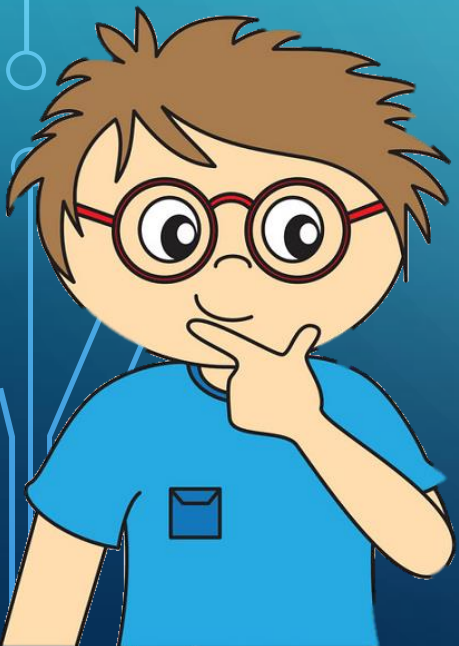


A decorative graphic on the left side of the slide, consisting of a network of white lines and circles on a blue background, resembling a circuit board or data flow diagram. The lines are vertical and horizontal, with some diagonal connections, and the circles are of varying sizes, some acting as nodes or endpoints.

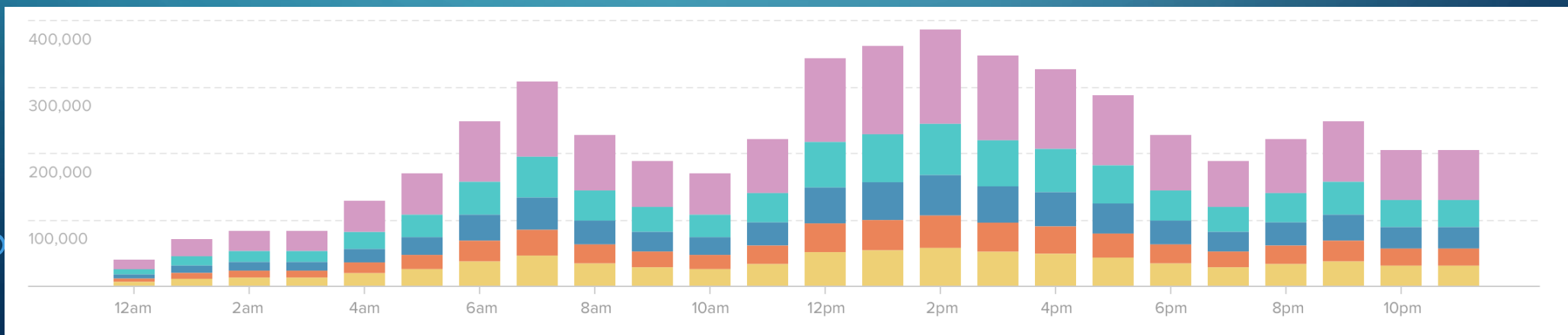
04. NUMERICAL PREDICTION

WHAT IS NUMERICAL PREDICTION?



NUMERICAL PREDICTION

- Numerical data prediction is the use of historical data to analyze trends in order to predict the future
- What can we predict?



You are traveling outside of the country but you don't know what type of clothes you should bring. What should you check?



What should farmers check to know how much of rain should they expect in specific months?



NUMERICAL PREDICTION

- The predictions can only be done by using historical data
- That data has to be analyzed so that we can know it's relationship between data sets and that we can create a prediction model that gives numerical results
- Prediction will be done using linear equations

NUMERICAL PREDICTION

- We can predict the data set of interest (y) when the other data set (x) is known to be related
- As an example we will use the data from previous class because we already know that they are related
- We have to draw a straight line through as many points in the distribution diagram as possible.
- We call this line a “Trendline”

PREDICTION FROM TRENDLINE

- We predict data in 3 steps
 - 1. Create a distribution diagram
 - 2. Draw the trendline
 - 3. Predict the value from trendline



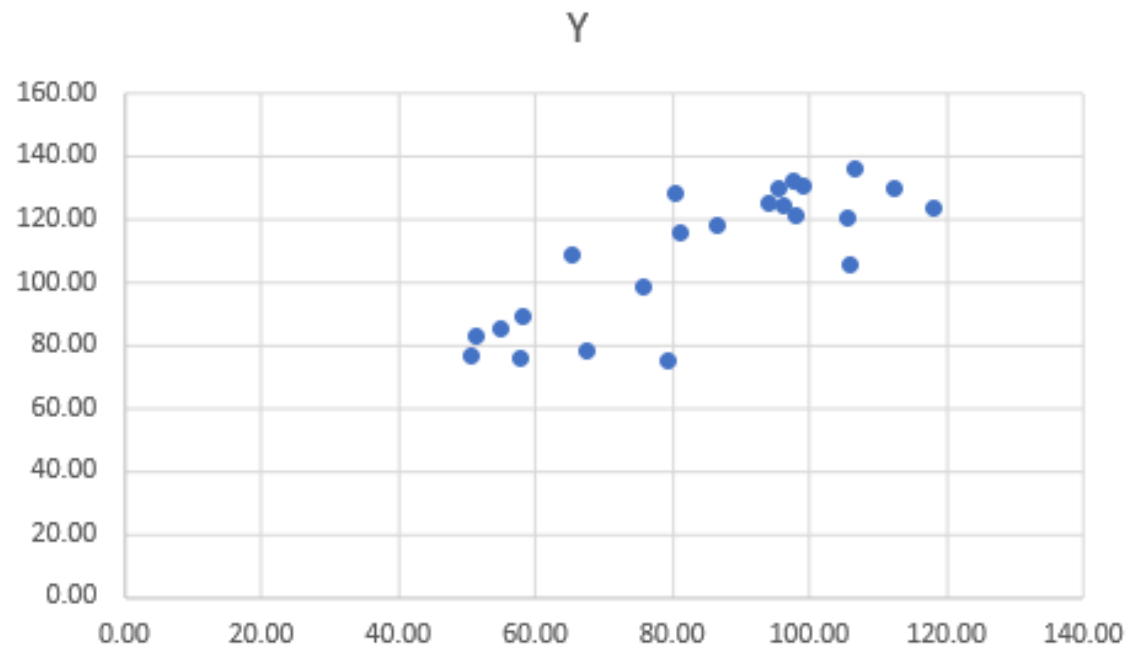
- Open this link:

