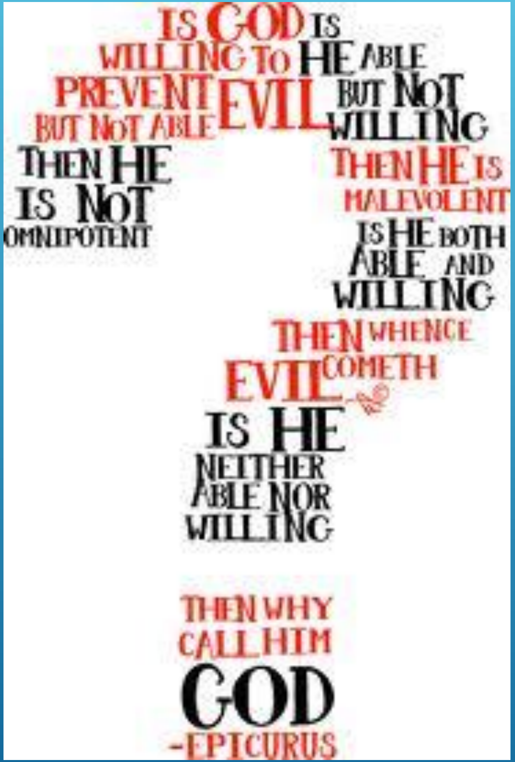
- Education is the only discipline in which it's acceptable to substitute an arbitrary value when data is incomplete, as illustrated by the practice of assigning a score of zero to assessments that are not submitted. This erroneously assumes that if we fail to measure something it doesn't exist. It would be more honest to admit we simply don't know about things for which we lack data. This presentation will explore some strategies to make more defensible decisions in the face of incomplete data, including simple methods to perform regressions that handle gaps in the data in a defensible way.

THE ABSENCE OF PROOF

Is not the proof of absence.

ANOTHER WAY TO THINK ABOUT ZEROS…

The Atheist: "I can't prove that God exists, therefore, there is no god."



IS GOD WILLING TO PREVENT EVIL BUT NOT ABLE
THEN HE IS NOT OMNIPOTENT
IS HE ABLE BUT NOT WILLING
THEN HE IS MALEVOLENT
IS HE BOTH ABLE AND WILLING
THEN WHENCE COMETH EVIL
IS HE NEITHER ABLE NOR WILLING
THEN WHY CALL HIM GOD
-EPICURUS

The Agnostic: "I can't prove that God exists, therefore, I don't know."

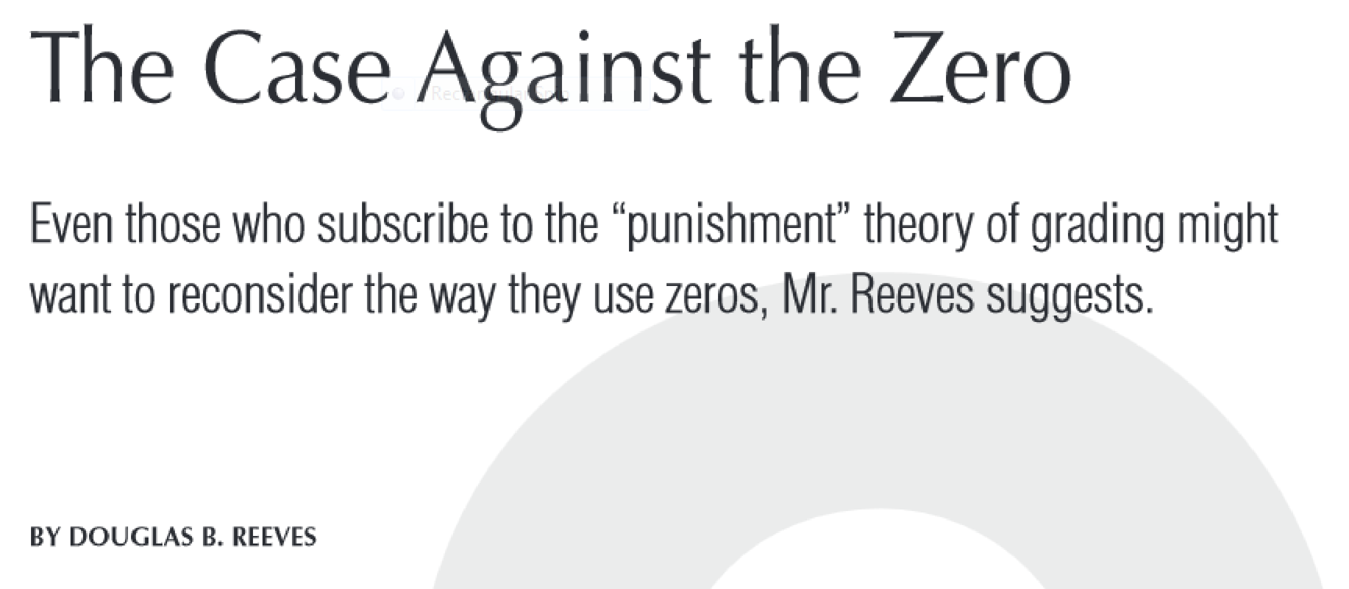
POP QUIZ: WHAT TYPE OF SCALE DOES THE CELSIUS SYSTEM USE?

Relative or Absolute

POP QUIZ: WHAT TYPE OF SCALE DOES THE KELVIN SYSTEM USE?

Relative or Absolute

IS YOUR GRADING SCALE ABSOLUTE OR RELATIVE?

