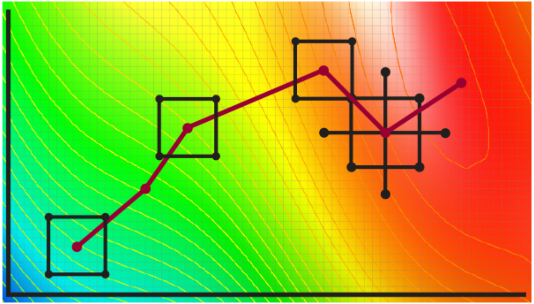
Experimentation for Improvement

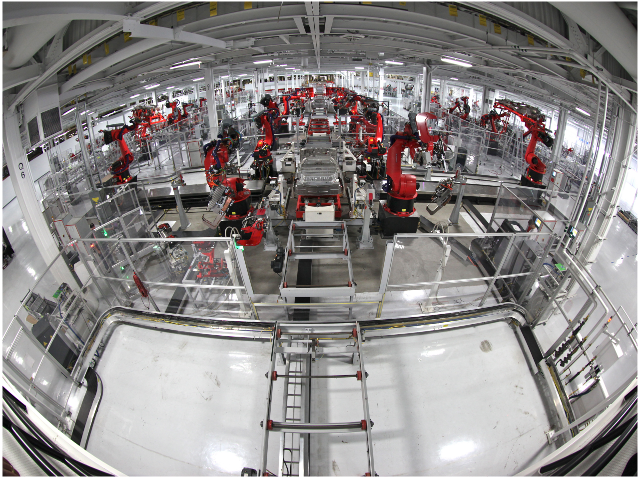


http://learnche.org/

Design and Analysis of Experiments

Copyright, sharing, and attribution notice

Case study: manufacturing of a mass produced product



[Flickr: jurvetson]

Tesla Motors assembly line

Case study: manufacturing of a mass produced product

[Flickr: archangel12]

Case study: manufacturing of a mass produced product

[Flickr: bensutherland]

Case study: manufacturing of a mass produced product

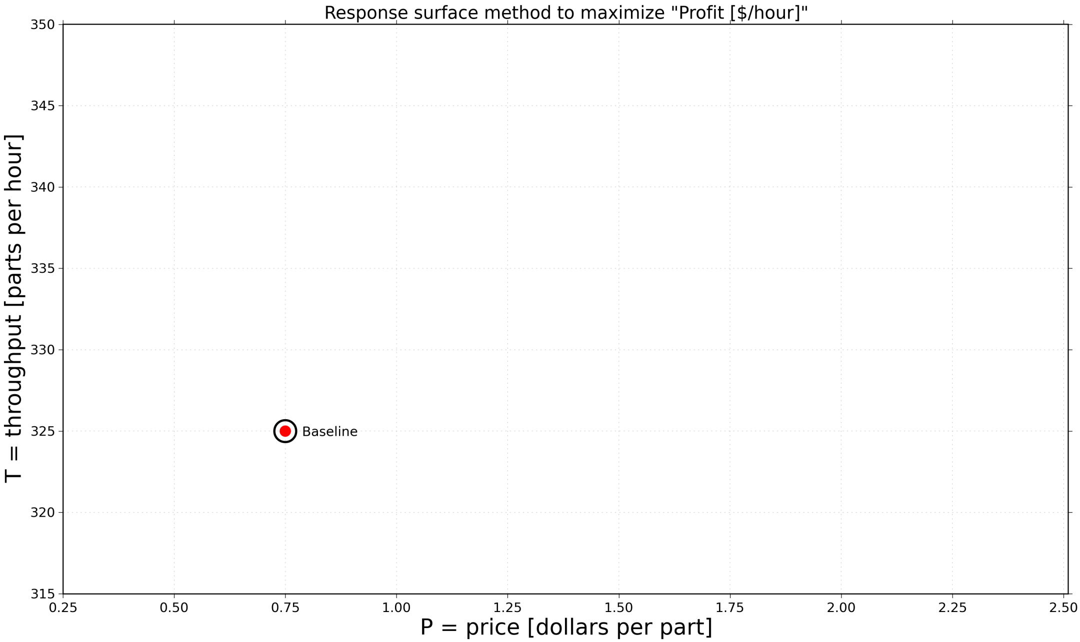
[Flickr: 6510472979]

Two factors are available to vary:

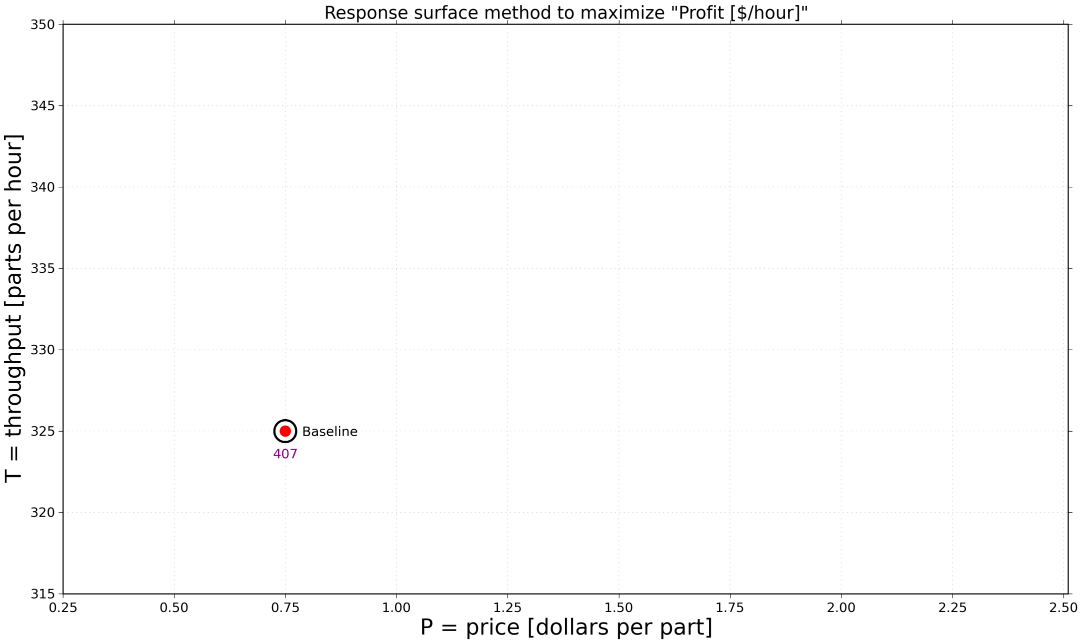
- **T**hroughput: number of parts per hour
- **P**rice: selling price per part produced

The outcome variable y = profit [$ per hour]

