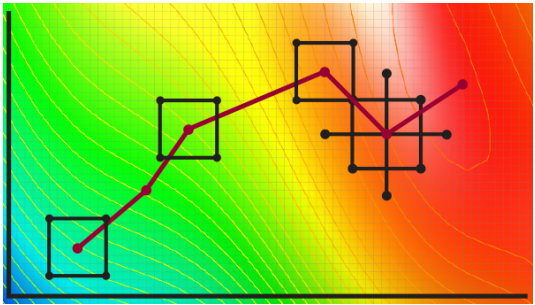
Experimentation for Improvement



http://learnche.org/

Design and Analysis of Experiments

Copyright, sharing, and attribution notice

The critical concepts covered in this video using the popcorn case study

1. Real-world units and coded units
2. **Linear vs nonlinear systems**
3. Prediction models are wrong, but still useful
4. **Noise and error** and the need for replicated experiments
5. How to systematically reach an optimum
6. Justifying the choice of every experiment

The connection between our model's coded units, and real-world units

The general formula for continuous variables

$$\text{coded value} = \frac{(\text{real value}) - (\text{center value})}{\frac{1}{2}\,(\text{range})}$$

$$\text{center value} = \frac{(\text{low value}) + (\text{high value})}{2}$$

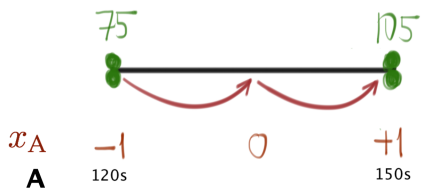
$$\text{range} = (\text{high value}) - (\text{low value})$$

Example: cooking time, factor **A**

$$x_{\mathrm{A}} = \frac{\mathrm{A} - 135}{\frac{1}{2}\,(30)} = \frac{\mathrm{A} - 135}{15}$$

$$\text{center}_{\mathrm{A}} = \frac{120 + 150}{2} = 135$$

$$\text{range}_{\mathrm{A}} = 150 - 120 = 30$$



Let's try it out

$$\text{coded value} = \frac{(\text{real value}) - (\text{center value})}{\frac{1}{2}\,(\text{range})} \qquad x_{\mathsf{A}} = \frac{\mathsf{A} - 135}{\frac{1}{2}\,(30)} = \frac{\mathsf{A} - 135}{15}$$

If cooking time is 135 seconds, what is x_{A}?

$$x_{\mathsf{A}} = \frac{135 - 135}{\frac{1}{2}\,(30)} = \frac{0}{15} = 0$$

If cooking time is 150 seconds, what is x_{A}?

$$x_{\mathsf{A}} = \frac{150 - 135}{\frac{1}{2}\,(30)} = \frac{15}{15} = +1$$

The connection between real-world units and coded units (in reverse!)

Going forwards:

$$\text{coded value} = \frac{(\text{real value}) - (\text{center value})}{\frac{1}{2}\,(\text{range})}$$

Going backwards:

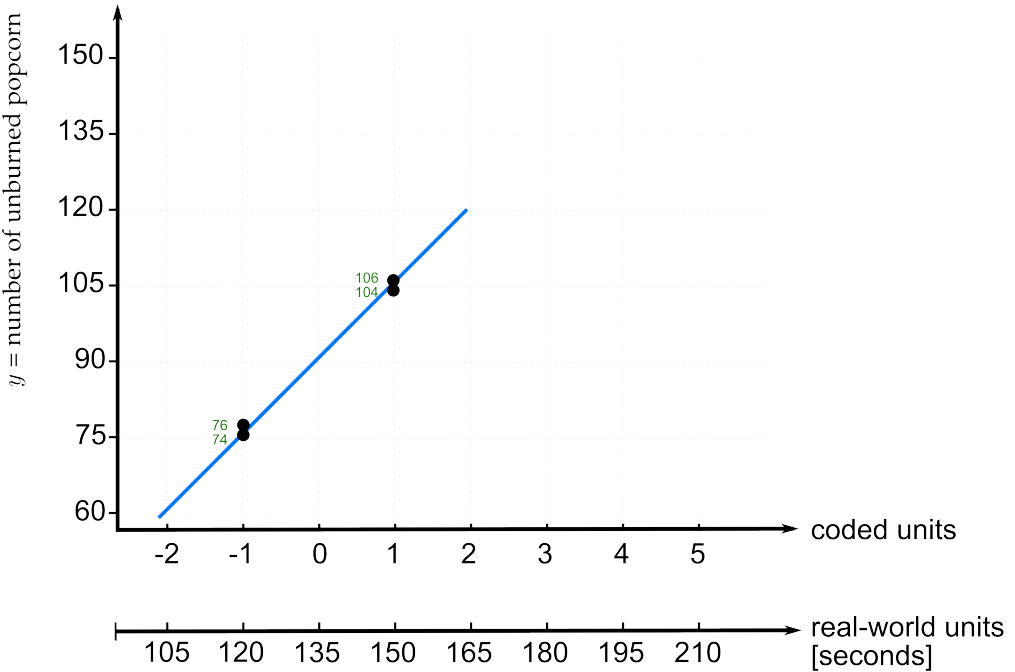
$$\text{real value} = (\text{coded value}) \times \tfrac{1}{2}\,(\text{range}) + (\text{center value})$$

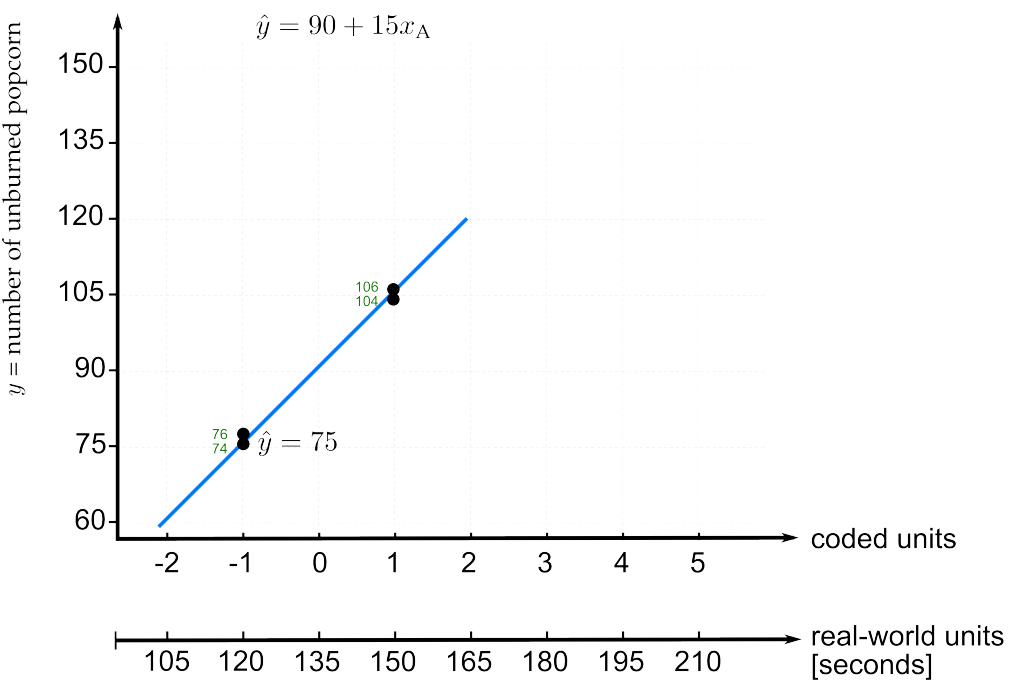
Example: cooking time, factor **A**

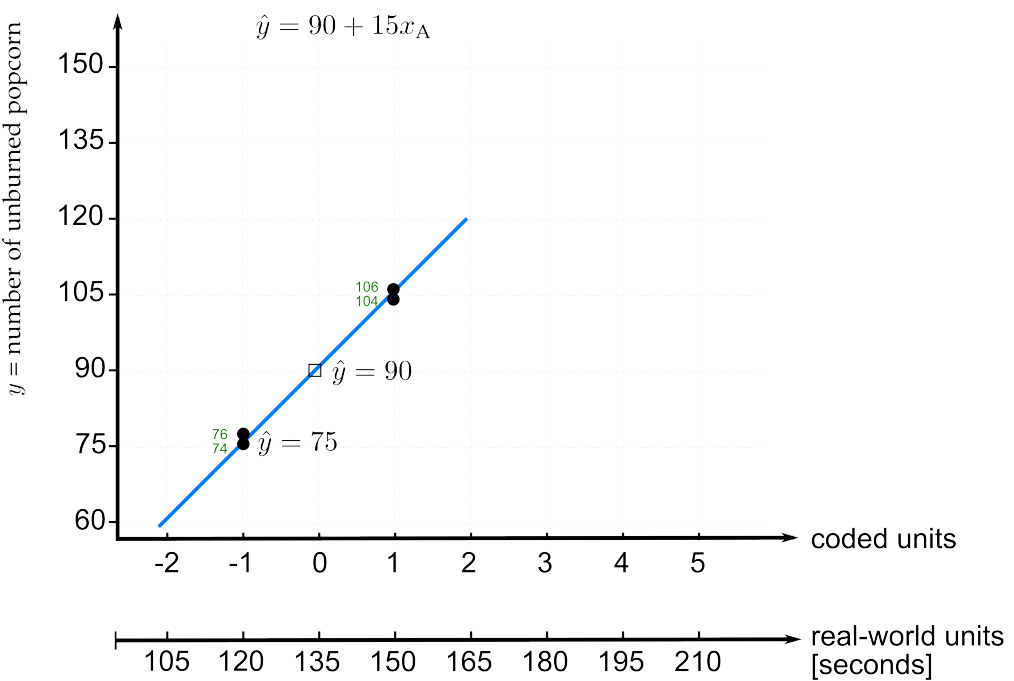
$$x_{\mathsf{A}} = \frac{\mathsf{A} - 135}{\frac{1}{2}\,(30)} = \frac{\mathsf{A} - 135}{15}$$