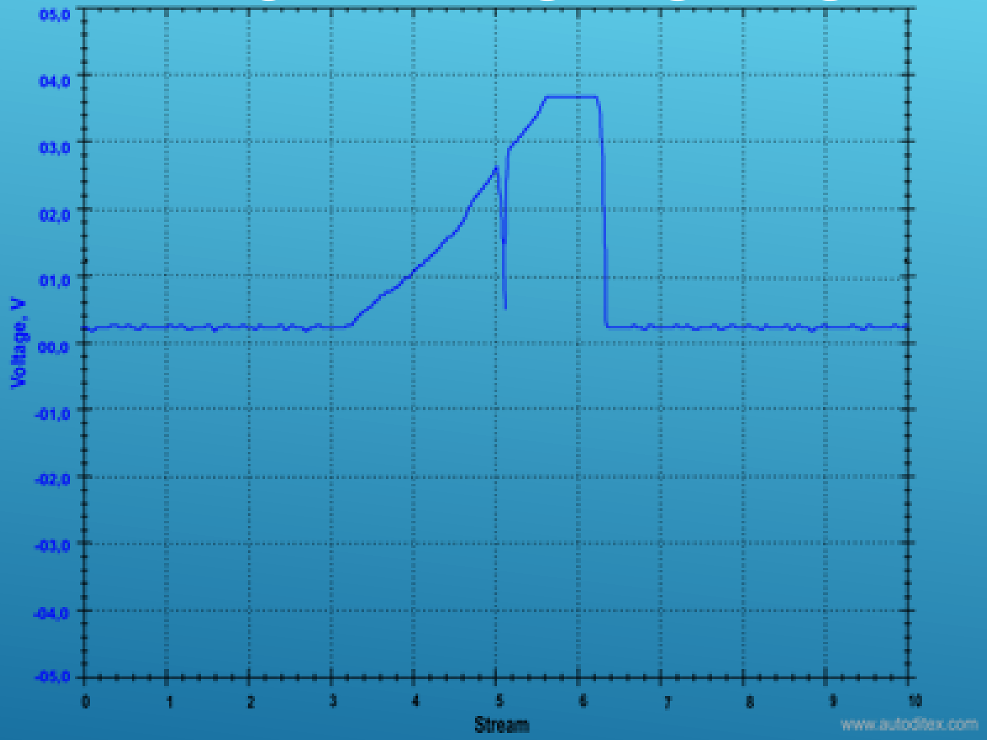


The Case Against the Zero

Even those who subscribe to the “punishment” theory of grading might want to reconsider the way they use zeros, Mr. Reeves suggests.

