Getting standard errors for estimates

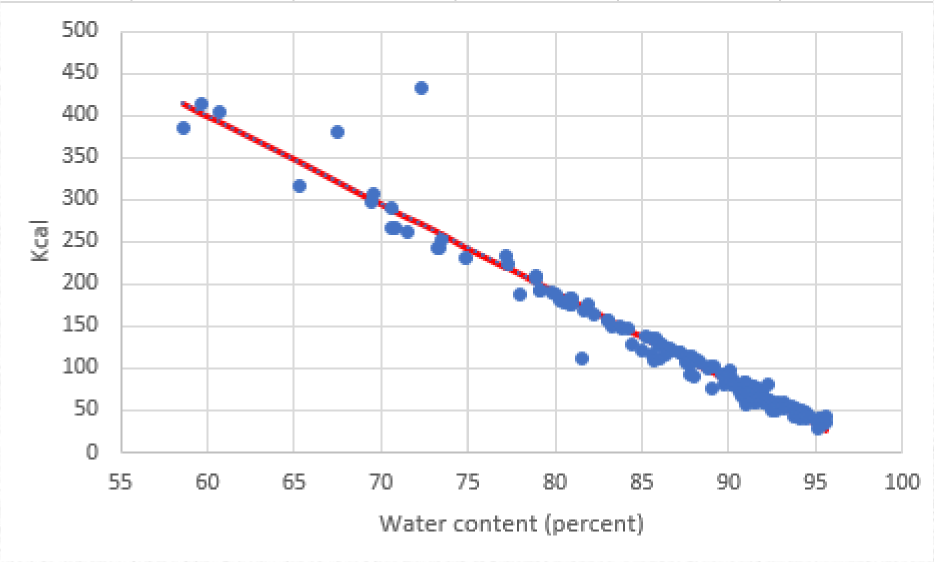
Problem: Solver does not provide standard errors

- Standard errors are measures of precision of estimates
  - A new set of data will give us different estimates
  - SE's used to measure how different we expect them to be
- They are also used for statistical hypothesis testing, and for calculating confidence intervals
- Slope and intercept estimates by themselves are not terribly useful without SE's
- We can estimate the SE's numerically with a little work, using **finite difference approximation**

Basis for numerical SE estimation

- The predicted values for the line are based on the estimated parameters
- If we vary one of the parameters at a time, we will change the predicted values by some amount
- Amount of change in predicted value per unit change in parameter value can be measured
  - Big changes in predicted value indicate precise estimates
  - Small changes in predicted value indicate imprecise estimates

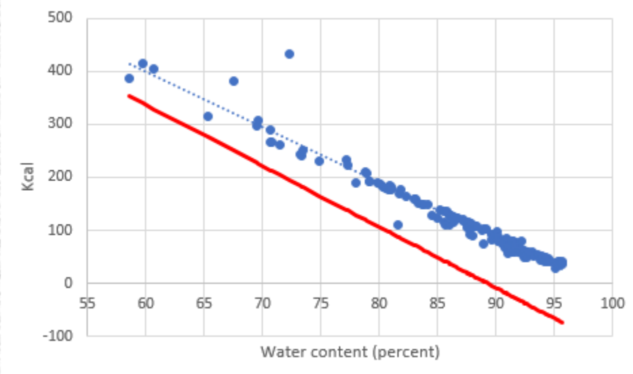
The linear data...



