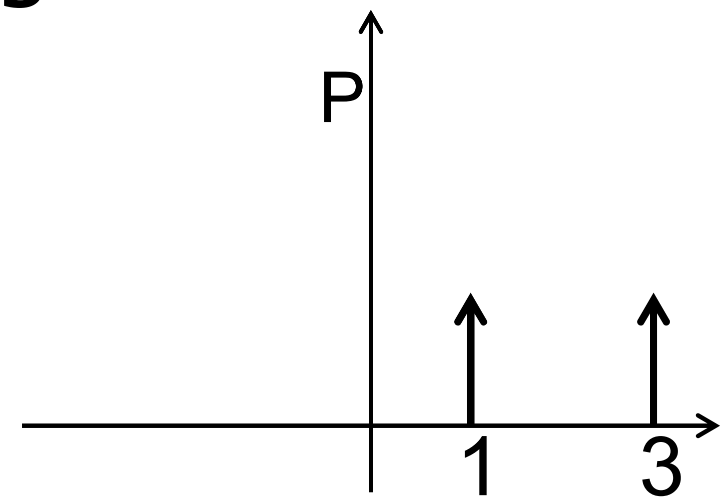
Coin tossing e.g 1

- Fair coin tossing
- Head → 1 & Tail → -1
- P(1) = P(-1) = 0.5
- Entropy = 1 bit



Coin tossing e.g 2

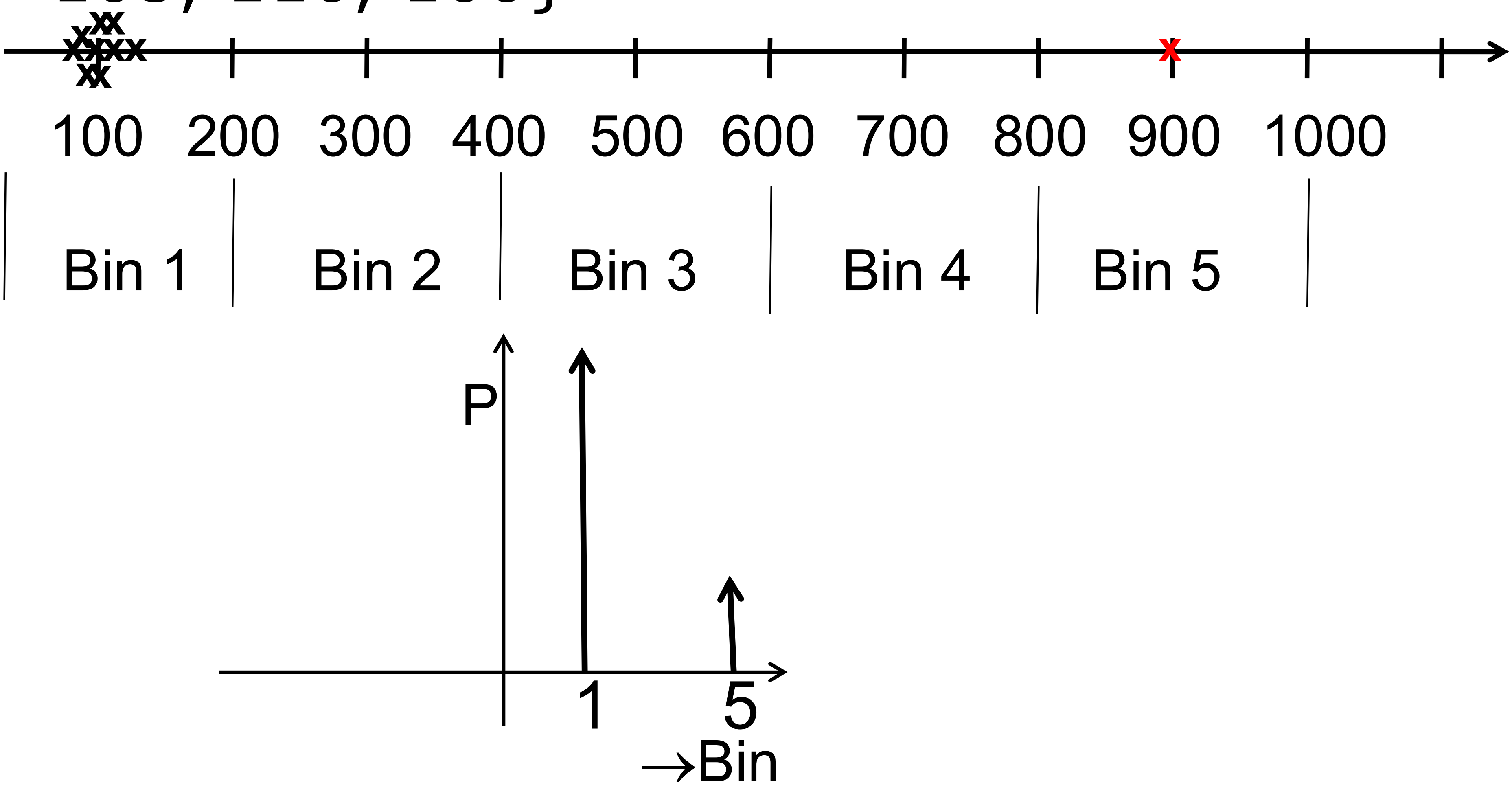
- Fair coin tossing
- Head → 1 & Tail → 3
- P(1) = P(3) = 0.5
- Entropy = 1 bit



Entropy does not depend on X-axis values

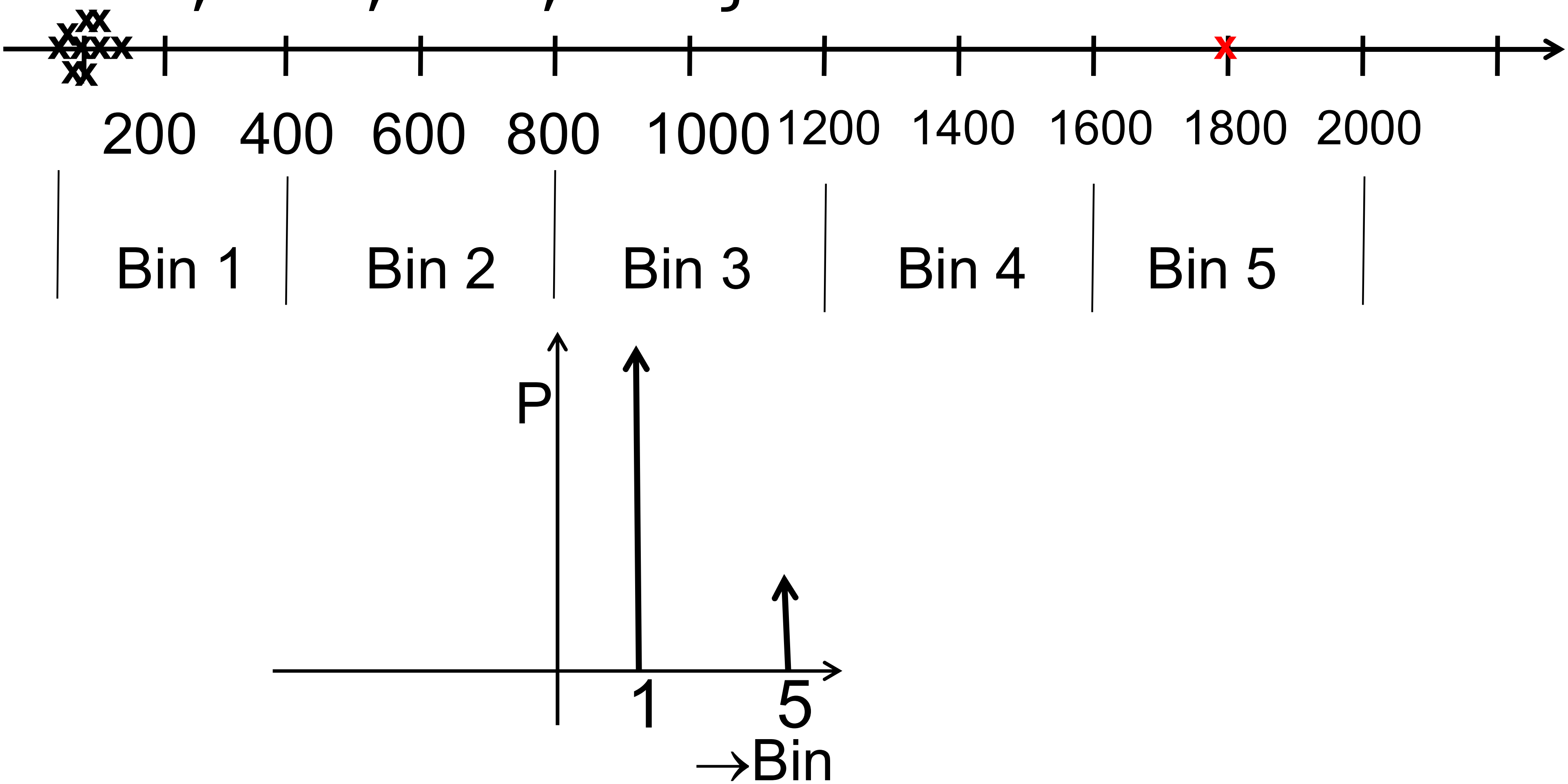
Data values do not affect[1]

- Data = {**<u>900</u>**, 120, 90, 110, 115, 125, 95, 105, 110, 100}



Data values do nor affect[2]