BY DOUGLAS B. REEVES

RATIONALITY CHECKING



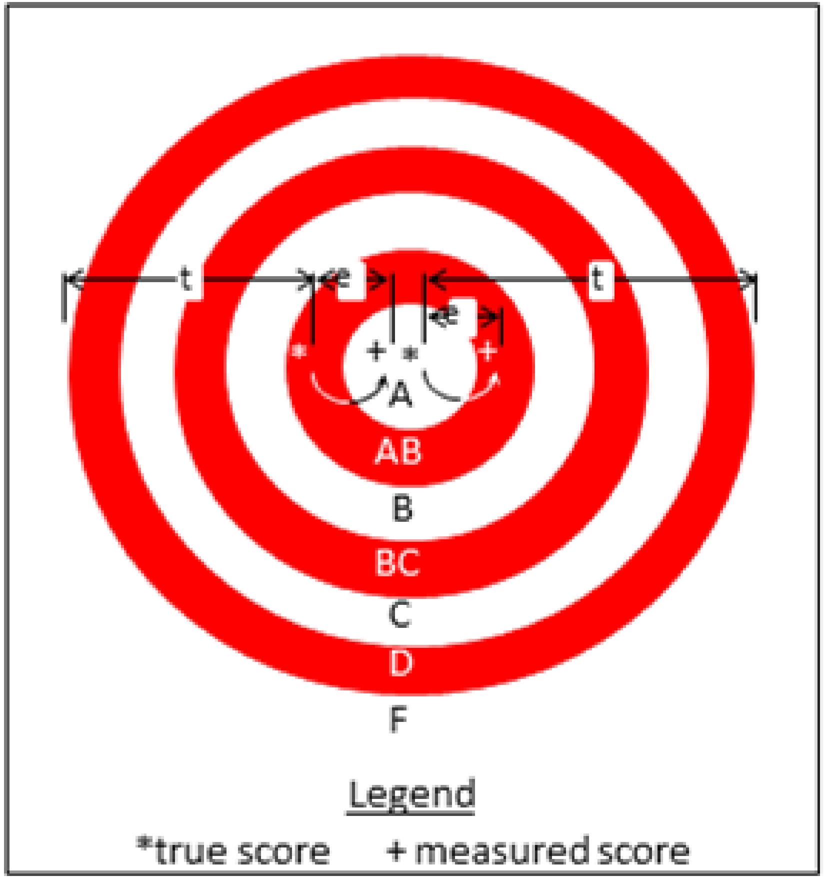
SIGNAL AND NOISE

True score

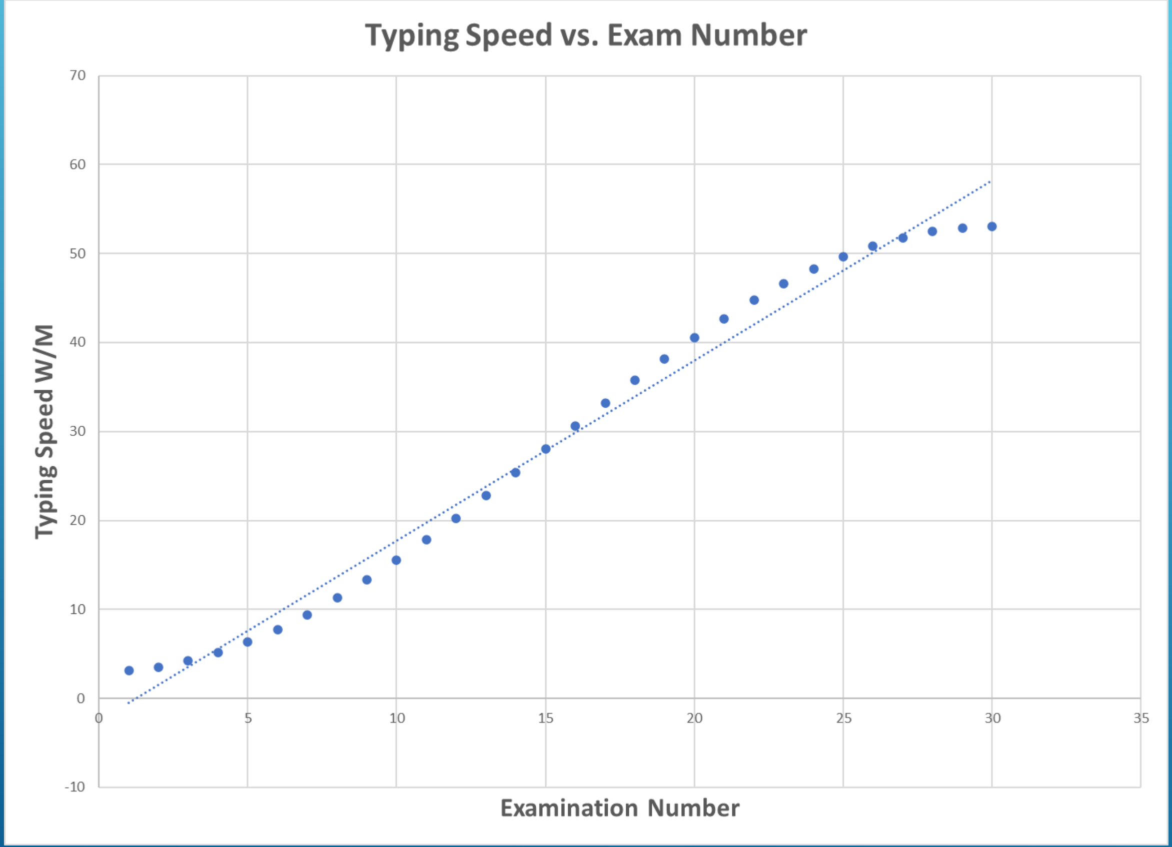
- A.K.A. Construct Relevant Variance

Error score

-A.K.A. Construct Irrelevant Variance



A VIEW WITH OMNISCIENCE – IS GROWTH LINEAR?