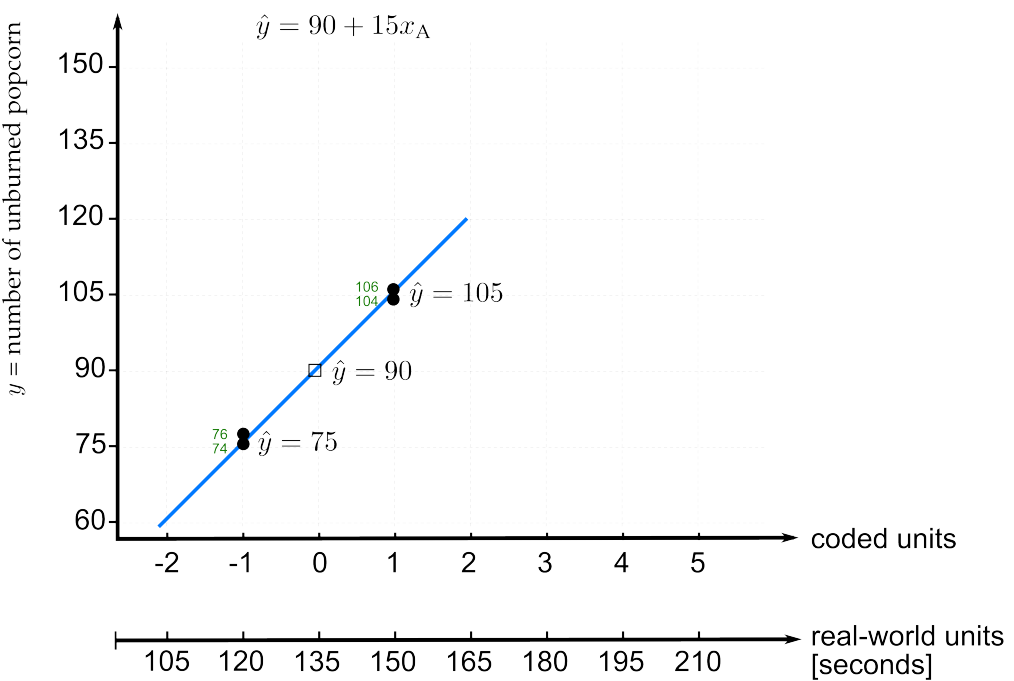
What is $x_{\mathsf{A}} = +2$?:

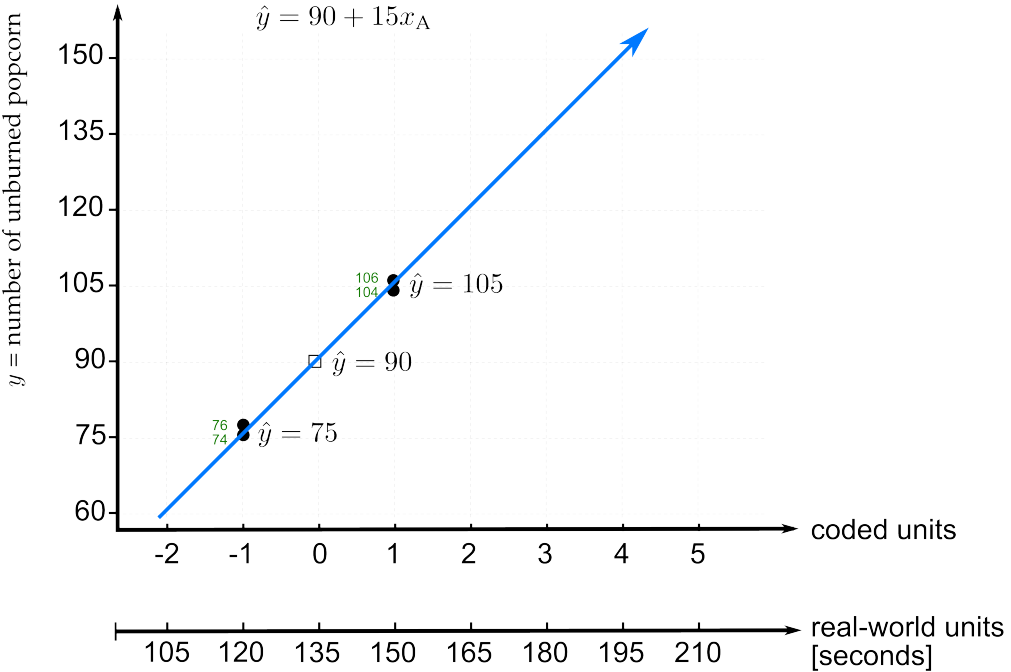
$$\begin{aligned}\text{real value} &= (+2) \times \tfrac{1}{2}\,(30) + (135) \\ &= (+2) \times 15 + (135) \\ &= 30 + 135 = 165\end{aligned}$$

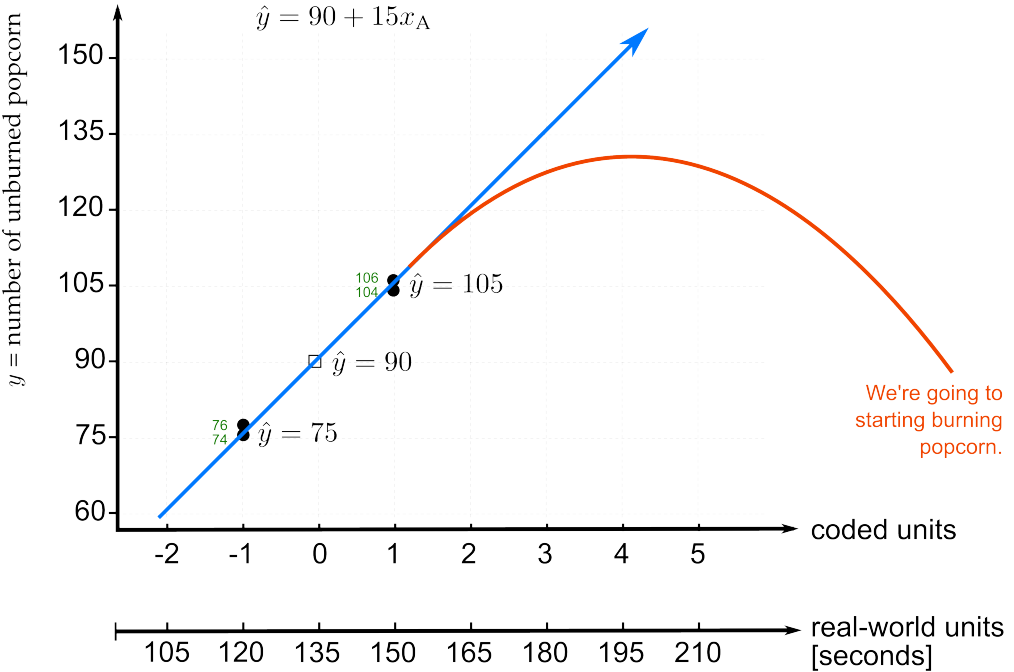

y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
76
74
$\hat{y} = 75$
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
$\hat{y} = 90$
76
74
$\hat{y} = 75$
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
$\hat{y} = 105$
$\hat{y} = 90$
76
74
$\hat{y} = 75$
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
$\hat{y} = 105$
$\hat{y} = 90$
76
74
$\hat{y} = 75$
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
$\hat{y} = 105$
$\hat{y} = 90$
76
74
$\hat{y} = 75$
We're going to starting burning popcorn.
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

The famous quote by George Box

"... all models are wrong, but some are useful."

"...the practical question is how wrong do they have to be [before they are] not useful?"

G. E. P. Box and N. R. Draper (1987), "*Empirical Model Building and Response Surfaces*", John Wiley & Sons, New York, NY.

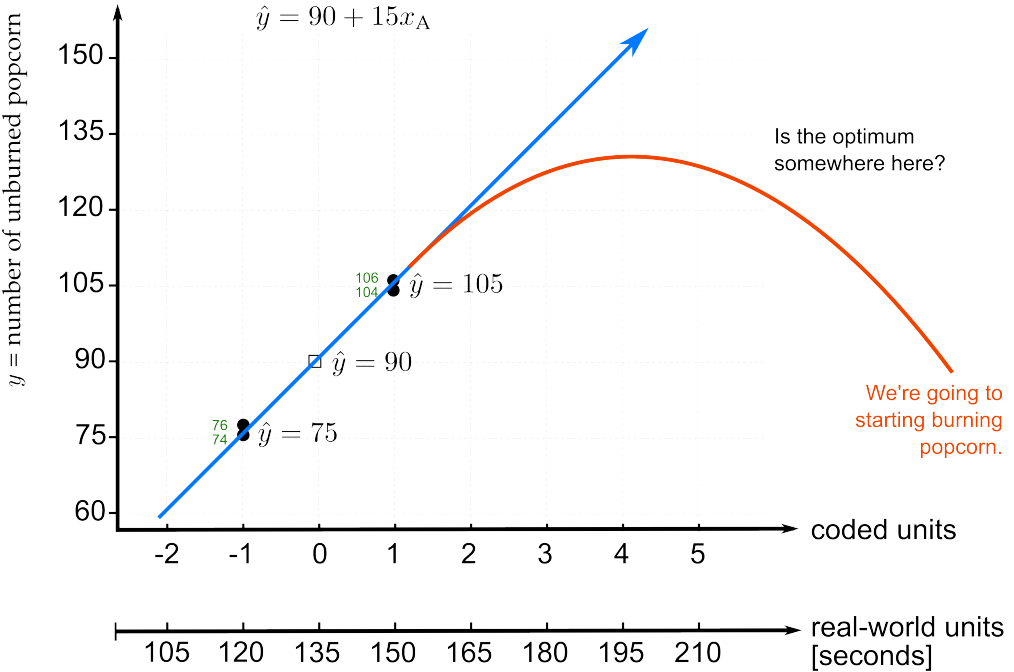
Answering George Box's question for the popcorn example

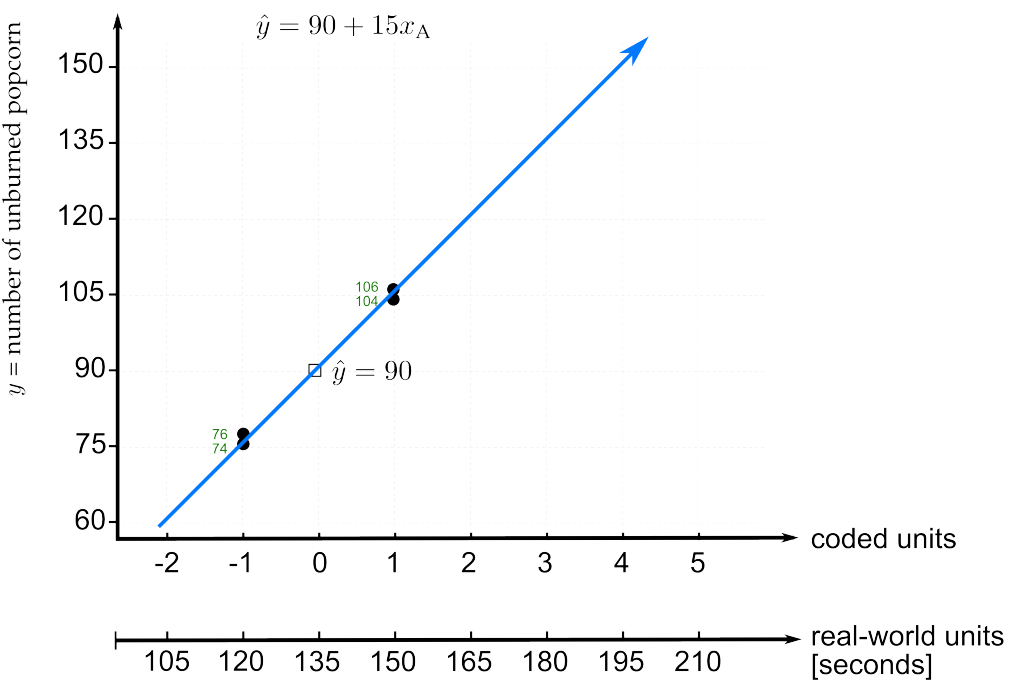
What is "*not useful*" in this case?

