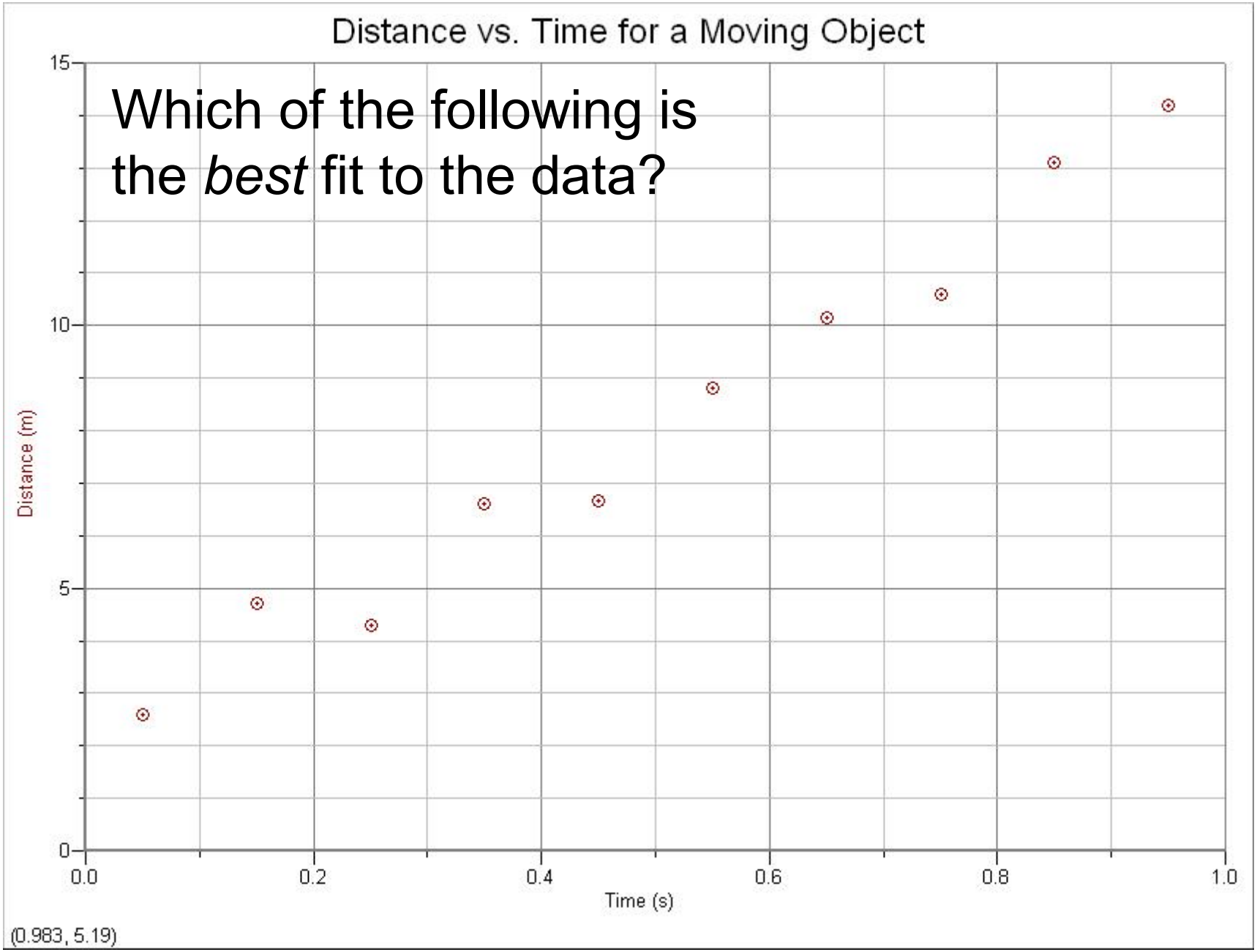
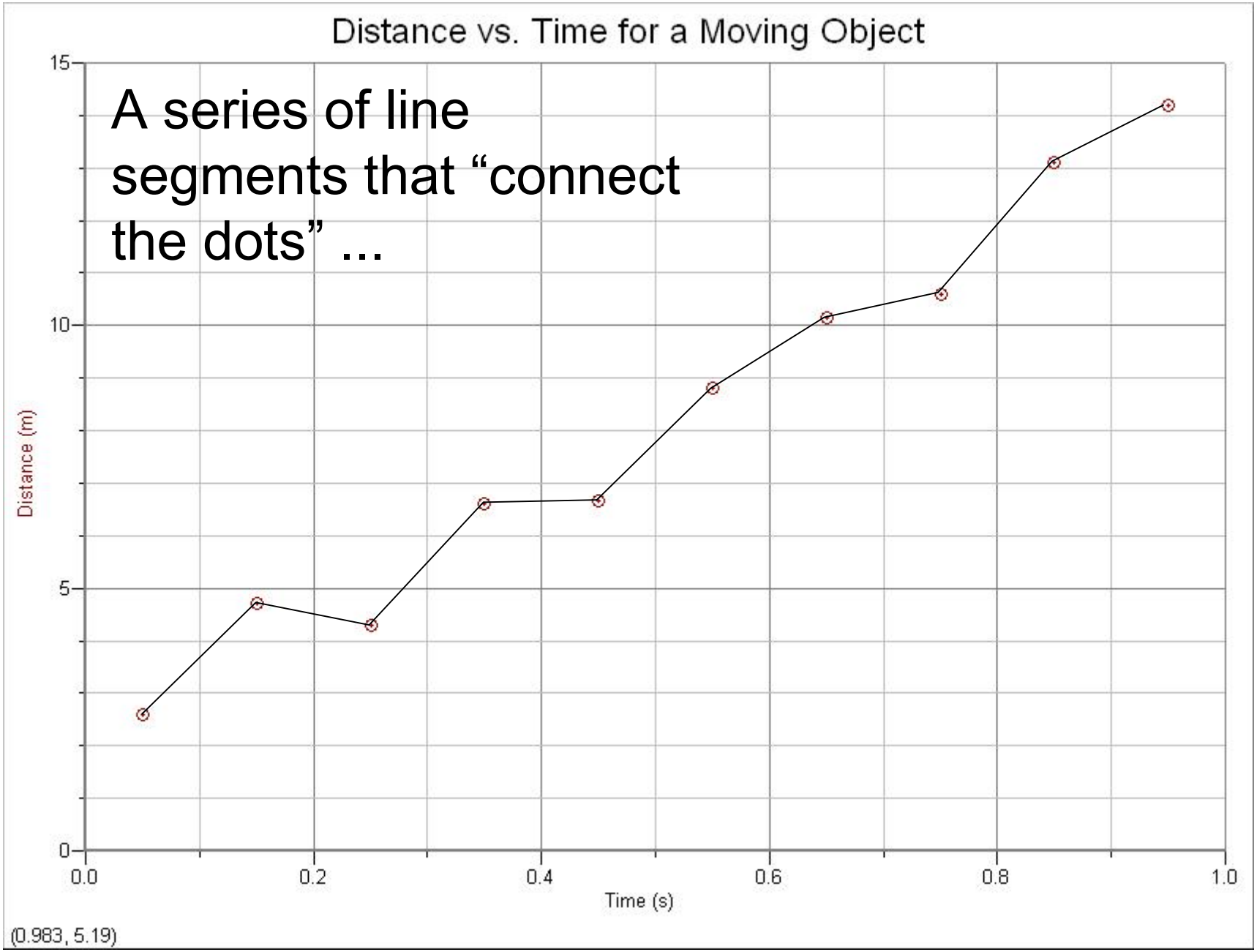
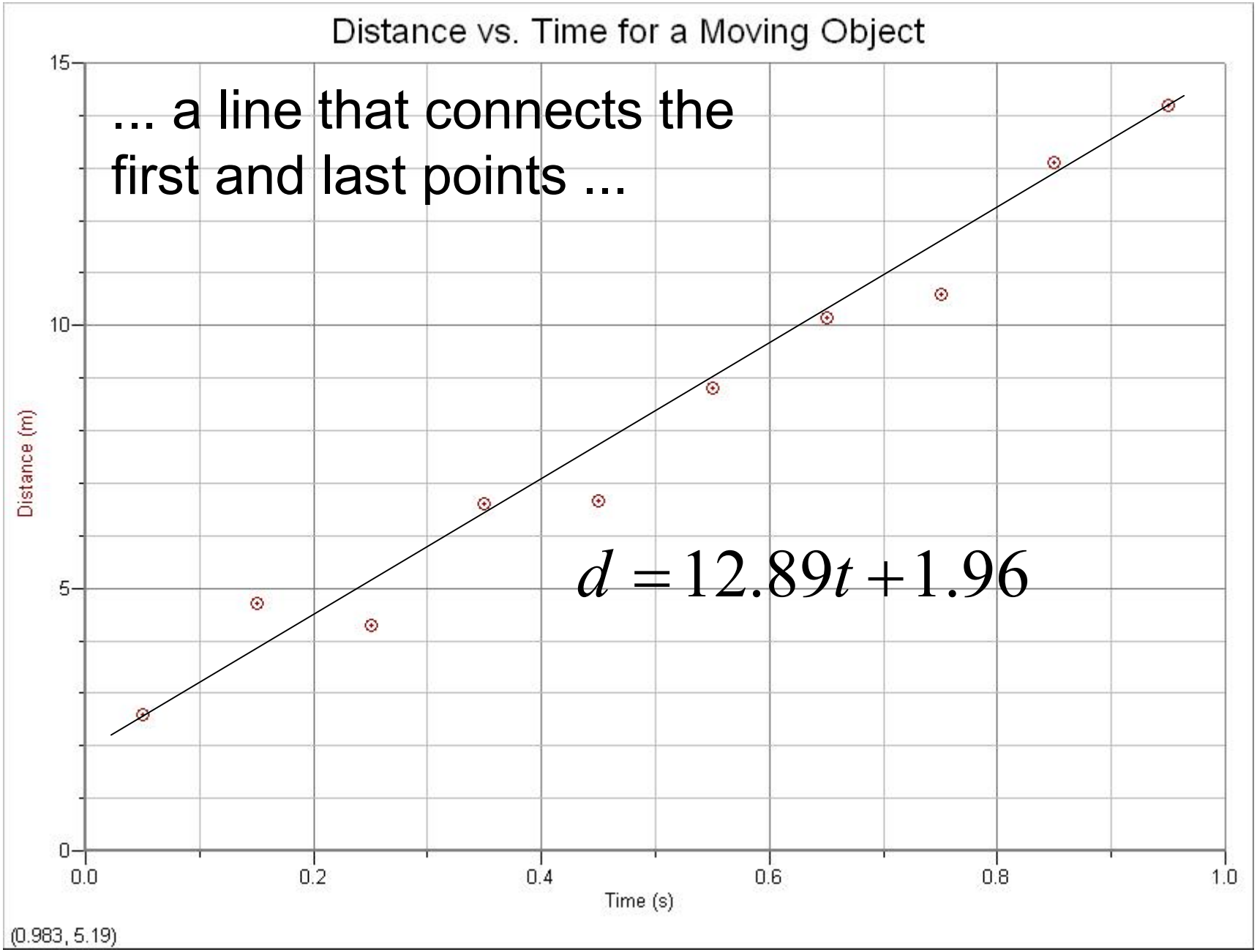
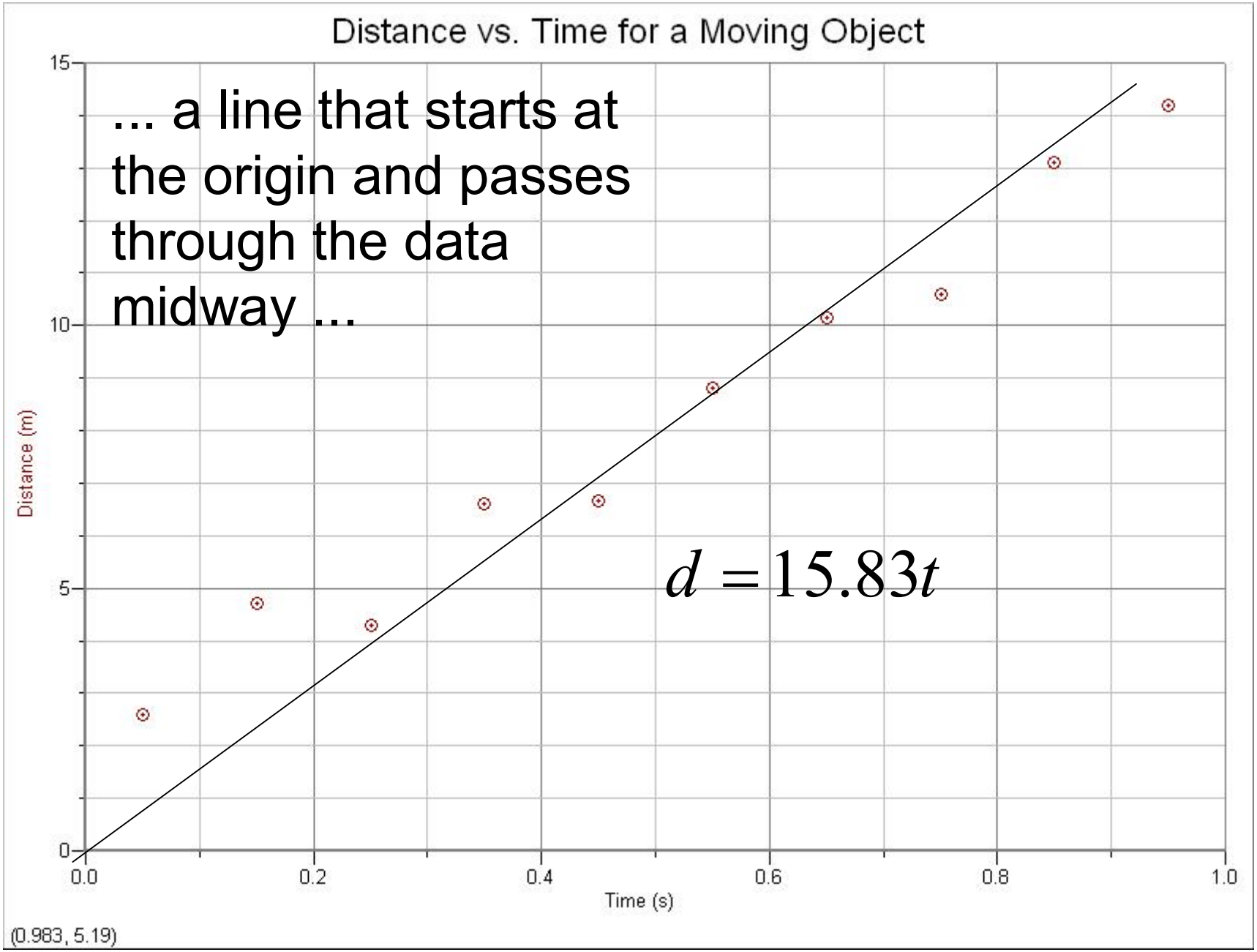
Curve Fitting