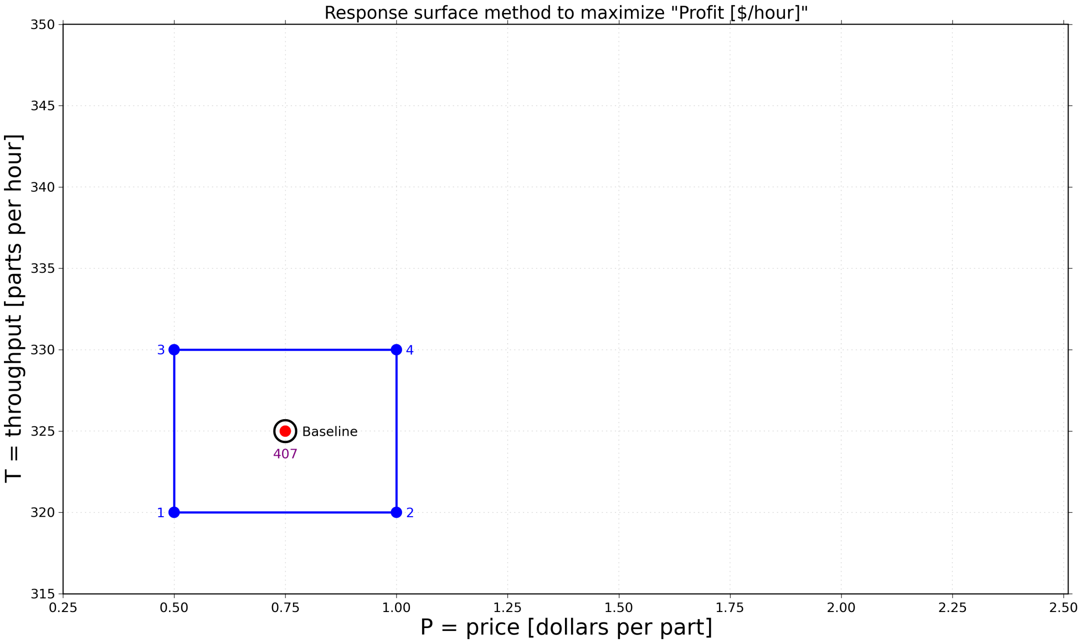
- profit = (all income) − (all expenses)
- both factors affect the profit
- profit is easily calculated



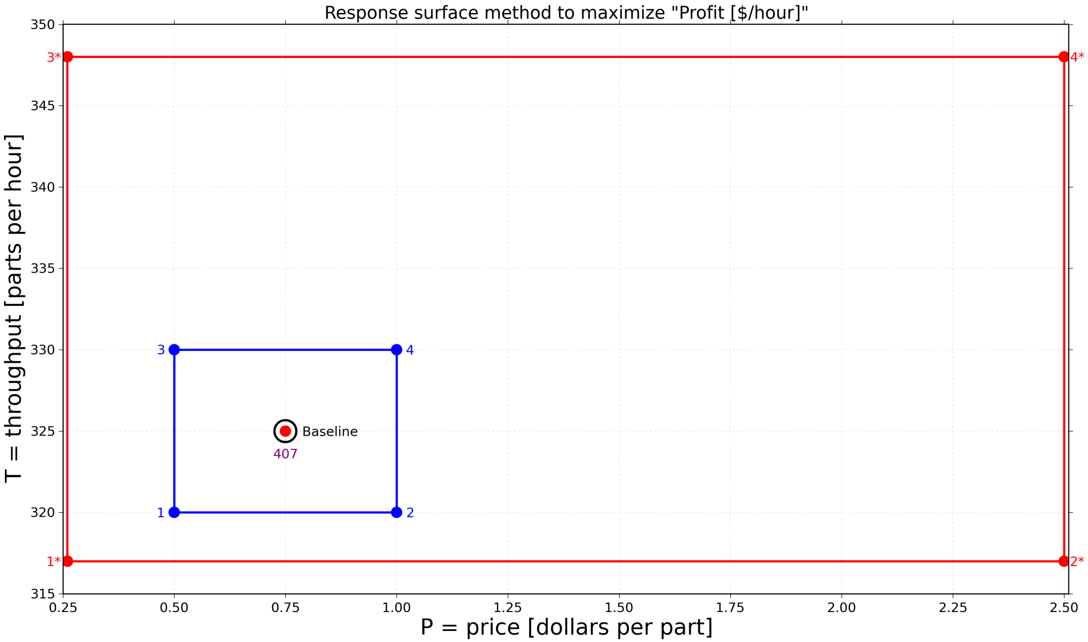
Response surface method to maximize "Profit [$/hour]"
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
Baseline
T = throughput [parts per hour]
P = price [dollars per part]



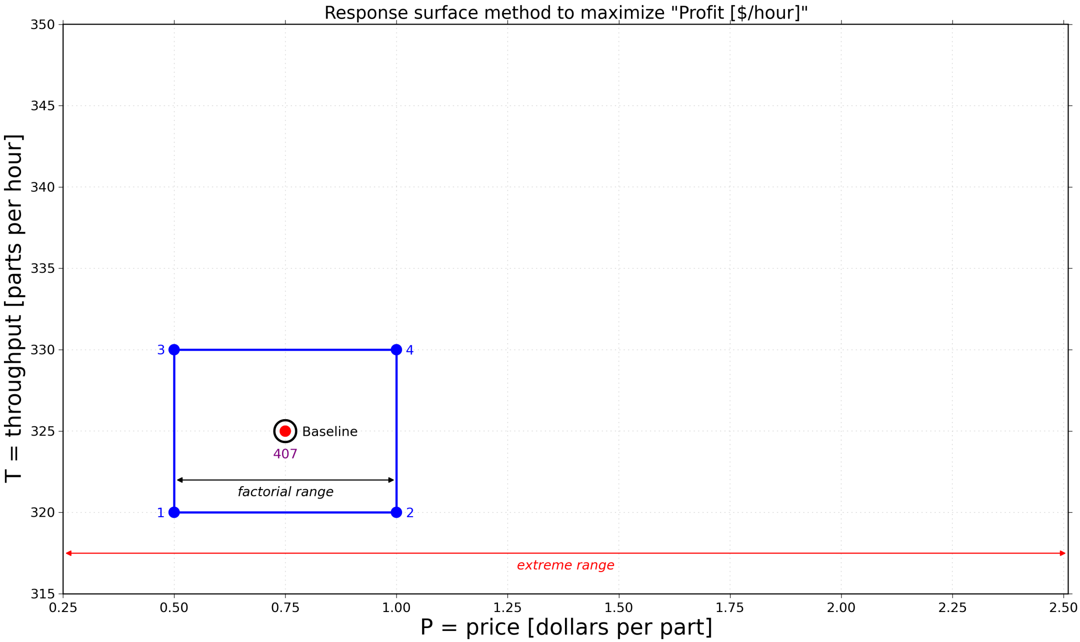
Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
Baseline
407
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50