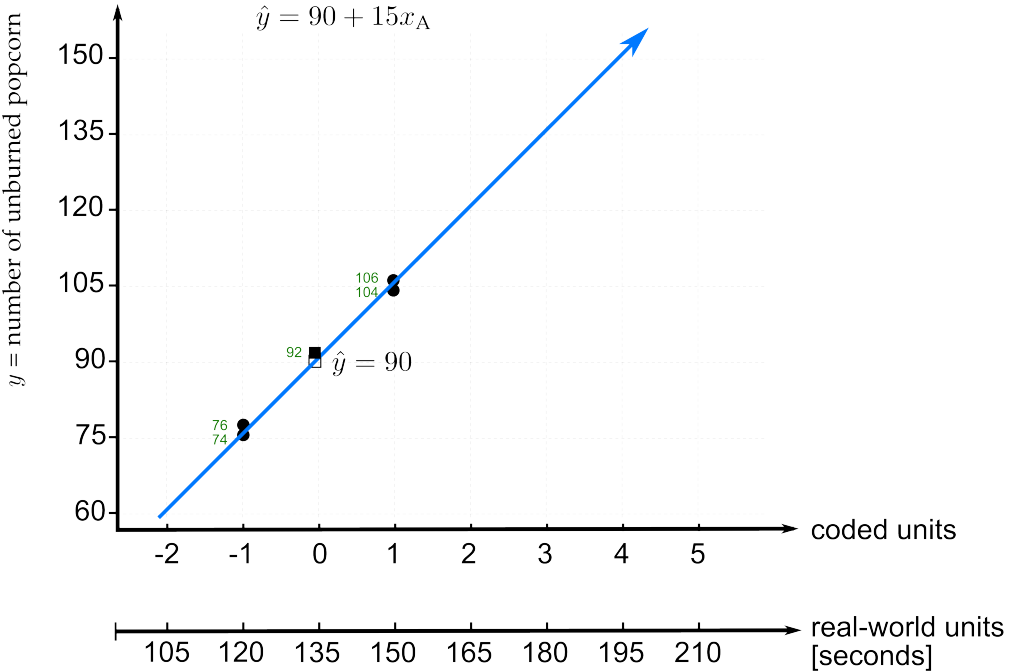
In an earlier video we said: "always have an objective in mind".

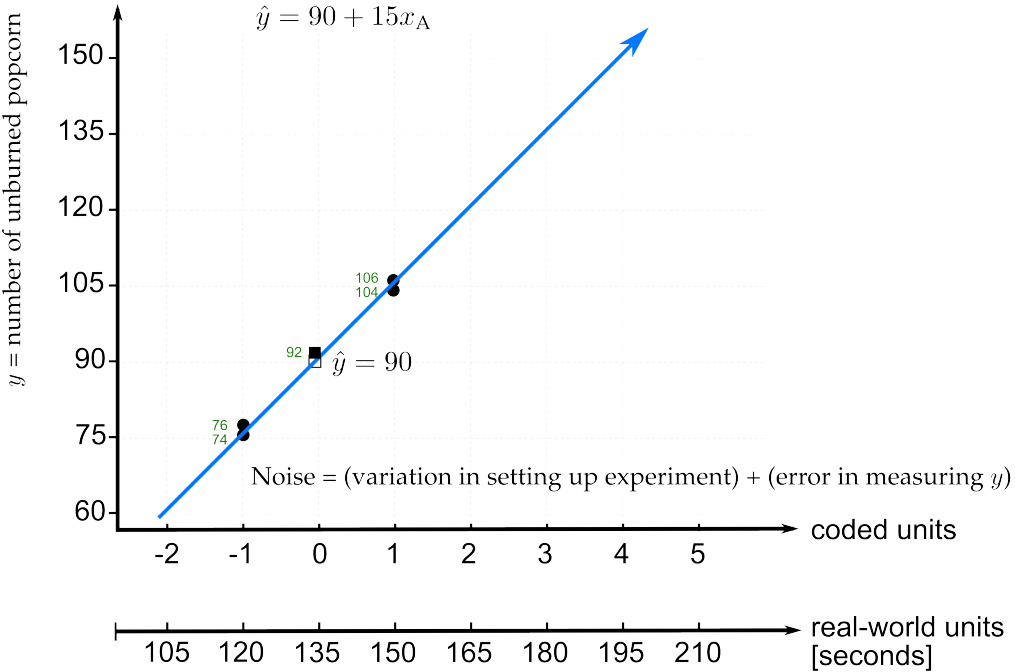
The answer to the question

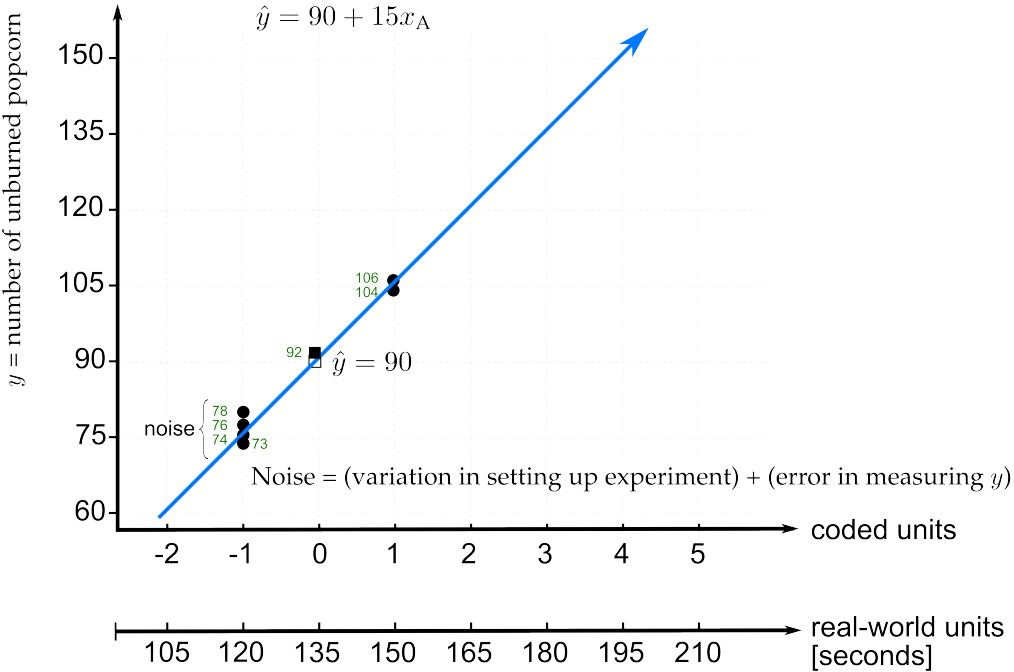
Our model is not useful when the predictions are not accurate; we need accurate predictions to optimize.

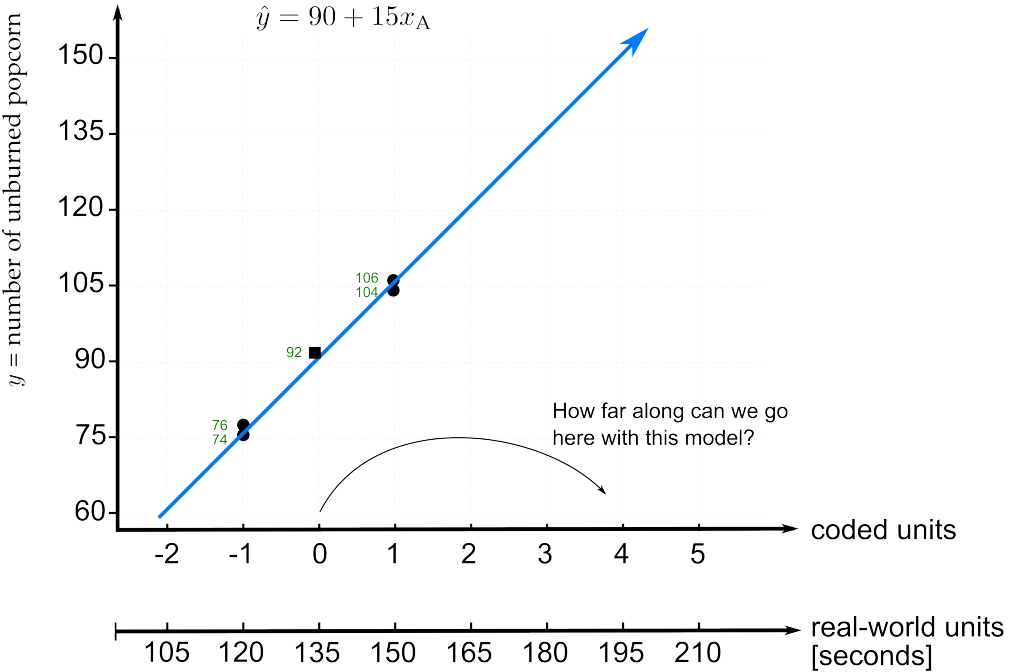

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
Is the optimum somewhere here?
106
104
$\hat{y} = 105$
$\hat{y} = 90$
76
74
$\hat{y} = 75$
We're going to starting burning popcorn.
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

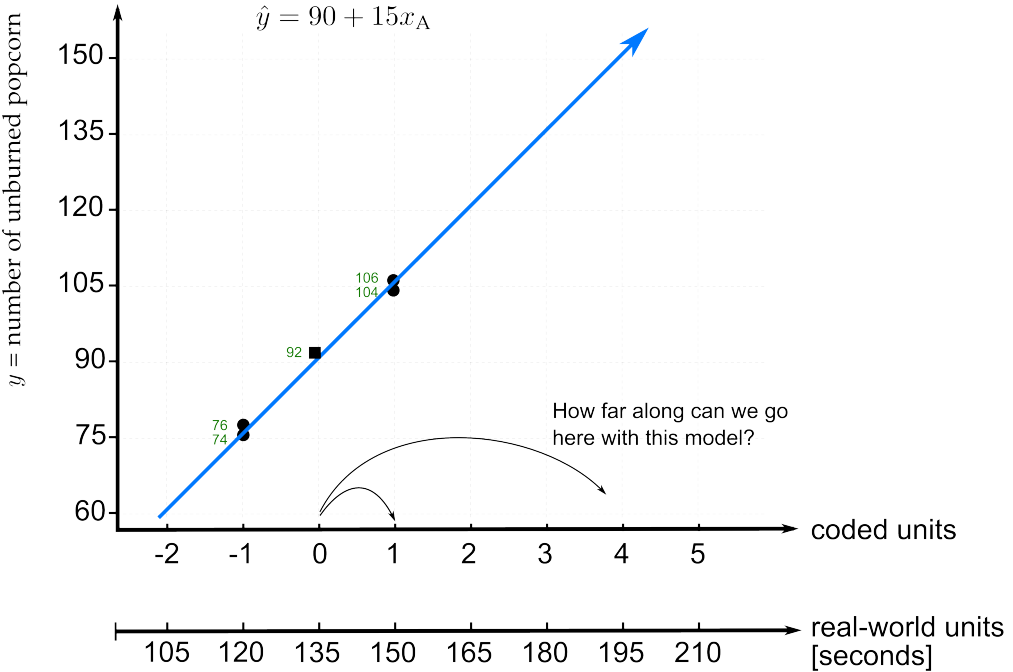

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
$\hat{y} = 90$
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