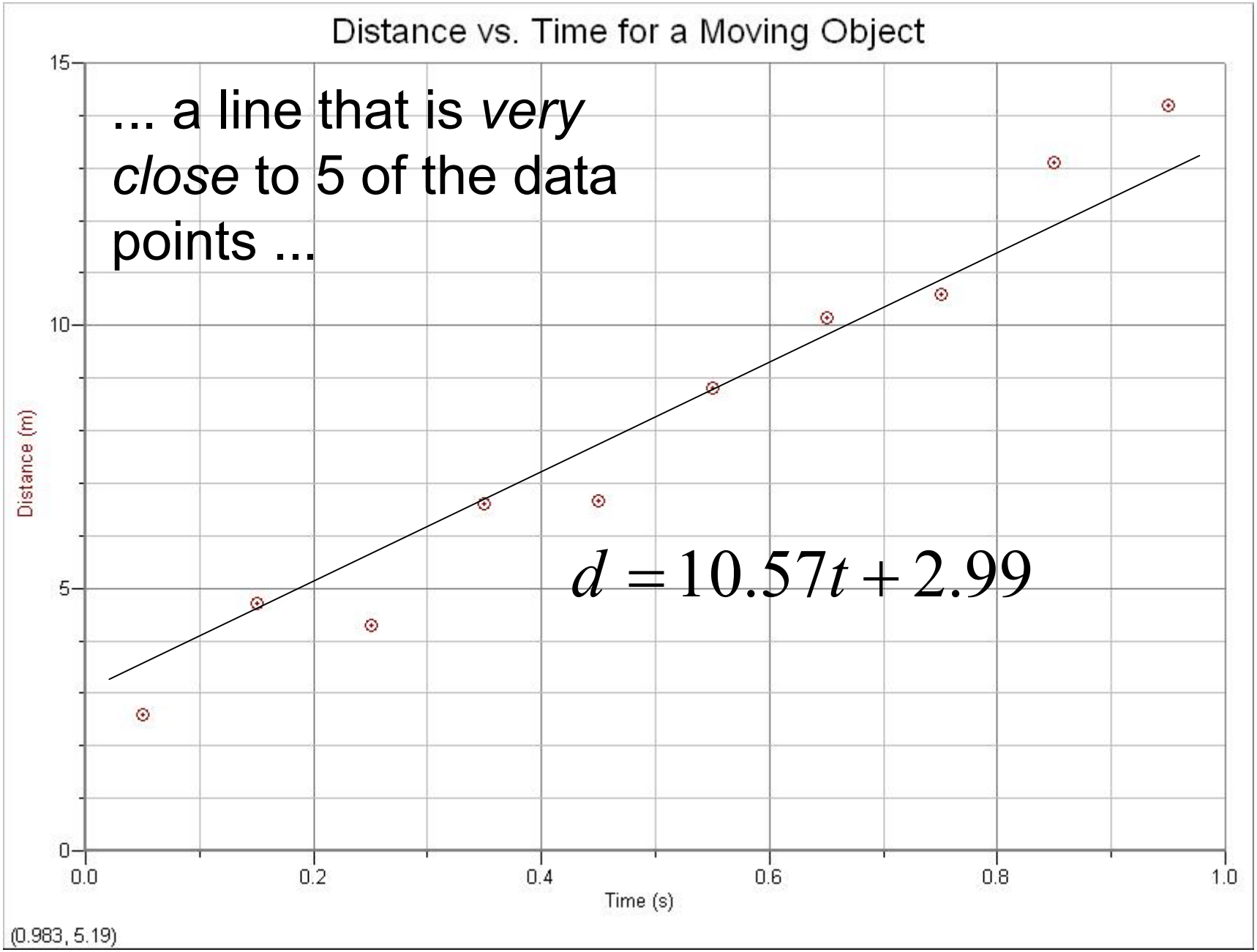
What is the “best fit”?