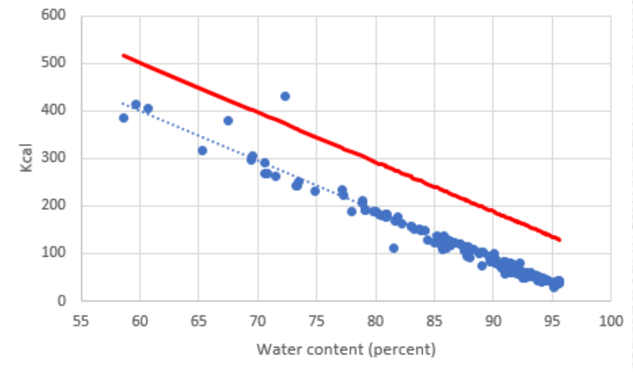
Changing slope or intercept

Slope

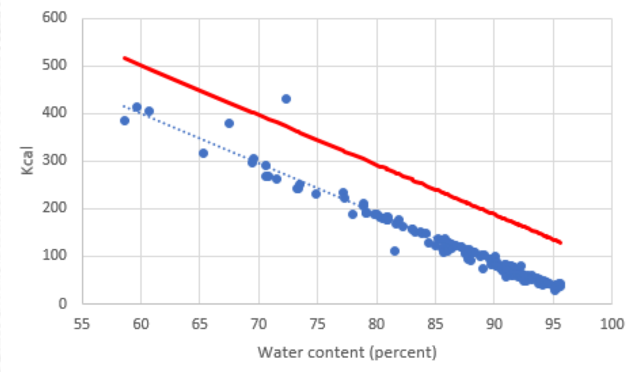
Increase by 10% →
(change by same relative amount)

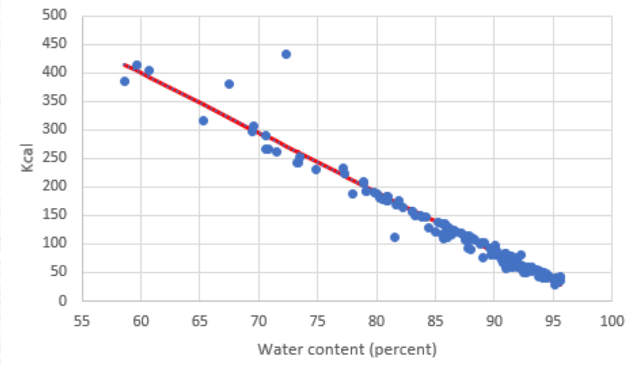


Add 1 →
(change by same absolute amount)