<http://bit.ly/DatasetRel>

1. CREATE A DISTRIBUTION DIAGRAM

- Create a scatter chart

Scatter

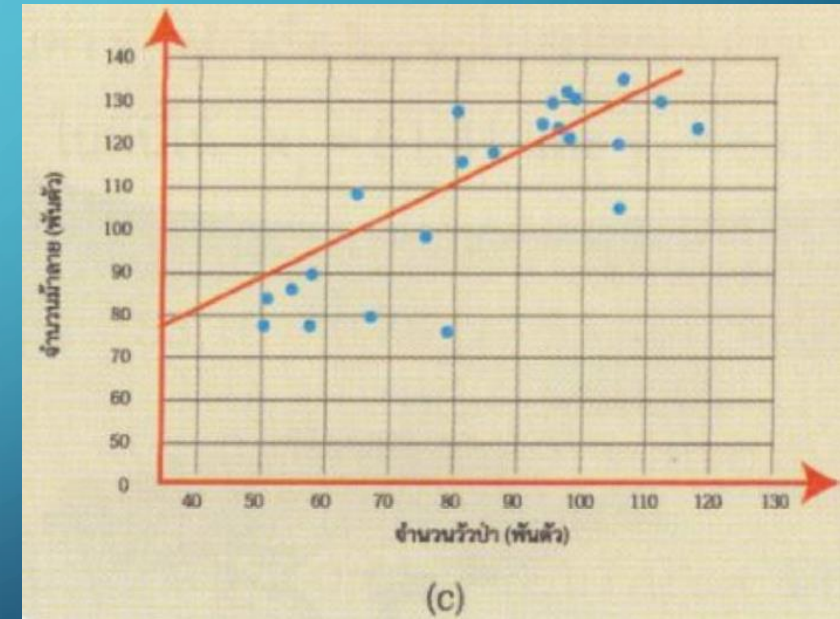
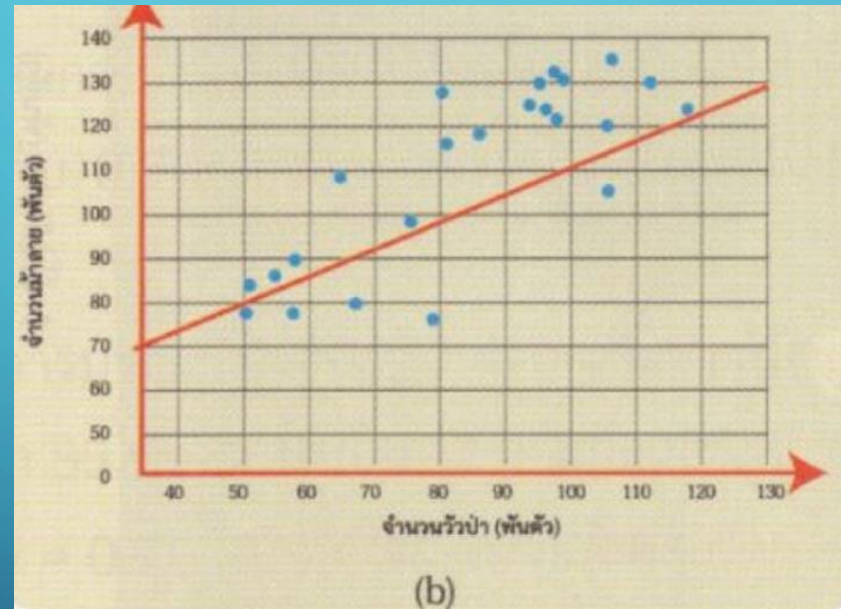
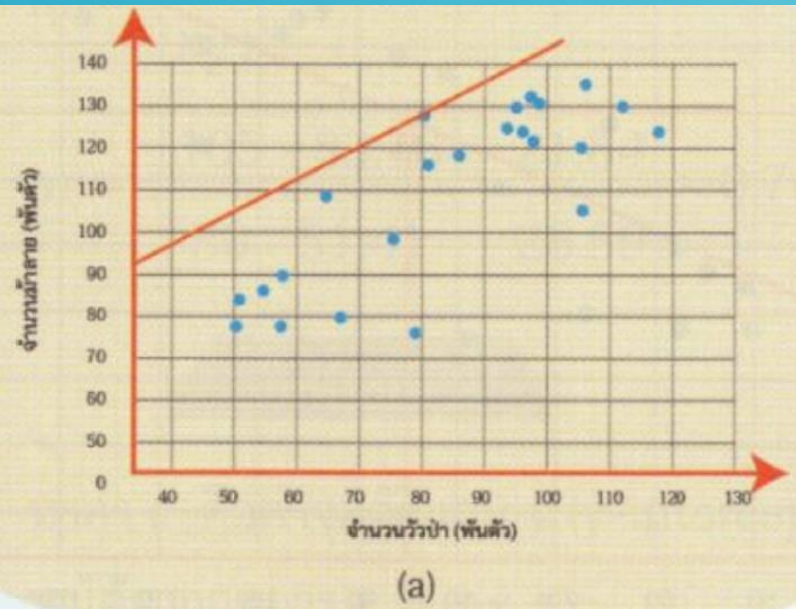


A scatter chart is used to compare at least two sets of values or pairs of data. Use it to show relationships between sets of values.

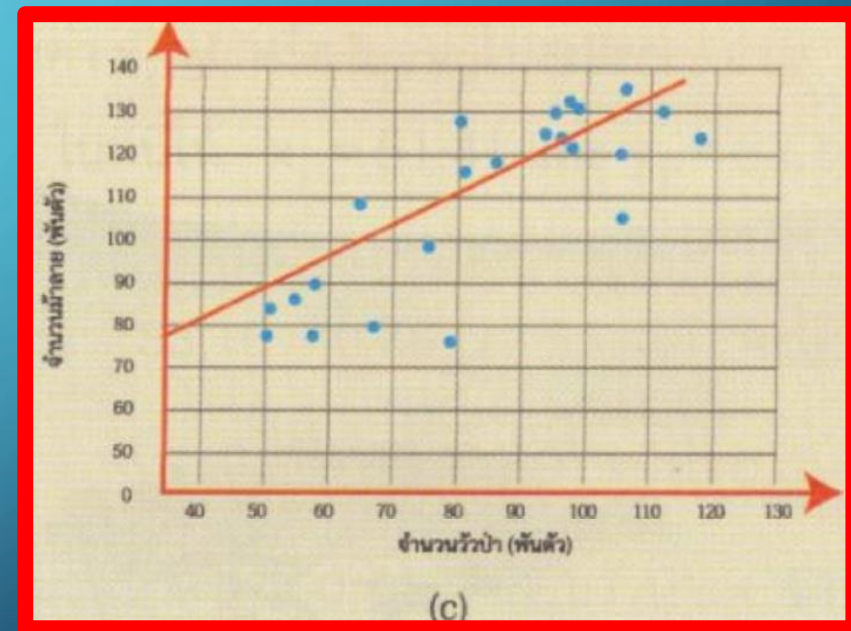
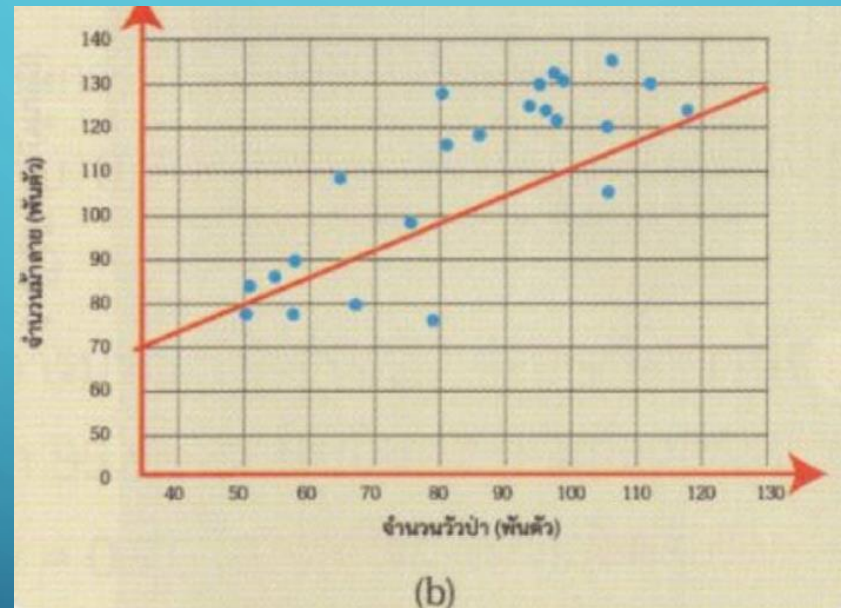
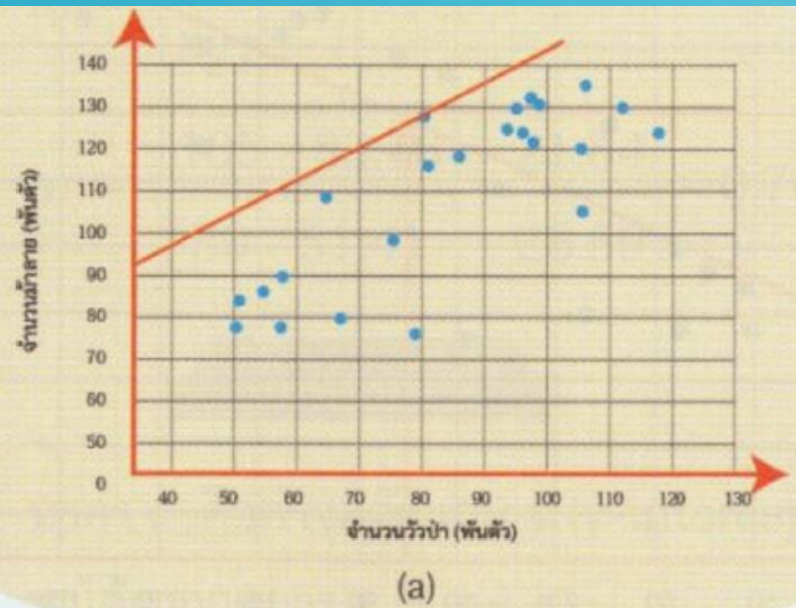
2. DRAW THE TRENDLINE