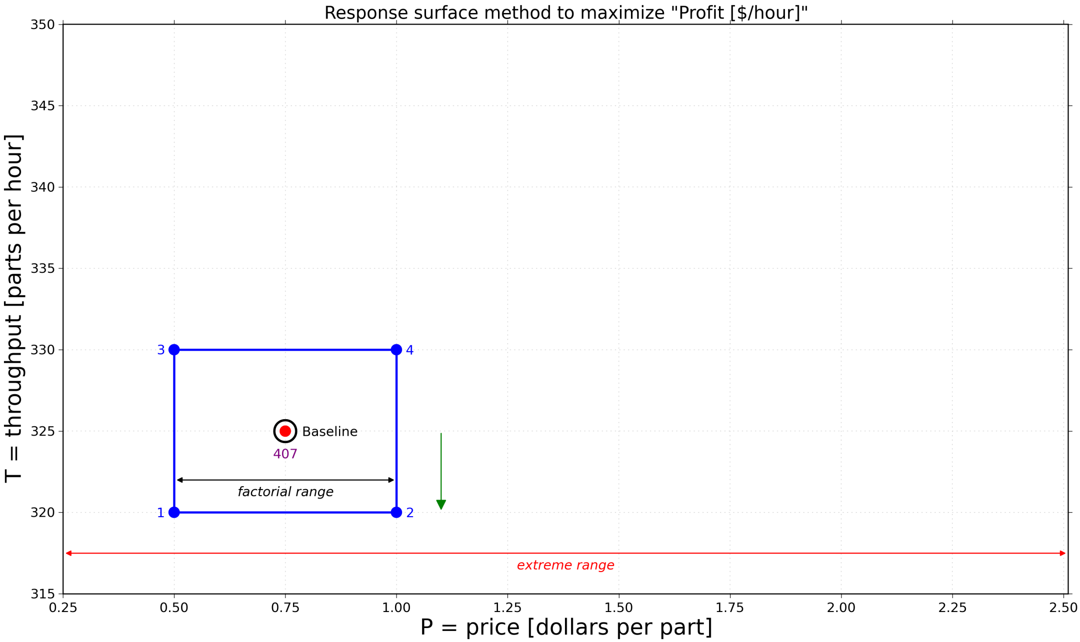


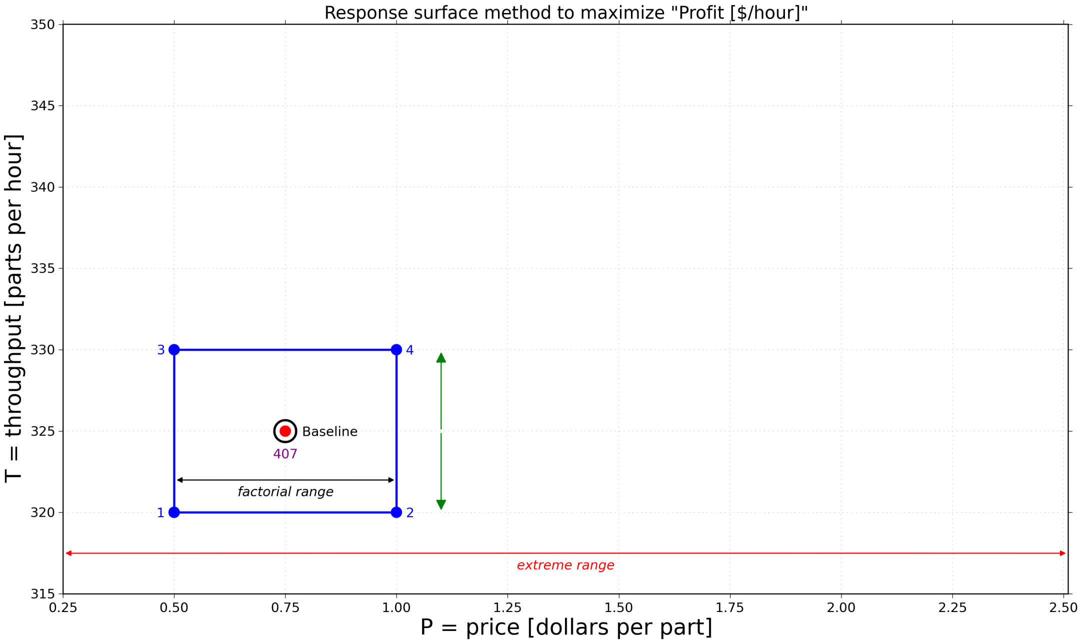
Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
Baseline
407



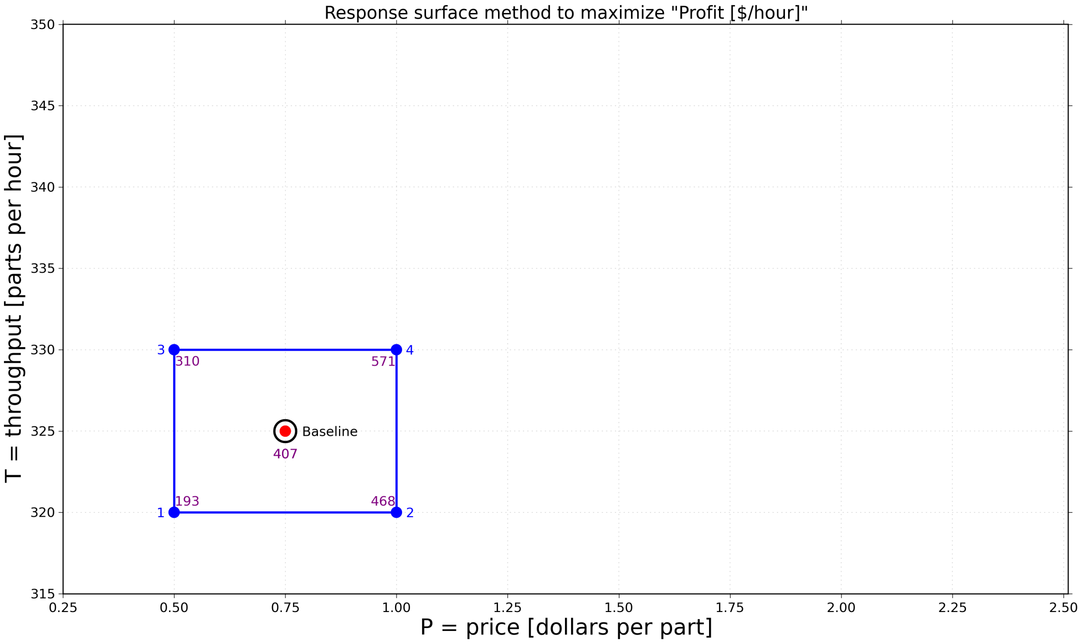
Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
3*
4*
1*
2*
3
4
1
2
Baseline
407

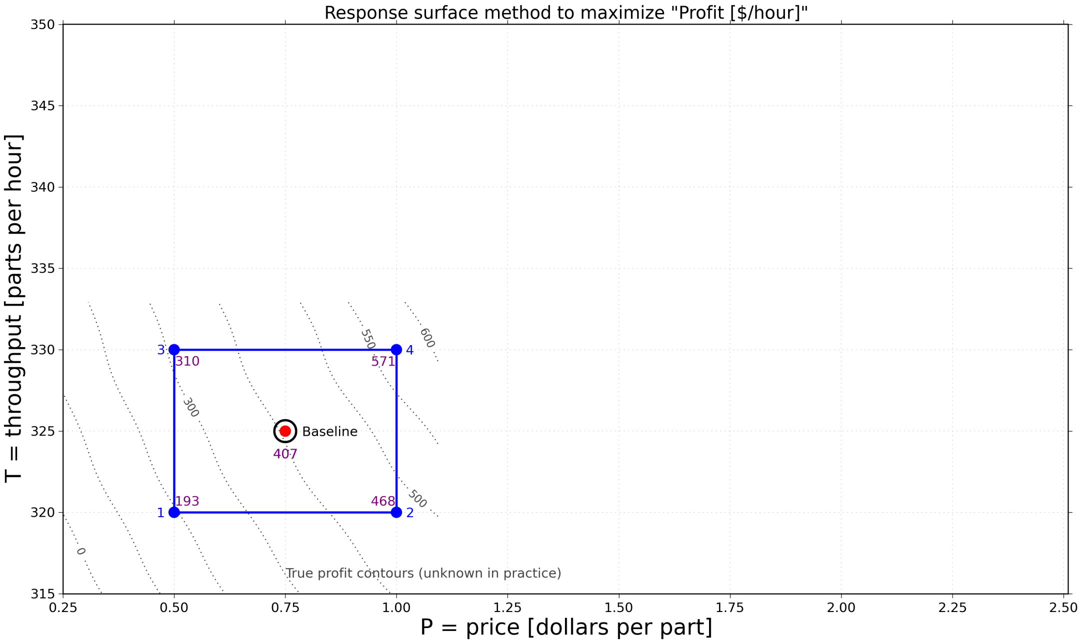

Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
Baseline
407
factorial range
extreme range


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
Baseline
407
factorial range
extreme range