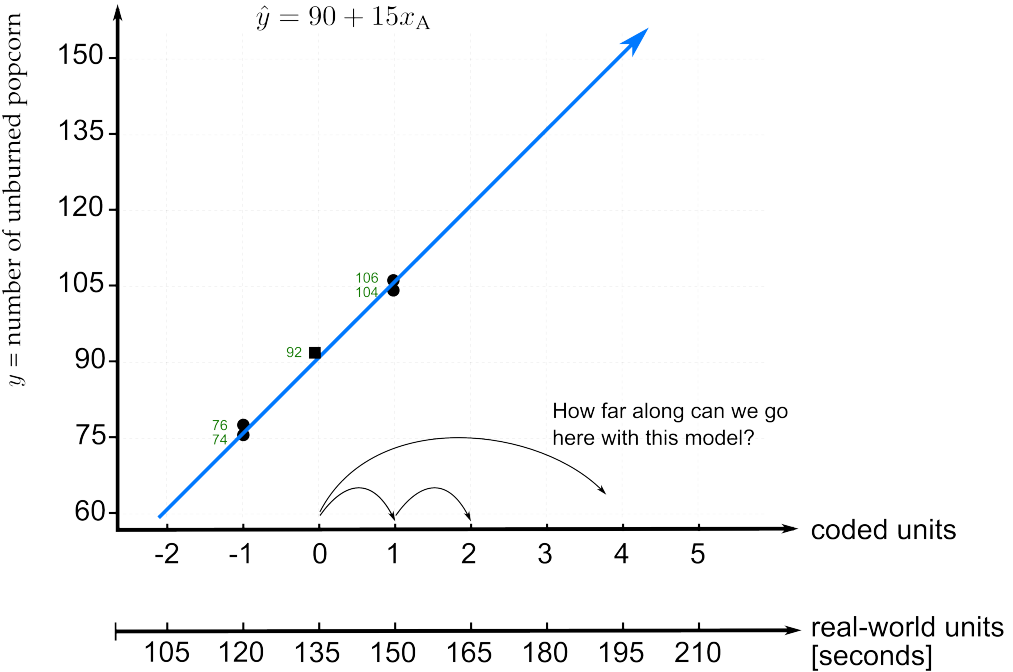

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
$\hat{y} = 90$
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

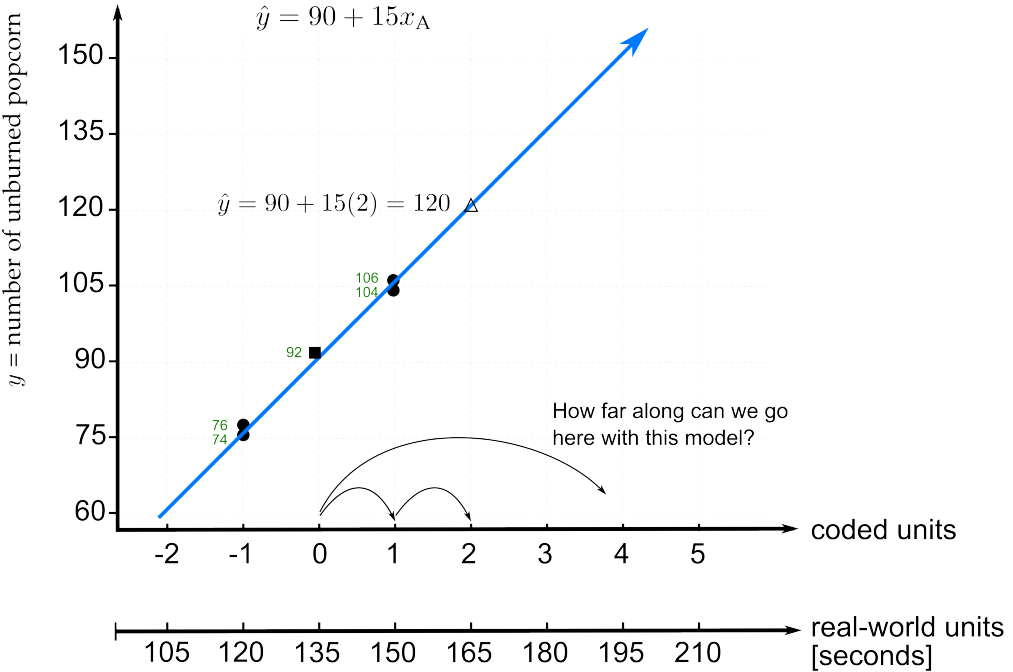

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
$\hat{y} = 90$
76
74
Noise = (variation in setting up experiment) + (error in measuring y)
coded units
-2
-1
0
1
2
3
4
5
real-world units
[seconds]
105
120
135
150
165
180
195
210

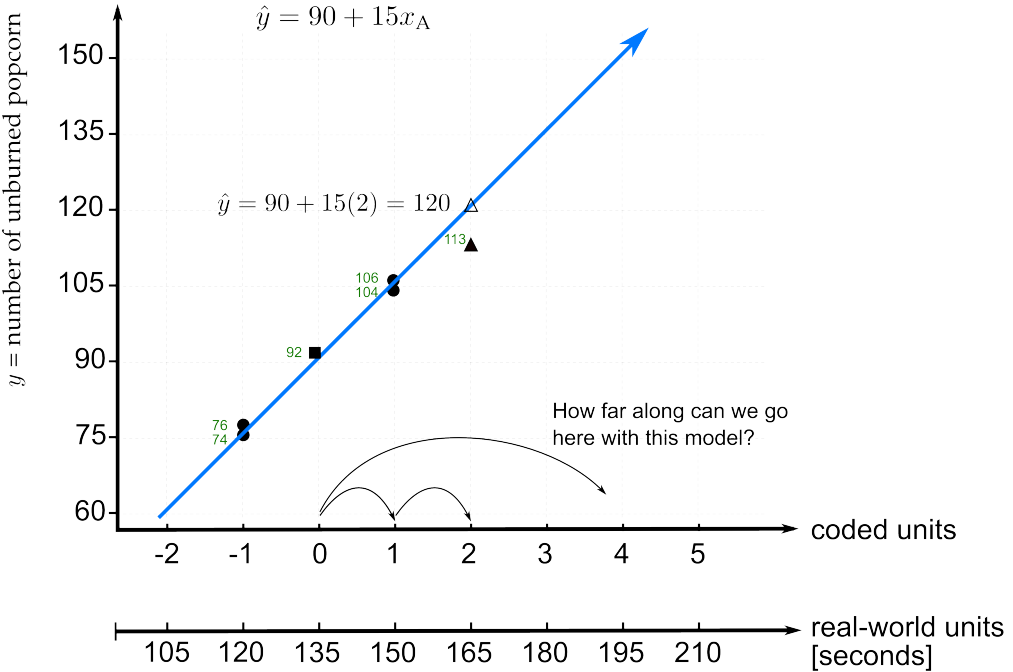

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
$\hat{y} = 90$
noise
78
76
74
73
Noise = (variation in setting up experiment) + (error in measuring y)
coded units
-2
-1
0
1
2
3
4
5
105
120
135
150
165
180
195
210
real-world units
[seconds]

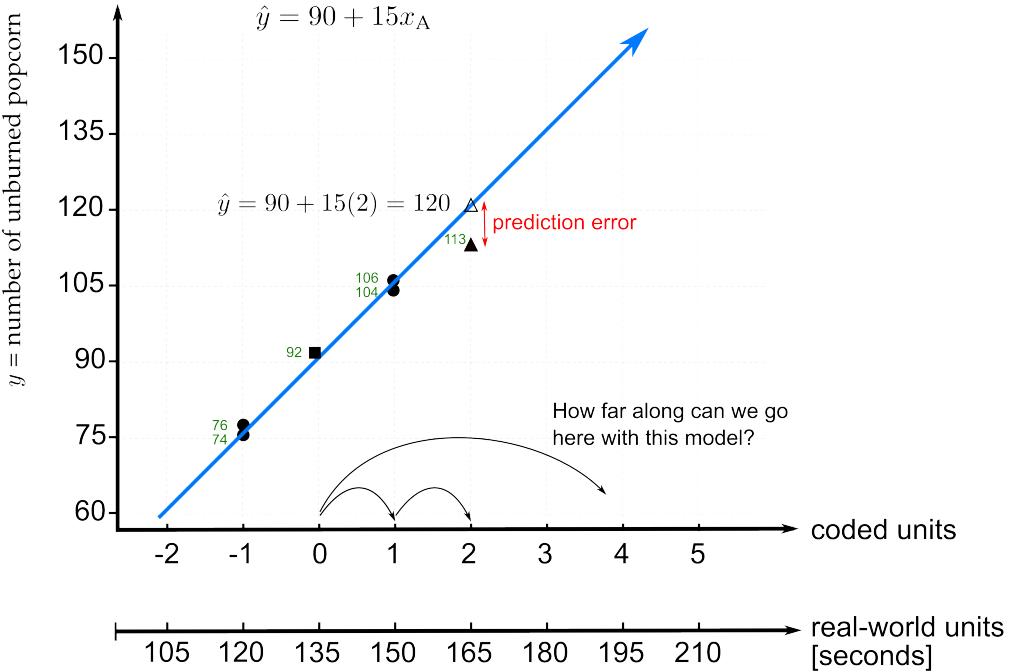

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
76
74
How far along can we go here with this model?
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

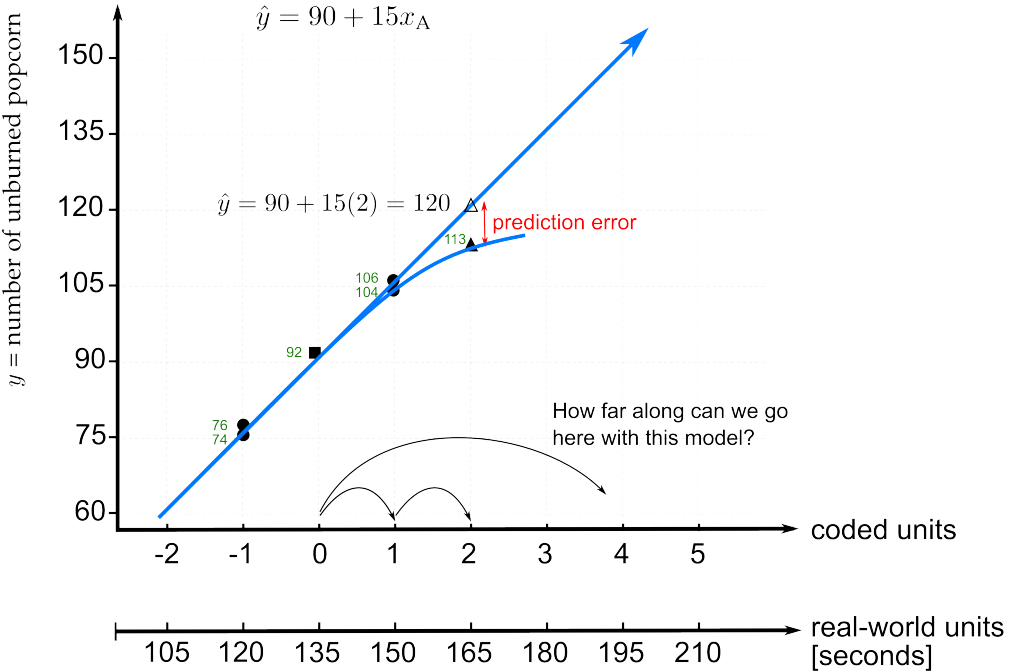

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
76
74
How far along can we go here with this model?
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