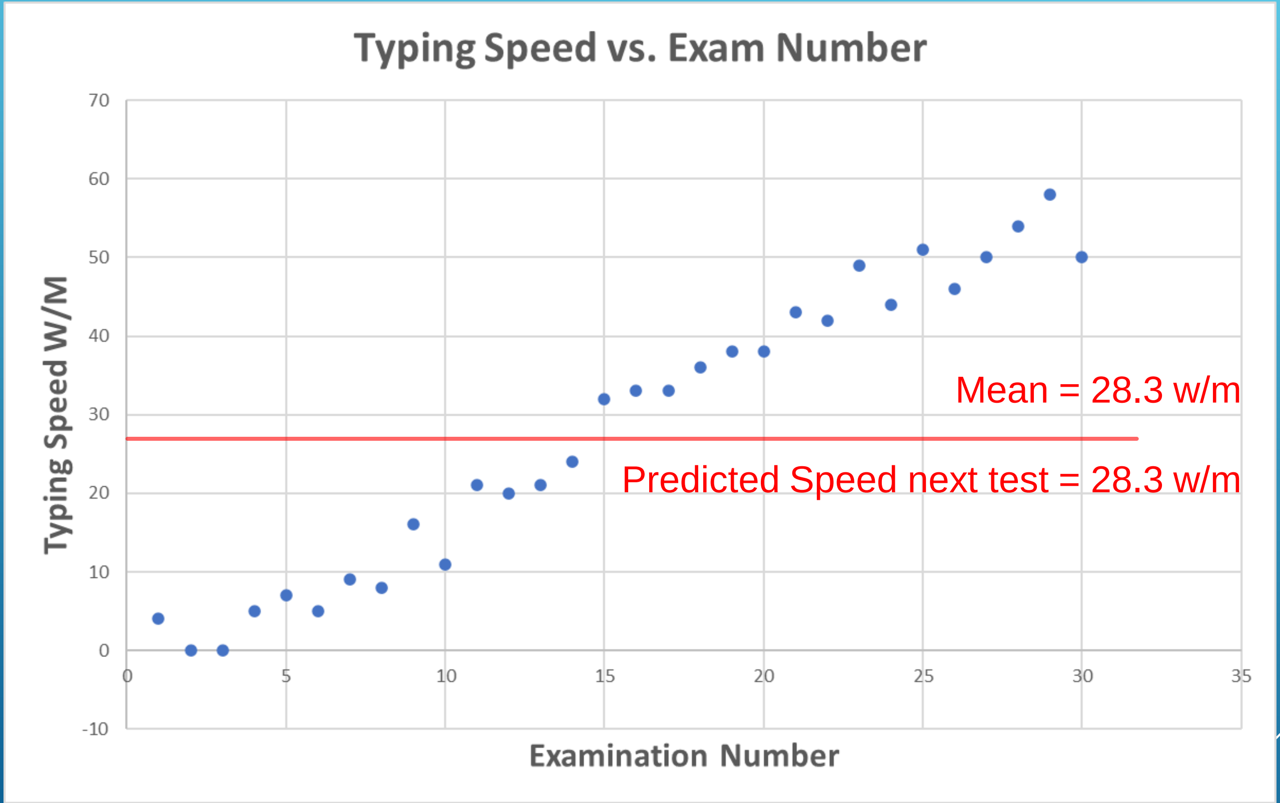


FIXES FOR PROBLEMS OF PRECISION

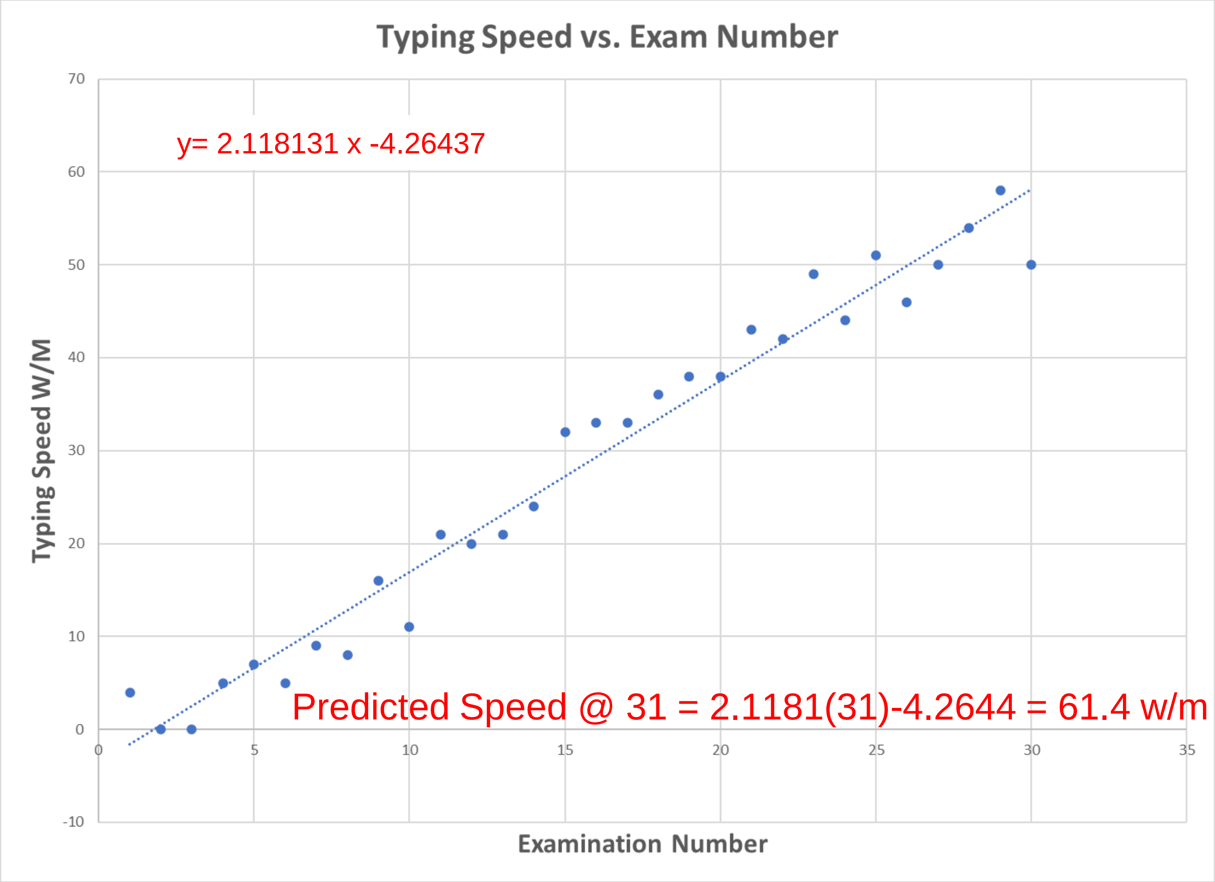
The law of large numbers - regression to the mean

Did you know that the most intelligent women tend to marry men who are less intelligent than they are?