- Data = {**1800**, 120, 90, 110, 115, 125, 95, 105, 110, 100}



Hypothesis

- Compared to mean square error, entropy can handle the outliers effectively

Adaptive system with cost function as MSE

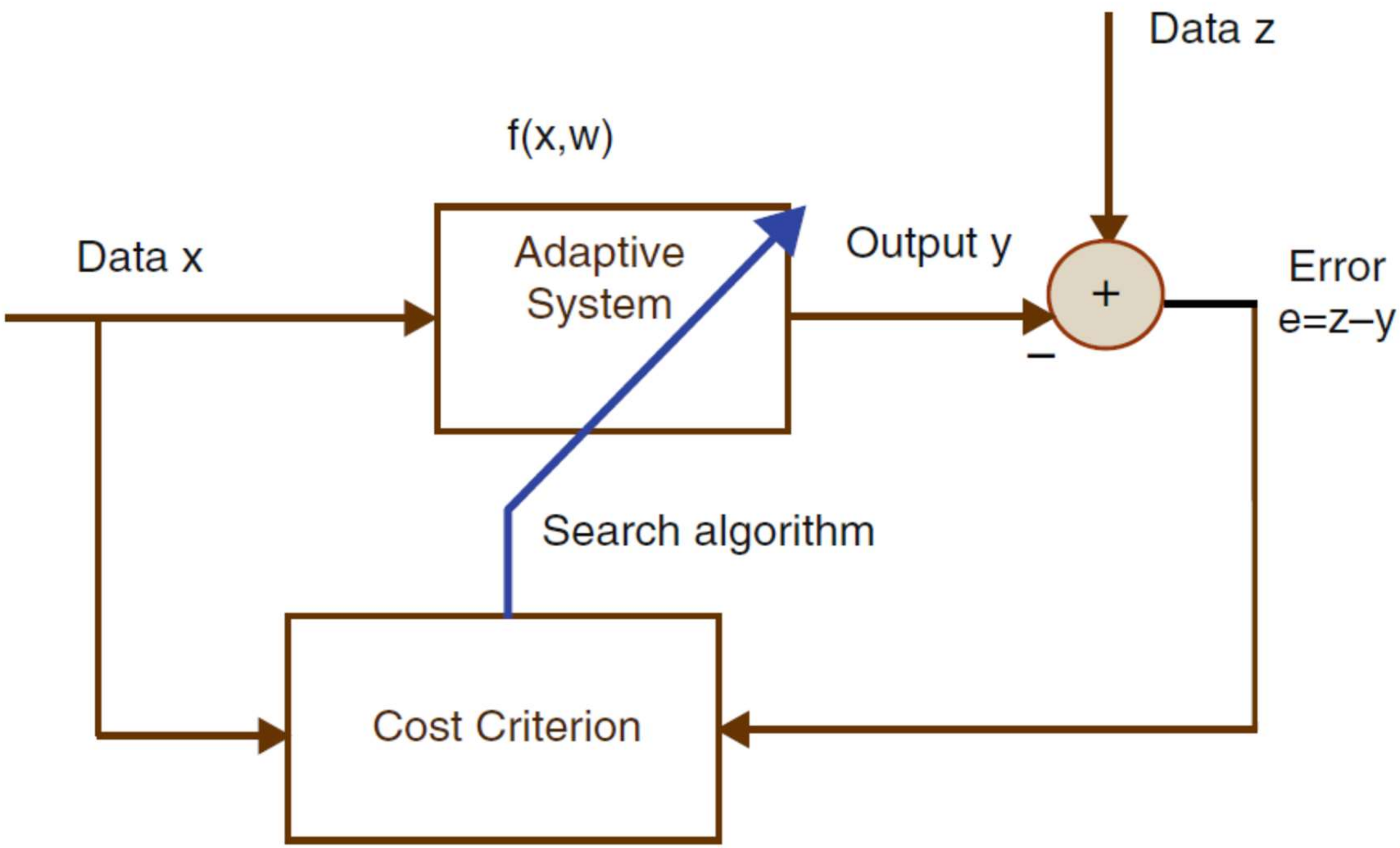
- Least mean square error

Adaptive system with cost function as entropy

- Minimum error entropy

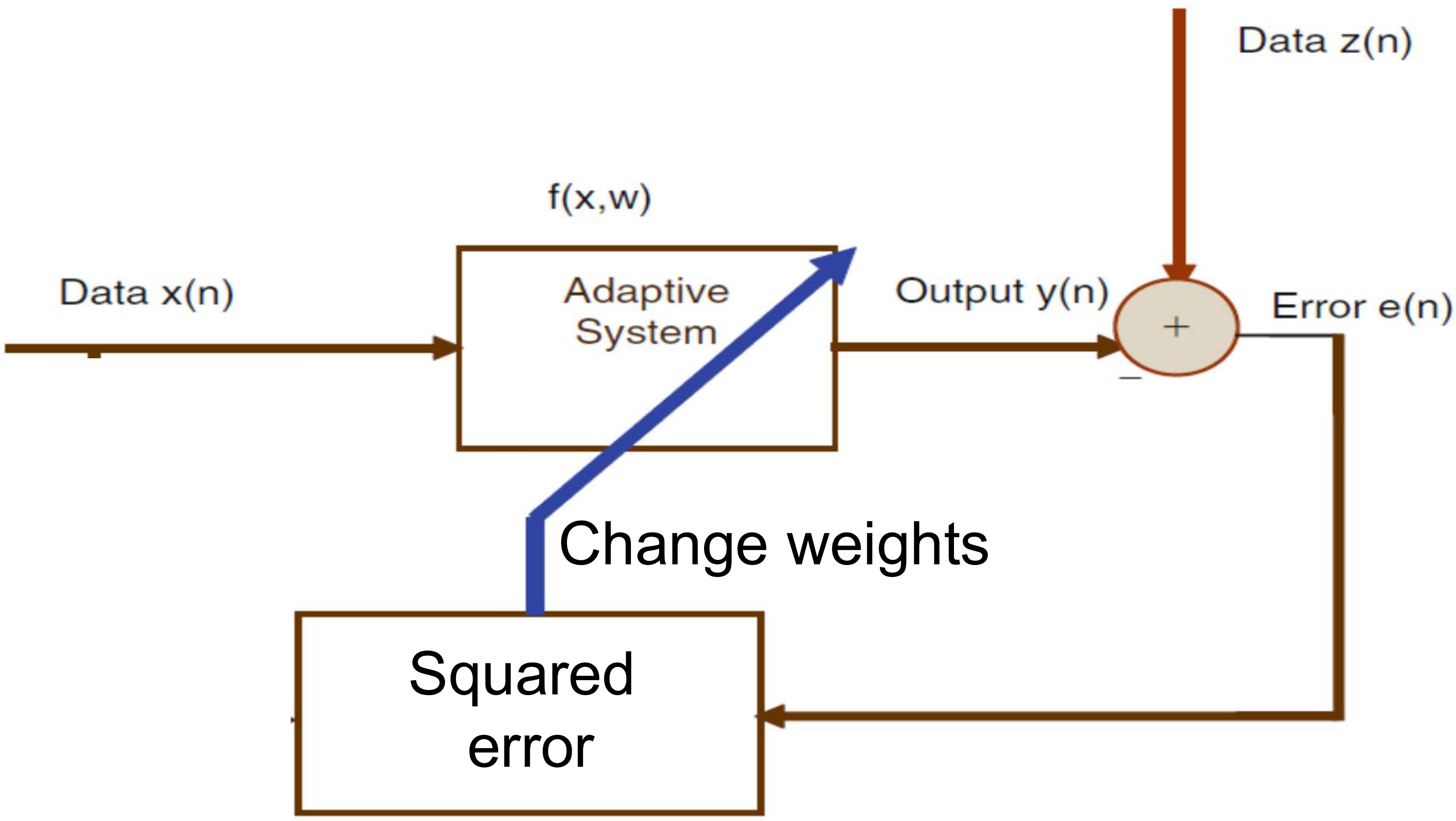
Error as a random variable

- X: explanatory variable
- Y = f(x,w): generated output
- Z: response or desired output
- Error variable = z – f(x)

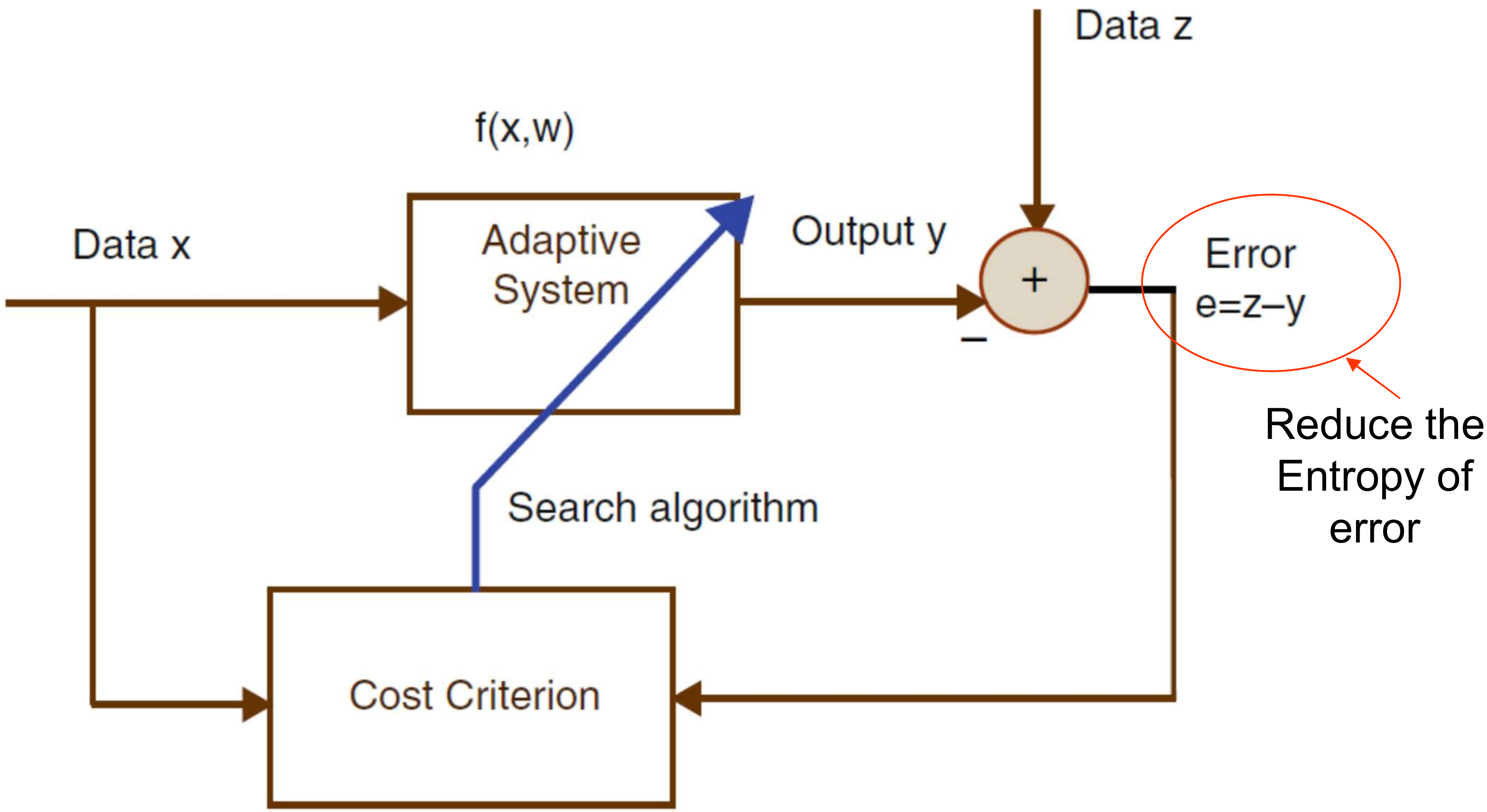


Least MSE - Schematic

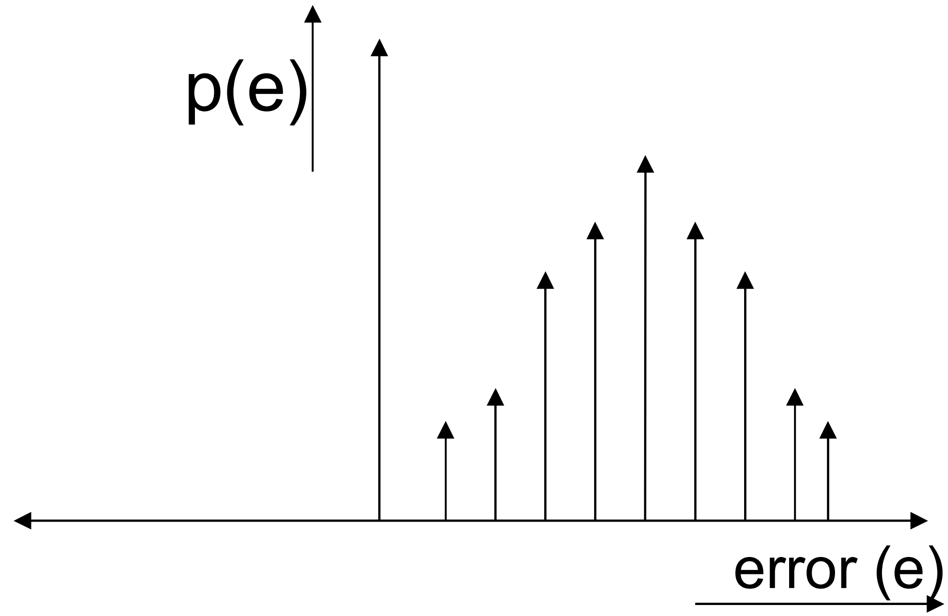
Squared error = $\left[(z(n) - f(\mathbf{w}, x(n)))^2\right]$