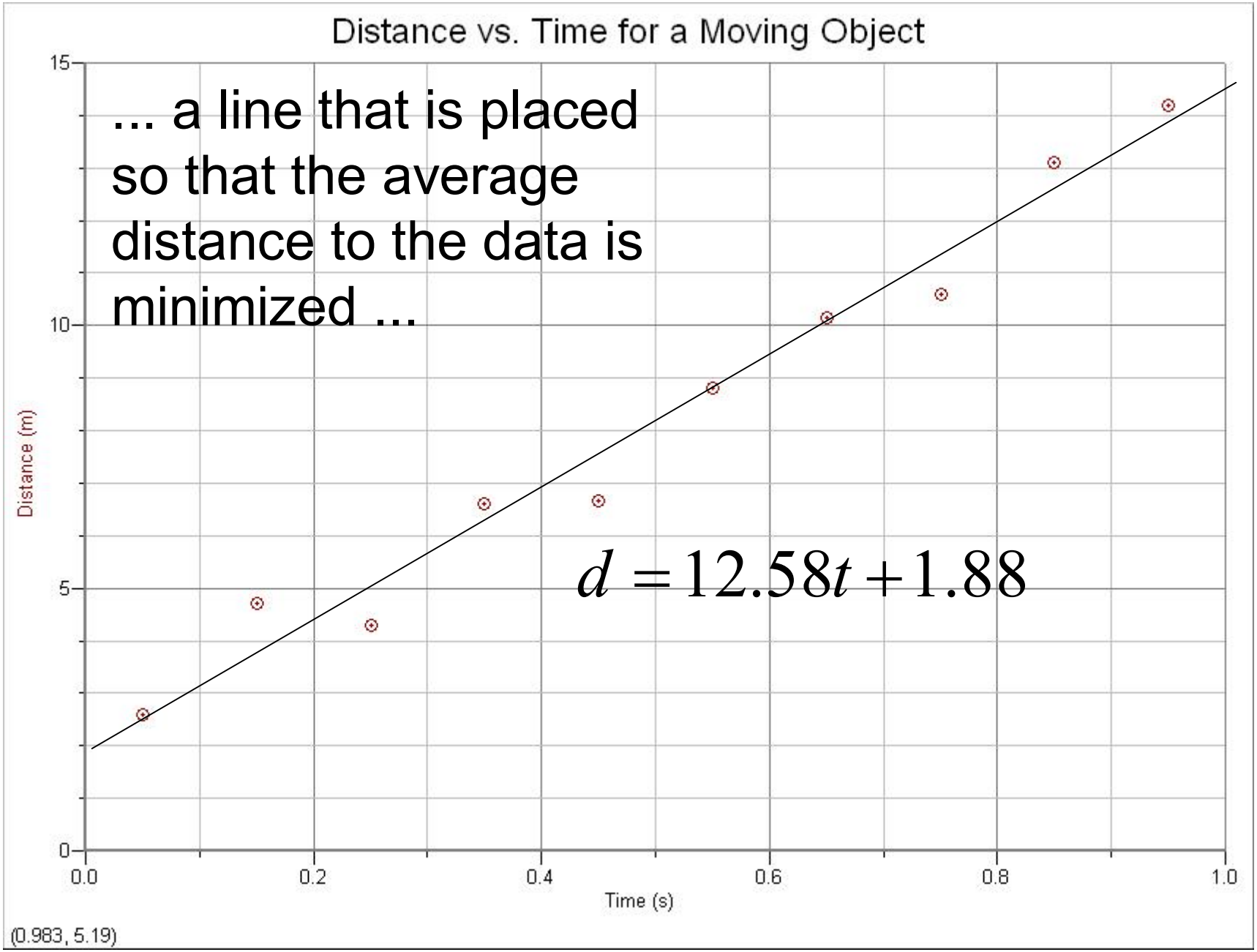

Distance vs. Time for a Moving Object
Which of the following is the *best* fit to the data?
Distance (m)
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
(0.983, 5.19)