REALLY SIMPLE REGRESSION - AVERAGING

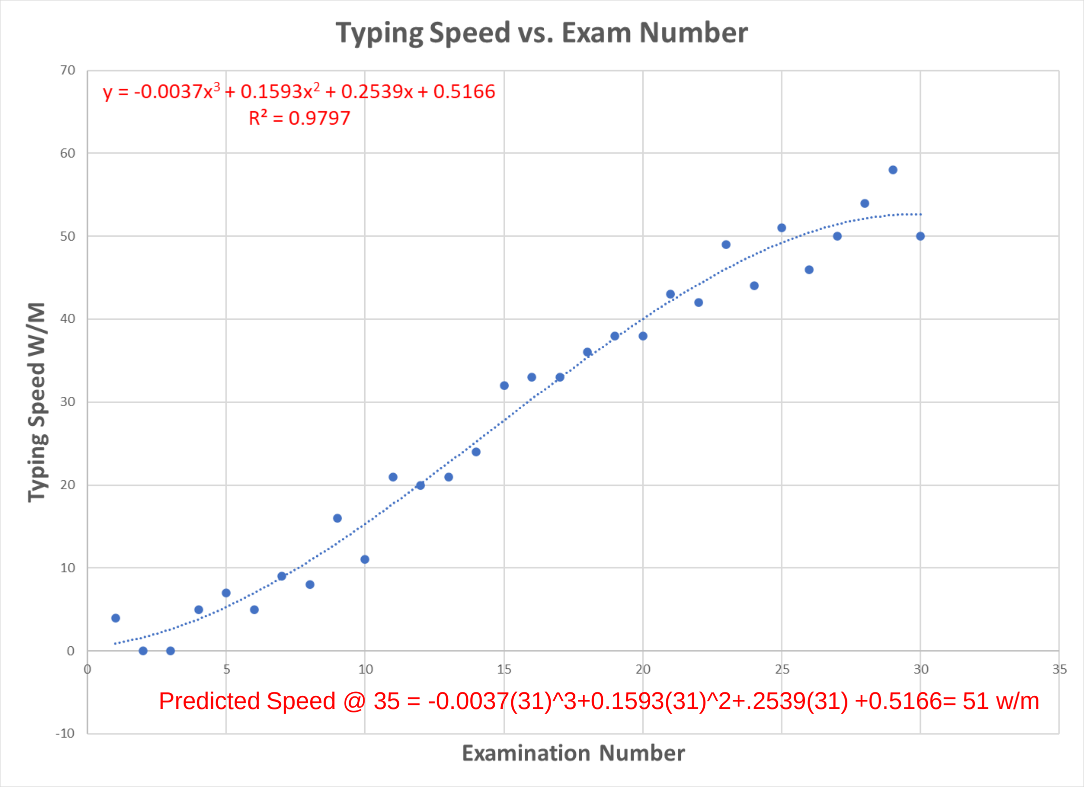


SLIGHTLY BETTER REGRESSION - LINEAR



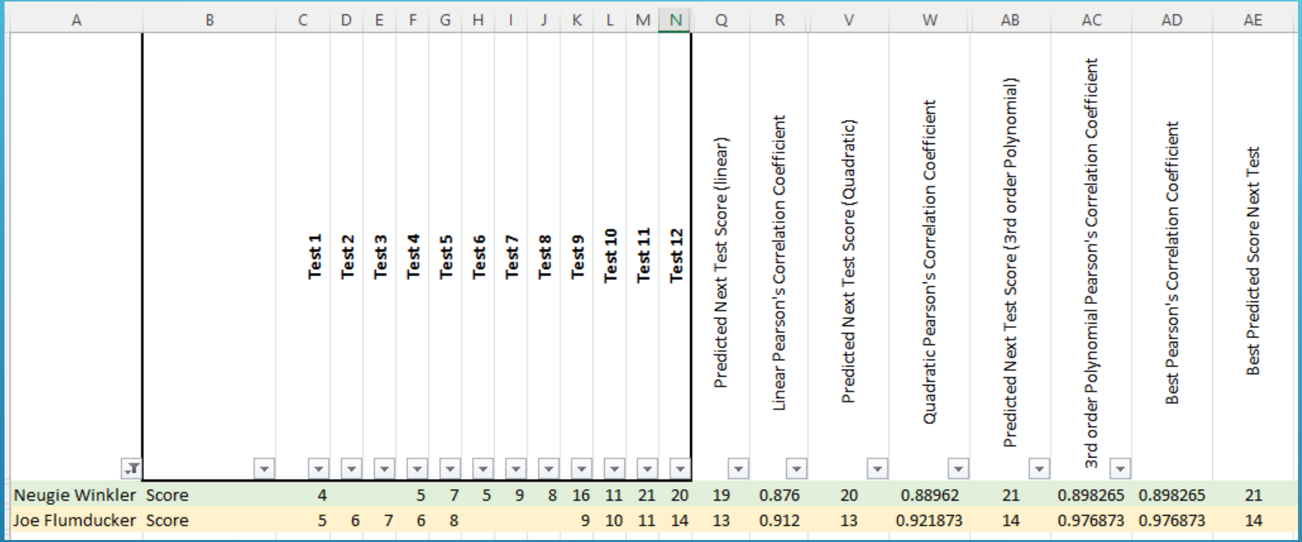
Predicted Speed @ 31 = 2.1181(31)-4.2644 = 61.4 w/m

BETTER REGRESSION - POLYNOMIAL



Predicted Speed @ 35 = -0.0037(31)^3+0.1593(31)^2+.2539(31) +0.5166= 51 w/m

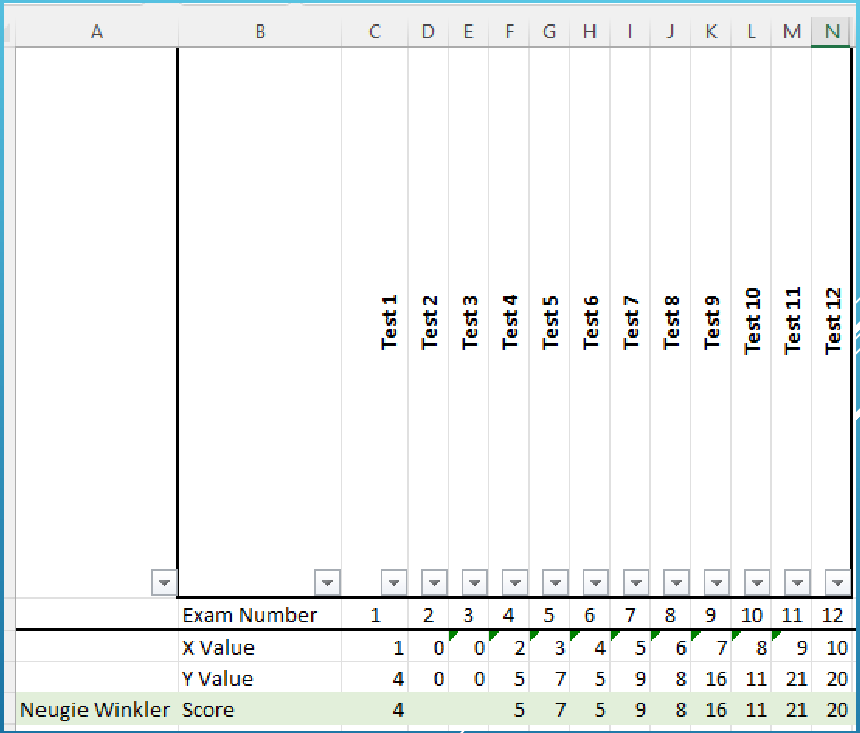
HOW TO SETUP A GRADEBOOK



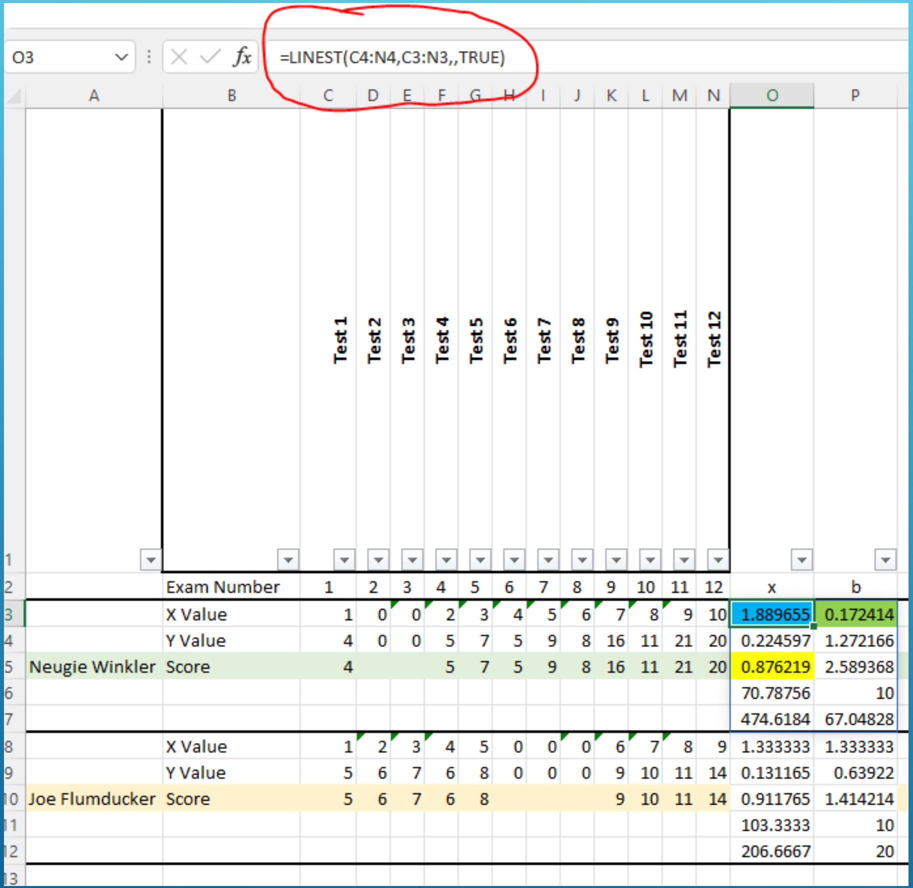
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Q	R	V	W	AB	AC	AD	AE
		Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12	Predicted Next Test Score (linear)	Linear Pearson's Correlation Coefficient	Predicted Next Test Score (Quadratic)	Quadratic Pearson's Correlation Coefficient	Predicted Next Test Score (3rd order Polynomial)	3rd order Polynomial Pearson's Correlation Coefficient	Best Pearson's Correlation Coefficient	Best Predicted Score Next Test
Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	19	0.876	20	0.88962	21	0.898265	0.898265	21
Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	13	0.912	13	0.921873	14	0.976873	0.976873	14

- Graphs
- LINEST function
  - Various Regressions
    - Linear
    - Polynomial
    - Log*
    - Power*
- Only appropriate for repeated measures of the same thing.