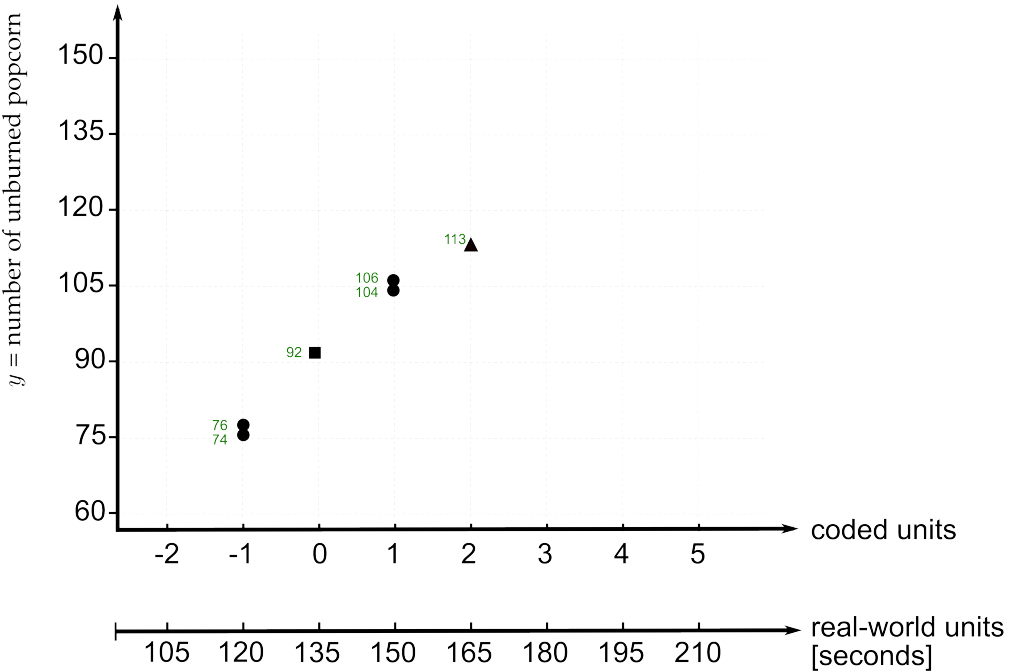

$\hat{y} = 90 + 15x_A$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
76
74
How far along can we go here with this model?
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

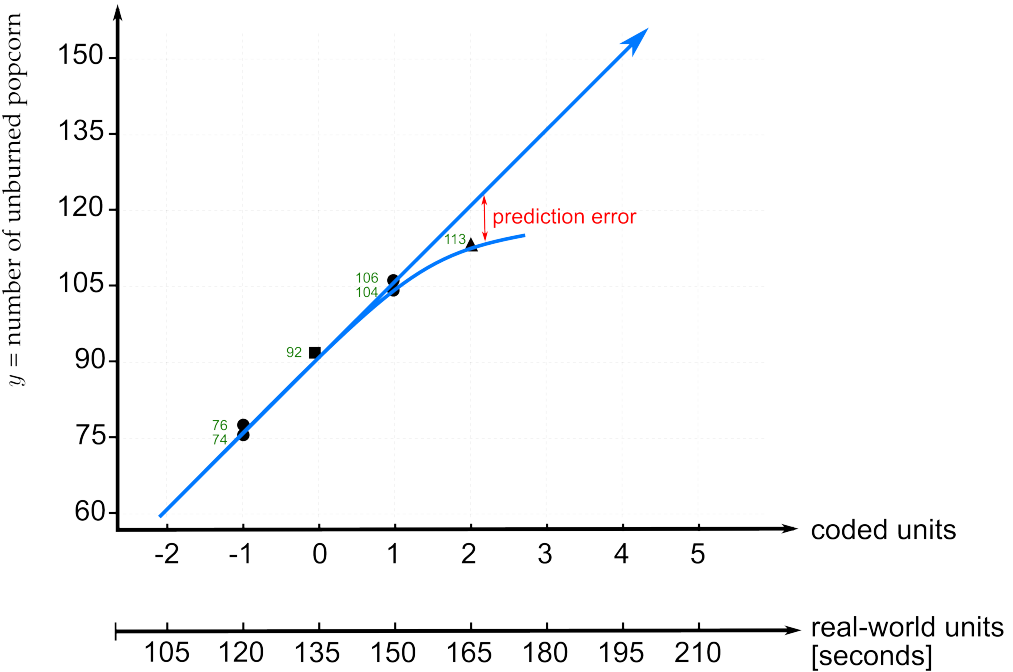

$\hat{y} = 90 + 15x_A$
$\hat{y} = 90 + 15(2) = 120$
y = number of unburned popcorn
150
135
120
105
90
75
60
106
104
92
76
74
How far along can we go here with this model?
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

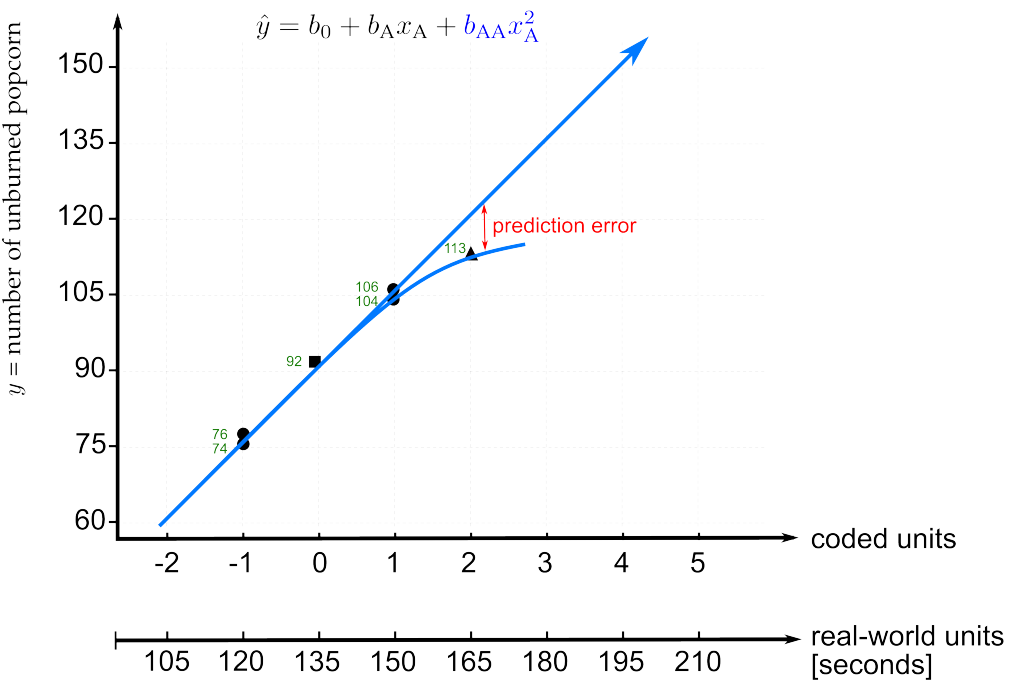

$\hat{y} = 90 + 15x_A$
$\hat{y} = 90 + 15(2) = 120$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
106
104
92
76
74
How far along can we go here with this model?
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

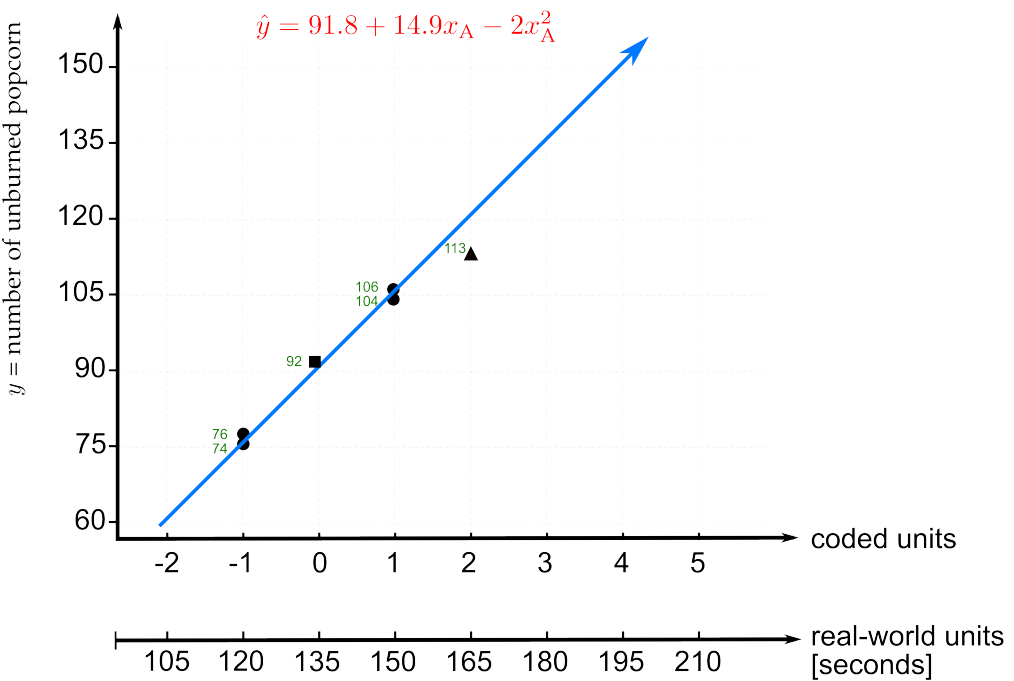

$\hat{y} = 90 + 15x_A$
$\hat{y} = 90 + 15(2) = 120$
prediction error
113
106
104
92
76
74
How far along can we go here with this model?
y = number of unburned popcorn
150
135
120
105
90
75
60
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