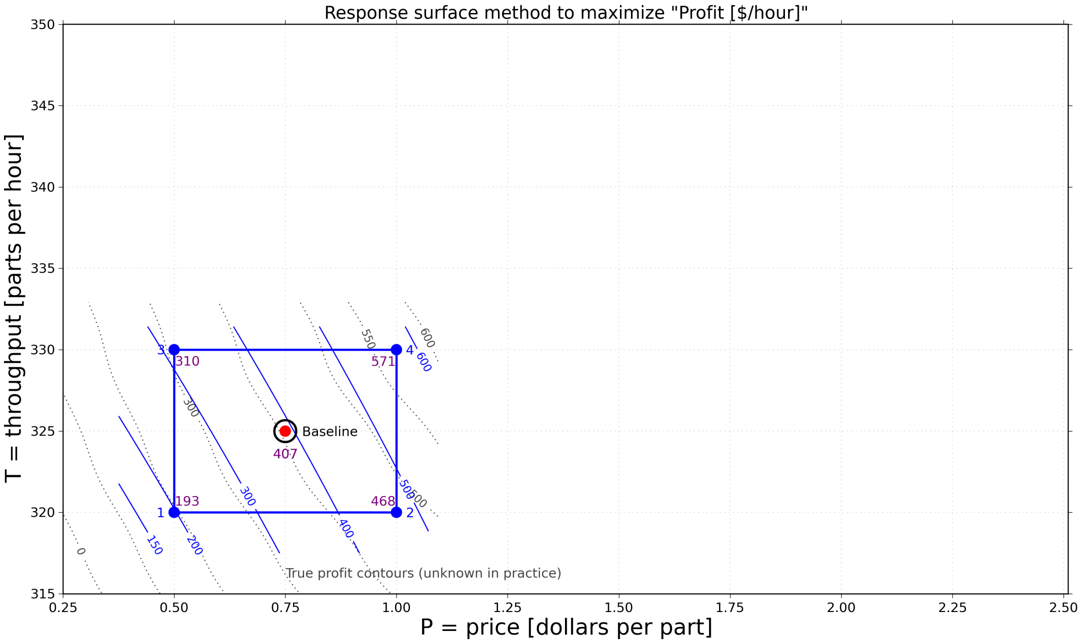


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
Baseline
407
factorial range
extreme range

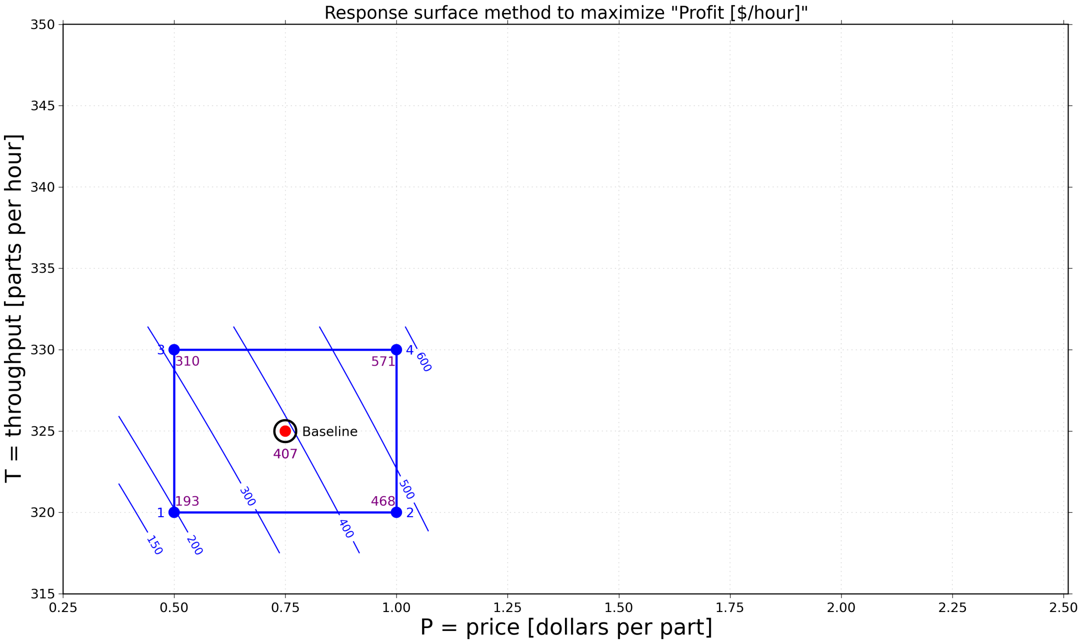

Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
3
310
4
571
Baseline
407
1
193
2
468



Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
193
468
310
571
407
Baseline
0
300
500
550
600
True profit contours (unknown in practice)



Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
Baseline
407
193
468
310
571
1
2
3
4
True profit contours (unknown in practice)



Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
193
468
310
571
407
Baseline
150
200
300
400
500
600

List of all experiments

Experiment	**P**	**T**	$x_{\mathbf{P}}$	$x_{\mathbf{T}}$	**Prediction** = $\hat{y}$	**Actual** = y	**Model**
Current point	$0.75	325	0	0	$ 390	$ 407	1
1	0.50	320	-1	-1	197	193	1
2	1.00	320	$+1$	-1	472	468	1
3	0.50	330	-1	$+1$	314	310	1
4	1.00	330	$+1$	$+1$	575	571	1
5	1.36	330	2.44	1.0	764	$ 669	1
5	1.36	330	2.44	1.0	764	$ 669	1

ADVICE for selecting factorial ranges on a response surface

1. You want to notice a difference between the low and high levels
   - Too close: and you just pick up noise
   - Too far: and you are misled by nonlinearities
2. Don't go to the extremes (a very common mistake)
3. No idea? Start with $\approx 25\%$ of the extreme range

The calculations from real-world units to coded units

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

Price

$$\text{center}_\text{P} = \$0.75$$

$$\text{range}_\text{P} = \$0.50$$

$$x_\text{P} = \frac{\text{P} - \text{center}_\text{P}}{\frac{1}{2}\text{range}_\text{P}}$$

Example: coded value for P $= \$1.00$?

$$x_\text{P} = \frac{1.00 - 0.75}{\frac{1}{2}(0.50)} = \frac{0.25}{0.25} = +1$$

Throughput

$$\text{center}_\text{T} = 325$$

$$\text{range}_\text{T} = 10$$

$$x_\text{T} = \frac{\text{T} - \text{center}_\text{T}}{\frac{1}{2}\text{range}_\text{T}}$$

Example: coded value for T $= 320$?

$$x_\text{T} = \frac{320 - 325}{\frac{1}{2}(10)} = \frac{-5}{5} = -1$$

Using the prediction model

$$y = 390 + 134x_{\mathsf{P}} + 55x_{\mathsf{T}} - 3.5x_{\mathsf{P}}x_{\mathsf{T}}$$

Check the model's "goodness of fit" at the center point:

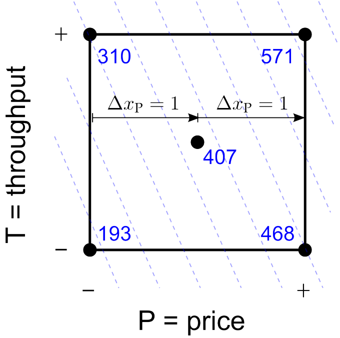
- At the center: $x_{\mathsf{P}} = 0$ and $x_{\mathsf{T}} = 0$
- Predicted $\hat{y} = 390 + 0 + 0 + 0 = \$390$
- Actual $y = \$407$
- That's a difference of $\$17$

Recall the concept of noise from the prior videos?
Perform replicate experiments to estimate noise.