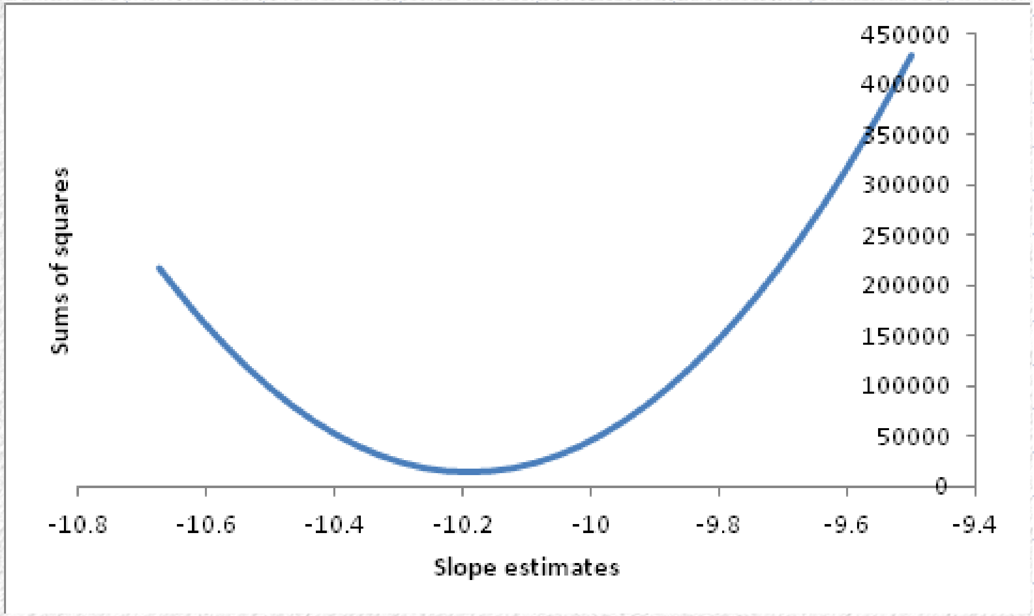
Intercept

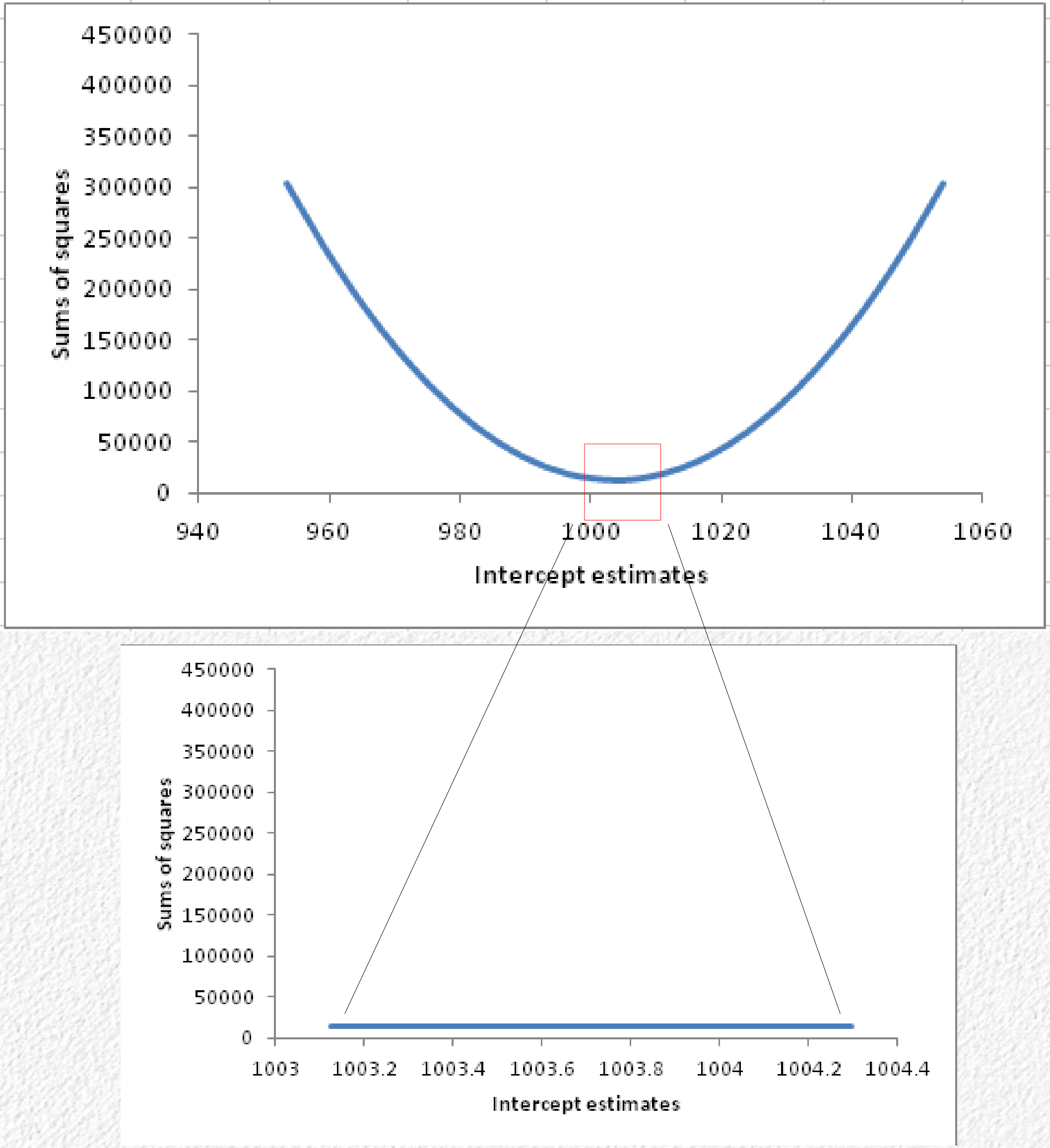




Changing the line away from best fit will make the sums of squared deviations ***larger***

On an absolute basis the fit gets much worse if you change the slope than the intercept by the same amount





Fit of the line is much more sensitive to change in slope than change in intercept

So, the estimate of the slope will be more precise – the range of possible slopes that are consistent with the data is narrow