- A trendline is a line that passes through the points in the distribution diagram
- Perfect trendline will go through all the points
- Data is never perfect so we have to use a “best fit line”
- Best fit trendline is a line that should have the same number of points above and below it

WHICH ONE OF THESE IS THE BEST FIT LINE

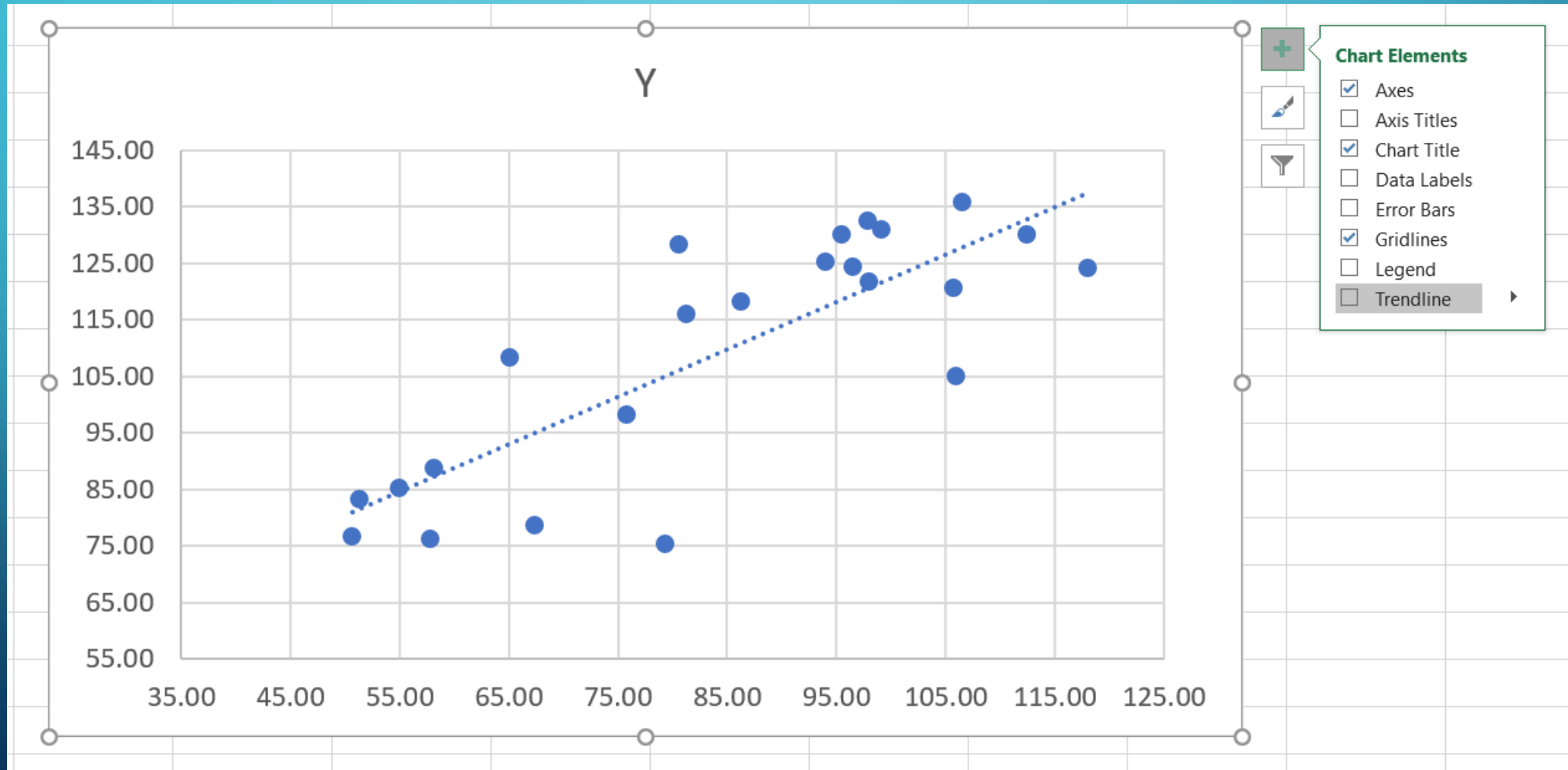


WHICH ONE OF THESE IS THE BEST FIT LINE

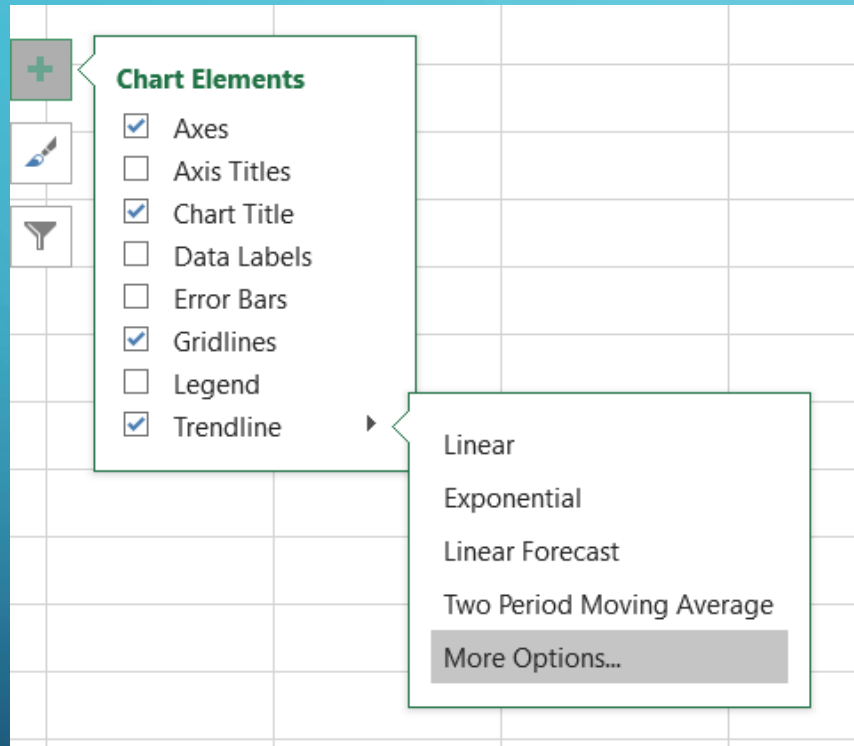


พิจารณาเส้นตรงในแต่ละภาพ ไม่มีเส้นตรงในภาพใดที่สามารถลากผ่านได้ครบทุกจุด แต่จะเห็นว่าเส้นตรงในภาพ (c) ผ่านจุดต่าง ๆ ในแผนภาพการกระจายมากกว่าภาพอื่น ๆ เส้นตรงในภาพ (a) มีจำนวนจุดที่อยู่ใต้เส้นค่อนข้างมาก ในขณะที่เส้นตรงในภาพ (b) มีจำนวนจุดที่อยู่เหนือเส้นค่อนข้างมาก ส่วนเส้นตรงในภาพ (c) มีจำนวนจุดทั้งเหนือเส้นและใต้เส้นใกล้เคียงกัน ดังนั้น เส้นตรงในภาพ (c) จึงเหมาะสมที่จะเป็นเส้นแนวโน้ม

HOW TO GET THE TRENDLINE IN EXCEL?