The steepest path of ascent using the local model of the system

$$
\begin{array}{lccccccc}
y = & b_0 & + & b_{\mathsf{P}} x_{\mathsf{P}} & + & b_{\mathsf{T}} x_{\mathsf{T}} & + & b_{\mathsf{PT}}\ x_{\mathsf{P}} x_{\mathsf{T}} \\
y = & 390 & + & 134 x_{\mathsf{P}} & + & 55 x_{\mathsf{T}} & + & (-3.5)\, x_{\mathsf{P}} x_{\mathsf{T}}
\end{array}
$$

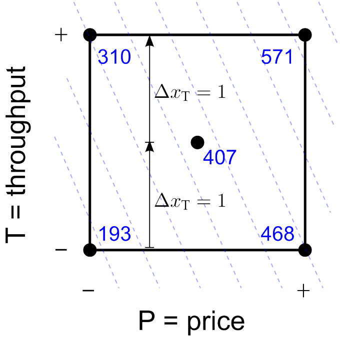
- $b_{\mathsf{P}} = 134$ interpretation:
  - each $\Delta x_{\mathsf{P}} = 1$ increase in x_{P} (coded value) improves y by $\$134$
- $b_{\mathsf{T}} = 55$ interpretation:
  - each $\Delta x_{\mathsf{T}} = 1$ increase in x_{T} (coded value) improves y by $\$55$



The steepest path of ascent using the local model of the system

$$
\begin{array}{lccccccc}
y = & b_0 & + & b_{\mathsf{P}} x_{\mathsf{P}} & + & b_{\mathsf{T}} x_{\mathsf{T}} & + & b_{\mathsf{PT}}\; x_{\mathsf{P}} x_{\mathsf{T}} \\
y = & 390 & + & 134 x_{\mathsf{P}} & + & 55 x_{\mathsf{T}} & + & (-3.5)\, x_{\mathsf{P}} x_{\mathsf{T}}
\end{array}
$$

- $b_{\mathsf{P}} = 134$ interpretation:
  - each $\Delta x_{\mathsf{P}} = 1$ increase in x_{P} (coded value) improves y by \$134
- $b_{\mathsf{T}} = 55$ interpretation:
  - each $\Delta x_{\mathsf{T}} = 1$ increase in x_{T} (coded value) improves y by \$55



A convenient link between coded unit **changes** and real-world **changes**

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

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

Example for throughput

$$\Delta x_{\mathsf{T}} = \frac{\Delta \mathsf{T}}{\frac{1}{2}\text{range}_{\mathsf{T}}}$$

What does the coded value of $\Delta x_{\mathsf{T}} = 1$ represent in the real-world?

$$\Delta x_{\mathsf{T}} = \frac{\Delta \mathsf{T}}{\frac{1}{2}\text{range}_{\mathsf{T}}}$$

$$+1 = \frac{\Delta \mathsf{T}}{\frac{1}{2}(10)}$$

$$\Delta \mathsf{T} = 5 \text{ parts per hour}$$

So $\Delta x_{\mathsf{T}} = +1$ is equivalent to $\Delta \mathsf{T} = +5$

A convenient link between coded unit **changes** and real-world **changes**

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

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

Example for price

$$\Delta x_{\text{P}} = \frac{\Delta \text{P}}{\frac{1}{2}\text{range}_{\text{P}}}$$

What does the coded value of $\Delta x_{\text{P}} = 1$ represent in the real-world?

$$\Delta x_{\text{P}} = \frac{\Delta \text{P}}{\frac{1}{2}\text{range}_{\text{P}}}$$

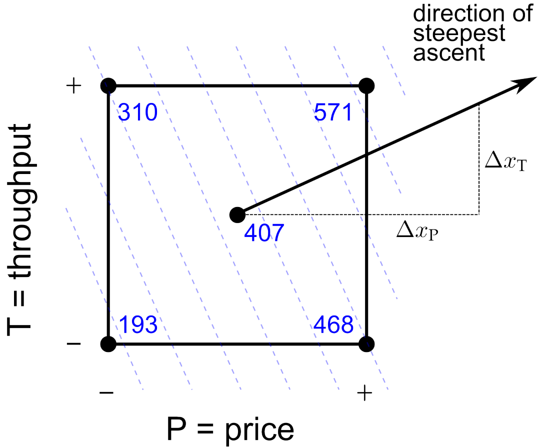
$$+1 = \frac{\Delta \text{P}}{\frac{1}{2}\,(0.50)}$$

$$\Delta \text{P} = \$0.25$$

So $\Delta x_{\text{P}} = +1$ is equivalent to $\Delta \text{P} = +\$0.25$

The steepest path of ascent using the local model of the system

$$
\begin{array}{lccccccc}
y = & b_0 & + & b_{\mathsf{P}} x_{\mathsf{P}} & + & b_{\mathsf{T}} x_{\mathsf{T}} & + & b_{\mathsf{PT}}\ x_{\mathsf{P}} x_{\mathsf{T}} \\
y = & 390 & + & 134 x_{\mathsf{P}} & + & 55 x_{\mathsf{T}} & + & (-3.5)\, x_{\mathsf{P}} x_{\mathsf{T}}
\end{array}
$$



Take a step of $b_{\mathsf{T}} = 55$ in throughput

for every $b_{\mathsf{P}} = 134$ steps in price

But, our actual step is Δx_{T}, so ratio it:

Take a step of $\dfrac{b_{\mathsf{T}}}{\Delta x_{\mathsf{T}}}$ in throughput

for every $\dfrac{b_{\mathsf{P}}}{\Delta x_{\mathsf{P}}}$ steps in price

$$\boxed{\frac{b_{\mathsf{T}}}{\Delta x_{\mathsf{T}}} = \frac{b_{\mathsf{P}}}{\Delta x_{\mathsf{P}}} \quad \Longrightarrow \quad \frac{\Delta x_{\mathsf{P}}}{\Delta x_{\mathsf{T}}} = \frac{b_{\mathsf{P}}}{b_{\mathsf{T}}}}$$

Systematic approach to take a step towards the optimum

	Price	**Throughput**
1. Pick change in coded units in one factor.		$\Delta x_{\mathsf{T}} = 1$ (this was chosen)
2. Find the ratios for the other factor(s).	$\Delta x_{\mathsf{P}} = \dfrac{b_{\mathsf{P}}}{b_{\mathsf{T}}} \cdot \Delta x_{\mathsf{T}}$ $\Delta x_{\mathsf{P}} = \dfrac{134}{55} \cdot 1$	
3. Calculate step size in coded units.	$\Delta x_{\mathsf{P}} = 2.44$	$\Delta x_{\mathsf{T}} = 1$
4. Convert these to real-world *changes*.	$\Delta \mathsf{P} = \Delta x_{\mathsf{P}} \cdot \frac{1}{2}(0.50)$ $\Delta \mathsf{P} = \$0.61$	$\Delta \mathsf{T} = \Delta x_{\mathsf{T}} \cdot \frac{1}{2}(10)$ $\Delta \mathsf{T} = 5$ parts per hour
5. Finally, take a step from the baseline! Get the real-world location of the next experiment.	$\mathsf{P}^{(5)} = \mathsf{P}^{(0)} + \Delta \mathsf{P}$ $\mathsf{P}^{(5)} = \$0.75 + 0.61$ $\mathsf{P}^{(5)} = \$1.36$	$\mathsf{T}^{(5)} = \mathsf{T}^{(0)} + \Delta \mathsf{T}$ $\mathsf{T}^{(5)} = 325 + 5$ $\mathsf{T}^{(5)} = 330$ parts per hour

Systematic approach to take a step towards the optimum

5. Get the real-world location of the next experiment.
6. Convert these back to coded-units.
7. Predict the next experiment's outcome.
8. Now run the next experiment, and record the values