MEE


Data z
f(x,w)
Data x
Adaptive System
Output y
+
−
Error
e=z−y
Reduce the Entropy of error
Search algorithm
Cost Criterion

Cost function (MSE and MEE)



$$MSE\ deals\ with\ E(error^2)$$

$$MEE\ deals\ with\ P(error)$$

$$i.e.\ entropy = -\sum p(error).\log\left(p(error)\right)$$

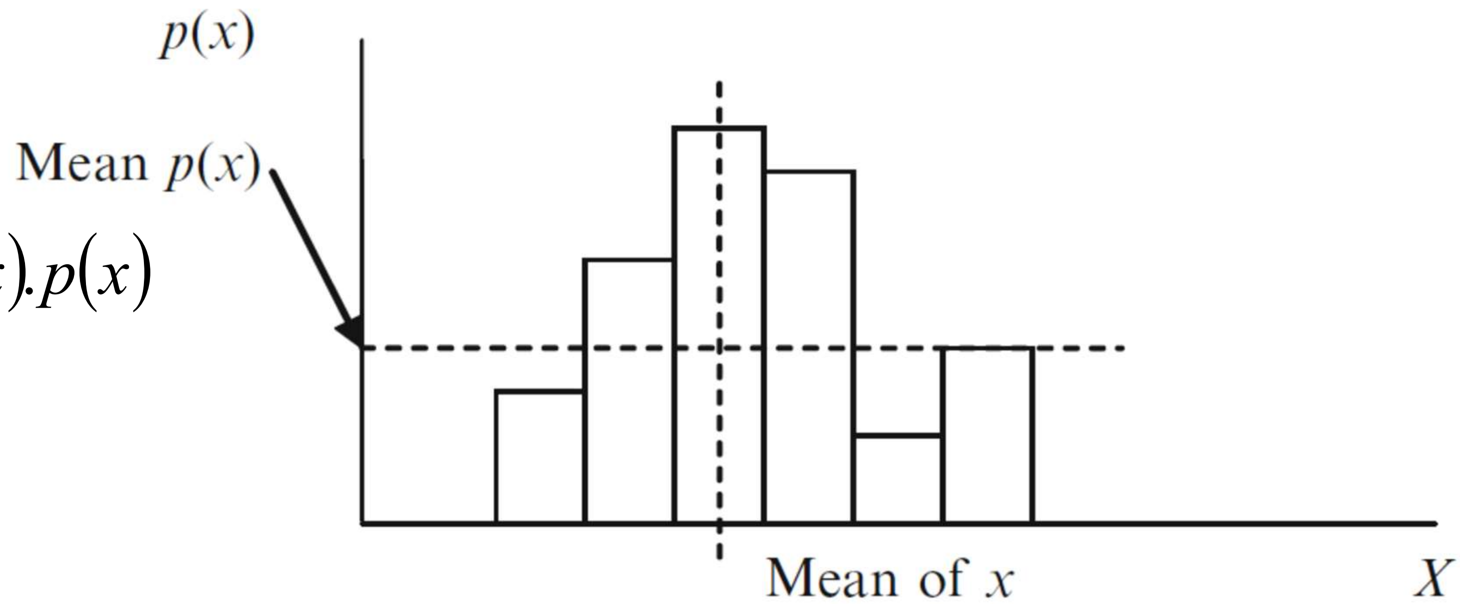
i.e. –E(log(P(error))

Expectation of data and PMF/PDF

- Not to confuse the moments of the PMF with the moments of the data

$$E[p(X)] = \sum_x p(x).p(x)$$

$$= \sum_x [p(x)]^2$$



$$E[X] = \sum_x x.p(x)$$

Minimum Error entropy

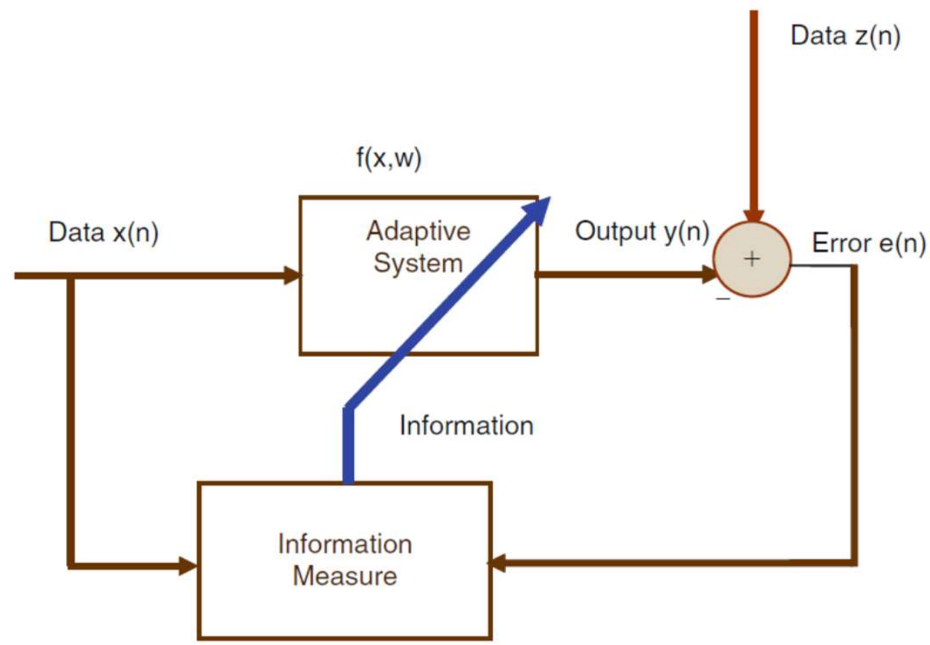
- We have decided to fit using y=m.x
- x=[---, ---, ---, ---, …., ---]
- t=[---, ---, ---, ---, …., ---]
- Choose m

1. ***Predict y=[---, ---, ---, ---, …., ---]***
2. ***Find out error***
3. ***e=[---, ---, ---, ---, …., ---]***
4. ***Estimate error pdf:***
   - ***p(e)=[---, ---, ---, ---, …., ---]***
5. ***Use p(e) to compute entropy (H(e))***
6. ***Use p(e) to compute expectation (E(e))***

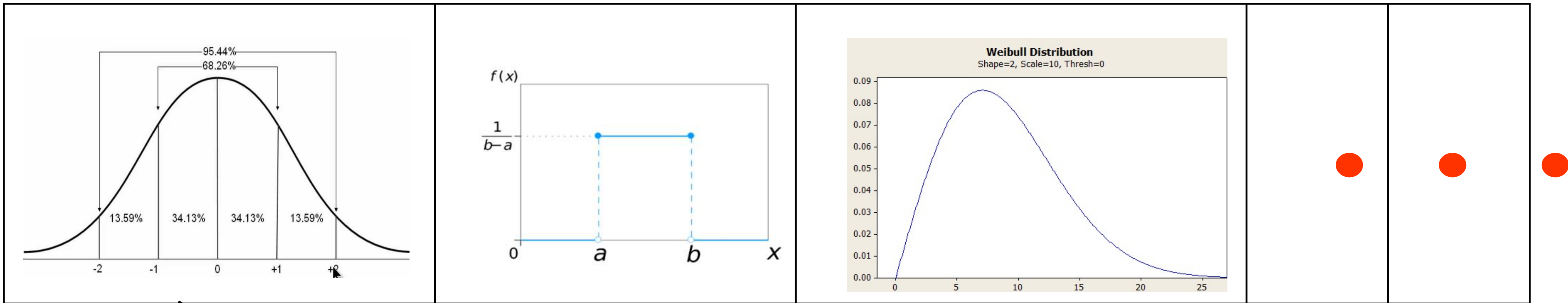
- Change m; repeat from 1 to 6
- Plot H(e) versus m and E(e) versus m

Error entropy criterion

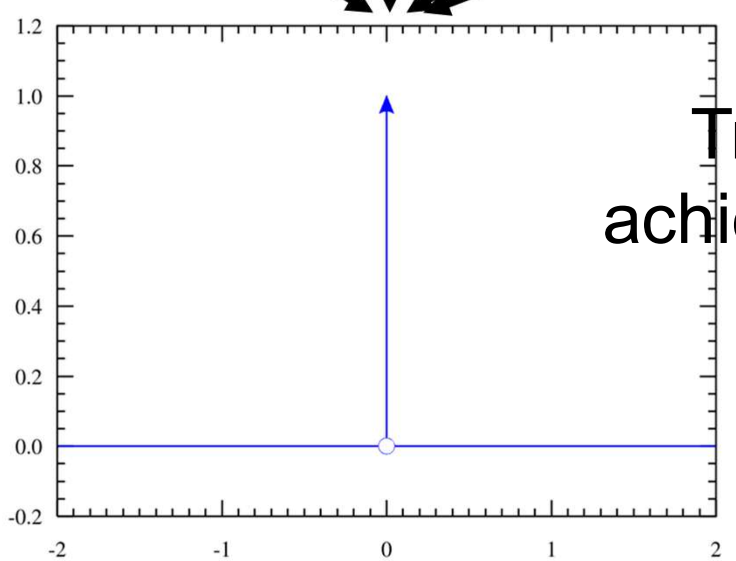
- Goal of adaptation is to remove as much uncertainty as possible from the error signal
  - Change y(n) in a way to reduce the entropy of e(n)
  - i.e. Adjust weights of adaptive system in a way to reduce H(e(n))