- The formulas in the hidden rows assign X and Y values of zero to assessments not taken.
- This tends to have a minimal effect on predicted score, because Y intercept values tend to be small
- Can be avoided all together with sorting



A	B	C	D	E	F	G	H	I	J	K	L	M	N
		Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12
	Exam Number	1	2	3	4	5	6	7	8	9	10	11	12
	X Value	1	0	0	2	3	4	5	6	7	8	9	10
	Y Value	4	0	0	5	7	5	9	8	16	11	21	20
Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20

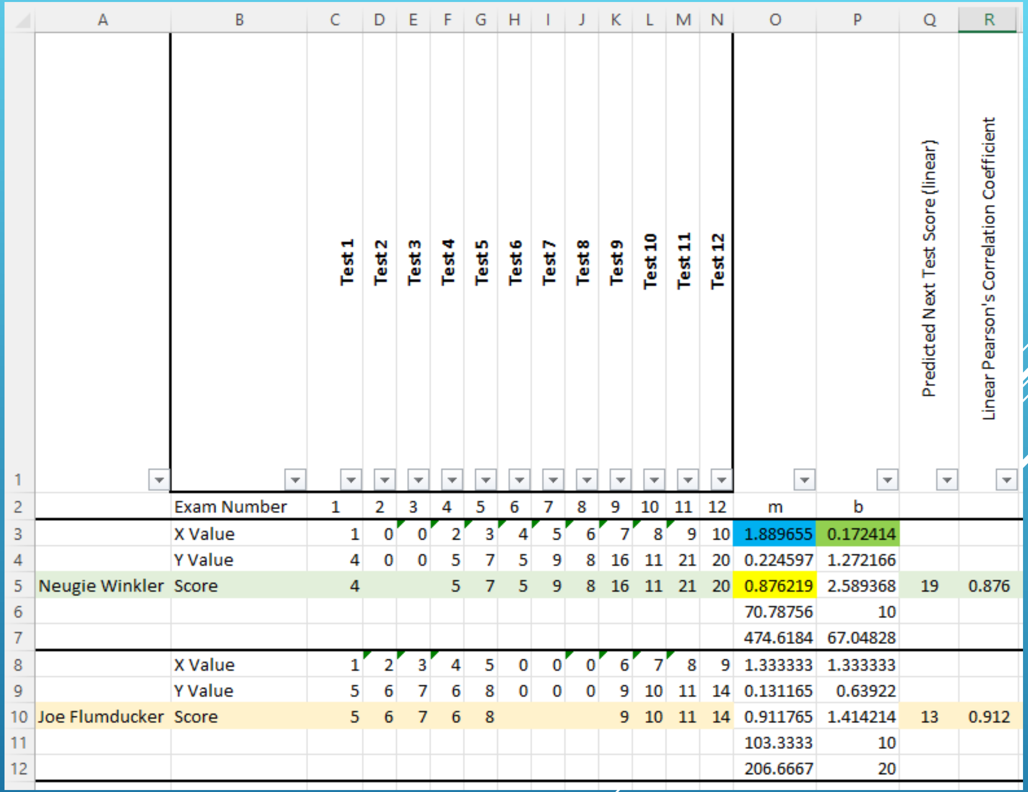


O3 =LINEST(C4:N4,C3:N3,,TRUE)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1			Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12		
2		Exam Number	1	2	3	4	5	6	7	8	9	10	11	12	x	b
3		X Value	1	0	0	2	3	4	5	6	7	8	9	10	1.889655	0.172414
4		Y Value	4	0	0	5	7	5	9	8	16	11	21	20	0.224597	1.272166
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	0.876219	2.589368
6															70.78756	10
7															474.6184	67.04828
8		X Value	1	2	3	4	5	0	0	0	6	7	8	9	1.333333	1.333333
9		Y Value	5	6	7	6	8	0	0	0	9	10	11	14	0.131165	0.63922
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	0.911765	1.414214
11															103.3333	10
12															206.6667	20
13																

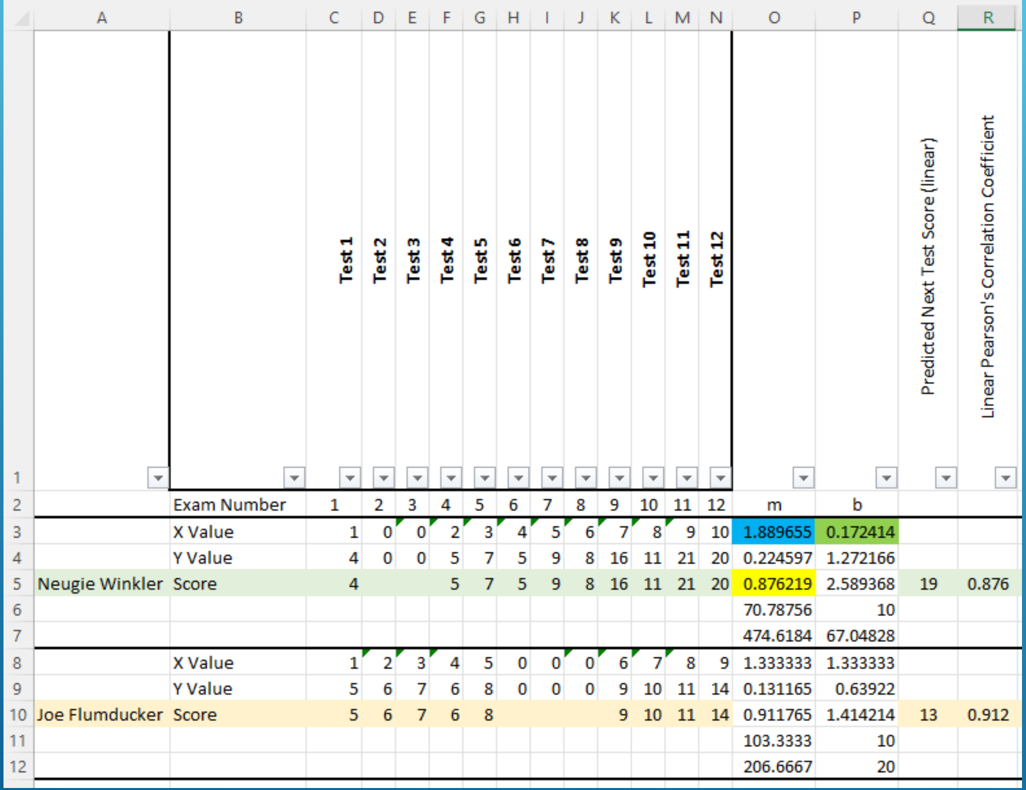
- Columns O and P are normally hidden.
- The LINEST function is in cell O3 (shown in blue)
  - Row 3 (C3:N3) represents the order in which the measurements were taken
    - Note that Test 2 and Test 3 have x and y values of zero.
  - Row 4 (C4:N4) represents the scores on the tests.