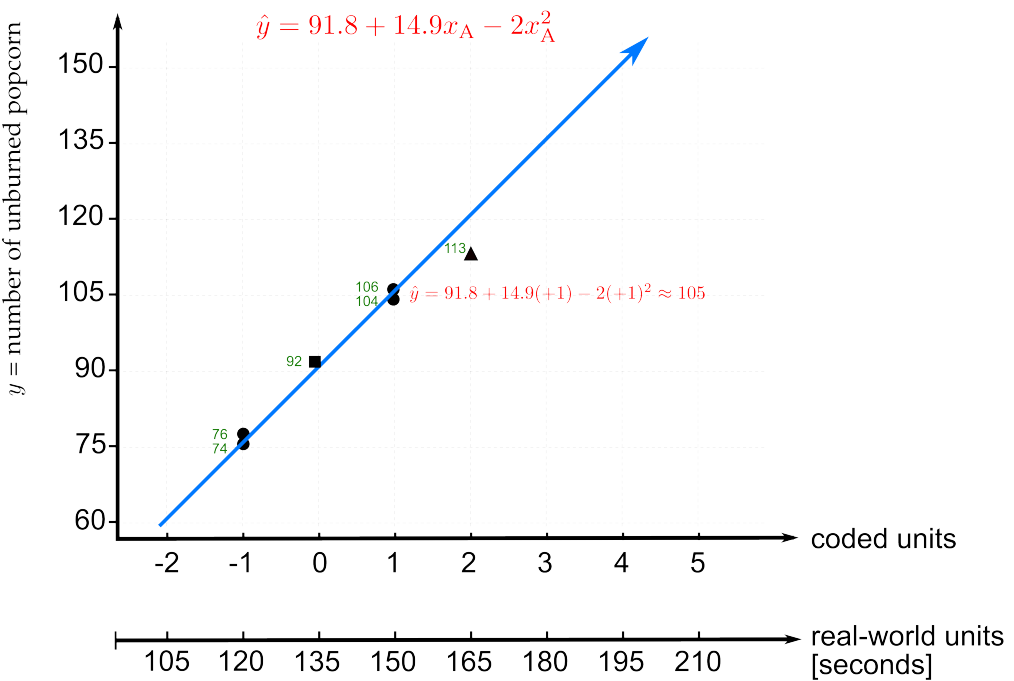

$\hat{y} = 90 + 15x_A$
$\hat{y} = 90 + 15(2) = 120$
prediction error
113
106
104
92
76
74
How far along can we go here with this model?
y = number of unburned popcorn
150
135
120
105
90
75
60
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units [seconds]

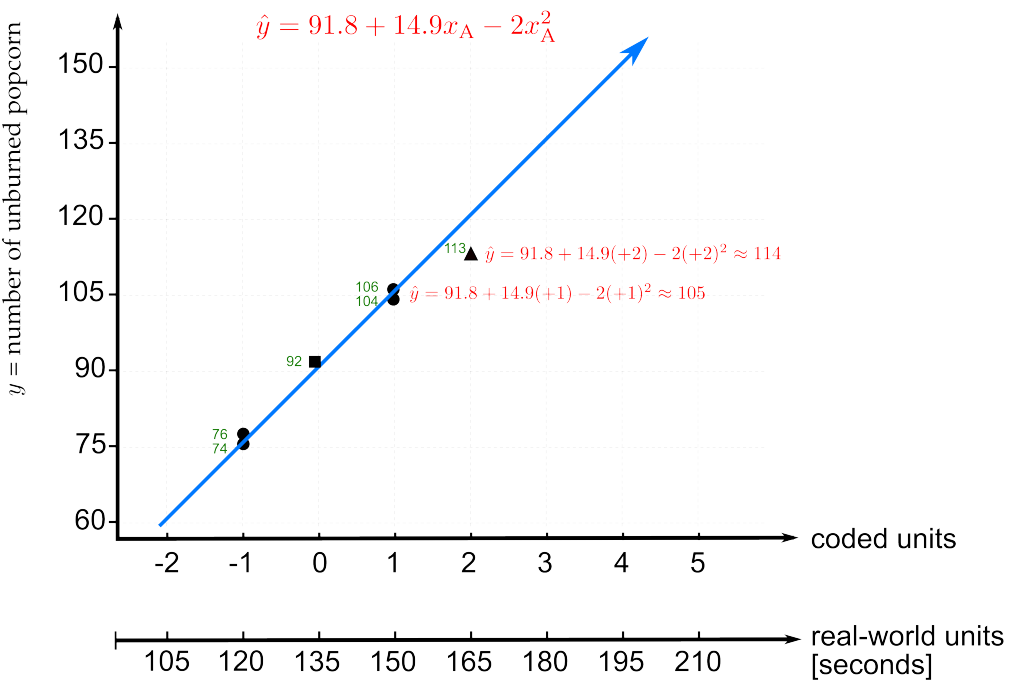

y = number of unburned popcorn
150
135
120
105
90
75
60
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

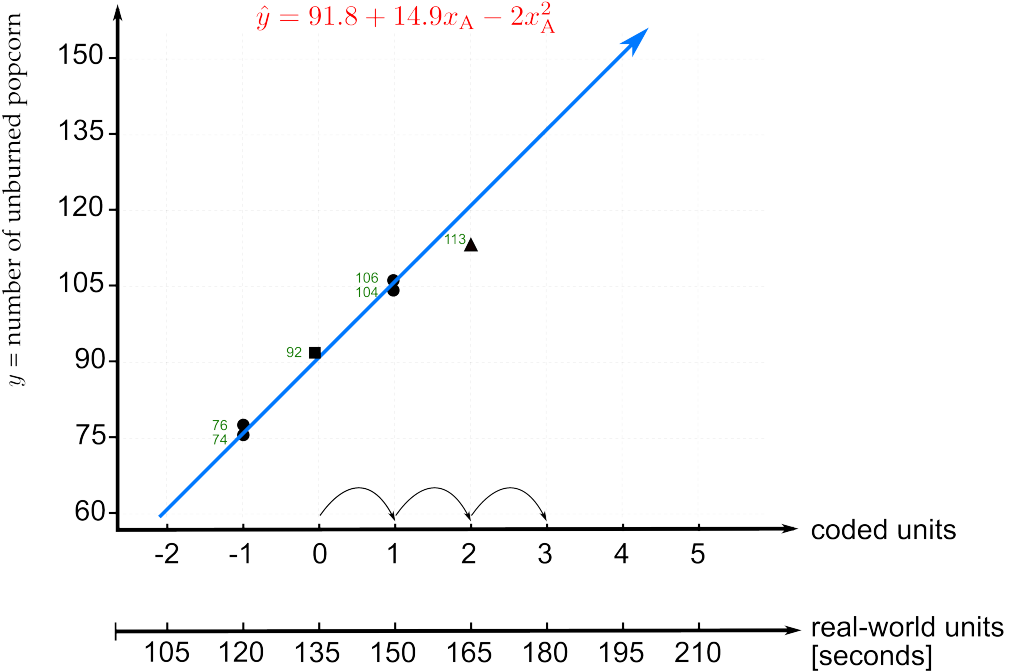

y = number of unburned popcorn
150
135
120
105
90
75
60
prediction error
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

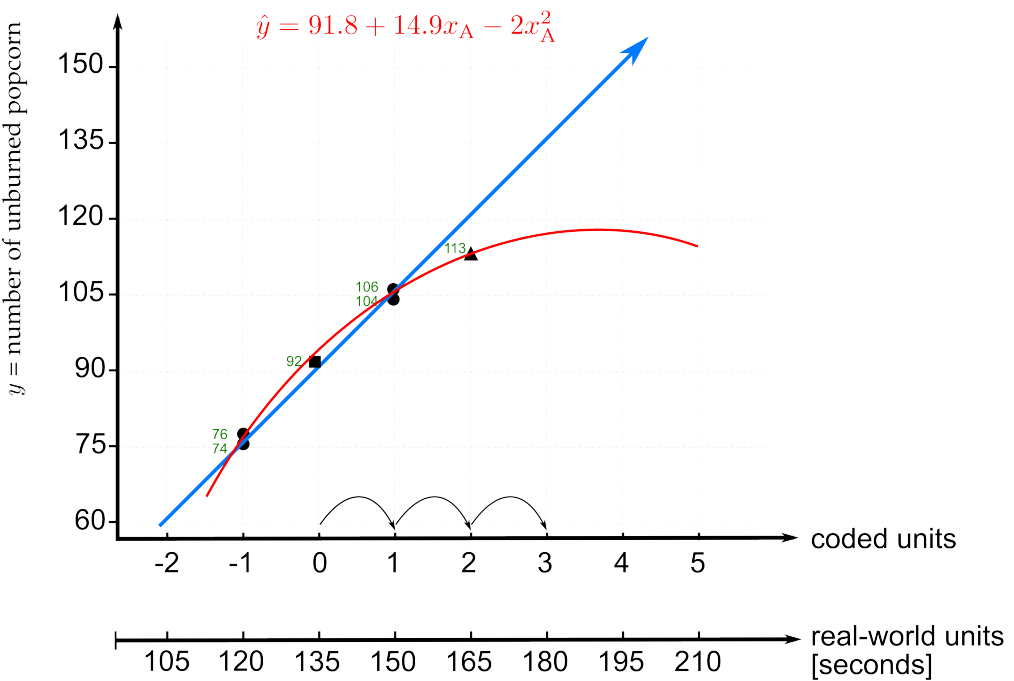

$\hat{y} = b_0 + b_A x_A + b_{AA} x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
prediction error
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