MEE – visual representation

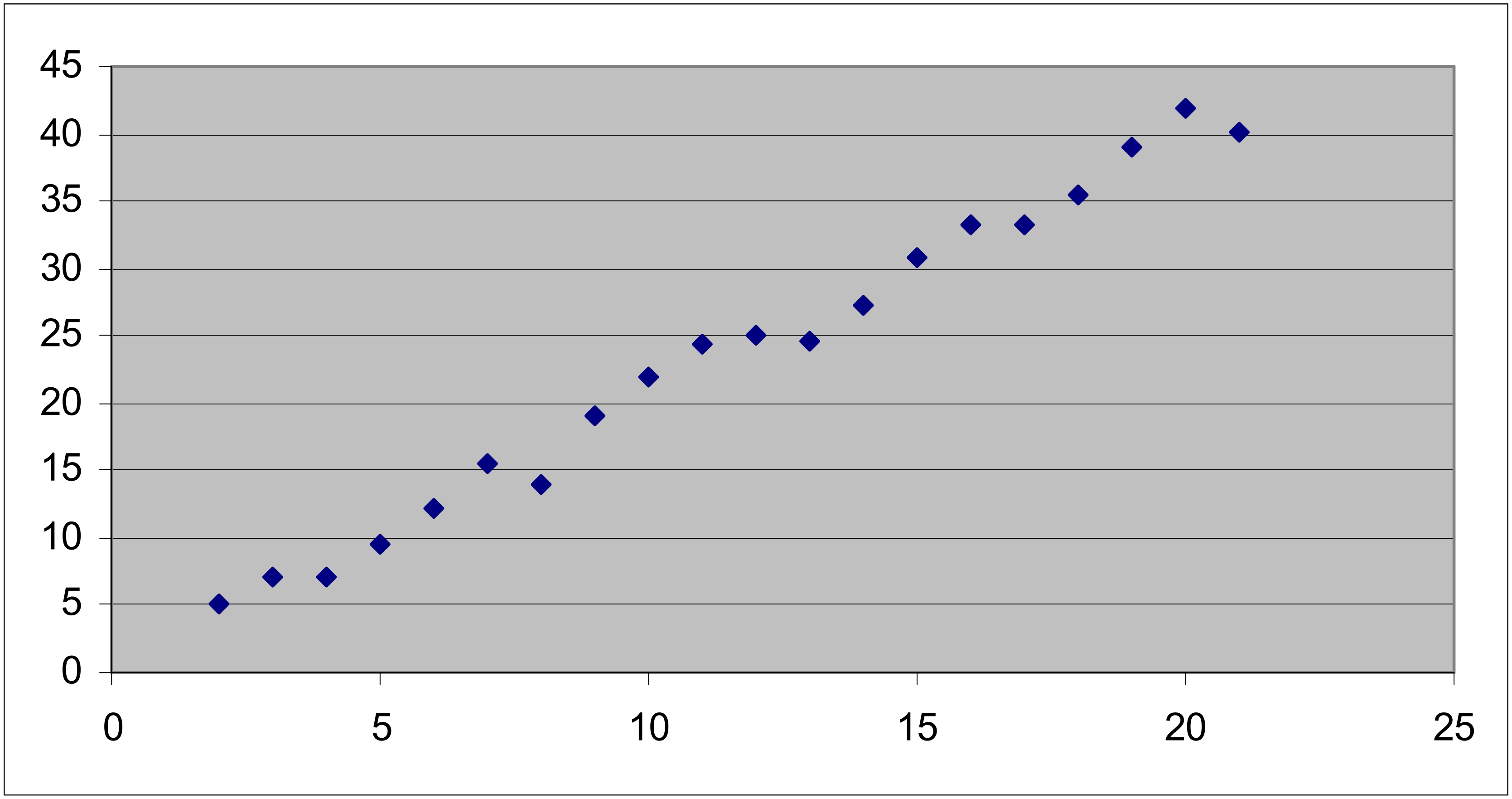


From these

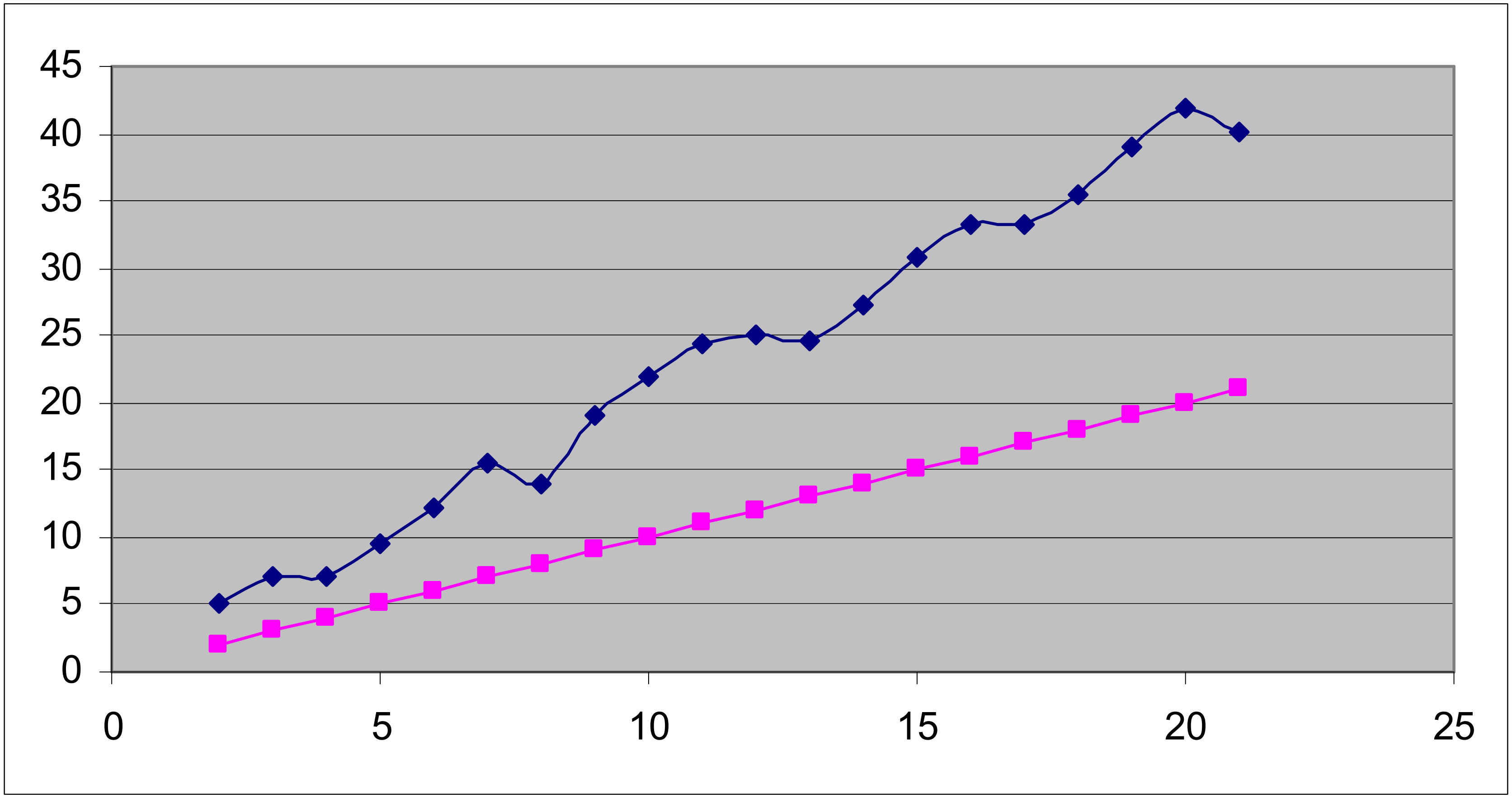


Try to achieve this

Example 1



Fit with $y = m.x + c$

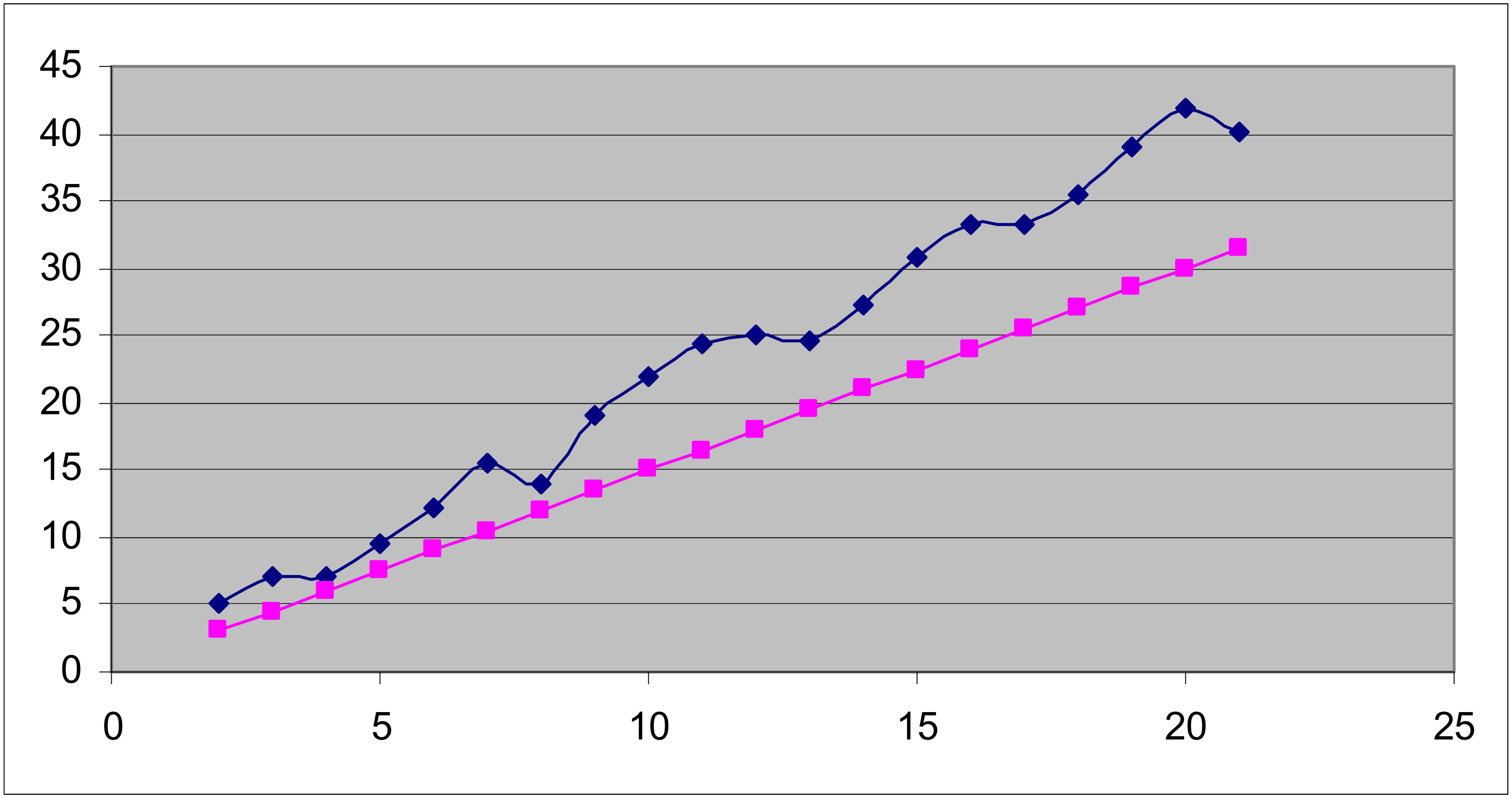


Fit with y = m.x + c

C = 0

m = 1

MSE = 173.97

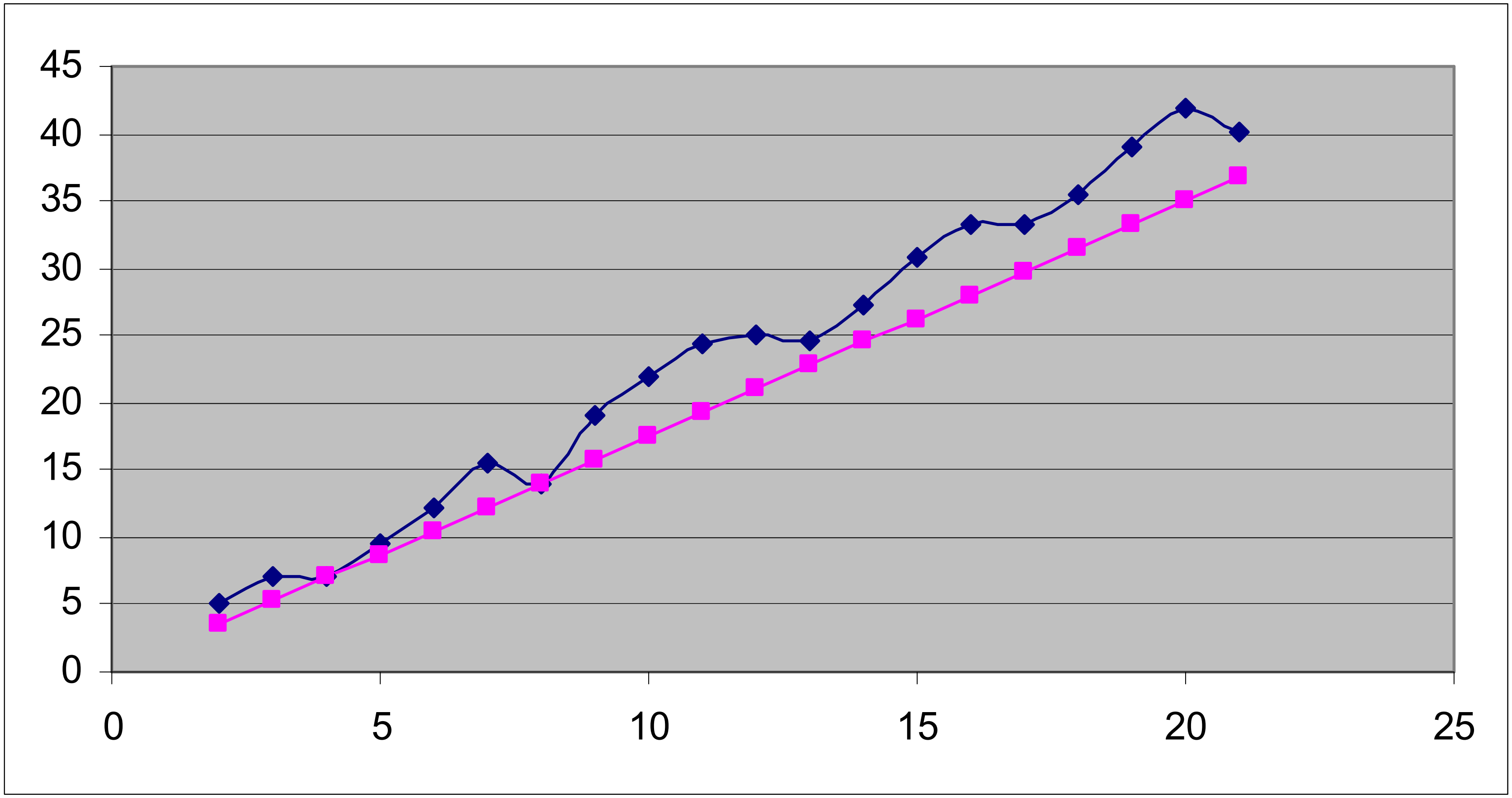


Fit with y = m.x + c

C = 0