IF WE WANT TO CHANGE SOME THINGS



Format Trendline

Trendline Options



Trendline Options

- Exponential
- Linear
- Logarithmic
- Polynomial Order 2
- Power
- Moving Average Period 2

Trendline Name

- Automatic Linear (Y)
- Custom

Forecast

- Forward 0.0 periods
- Backward 0.0 periods

- Set Intercept 0.0
- Display Equation on chart



WHAT IS STEP 3?

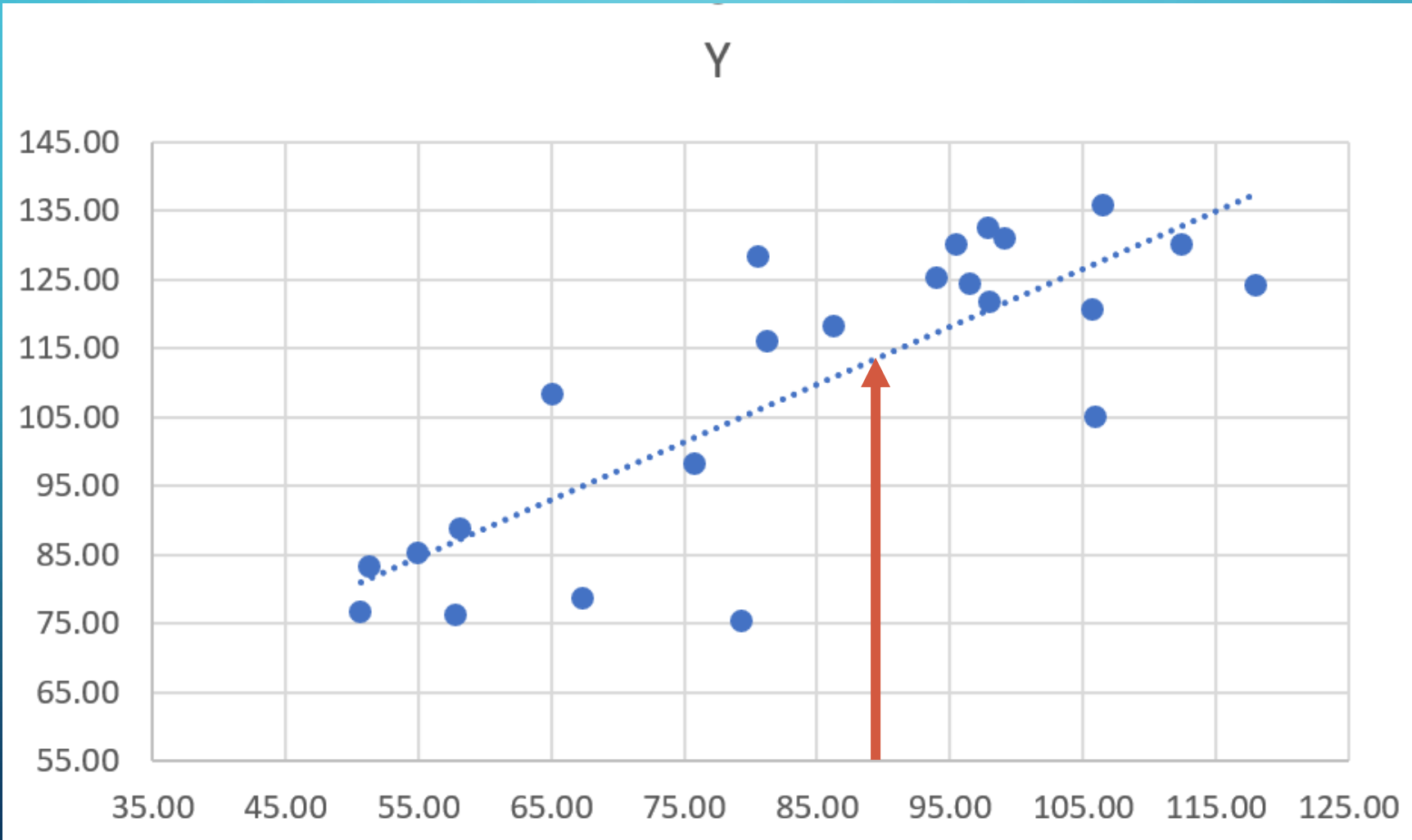
3. PREDICT THE VALUE FROM TRENDLINE

- There are two ways to predict the value
 - Estimate from the trendline
 - Predict values using linear equations