Price	Throughput
$\mathsf{P}^{(5)} = \$1.36$	$\mathsf{T}^{(5)} = 330$ parts per hour
$x_{\mathsf{P}}^{(5)} = 2.44$	$x_{\mathsf{T}}^{(5)} = 1.0$

$$\hat{y} = 390 + 134x_{\mathsf{P}} + 55x_{\mathsf{T}} - 3.5x_{\mathsf{P}}x_{\mathsf{T}}$$

$$\hat{y}^{(5)} = 390 + 134(2.44) + 55(1.0) - 3.5(2.44)(1.0)$$

$$\hat{y}^{(5)} = 390 + 327 + 55 - 8.50$$

$$\hat{y}^{(5)} = 764 \text{ profit per hour}$$

$$y^{(5)} = \$669 \text{ profit per hour}$$

A convenient link between coded unit **changes** and real-world **changes**

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

$$\Delta(\text{coded change}) = \frac{\Delta(\text{real-world change})}{\frac{1}{2}\text{range}}$$

Example for throughput

$$\Delta x_{\mathsf{T}} = \frac{\Delta\mathsf{T}}{\frac{1}{2}\text{range}_{\mathsf{T}}}$$

What does the coded value of $\Delta x_{\mathsf{T}} = 1$ represent in the real-world?

$$\Delta x_{\mathsf{T}} = \frac{\Delta\mathsf{T}}{\frac{1}{2}\text{range}_{\mathsf{T}}}$$

$$+1 = \frac{\Delta\mathsf{T}}{\frac{1}{2}(10)}$$

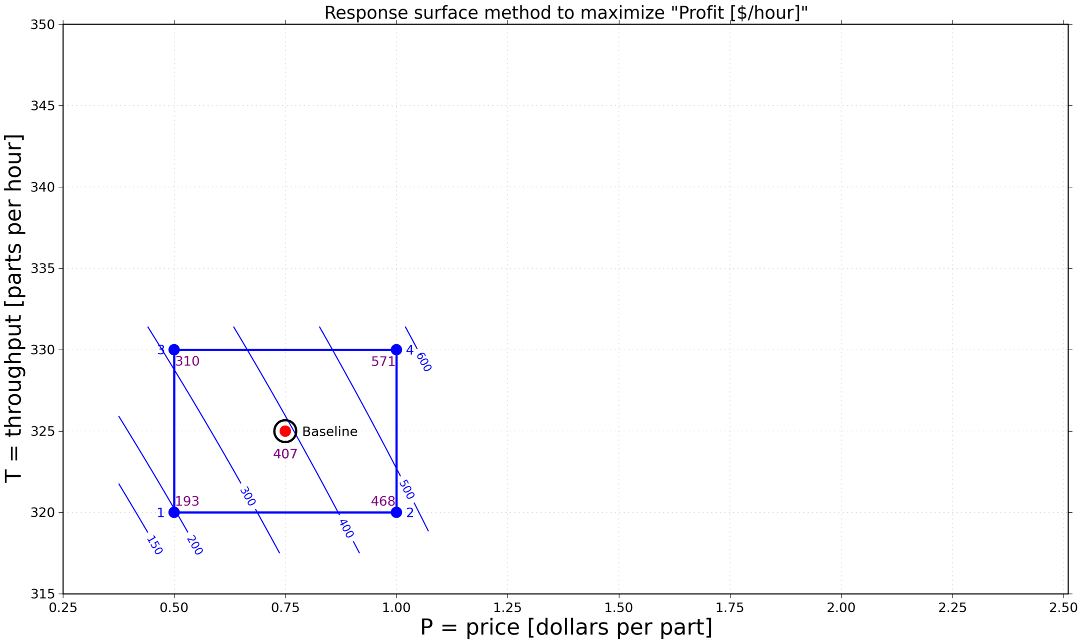
$$\Delta\mathsf{T} = 5 \text{ parts per hour}$$

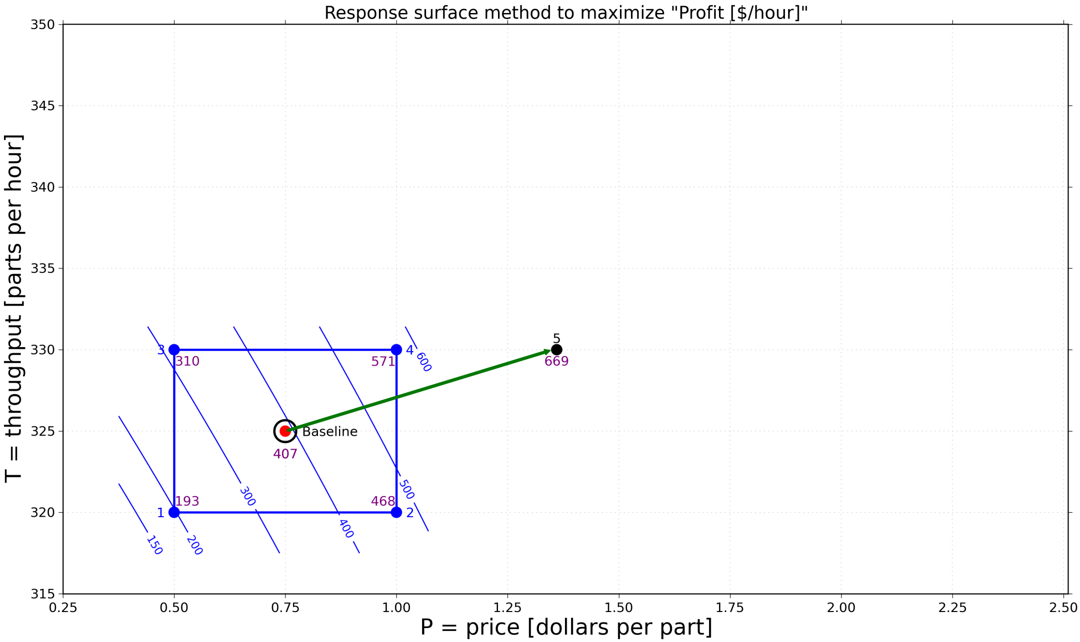
So $\Delta x_{\mathsf{T}} = +1$ is equivalent to $\Delta\mathsf{T} = +5$

Judging the predictive ability from the model

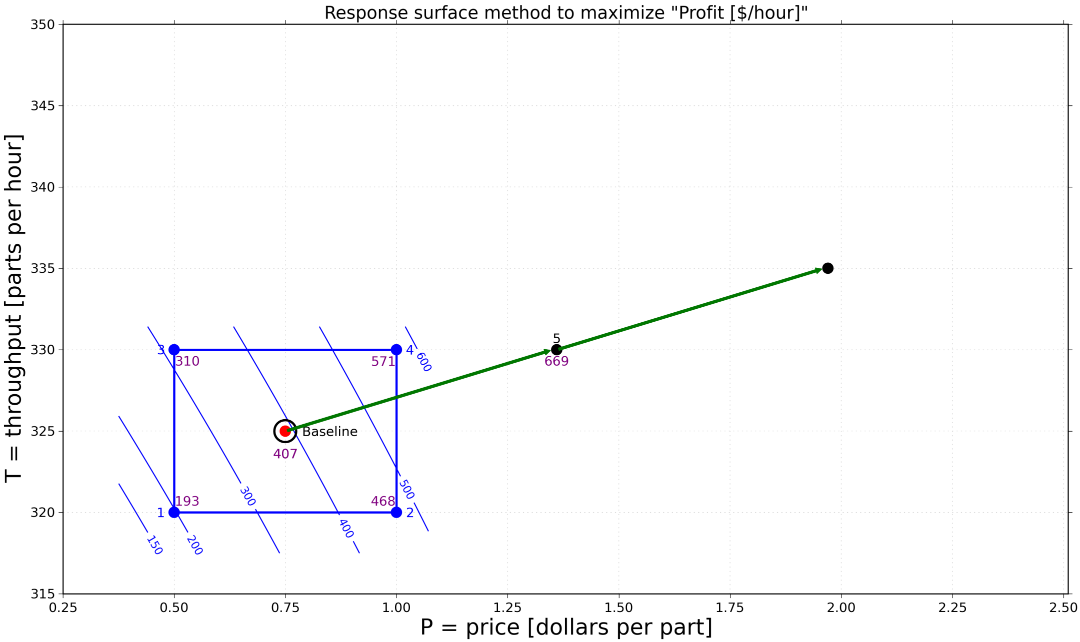
An approximate way to judge the model's prediction ability:

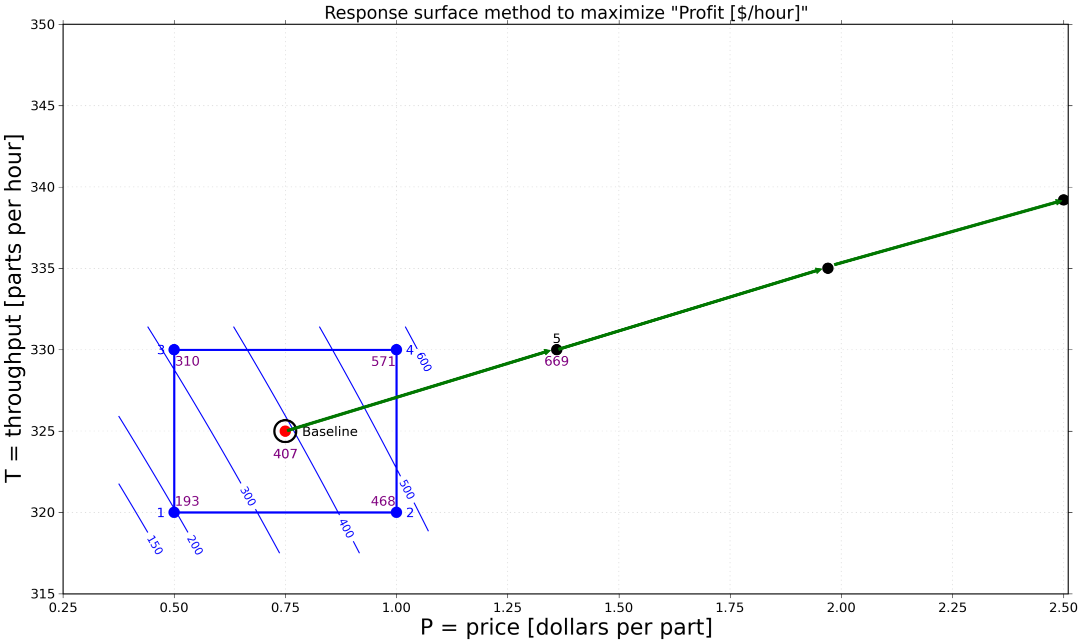
- $\hat{y} = \$764$
- $y = \$669$
- The prediction error is $\$95$.
- Note that the two main effects are: $b_{\mathsf{P}} = 134$ and $b_{\mathsf{T}} = 55$
- So this error is comparable to these;
  - it's smaller than the effect of a $\Delta x_{\mathsf{P}} = 1$
  - it's larger than the effect of a $\Delta x_{\mathsf{T}} = 1$
  - so that error is substantial


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
193
468
310
571
407
Baseline
150
200
300
400
500
600

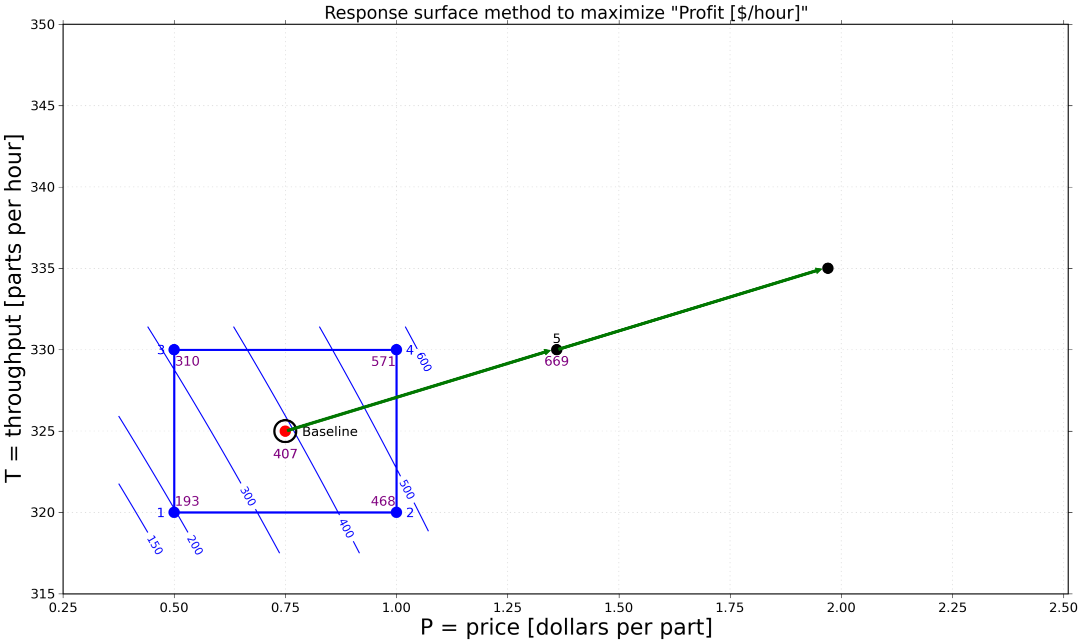


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
Baseline
407
193
468
310
571
669
1
2
3
4
5
150
200
300
400
500
600
315
320
325
330
335
340
345
350
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50

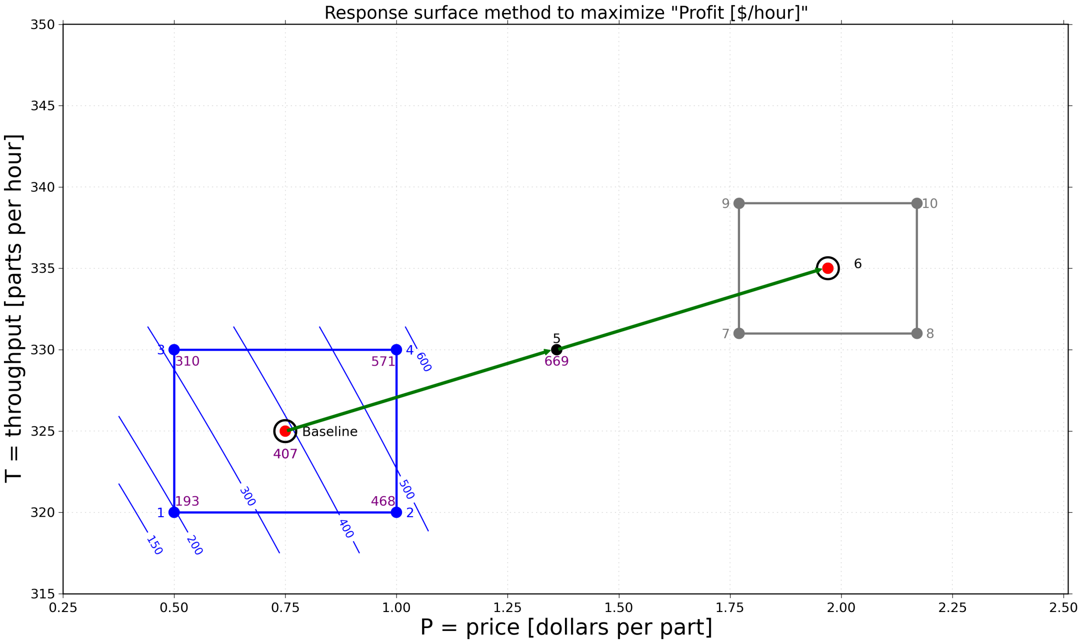

Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
5
193
468
310
571
407
669
Baseline
150
200
300
400
500
600