Finite difference approximation of SE

- Standard errors can be calculated using a matrix (P) of summed squared differences in *predicted value* per unit change in the estimates
- These approximate the first partial derivative of the line with respect to the estimate
  - Derivatives = slopes of lines tangent to a curve
  - Partial derivatives = derivative with respect to just one term, treating all others as constants
- The inverse of P can be used to estimate standard errors
- We can calculate P using tiny, finite changes to the parameters

The P matrix for the coefficients of a line

s = slope
i = intercept
f = predicted value for the line

$$P = \begin{bmatrix} \Sigma\left(\frac{\Delta f}{\Delta s}\right)^2 & \Sigma \frac{\Delta f}{\Delta s}\frac{\Delta f}{\Delta i} \\ \Sigma \frac{\Delta f}{\Delta s}\frac{\Delta f}{\Delta i} & \Sigma\left(\frac{\Delta f}{\Delta i}\right)^2 \end{bmatrix}$$

Cross products

Squared differences

- Derivatives used for continuous functions, instantaneous change
- We will use deltas as an approximation

Finite difference approximation

1) Start with the Solver estimates
2) Change the slope by a tiny amount
3) Calculate the change in the predicted values, divided by the change in the parameter
4) Return the slope to its Solver-estimated value
5) Repeat with the intercept
6) Calculate squares and cross-products of the deviations, and sum them to estimate P

1. Predicted values from Solver estimates

	A	B	C	D
1	food	water	kcal	Predicted kcal
2	watercress	95.11	28	34.499
3	pak-choi cabbage	95.32	34	32.359
4	iceberg lettuce	95.64	36	29.098
5	white gourd	95.54	36	30.117
6	green leaf lettuce	95.07	38	34.907
7	cucumber	95.23	40	33.276
110	taro root	70.64	290	283.859
111	palm hearts	69.5	298	295.476
112	yam	69.6	306	294.457
113	plantains	65.28	316	338.479
114	soybeans	67.5	381	315.856
115	garlic	58.58	386	406.755
116	prairie turnips	60.69	405	385.253
117				
118	Slope	-10.1904		
119	Intercept	1003.709		

Calculated as:

-10.1904 (water) + 1003.709

2. Change the slope – multiply slope by 1.000001

	food	water	kcal	Predicted kcal
2	watercress	95.11	28	34.489
3	pak-choi cabbage	95.32	34	32.349
4	iceberg lettuce	95.64	36	29.088
5	white gourd	95.54	36	30.107
6	green leaf lettuce	95.07	38	34.897
7	cucumber	95.23	40	33.267
110	taro root	70.64	290	283.851
111	palm hearts	69.5	298	295.469
112	yam	69.6	306	294.449
113	plantains	65.28	316	338.472
114	soybeans	67.5	381	315.850
115	garlic	58.58	386	406.749
116	prairie turnips	60.69	405	385.247
117				
118	Slope	-10.1905		
119	Intercept	1003.709		

Slight change in predicted values

Slope changed

Intercept kept constant

3. Change in predicted value divided by change in slope

D	E	F
Predicted kcal	Predicted kcal from Solver estimates	Differences (change in F)
34.489	34.4990905	0.010
32.349	32.3591036	0.010
29.088	29.0981712	0.010
30.107	30.1172126	0.010
34.897	34.906707	0.010
33.267	33.2762408	0.010
283.851	283.858514	0.007
295.469	295.475585	0.007
294.449	294.456544	0.007
338.472	338.479131	0.007
315.850	315.856413	0.007
406.749	406.754903	0.006
385.247	385.25313	0.006

÷

118	***Solver's solutions (best fit)***	
119	Slope	-10.1904
120	Int	1003.709
121		
122	***Altered solutions***	
123	Slope	-10.1905
124	Intercept	1003.709
125		
126	Change in slope	0.000102
127		

=

dY/ds
-95.11
-95.32
-95.64
-95.54
-95.07
-95.23
-70.64
-69.5
-69.6
-65.28
-67.5
-58.58
-60.69