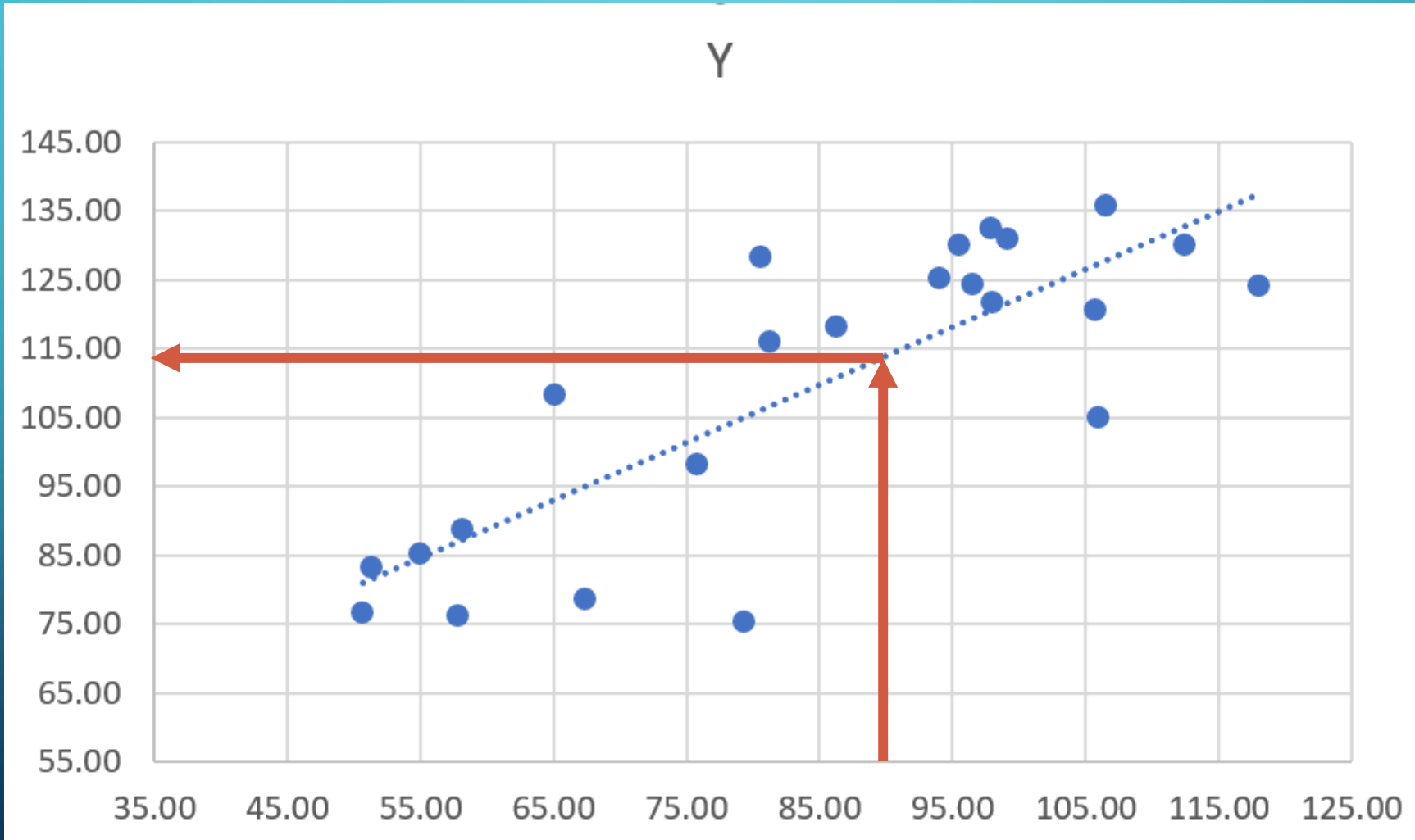
PREDICTING THE VALUE FROM THE TRENDLINE

- When we are predicting the value from the trendline usually we will be looking for Y value and we will already know the X value
- From our X value we draw a line to the trendline and then from the trendline we draw a line to Y axis where we will find Y value