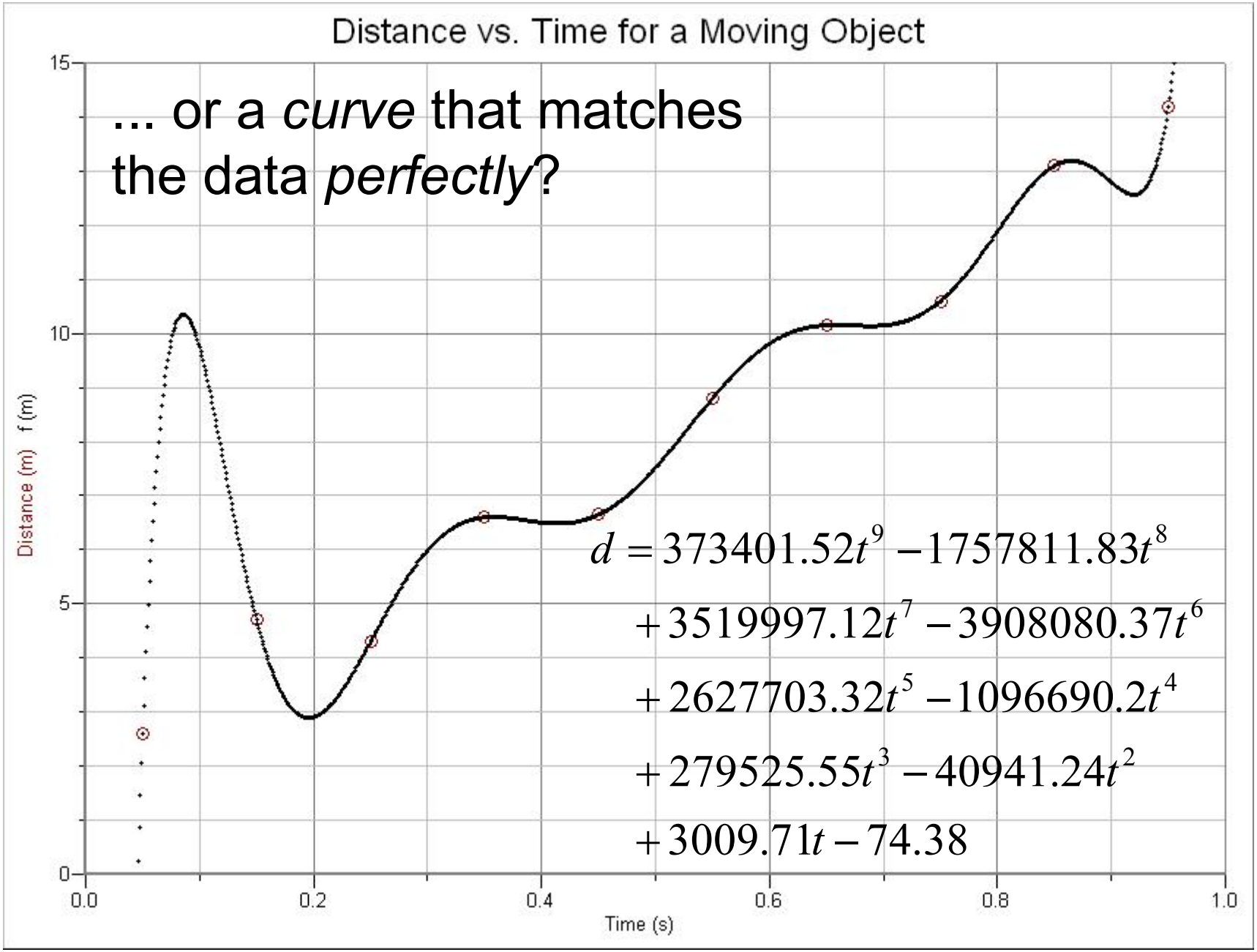

Distance vs. Time for a Moving Object
A series of line segments that “connect the dots” ...
Distance (m)
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
(0.983, 5.19)


Distance vs. Time for a Moving Object
... a line that connects the first and last points ...
$d = 12.89t + 1.96$
Distance (m)
Time (s)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
(0.983, 5.19)


Distance vs. Time for a Moving Object
... a line that starts at the origin and passes through the data midway ...
$d = 15.83t$
Distance (m)
Time (s)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
(0.983, 5.19)

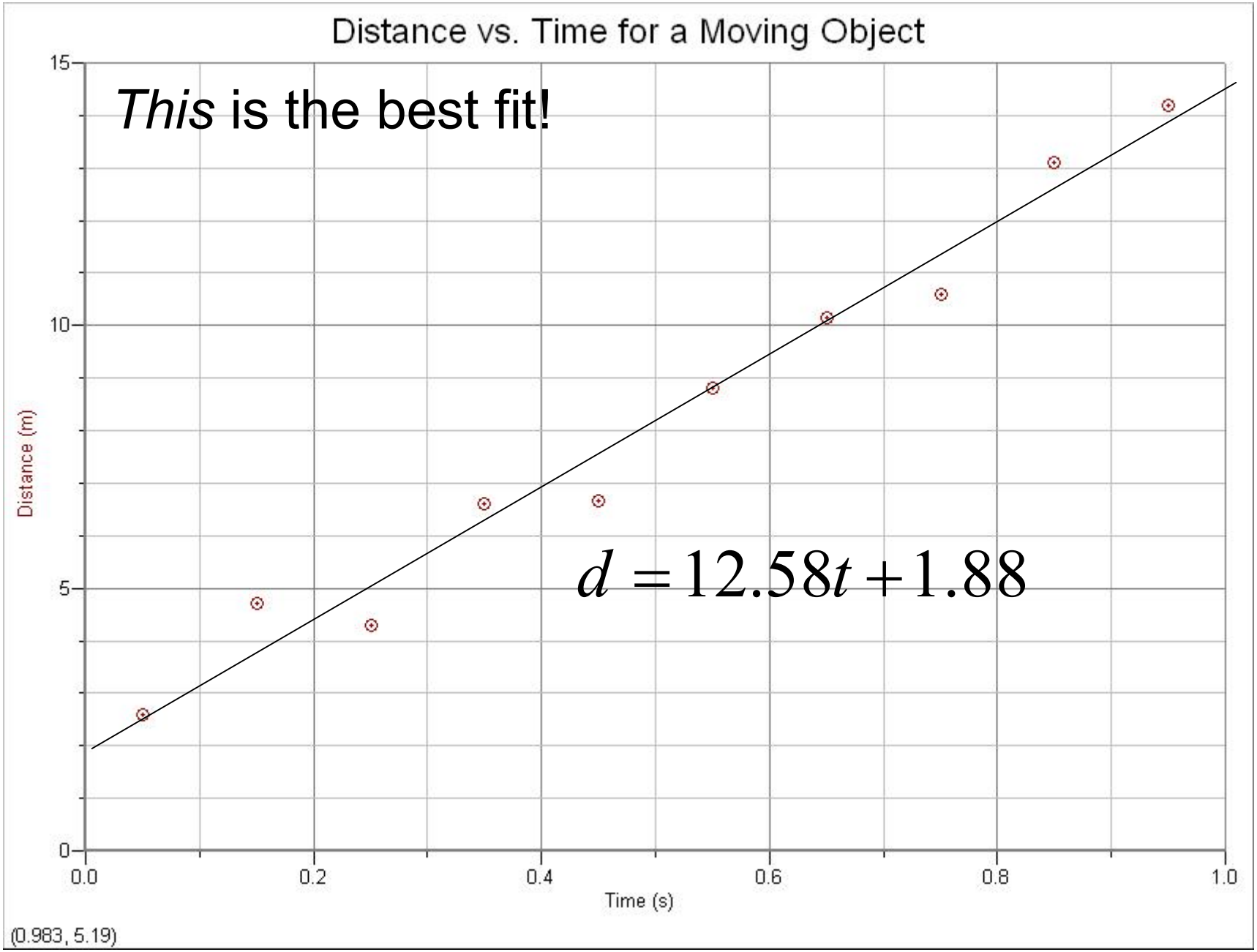

Distance vs. Time for a Moving Object
... a line that is *very close* to 5 of the data points ...
$d = 10.57t + 2.99$
Distance (m)
Time (s)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
(0.983, 5.19)


Distance vs. Time for a Moving Object
... a line that is placed so that the average distance to the data is minimized ...
$d = 12.58t + 1.88$
Distance (m)
Time (s)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
(0.983, 5.19)


Distance vs. Time for a Moving Object
... or a *curve* that matches the data *perfectly*?
$d = 373401.52t^9 - 1757811.83t^8 + 3519997.12t^7 - 3908080.37t^6 + 2627703.32t^5 - 1096690.2t^4 + 279525.55t^3 - 40941.24t^2 + 3009.71t - 74.38$
Distance (m) f (m)
Time (s)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0

Curve Fitting Basics:

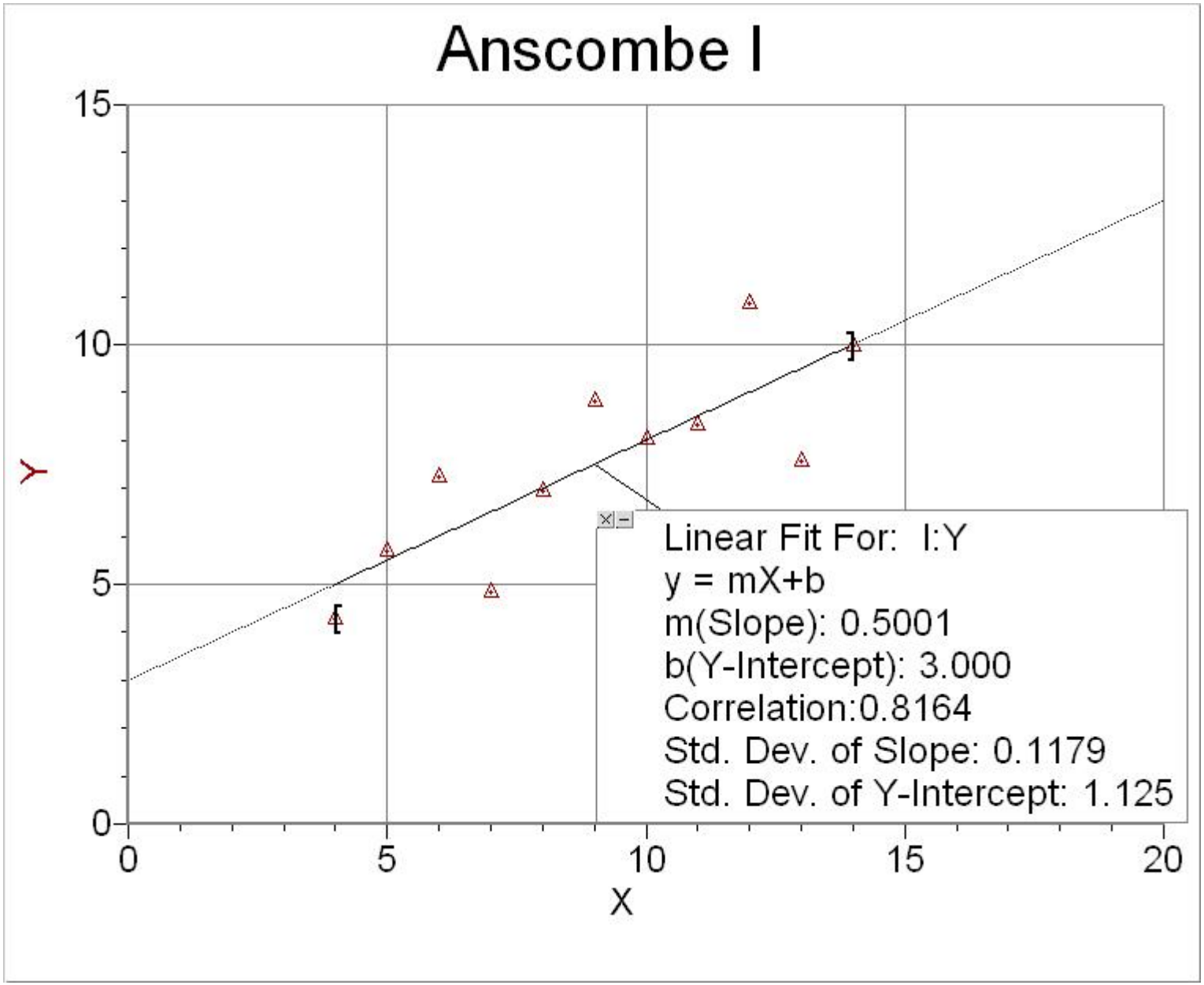
- Because experimental data *always* contain error, “connecting the dots” on a graph is seldom done in science.
- Instead a line or curve is chosen that shows the most likely *actual* relation between variables (allowing for error or scattering of the data).
- The line or curve may or *may not* actually pass through data but rather “fits” the data as well as possible.

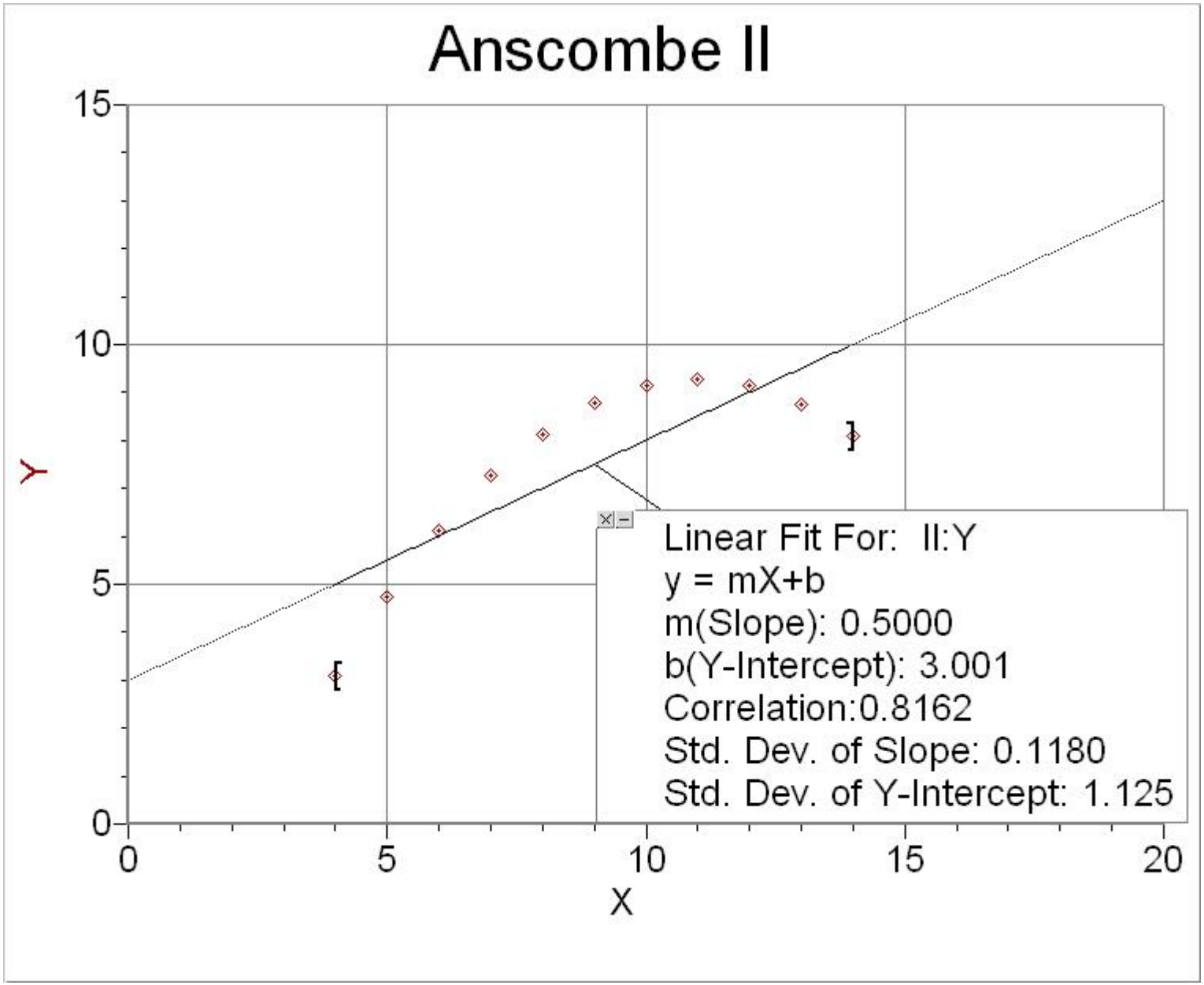

Distance vs. Time for a Moving Object
This is the best fit!
$d = 12.58t + 1.88$
Distance (m)
Time (s)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
(0.983, 5.19)