- Note the labels m and b in cells O3 (blue) and P3 (green)
- You might recall from high school algebra: y=mx+b
- Where:
  - Y is the predicted score after some number of practices x
  - X is the number of practices
  - B is the y intercept. (essentially native ability that was present before training started)
- The score on the next test (test13) can be predicted by plugging 13 in for x:
- Y=1.89(13)+0.17 = approximately19



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12			Predicted Next Test Score (linear)	Linear Pearson's Correlation Coefficient
2		Exam Number	1	2	3	4	5	6	7	8	9	10	11	12	m	b		
3		X Value	1	0	0	2	3	4	5	6	7	8	9	10	1.889655	0.172414		
4		Y Value	4	0	0	5	7	5	9	8	16	11	21	20	0.224597	1.272166		
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	0.876219	2.589368	19	0.876
6															70.78756	10		
7															474.6184	67.04828		
8		X Value	1	2	3	4	5	0	0	0	6	7	8	9	1.333333	1.333333		
9		Y Value	5	6	7	6	8	0	0	0	9	10	11	14	0.131165	0.63922		
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	0.911765	1.414214	13	0.912
11															103.3333	10		
12															206.6667	20		

PEARSON'S CORRELATION COEFFICIENT

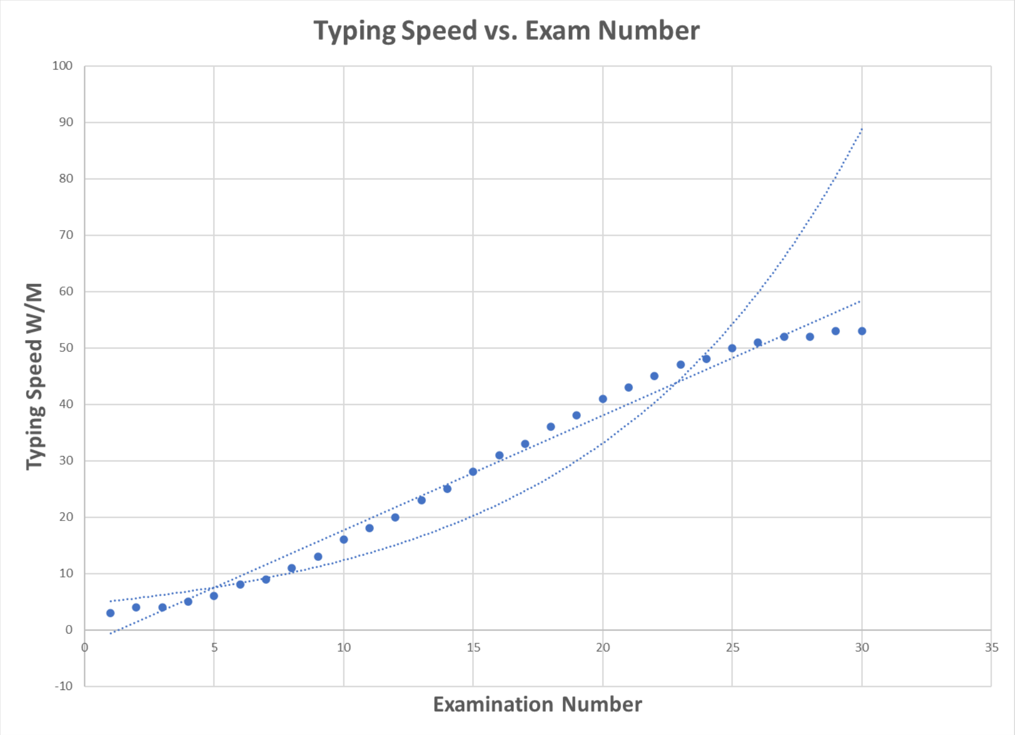


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12			Predicted Next Test Score (linear)	Linear Pearson's Correlation Coefficient
2		Exam Number	1	2	3	4	5	6	7	8	9	10	11	12	m	b		
3		X Value	1	0	0	2	3	4	5	6	7	8	9	10	1.889655	0.172414		
4		Y Value	4	0	0	5	7	5	9	8	16	11	21	20	0.224597	1.272166		
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	0.876219	2.589368	19	0.876
6															70.78756	10		
7															474.6184	67.04828		
8		X Value	1	2	3	4	5	0	0	0	6	7	8	9	1.333333	1.333333		
9		Y Value	5	6	7	6	8	0	0	0	9	10	11	14	0.131165	0.63922		
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	0.911765	1.414214	13	0.912
11															103.3333	10		
12															206.6667	20		