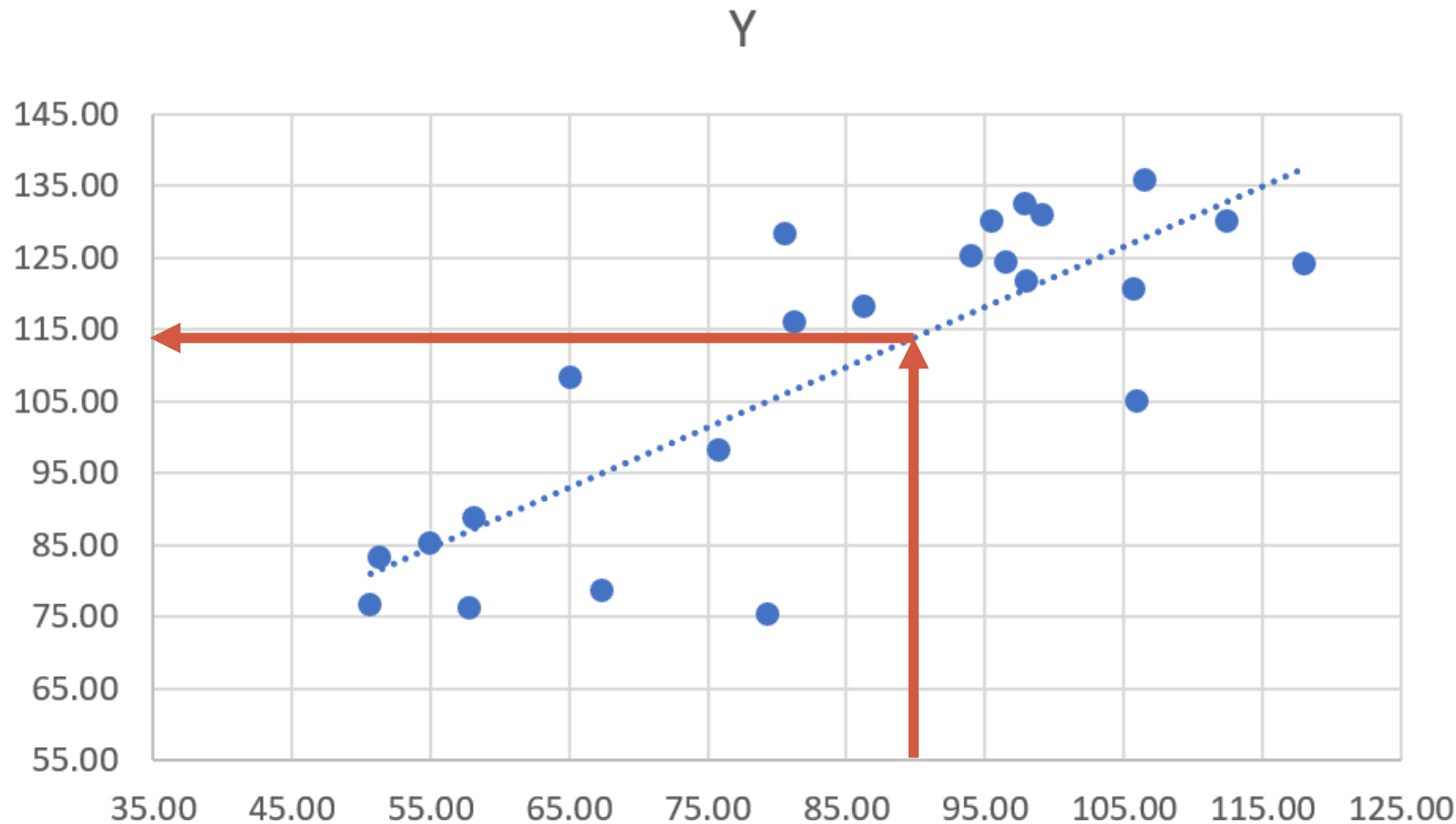
PREDICTING THE VALUE FROM THE TRENDLINE



PREDICTING THE VALUE FROM THE TRENDLINE



PREDICTING THE VALUE FROM THE TRENDLINE



From this we can see that when X is 90 the value of Y is around 114

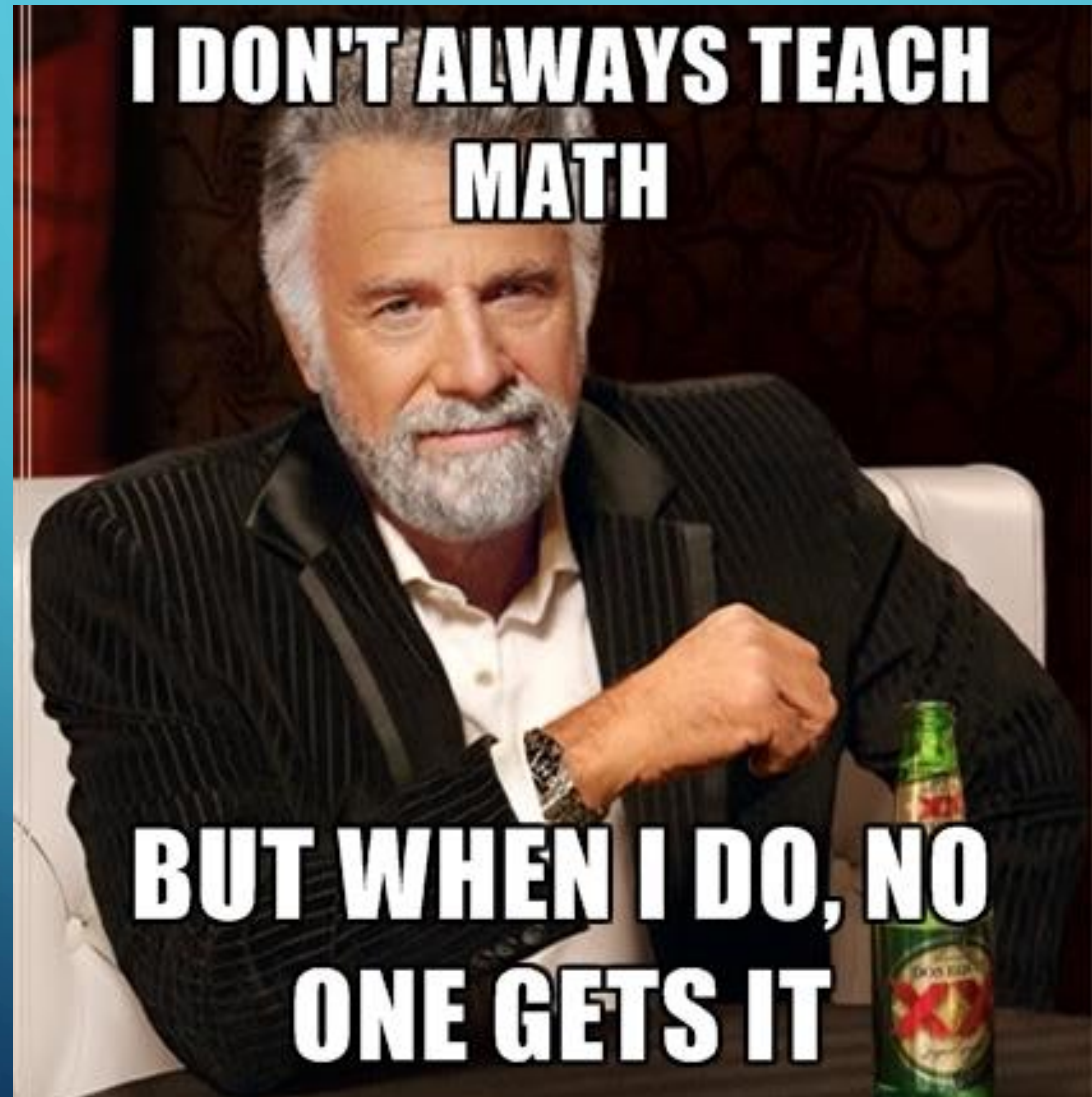
(we can not know the exact value)

PREDICT VALUES USING LINEAR EQUATIONS

- When the trendline is obtained we will find the linear equation of this trendline in the form

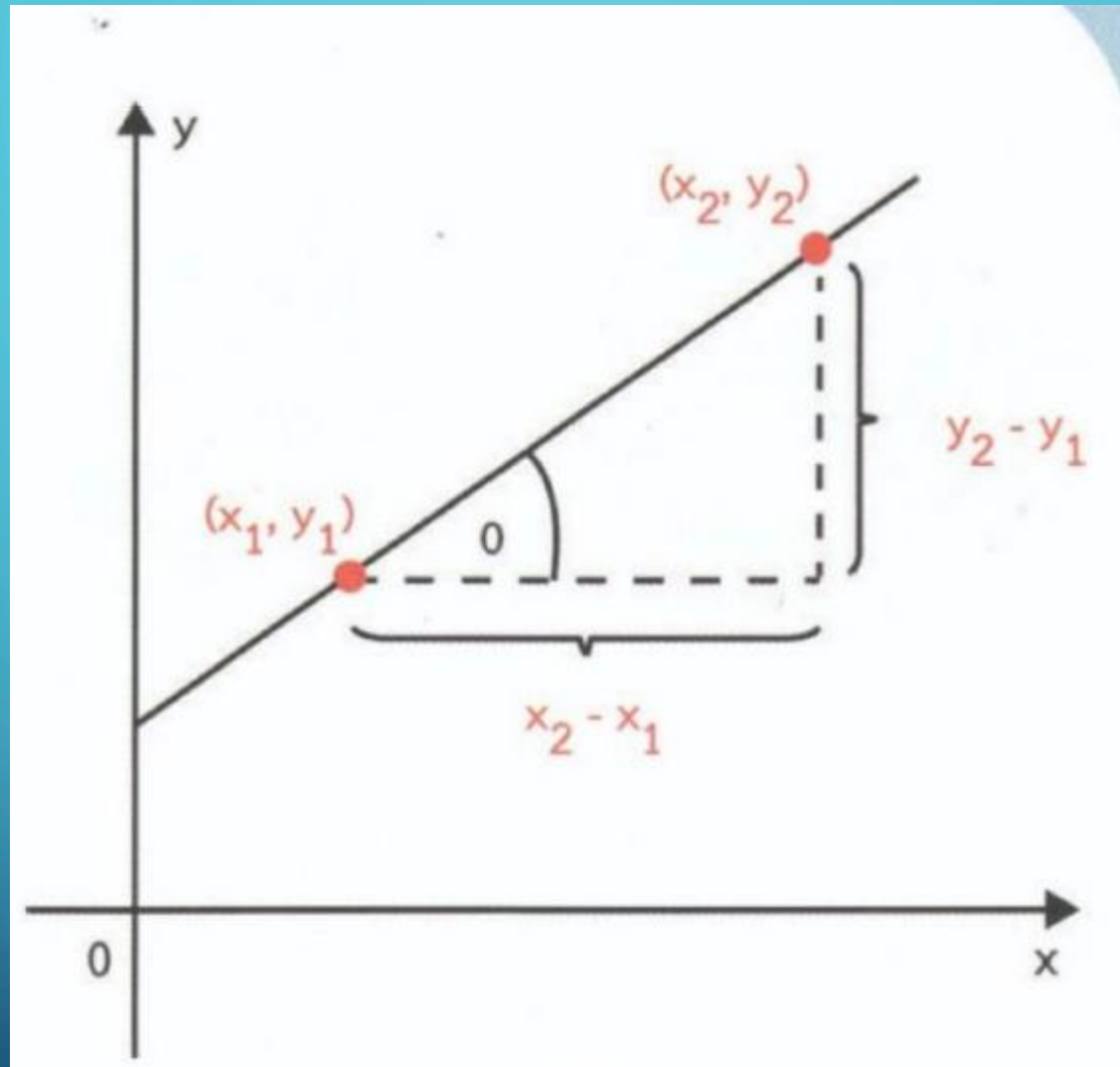
$$y = mx + C$$

MATH TIME



**I DON'T ALWAYS TEACH
MATH**

**BUT WHEN I DO, NO
ONE GETS IT**



ขั้นตอนที่ 1

หาค่าความชัน (slope) ของสมการเชิงเส้น ใน
ที่นี้คือ m

- 1 เลือกจุด 2 จุดบนเส้นตรง โดยใช้ชุดข้อมูลใน
ตารางที่ 3.1

สมมติ จุดที่ 1 (x_1, y_1) คือ (51.41, 83.16)

และ จุดที่ 2 (x_2, y_2) คือ (80, 105)

- 2 แทนค่า x และ y เพื่อหาค่าความชัน

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m = \frac{105 - 83.16}{80 - 51.41} = \frac{21.84}{28.59} = 0.76$$

ขั้นตอนที่ 2

หาค่า c คือค่าจุดตัดแกน y (y -intercept)

หาได้จากสมการ $y = mx + c$

โดยแทนค่า m ที่คำนวณได้ในขั้นตอนที่ 1 และ

แทนค่า x_1, y_1 หรือ x_2, y_2 ในสมการ

ในที่นี้ให้ $x_1 = 51.41$ และ $y_1 = 83.16$

$$y - y_1 = m(x - x_1)$$

$$y - 83.16 = 0.76(x - 51.41) \text{ จะได้}$$

$$y = 0.76x + 44.09 \leftarrow \text{ค่า } c$$

$$\text{จำนวนมัลลาย} = 0.76 (\text{จำนวนวัวป่า}) + 44.09$$

สมการเส้นตรงที่ได้ อธิบายว่า จำนวนมัลลาย (y)
เท่ากับ 0.76 เท่าของจำนวนวัวป่า (x) บวกกับ
44.09