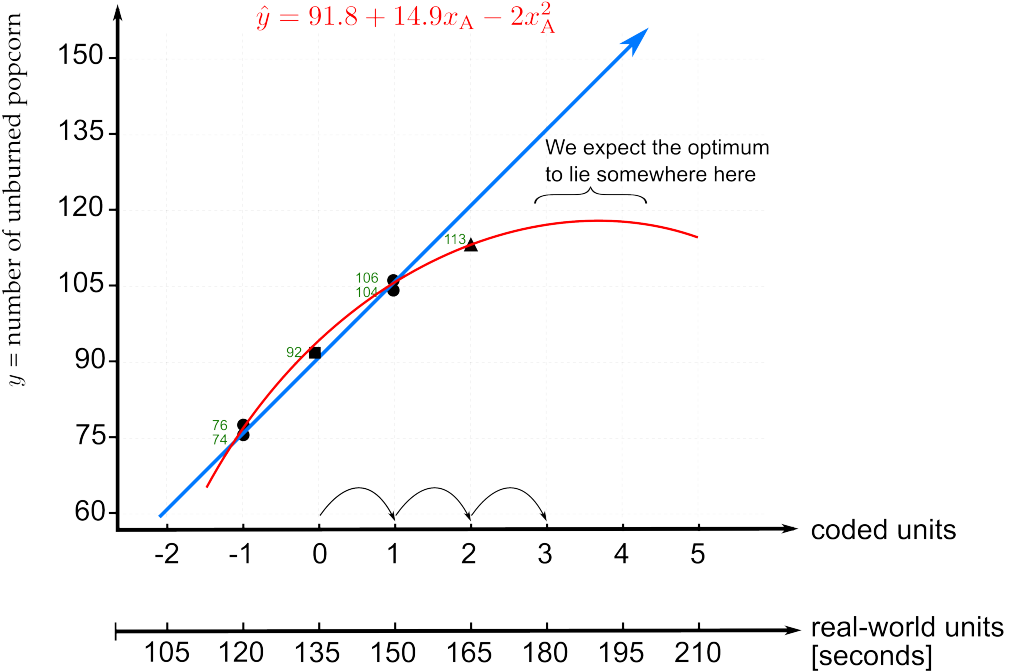

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

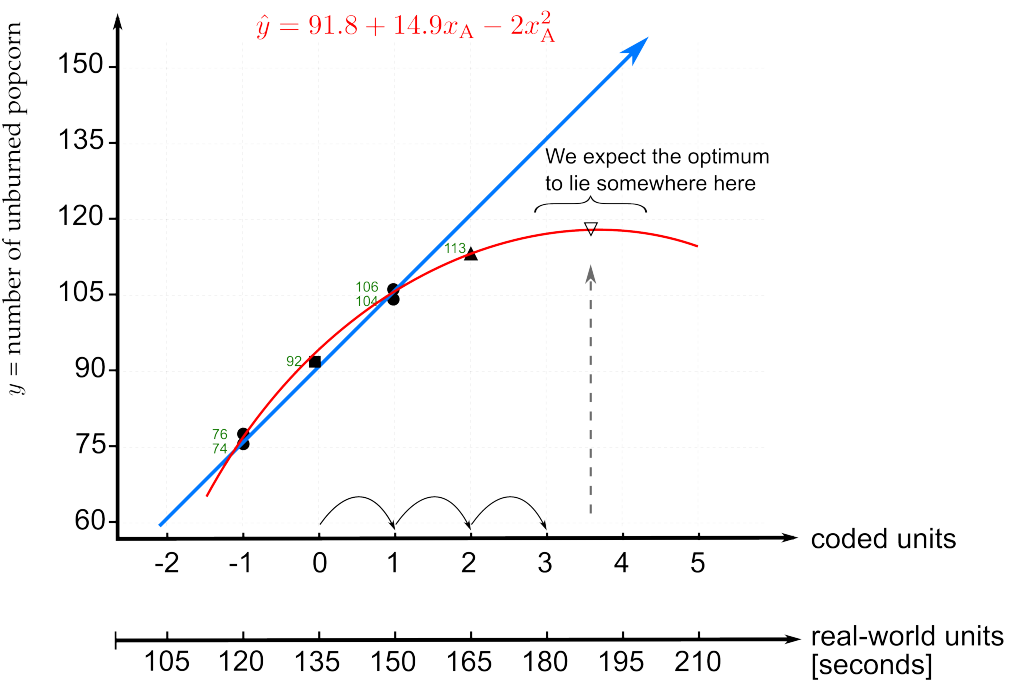

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
106
104
$\hat{y} = 91.8 + 14.9(+1) - 2(+1)^2 \approx 105$
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

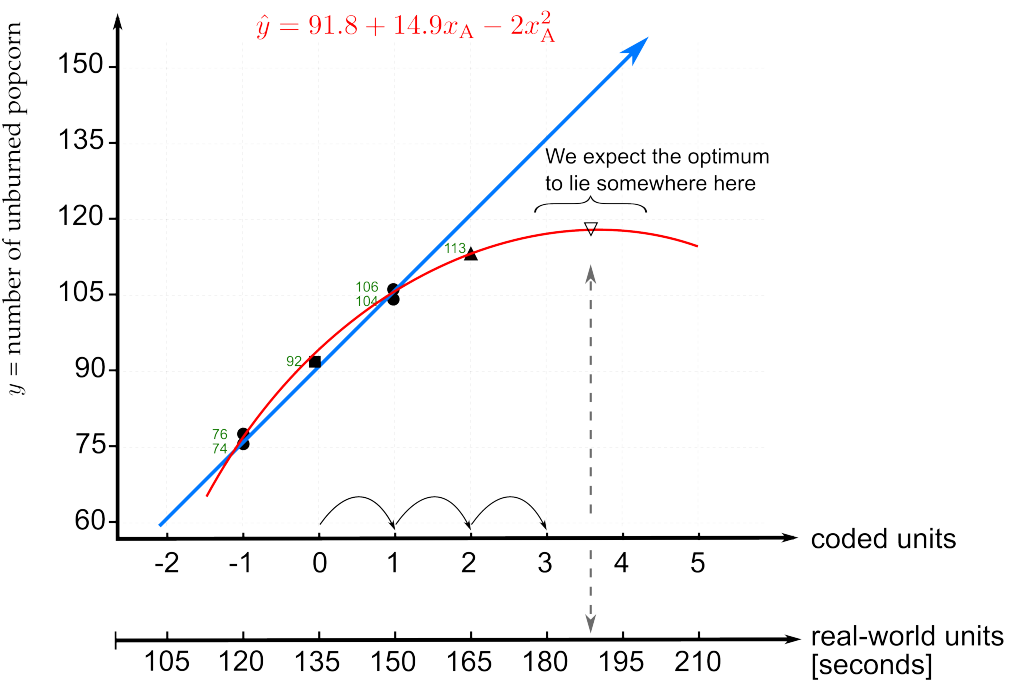

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
$\hat{y} = 91.8 + 14.9(+2) - 2(+2)^2 \approx 114$
106
104
$\hat{y} = 91.8 + 14.9(+1) - 2(+1)^2 \approx 105$
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

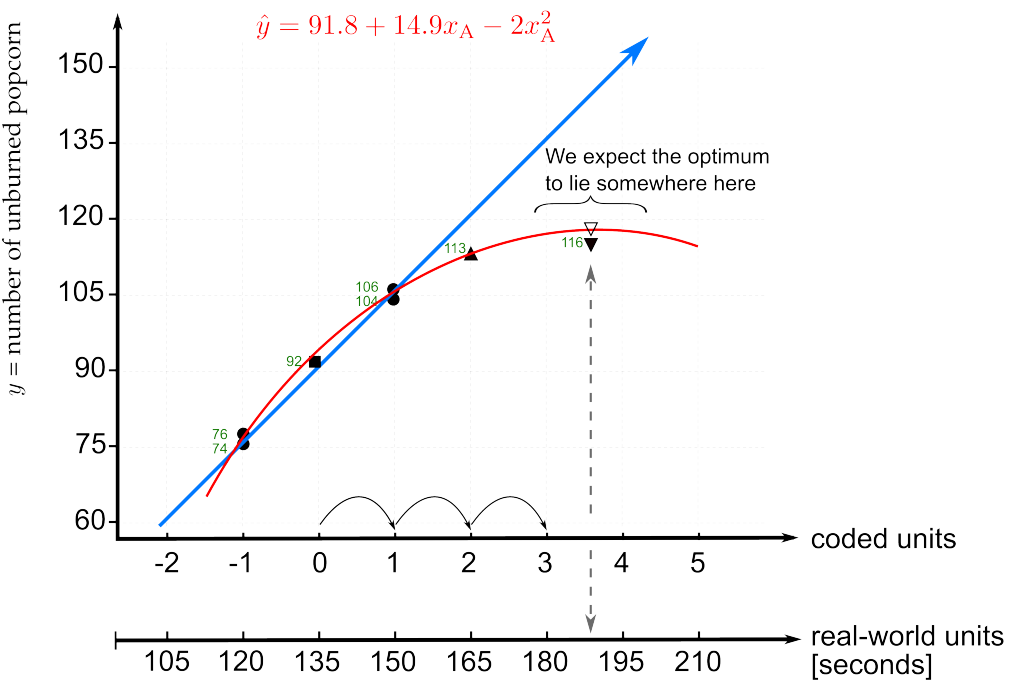

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

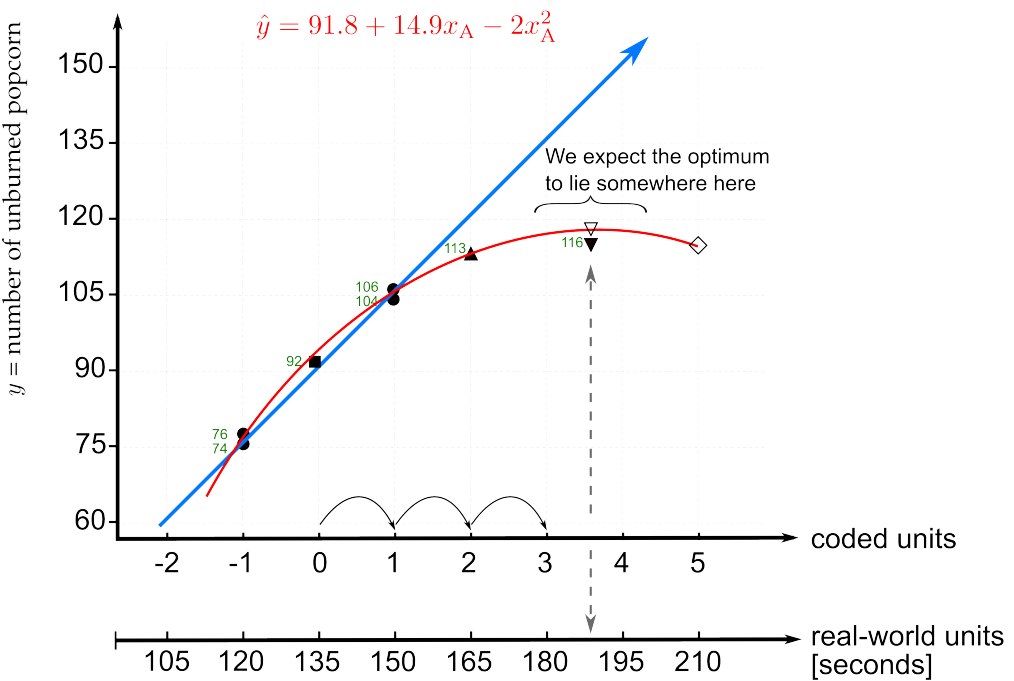

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

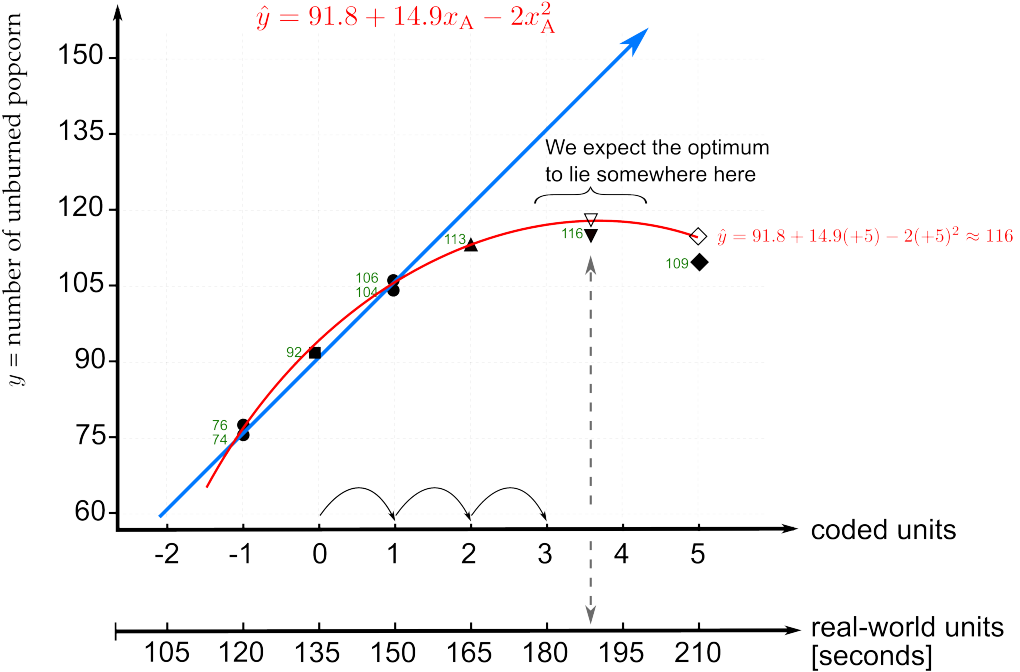

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
We expect the optimum
to lie somewhere here
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