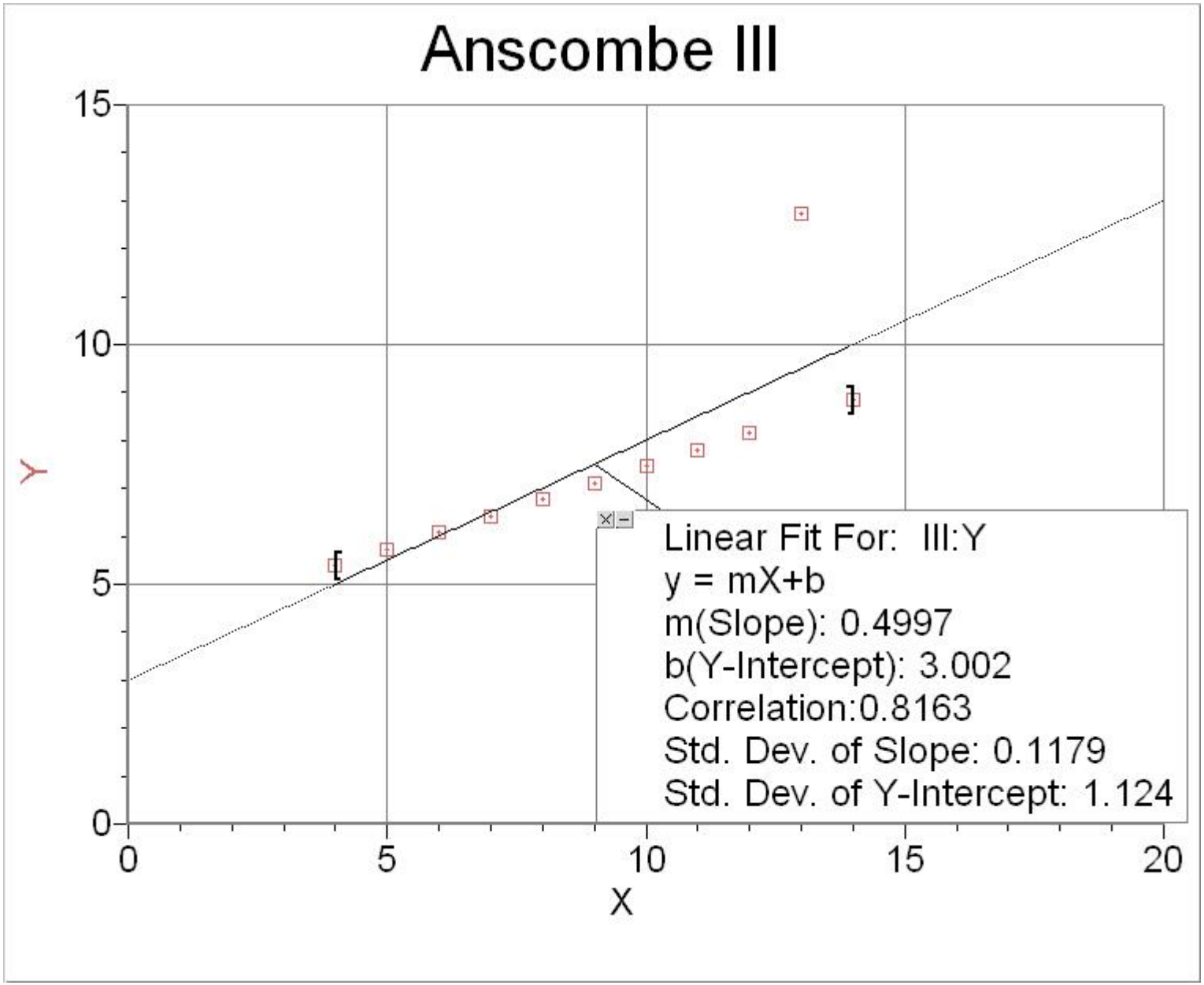
Is it better to judge the best fit by eye **or** by calculation?

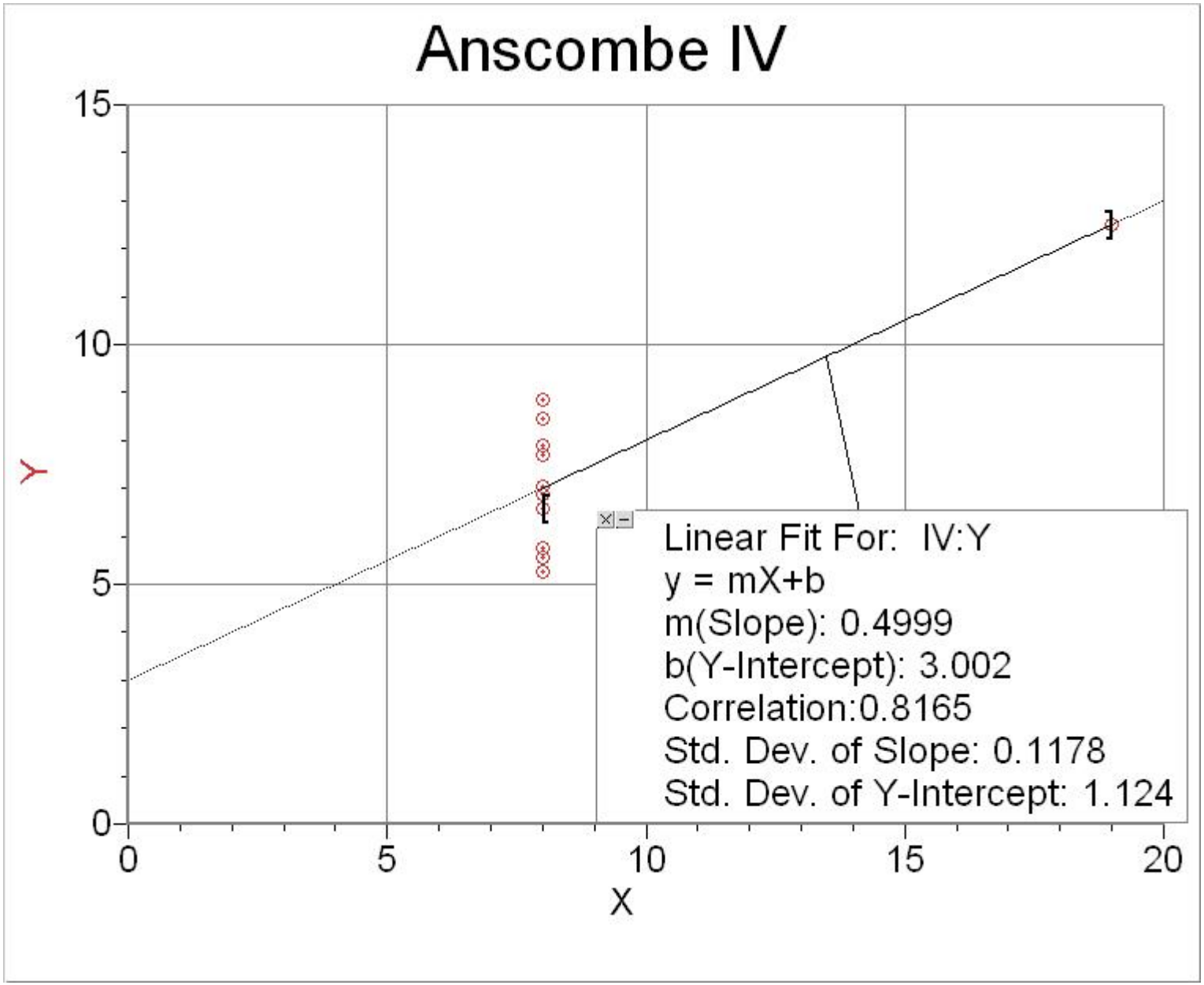
(*i.e.* by human judgment or by statistical analysis?)

Consider the following four sets of data that all have the *same* regression and statistical results. (Developed by a statistician named Anscombe to make a point...)


Anscombe I
15
10
5
0
Y
0
5
10
15
20
X
Linear Fit For: I:Y
y = mX+b
m(Slope): 0.5001
b(Y-Intercept): 3.000
Correlation:0.8164
Std. Dev. of Slope: 0.1179
Std. Dev. of Y-Intercept: 1.125


Anscombe II
15
10
5
0
Y
0
5
10
15
20
X
Linear Fit For: II:Y
y = mX+b
m(Slope): 0.5000
b(Y-Intercept): 3.001
Correlation:0.8162
Std. Dev. of Slope: 0.1180
Std. Dev. of Y-Intercept: 1.125