การประมาณจำนวนมัลลาย เมื่อทราบจำนวน
วัวป่า 85 พันตัว แทนค่าจากสมการจะได้ว่า

$$y = 0.76(85) + 44.09$$

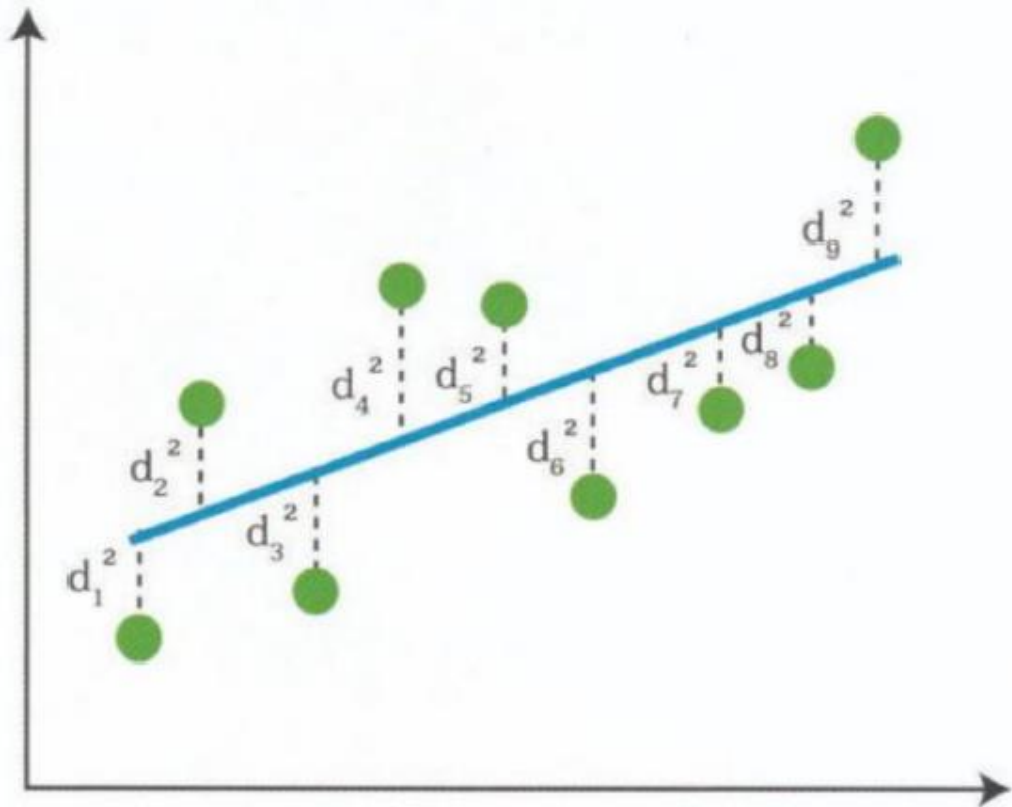
$$y = 108.69$$

จะได้ว่าจำนวนมัลลายจากการประมาณมีค่า
108.69 พันตัว เท่ากับ 108,690 ตัว

PREDICT VALUES USING LINEAR EQUATIONS

- Before applying the trendline equation that will be used for prediction we have to check the deviation of the trendline equation
- The error checking is calculated by taking the sum of the squares of the real and predicted differences (sum of squared errors)
- Subtracting the real value from the predicted value of each point and then find the sum of these squares

การคำนวณค่าความคลาดเคลื่อน



รูป 3.7 ความคลาดเคลื่อนของสมการเส้นแนวโน้ม

D แทนความคลาดเคลื่อนของสมการเส้นแนวโน้ม

d แทนผลต่างระหว่างค่าจริงและค่าทำนาย ซึ่งคำนวณได้จาก

$$y - \hat{Y}$$

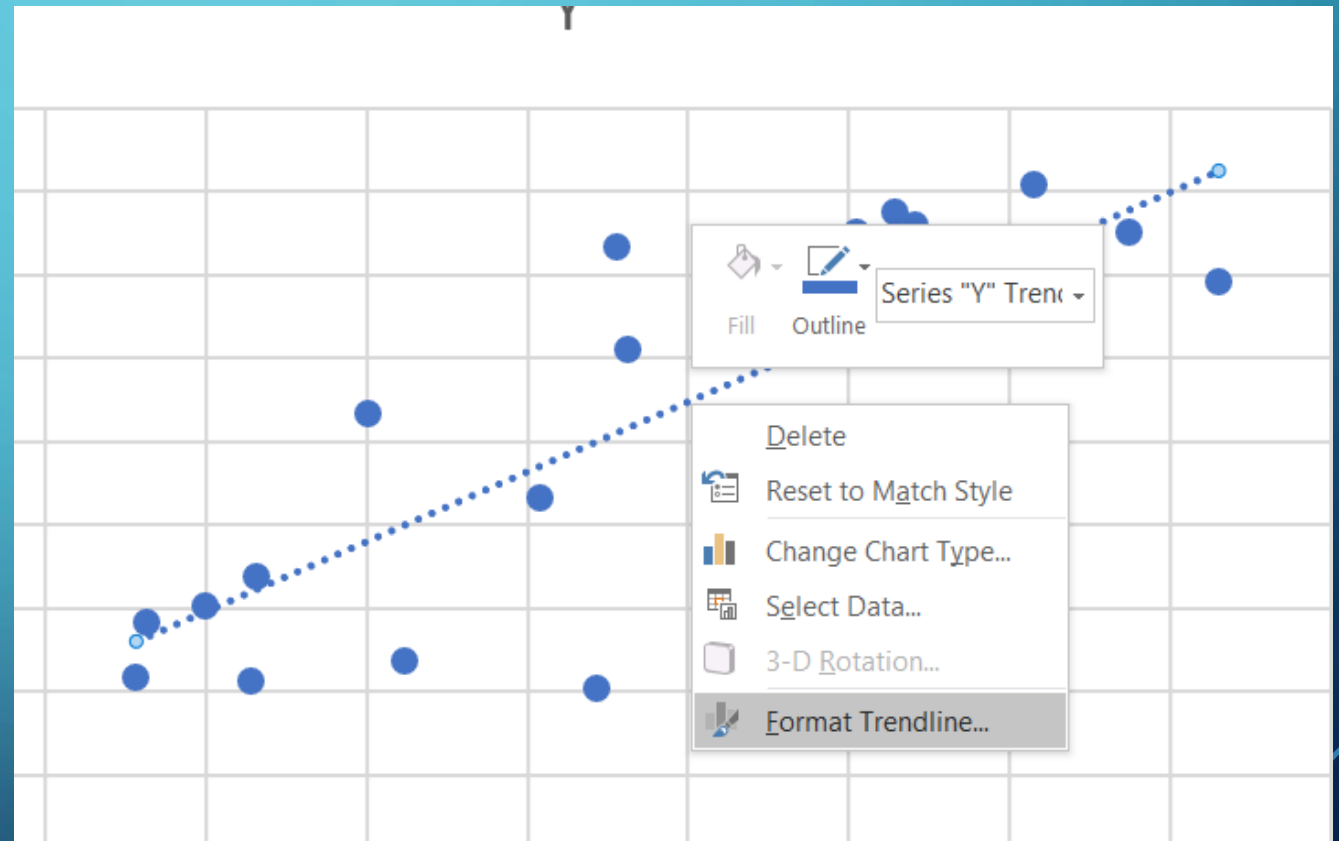
เมื่อ y คือ ค่าจริงที่อยู่ในชุดข้อมูล และ \hat{Y} (อ่านว่า วายแฮท) คือ ค่าที่ได้จากการแทนค่า x ในสมการเส้นแนวโน้ม