m = 1.5

MSE = 46.49

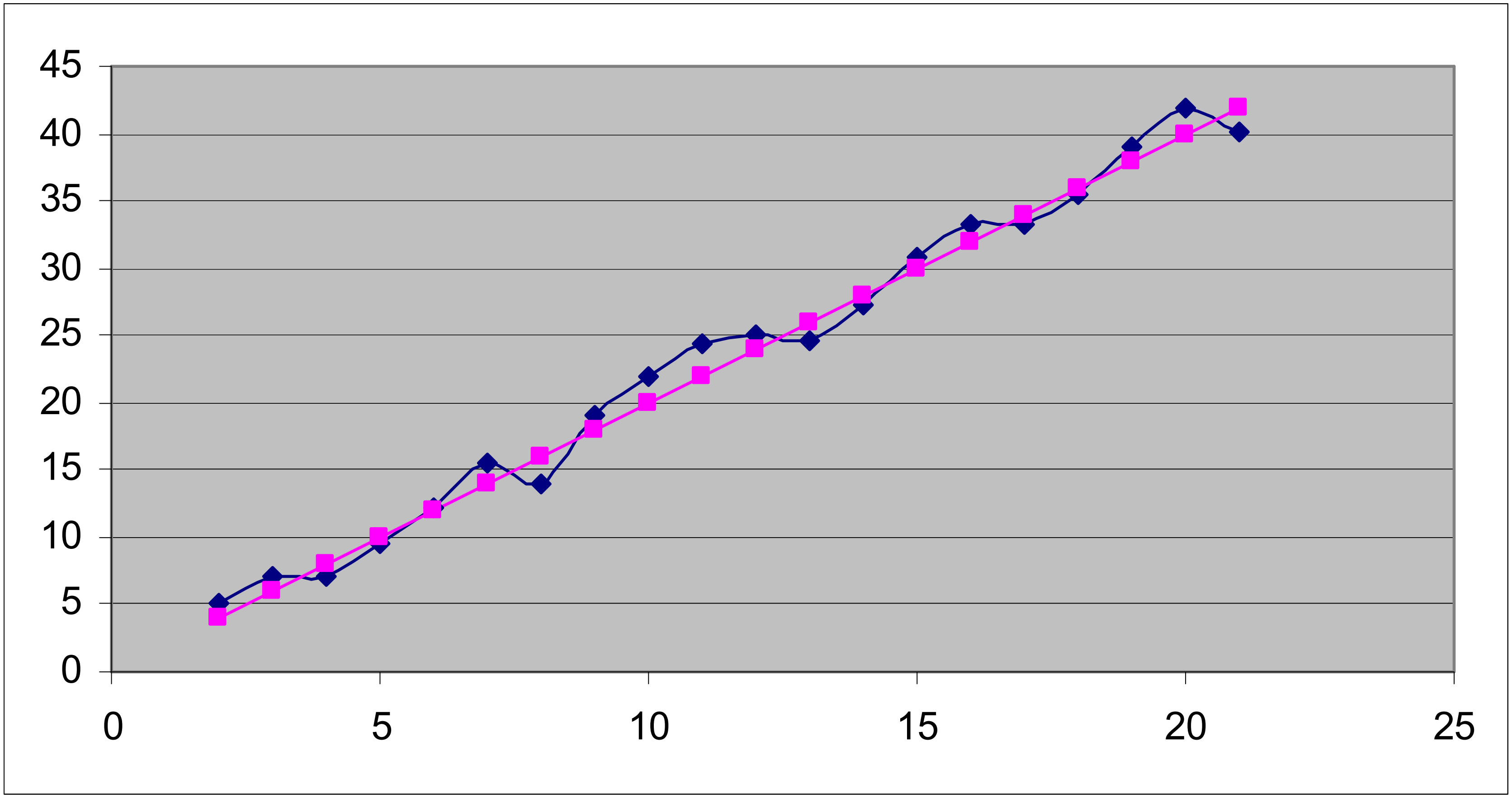


Fit with y = m.x + c

C = 0

m = 1.75

MSE = 13.78

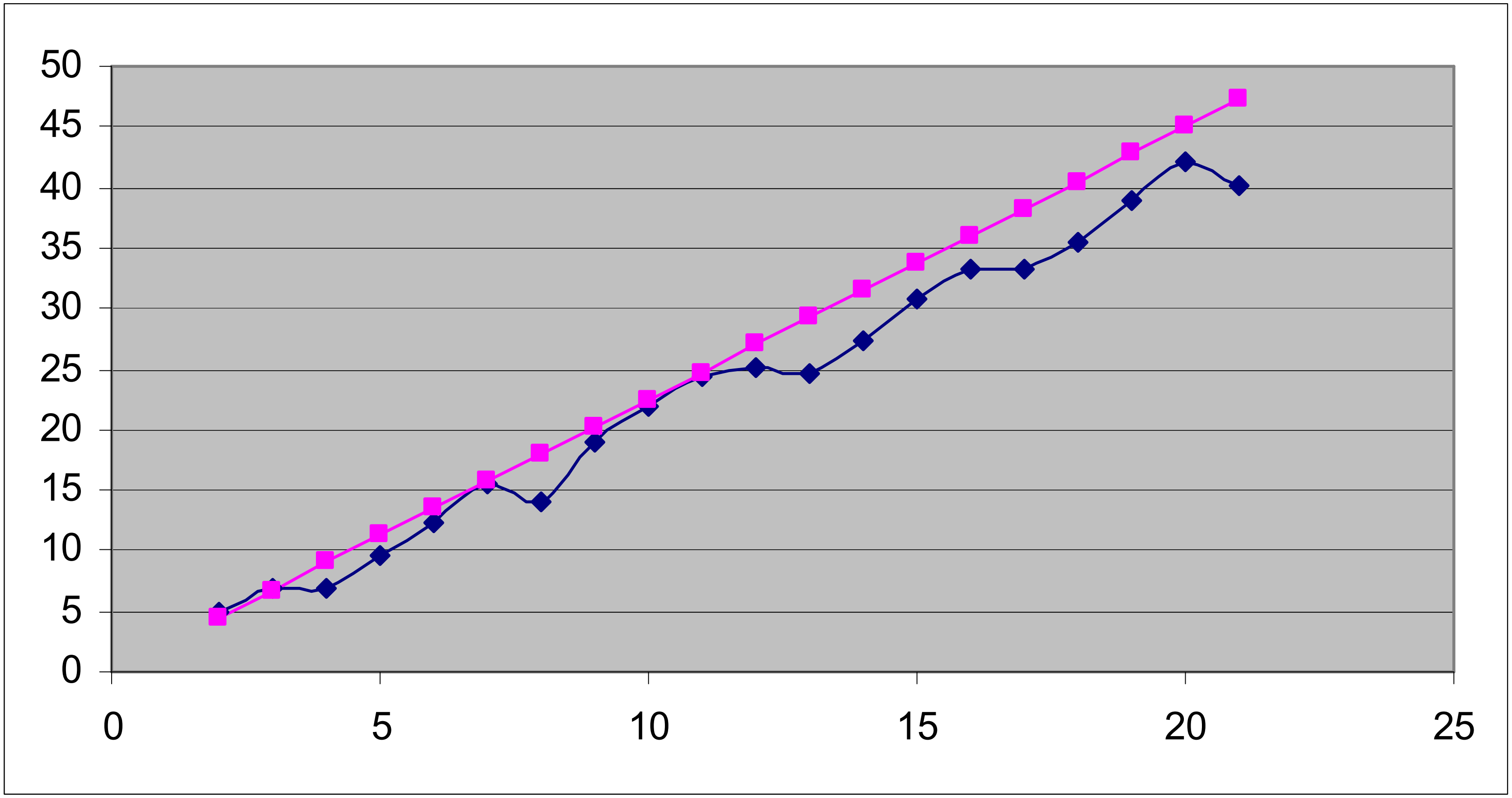


Fit with $y = m.x + c$

$C = 0$

$m = 2$

$MSE = 1.76$



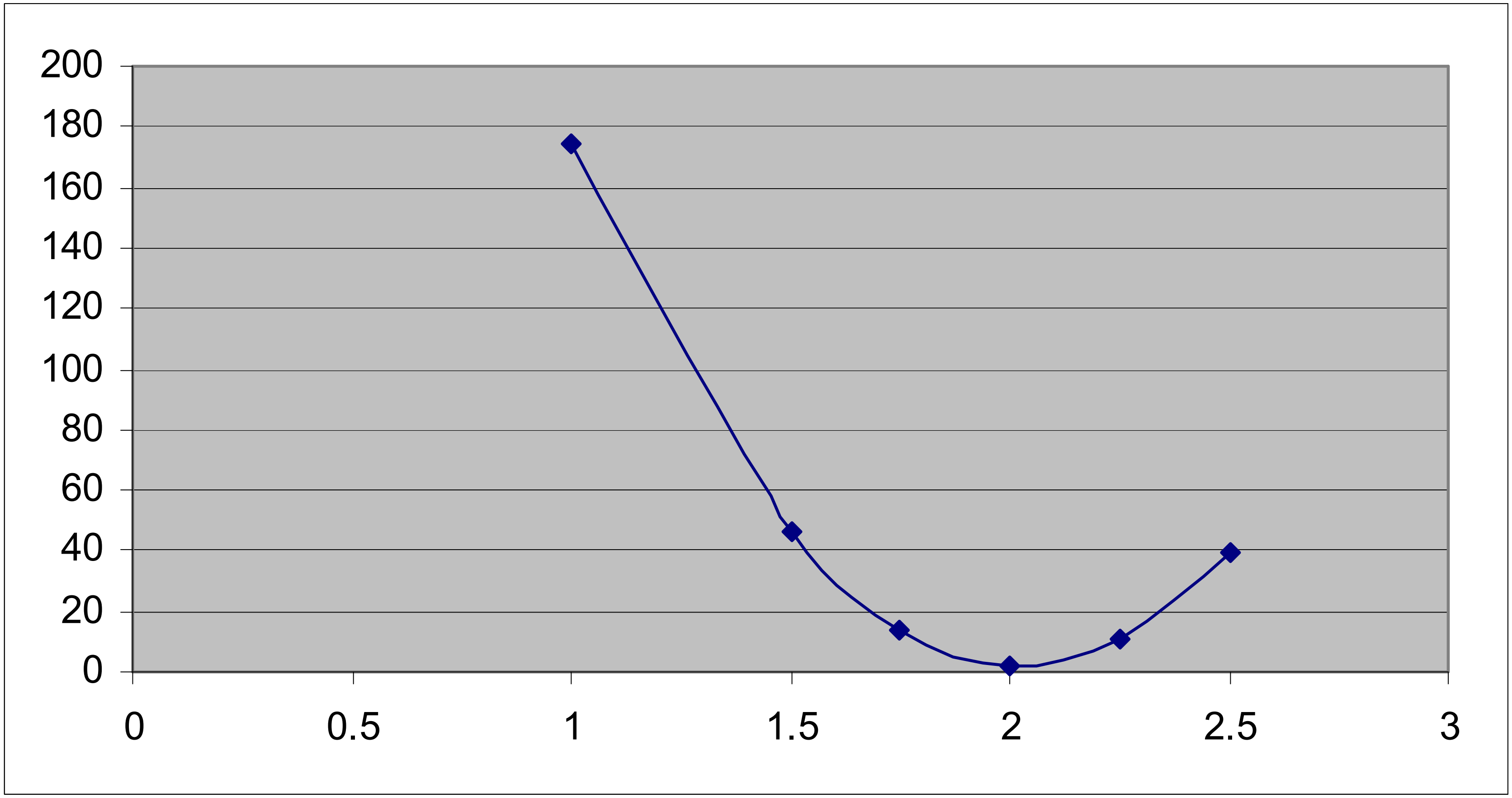
Fit with y = m.x + c

C = 0

m = 2.25

MSE = 10.42

Squared error versus m

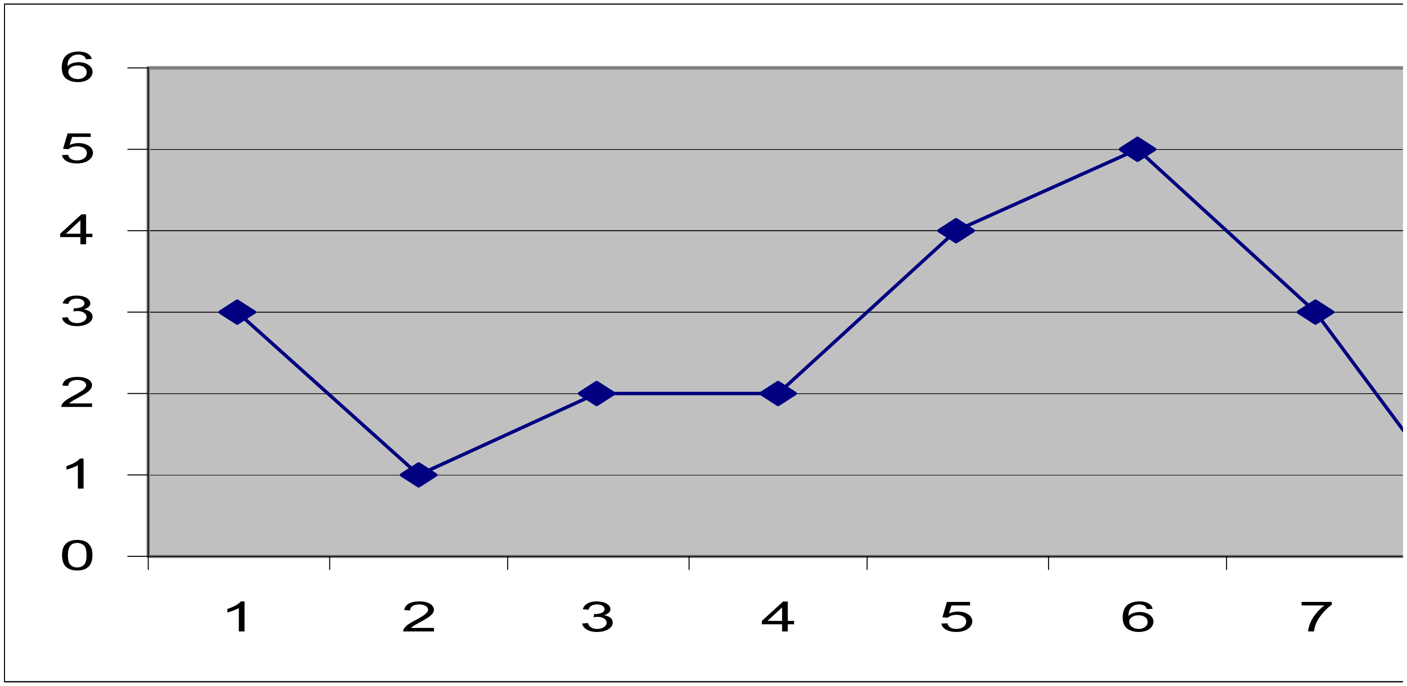