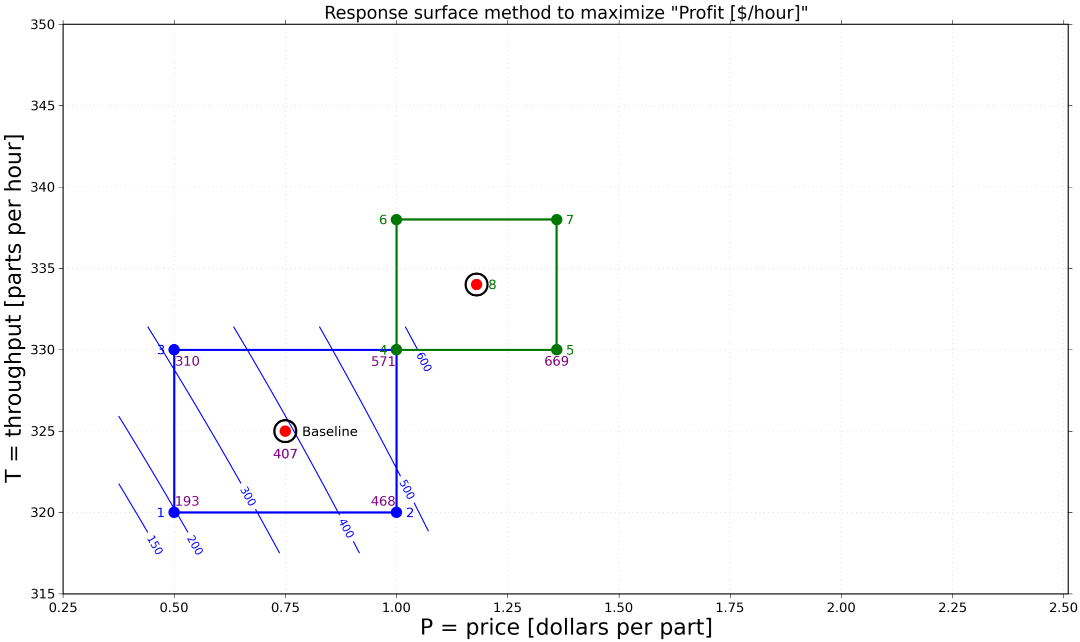


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
5
193
468
310
571
407
669
Baseline
150
200
300
400
500
600



Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
5
193
468
310
571
407
669
Baseline
150
200
300
400
500
600


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
Baseline
407
1
193
2
468
3
310
4
571
5
669
6
7
8
9
10
150
200
300
400
500
600



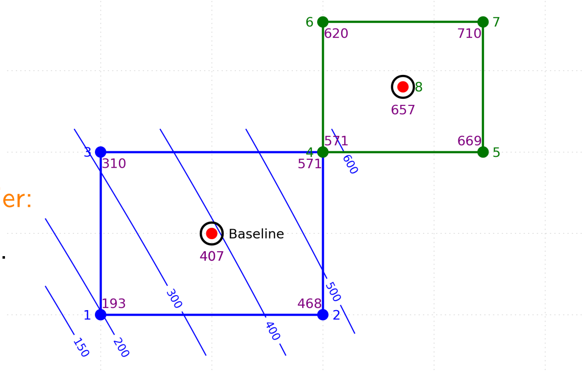
Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
5
6
7
8
193
468
310
571
669
407
Baseline
150
200
300
400
500
600

List of all experiments

Experiment	P	T	x_{P}	x_{T}	Prediction $= \hat{y}$	Actual $= y$	Model
4	1.00	330	-1	-1		\$ 571	2
5	1.36	330	$+1$	-1		669	2
6	1.00	338	-1	$+1$		620	2
7	1.36	338	$+1$	$+1$		710	2
8	1.18	334	0	0		657	2

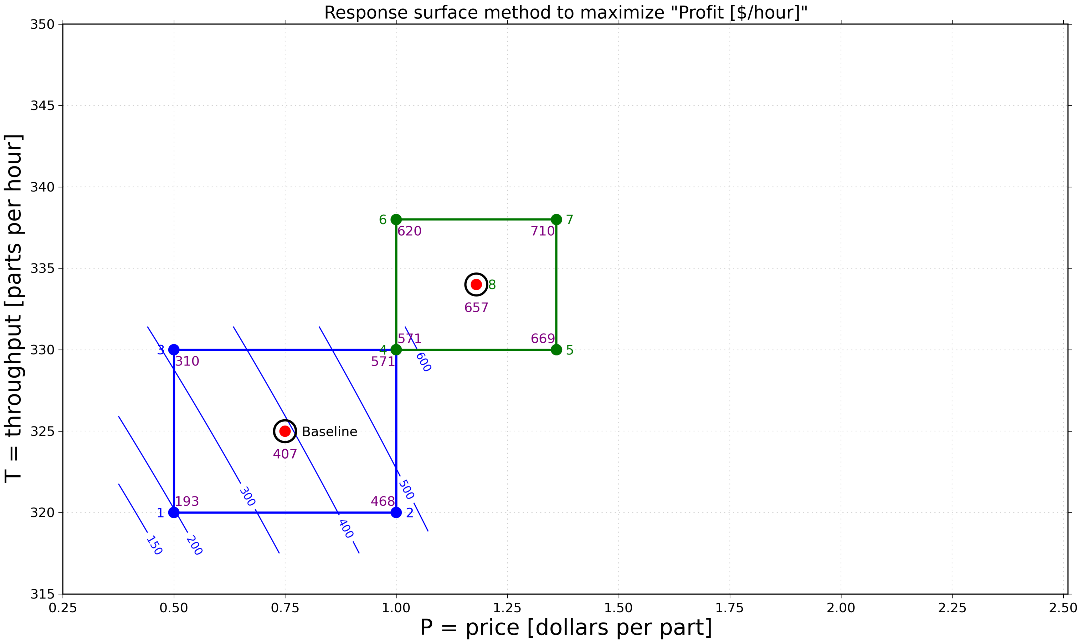
Try taking the next step up the mountain on your own

- Visualize the results first
- Build a model using computer software
- Sketch a contour plot by hand, or with software



Now use the 8 step approach we showed earlier:

1. Pick change in coded units in one factor.
   - Use $\Delta x_{\text{P}} = 1.5$
2. Find the ratios for the other factor(s).
3. Calculate step size in coded units.
4. Convert these to real-world *changes*.
5. Get the real-world location of the next experiment.
6. Convert these back to coded-units.
7. Predict the next experiment's outcome.
8. Run the next experiment, and record the outcome value. http://yint.org/run-expt


Response surface method to maximize "Profit [$/hour]"
T = throughput [parts per hour]
P = price [dollars per part]
350
345
340
335
330
325
320
315
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
1
2
3
4
5
6
7
8
193
468
310
571
571
669
620
710
657
407
Baseline
150
200
300
400
500
600