ดังนั้น ค่าความคลาดเคลื่อนของสมการเส้นแนวโน้มในรูป 3.7 จะเท่ากับ

$$D = d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 + d_7^2 + d_8^2 + d_9^2$$

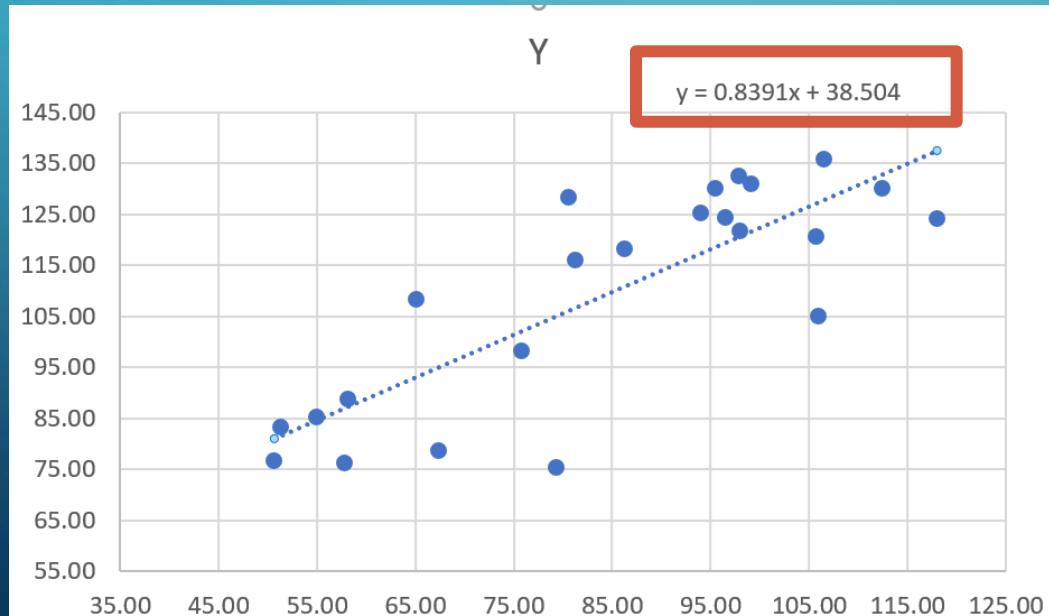
LINEAR EQUATION IN EXCEL

- We can show the linear equation ($y = mx + C$) in excel also by right clicking on the trendline and selecting format trendline



LINEAR EQUATION IN EXCEL

- On the right side we should get options for formatting
- We need to check **Display equation on chart**
- If you did correctly you will get the equation on the chart



Format Trendline

Trendline Options

- Exponential
- Linear
- Logarithmic
- Polynomial Order 2
- Power
- Moving Average Period 2

Trendline Name

- Automatic Linear (Y)
- Custom

Forecast

Forward 0.0 periods

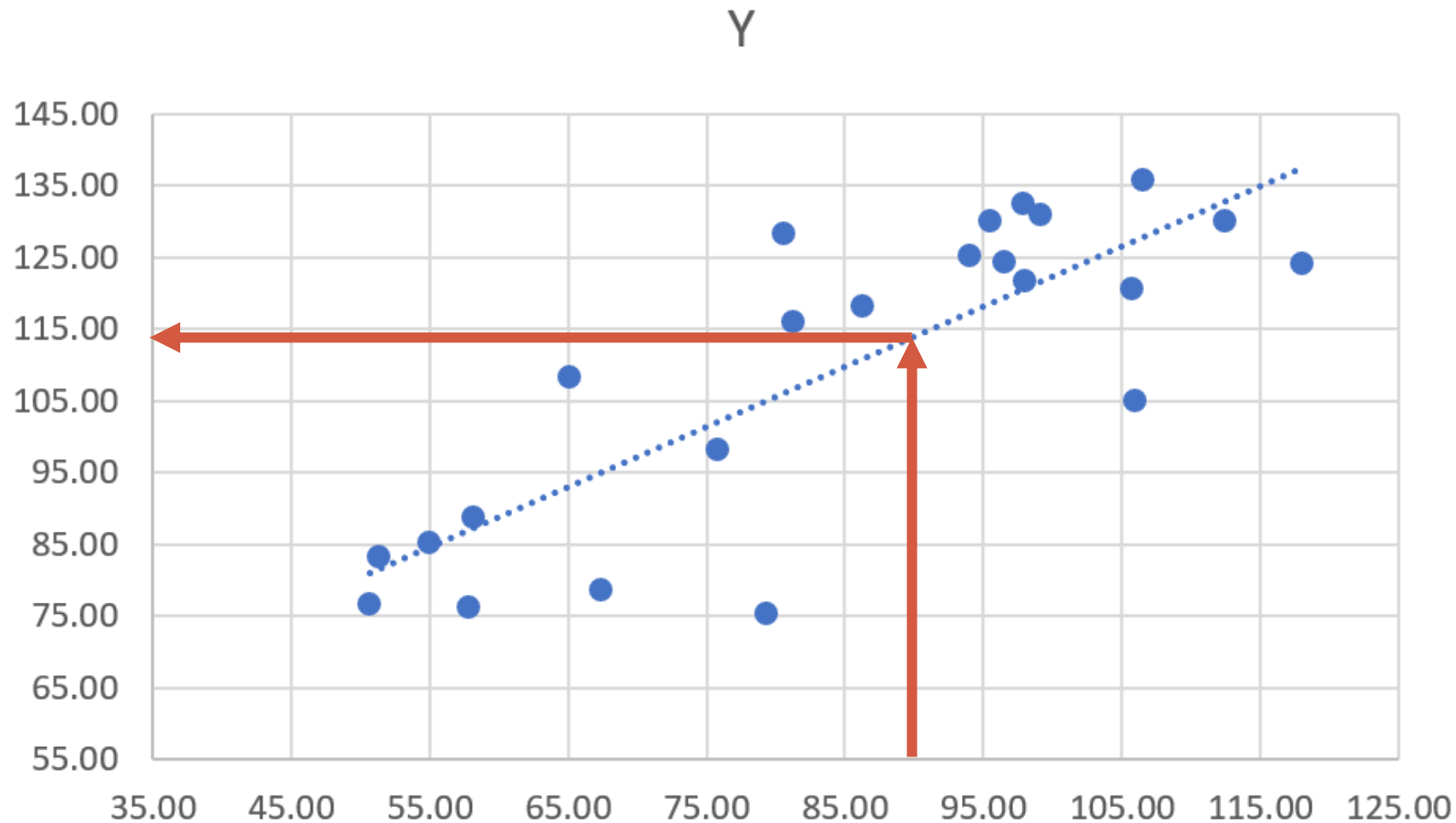
Backward 0.0 periods

Set Intercept 0.0

Display Equation on chart

Display R-squared value on chart

LET'S CHECK WHAT IS THE EXACT VALUE PREDICTING THE VALUE FROM THE TRENDLINE



From this we can see that when X is 90 the value of Y is around 114

(we can not know the exact value)

$$x = 90$$

$$y = 0.8391x + 38.504$$

$$= 0.8391 * 90 + 38.504$$

$$= 75.519 + 38.504$$

$$= 114.023$$

Our answer was only different by 0.023 which is a small difference



ANY QUESTIONS?



HOMEWORK



THE END