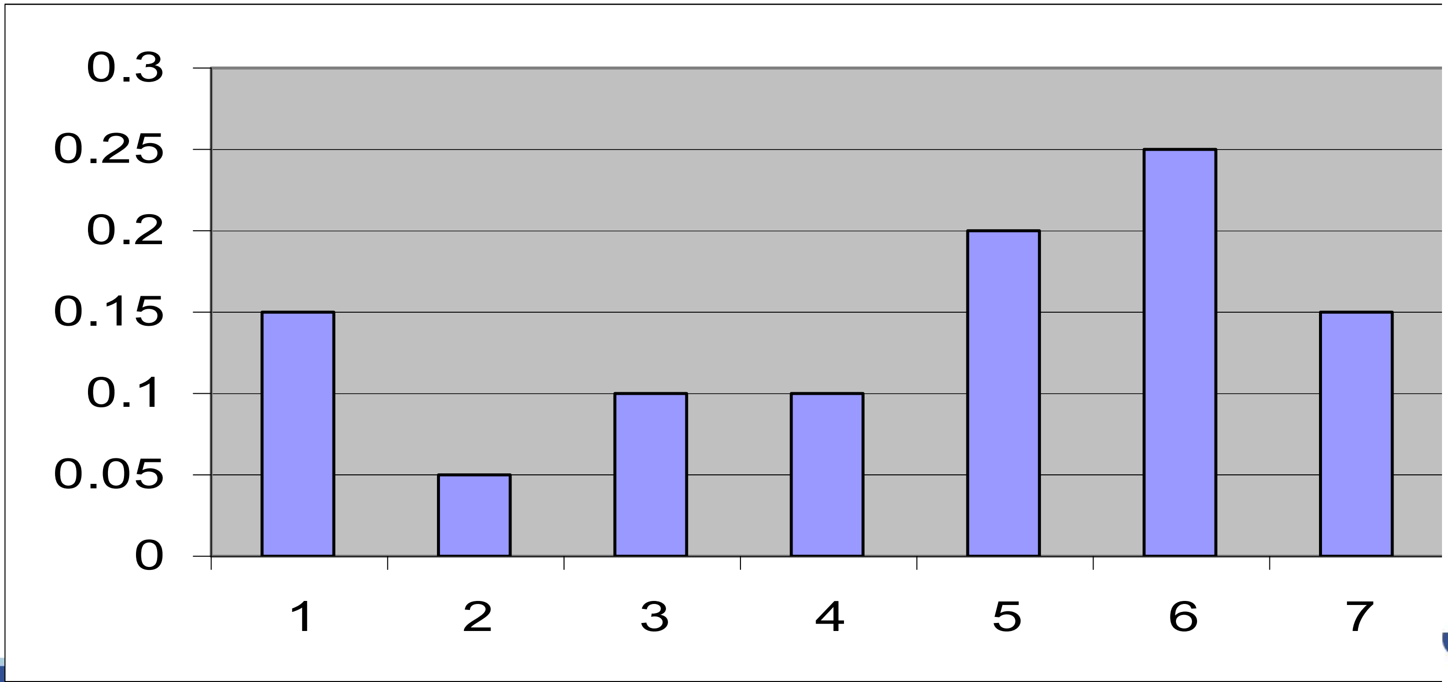
Error frequencies

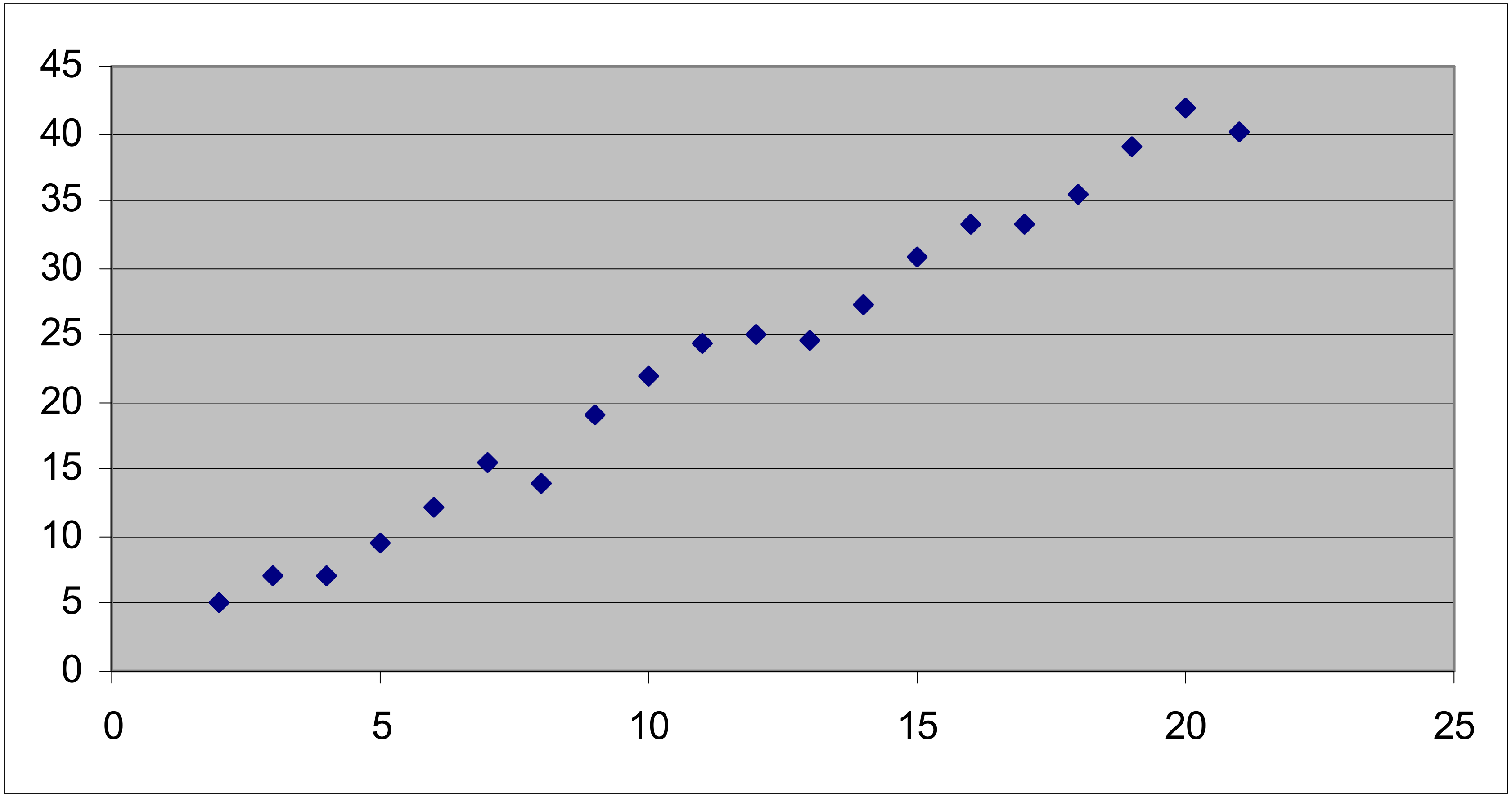
13	-157
18	-158
3	-202
-7	-380
-3	-278
-48	-478
-52	-356
-73	-543
-122	-500
-108	-747

-599	-500	3
-499	-400	1
-399	-300	2
-299	-200	2
-199	-100	4
-99	0	5
1	100	3
101	200	0
201	300	0
301	400	0
401	500	0
501	600	0
601	700	0
701	800	0