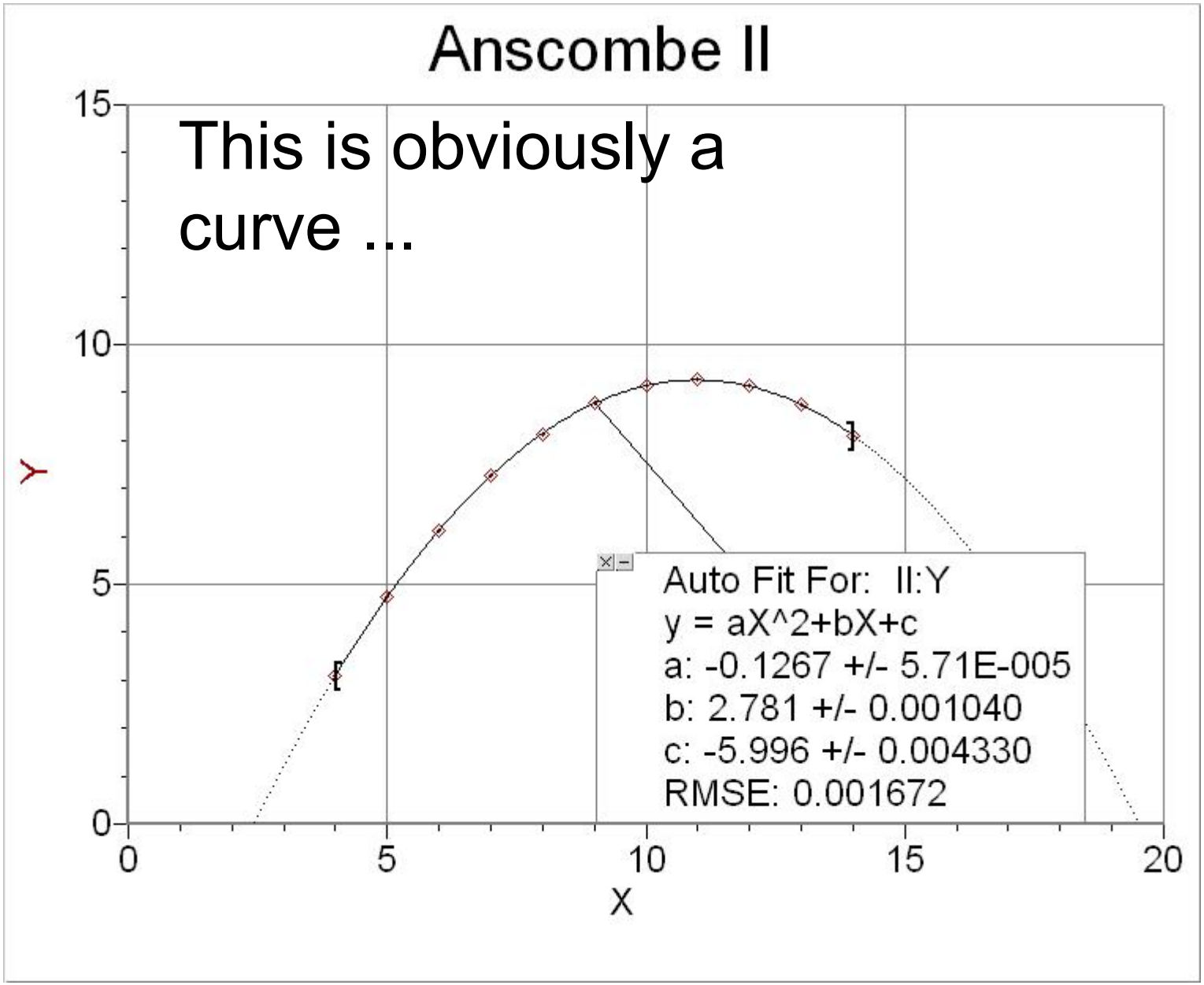

Anscombe III
15
10
5
0
Y
0
5
10
15
20
X
Linear Fit For: III:Y
y = mX+b
m(Slope): 0.4997
b(Y-Intercept): 3.002
Correlation:0.8163
Std. Dev. of Slope: 0.1179
Std. Dev. of Y-Intercept: 1.124

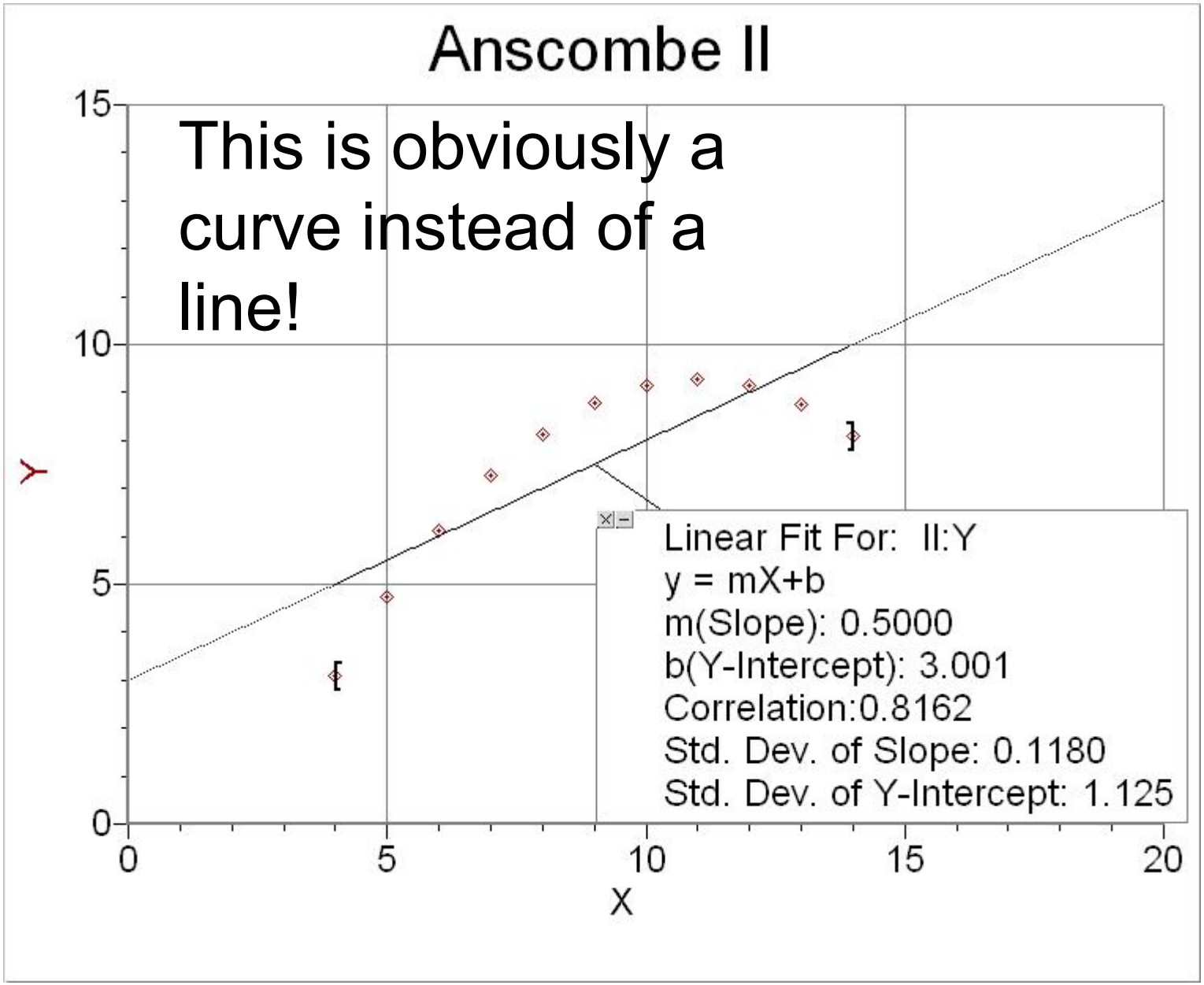

Anscombe IV
15
10
5
0
Y
0
5
10
15
20
X
Linear Fit For: IV:Y
y = mX+b
m(Slope): 0.4999
b(Y-Intercept): 3.002
Correlation:0.8165
Std. Dev. of Slope: 0.1178
Std. Dev. of Y-Intercept: 1.124