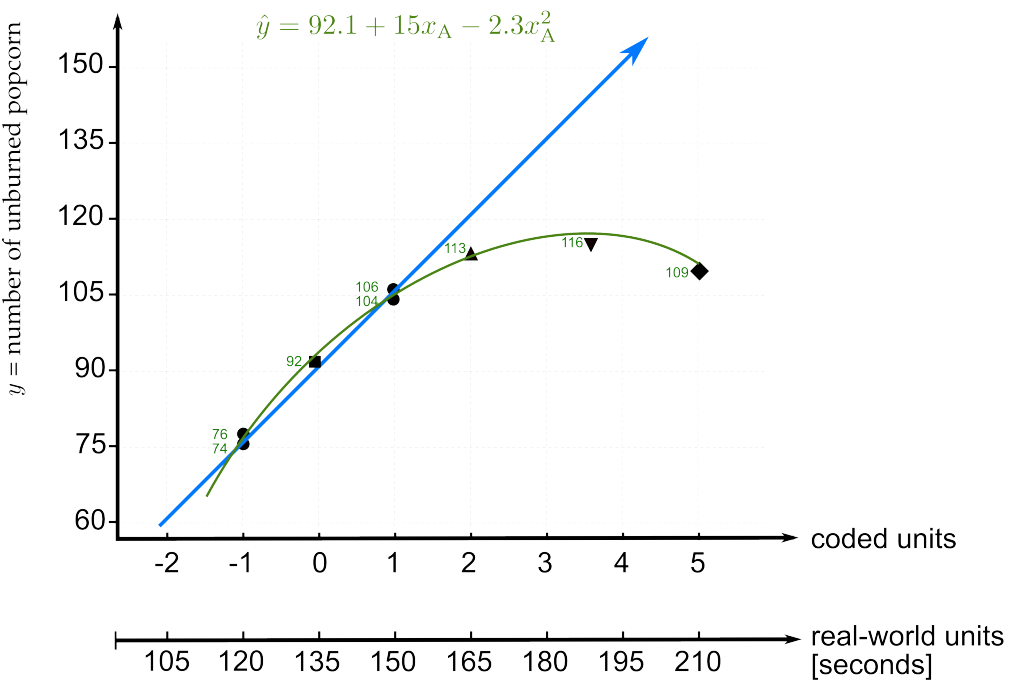

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
We expect the optimum
to lie somewhere here
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

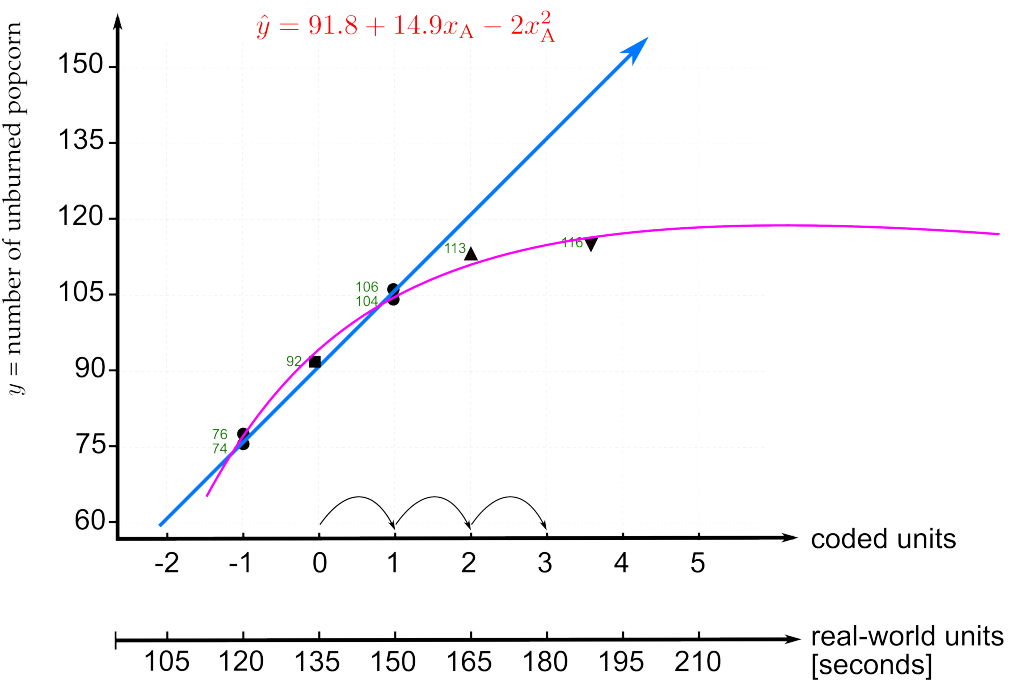

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
We expect the optimum
to lie somewhere here
113
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

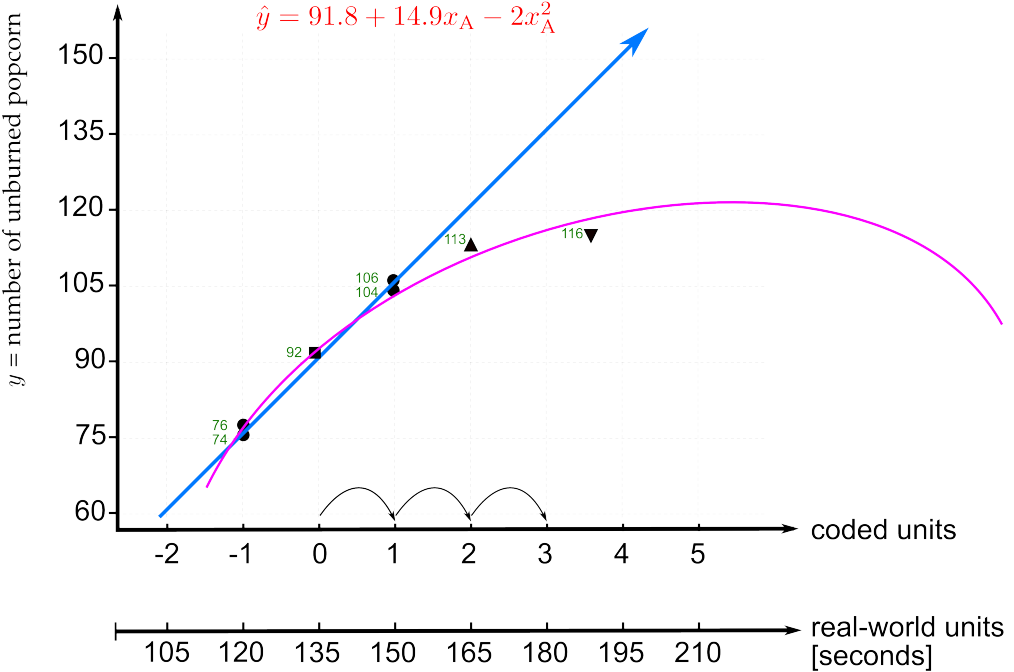

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
We expect the optimum
to lie somewhere here
113
116
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]

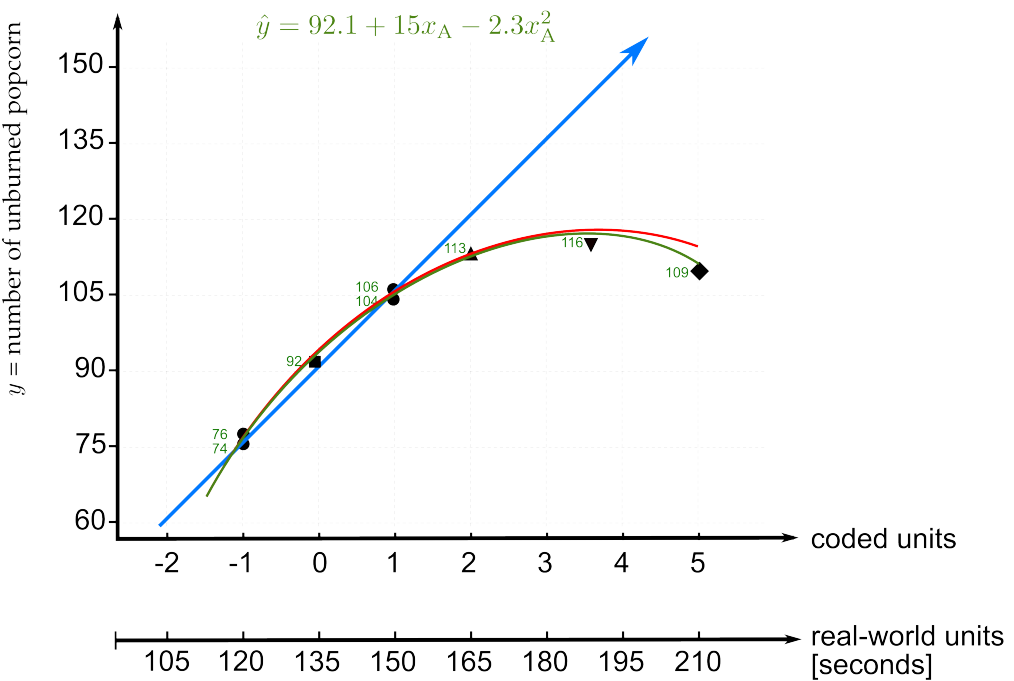

$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
We expect the optimum
to lie somewhere here
113
116
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
We expect the optimum
to lie somewhere here
$\hat{y} = 91.8 + 14.9(+5) - 2(+5)^2 \approx 116$
116
113
109
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 92.1 + 15x_A - 2.3x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
76
74
92
106
104
113
116
109
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
116
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 91.8 + 14.9x_A - 2x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
113
116
106
104
92
76
74
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]


$\hat{y} = 92.1 + 15x_A - 2.3x_A^2$
y = number of unburned popcorn
150
135
120
105
90
75
60
-2
-1
0
1
2
3
4
5
coded units
105
120
135
150
165
180
195
210
real-world units
[seconds]
76
74
92
106
104
113
116
109

Recap of each experiment used so far

- **1, 2, 3, 4**
  the used to rule out factor B (oil type)
  used to provide an initial model
- **5**
  center point: confirmed our model was linear
  used to check prediction quality
- **6**
  exploratory first step outside the initial factorial
  to test the model's prediction quality
  it suggested we rebuild the model (add quadratic term)
- **7**
  seems like we are near an optimum

are you able to justify the need for each experiment?

The critical concepts covered in this video using the popcorn case study

1. Real-world units and coded units
2. Linear vs nonlinear systems
3. Prediction models are wrong, but still useful
4. Noise and error and the need for replicated experiments
5. How to systematically reach an optimum
6. Justifying the choice of every experiment