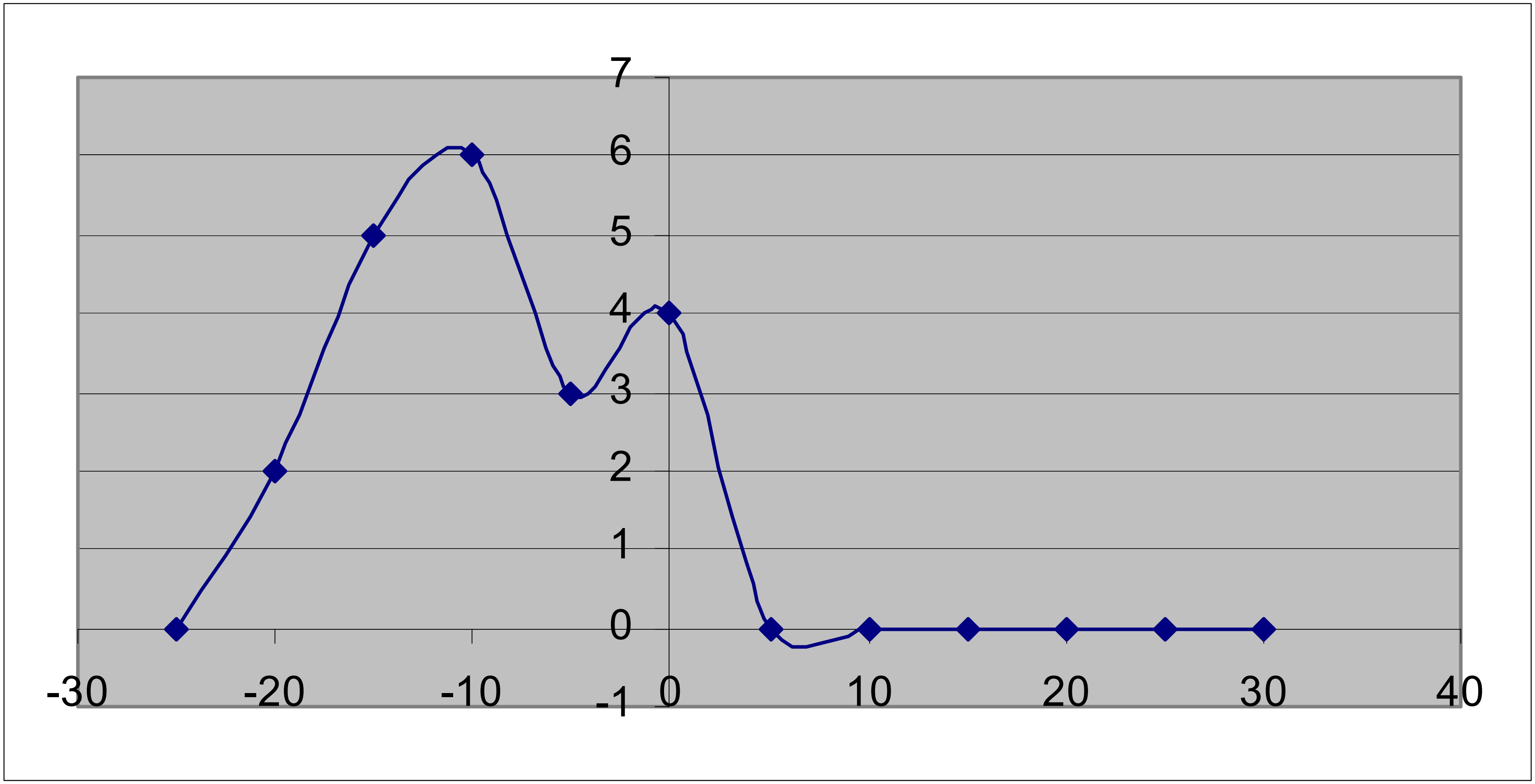
Plotting error distribution






45
40
35
30
25
20
15
10
5
0
0
5
10
15
20
25

Error distribution

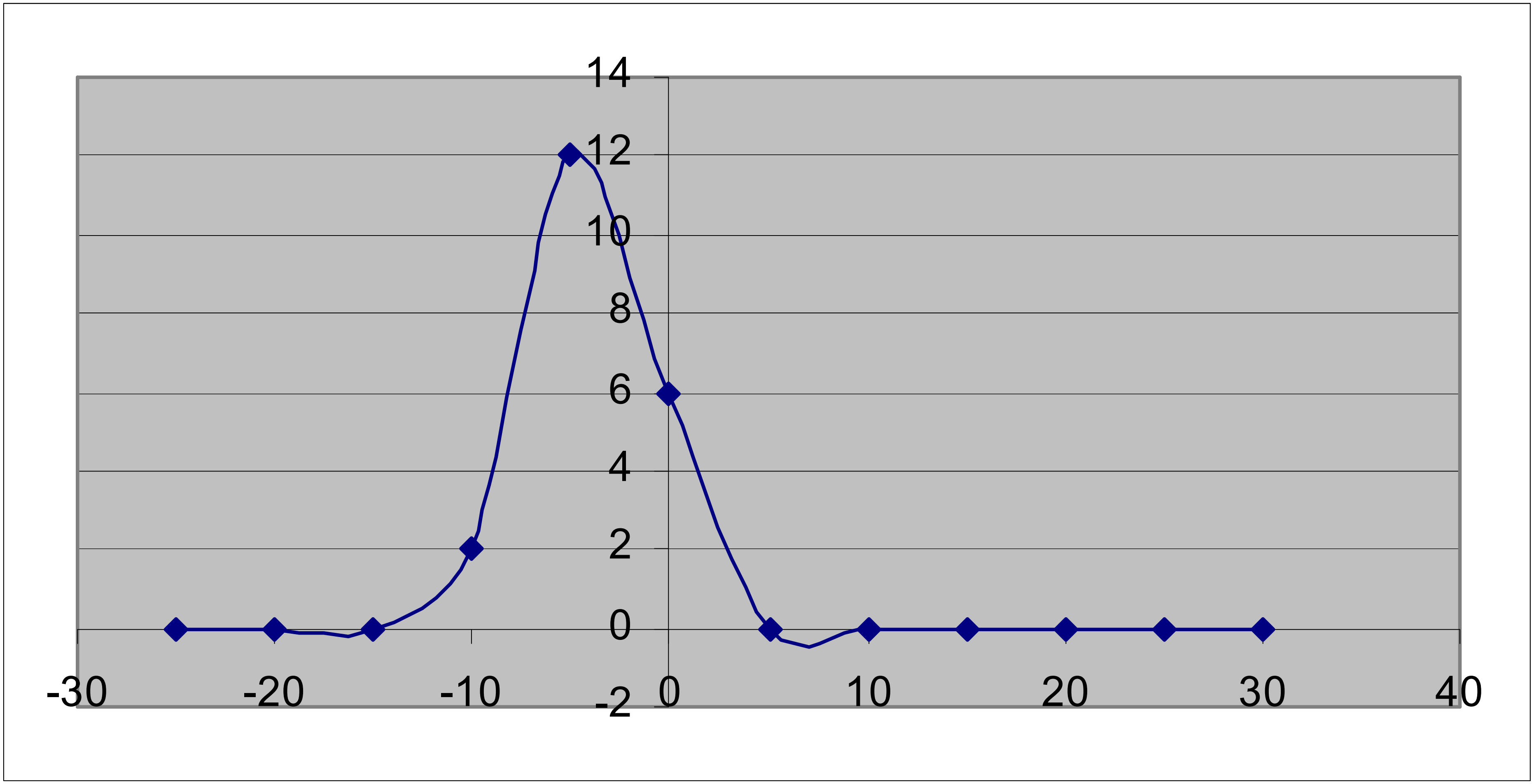


Fit with y = m.x + c

C = 0

m = 1

Entropy = 2.228

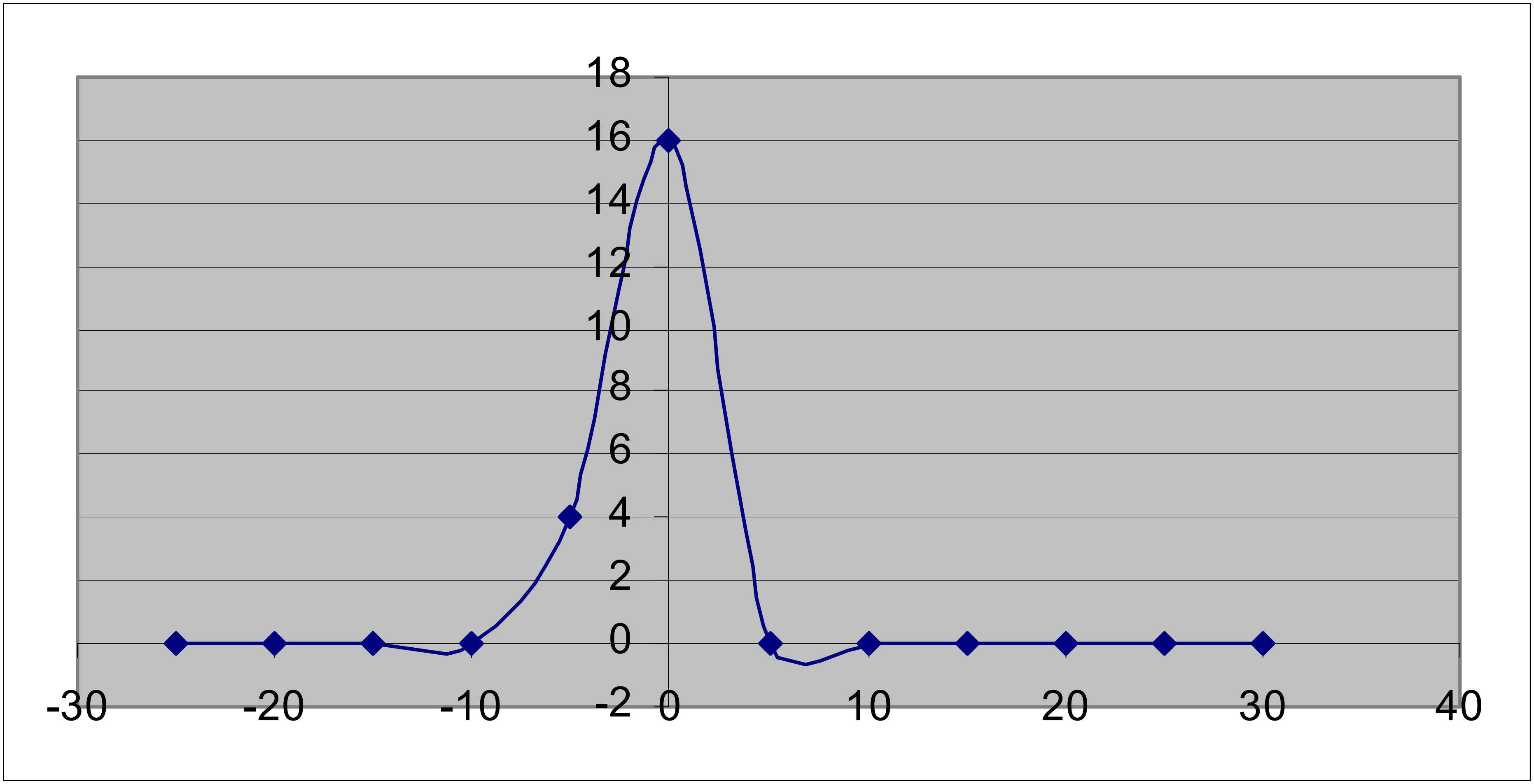


Fit with y = m.x + c

C = 0

m = 1.5

Entropy = 1.2954

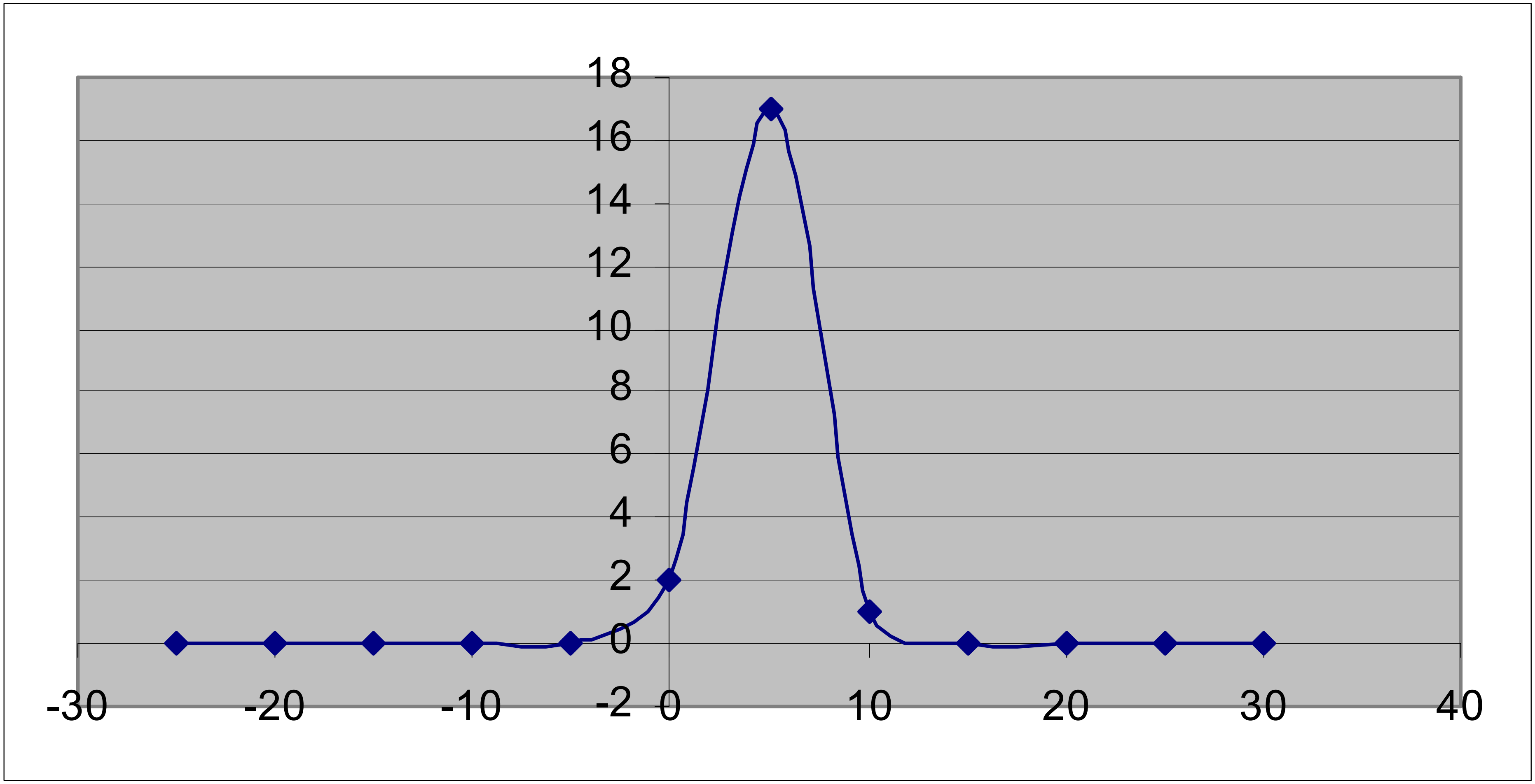


Fit with y = m.x + c

C = 0

m = 1.75

Entropy = 0.721928

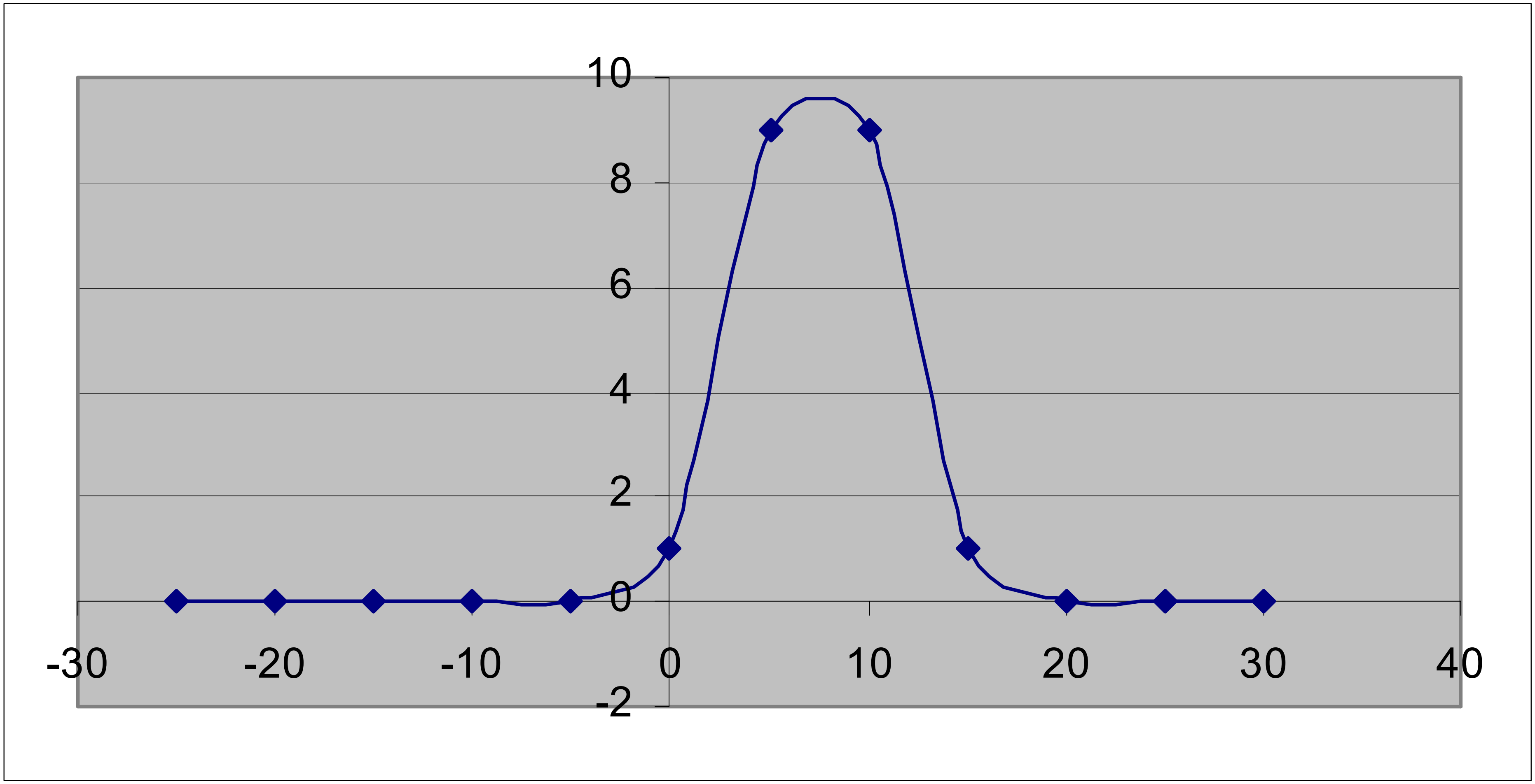


Fit with y = m.x + c

C = 0

m = 2.25

Entropy = 0.747584



Fit with y = m.x + c

C = 0

m = 2.5

Entropy = 1.468994

Validation (not correlation – correntropy)

- X: input; Z: target; and Y:prediction
- Through MEE we got Y
- How close Z and Y?
- If it is MSE then correlation
  - We want more correlation
- Since it is MEE we use correntropy
  - We want more correntropy

Correntropy between two random variables,Y and Z

$$v(Z,Y) = E_{ZY}[G_\sigma(Z - Y)] = \int\int G_\sigma(z - y)p(z,y)\,dz\,dy$$