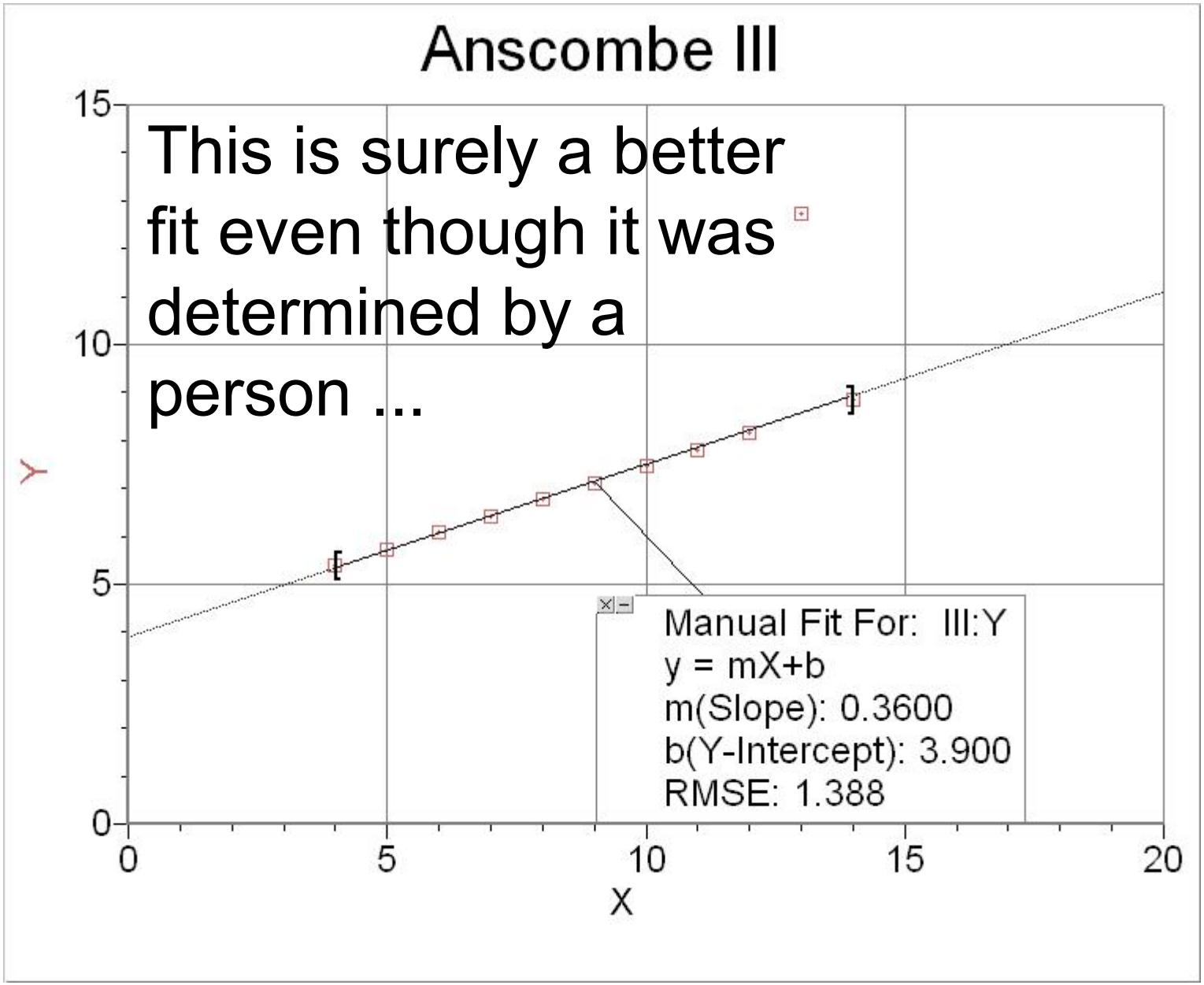
Do not underestimate the ability of your mind's eye to recognize patterns!

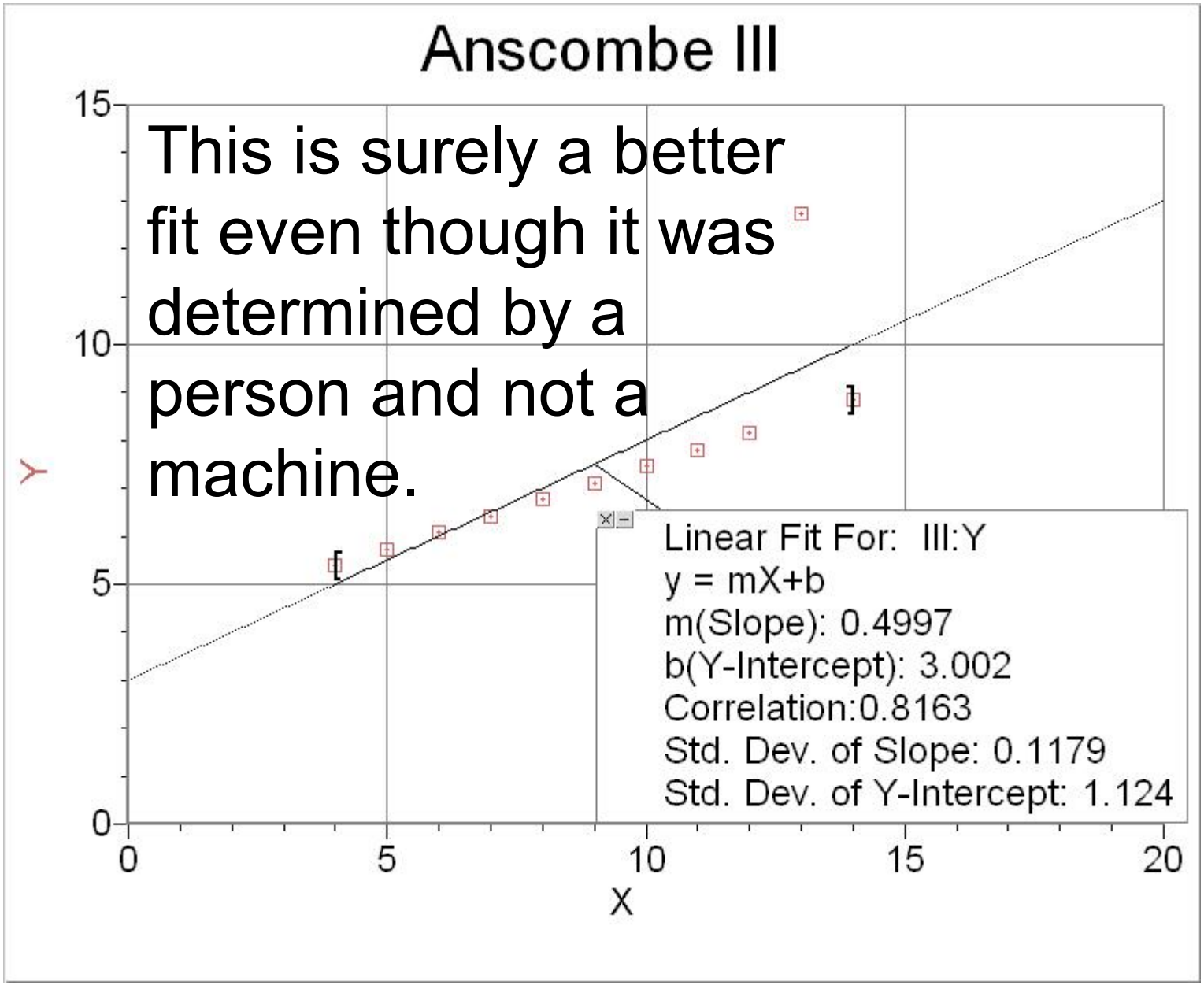
Based on statistics and number crunching all four fits are equal.

But, the "mechanical" statistical result would be appropriate only for the first of the "Anscombe Quartet".


Anscombe II
This is obviously a curve ...
Auto Fit For: II:Y
y = aX^2+bX+c
a: -0.1267 +/- 5.71E-005
b: 2.781 +/- 0.001040
c: -5.996 +/- 0.004330
RMSE: 0.001672
Y
X
0
5
10
15
20


Anscombe II
This is obviously a curve instead of a line!
Linear Fit For: II:Y
y = mX+b
m(Slope): 0.5000
b(Y-Intercept): 3.001
Correlation:0.8162
Std. Dev. of Slope: 0.1180
Std. Dev. of Y-Intercept: 1.125
Y
X
0
5
10
15
20


Anscombe III
This is surely a better fit even though it was determined by a person ...
Manual Fit For: III:Y
y = mX+b
m(Slope): 0.3600
b(Y-Intercept): 3.900
RMSE: 1.388
Y
X
0
5
10
15
20


Anscombe III
This is surely a better fit even though it was determined by a person and not a machine.
Linear Fit For: III:Y
y = mX+b
m(Slope): 0.4997
b(Y-Intercept): 3.002
Correlation:0.8163
Std. Dev. of Slope: 0.1179
Std. Dev. of Y-Intercept: 1.124
Y
X
0
5
10
15
20