4,5. Return slope to Solver value, repeat with intercept

Predicted kcal	Predicted kcal from Solver estimates	Differences (change in F)
34.500	34.4990905	-0.001
32.360	32.3591036	-0.001
29.099	29.0981712	-0.001
30.118	30.1172126	-0.001
34.908	34.906707	-0.001
33.277	33.2762408	-0.001
283.860	283.858514	-0.001
295.477	295.475585	-0.001
294.458	294.456544	-0.001
338.480	338.479131	-0.001
315.857	315.856413	-0.001
406.756	406.754903	-0.001
385.254	385.25313	-0.001

÷

118	***Solver's solutions (best fit)***	
119	Slope	-10.1904
120	Int	1003.709
121		
122	***Altered solutions***	
123	Slope	-10.1904
124	Intercept	1003.71
125		
126	Change in intercept	-0.001

=

dY/di
1
1
1
1
1
1
1
1
1
1
1
1
1

Calculating sums of squares and cross products the “easy” way

- We just made a matrix with columns dY/ds and dY/di
- Want to square these and sum them for the main diagonal of P
- We want to multiply them together and sum them for the off-diagonals
- We can do this in one calculation using matrix multiplication – multiply the matrix by its transpose
- What's a transpose?

Transpose of a matrix

- A matrix is “transposed” by swapping the rows and columns

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \boldsymbol{A}' = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

$$\boldsymbol{A}' \times \boldsymbol{A} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} aa + cc & ab + cd \\ ba + dc & bb + dd \end{bmatrix}$$

Pre-multiplying a matrix by its transpose gives sums of squares and cross products

6. Calculate sums of squares and cross-products of dY/ds and dY/di

Array formula for matrix multiplication

{=MMULT(R2:EB3,N2:O116)}

M	N	O	P	Q	R	S	T	U	V
	dY/ds	dY/di							
	-95.11	1		dY/ds	-95.11	-95.32	-95.64	-95.54	-95.07
	-95.32	1		dY/di	1	1	1	1	1
	-95.64	1							
	-95.54	1							
	-95.07	1							
	-95.23	1			P Matrix, by finite difference approximation				
	-95.27	1							
	-94.12	1				Slope	Intercept		
	-95.43	1			Slope	870102.3	-9965.21		
	-94.39	1			Intercept	-9965.21	115		
	-95.64	1							
	-94.64	1							
	-93.79	1							

dY/ds and dY/di, transposed

dY/ds and dY/di calculated by altering slopes and intercepts

Matrix multiplication of dY/ds, dY/di by transposed dY/ds, dY/di

Finally, calculate the standard errors

- To calculate the standard errors from the P matrix, we need to:
  - Invert the matrix
  - Multiply the square root of each of the values on the main diagonal by the standard error of Y
- This will give us both the standard errors we need
- What's a matrix inverse?

Matrix inverse

For a single number, a, the inverse is 1/a
a x 1/a = 1

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \boldsymbol{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\boldsymbol{A} \times \boldsymbol{A}^{-1} = \boldsymbol{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{A} \times \boldsymbol{I} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = ?$$

We will let the computer solve inverses for us...

The inverse of our P matrix

{=MINVERSE(S10:T11)}

P Matrix, by finite difference approximation

	Slope	Intercept
Slope	870102.3	-9965.21
Intercept	-9965.21	115

Inverse of P matrix

	Slope	Intercept
Slope	0.000152	0.013175
Intercept	0.013175	1.150391

Standard errors

=SQRT(J122)*G120

F	G	H	I	J	K
	Squared deviations				
	36.0				
	2.7				
	47.6				
	40.7				
	9.6				
Sum Sq.	13641.3		Inverse of the P matrix		
SE(y)	10.98725282				
				Slope	Intercept
SE(s)	0.135480164		Slope	0.000152	0.013175
SE(i)	11.7845214		Intercept	0.013175	1.150391

Standard error of y is:

$$SE(Y)=\sqrt{\frac{SSY}{n-2}}=10.987$$

SE of slope is:

$$SE(s)=\sqrt{P^{-1}{}_{11}}\,SE(Y)=\sqrt{0.000152}\,(10.987)$$

SE of intercept is:

$$SE(s)=\sqrt{P^{-1}{}_{22}}\,SE(Y)=\sqrt{1.150391}\,(10.987)$$

Today...

- We will calculate the standard errors for the fitted lines from last time