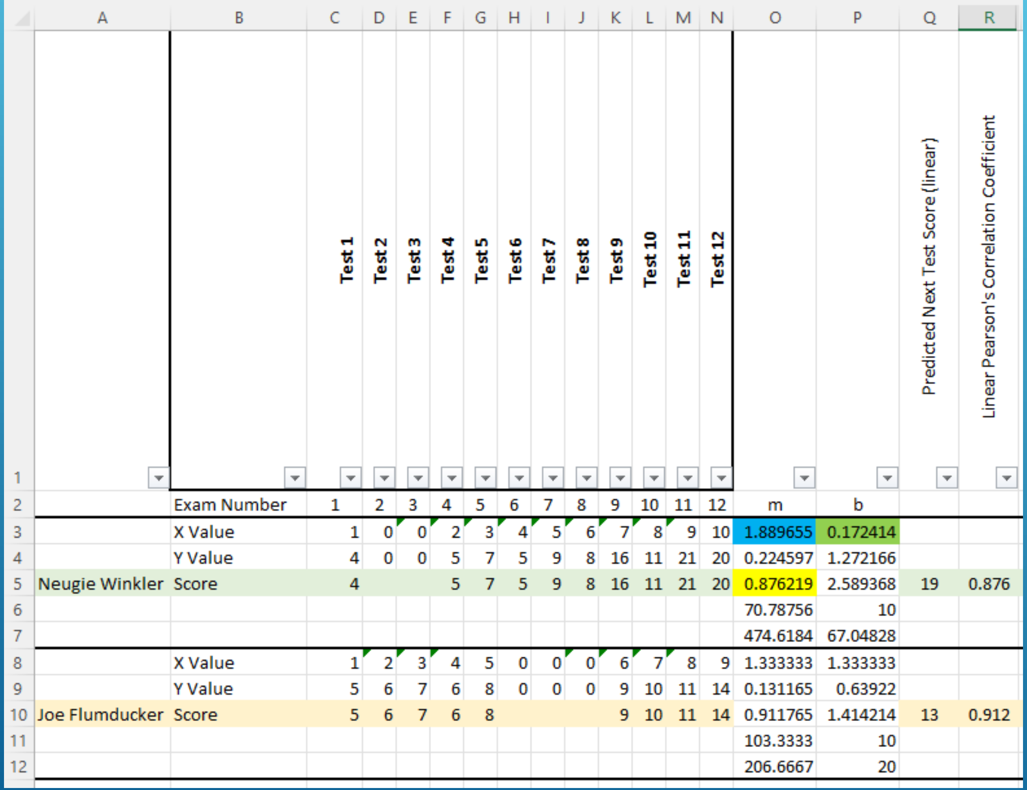
- Pearson's Correlation Coefficient is highlighted in yellow
  - It describes how well the regression correlates to the actual data
  - A value of 1 is perfect, and higher is better.
  - Pearson's correlation is an objective way to compare different regressions
- Note that Joe's progress is more linear than Neugie's, based on the higher Pearson Correlation Coefficient

A WORD OF CAUTION…

- Linear and Quadratic regressions assume that ability will continue to grow without limit
- The real learning curve is sigmoid shaped
- Linear regression makes reasonable predictions in the short run, but is not appropriate for longer term predictions



PEARSON'S CORRELATION COEFFICIENT

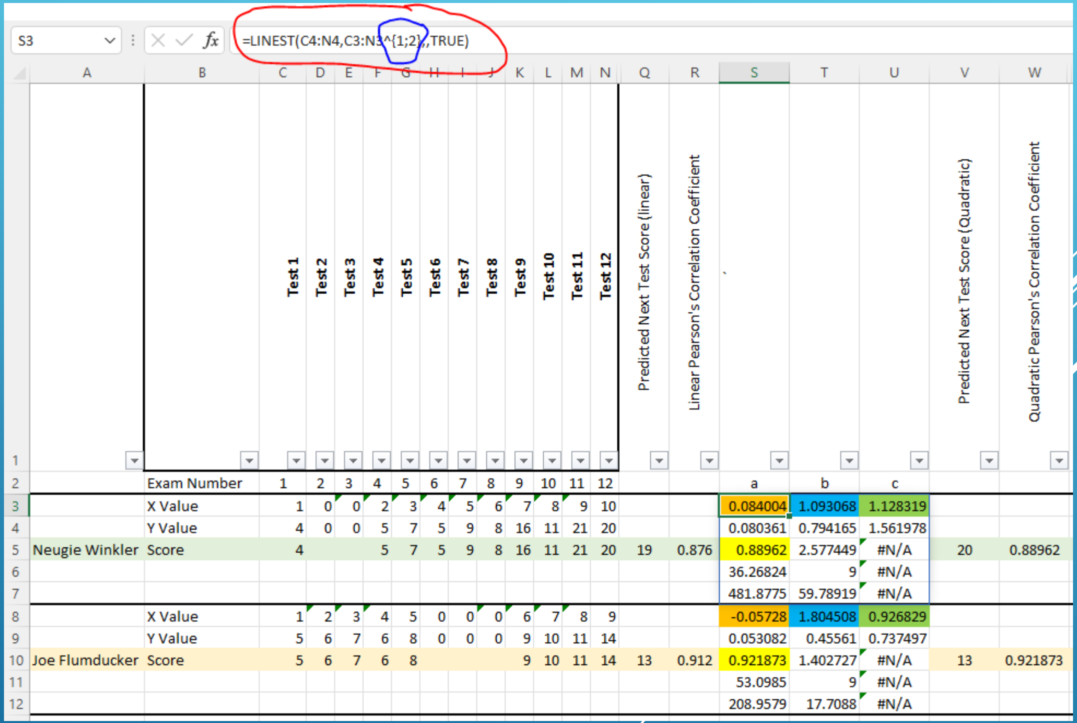


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12			Predicted Next Test Score (linear)	Linear Pearson's Correlation Coefficient
2		Exam Number	1	2	3	4	5	6	7	8	9	10	11	12	m	b		
3		X Value	1	0	0	2	3	4	5	6	7	8	9	10	1.889655	0.172414		
4		Y Value	4	0	0	5	7	5	9	8	16	11	21	20	0.224597	1.272166		
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	0.876219	2.589368	19	0.876
6															70.78756	10		
7															474.6184	67.04828		
8		X Value	1	2	3	4	5	0	0	0	6	7	8	9	1.333333	1.333333		
9		Y Value	5	6	7	6	8	0	0	0	9	10	11	14	0.131165	0.63922		
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	0.911765	1.414214	13	0.912
11															103.3333	10		
12															206.6667	20		

- Pearson's Correlation Coefficient is highlighted in yellow
  - It describes how well the regression correlates to the actual data
  - A value of 1 is perfect, and higher is better.
  - Pearson's correlation is an objective way to compare different regressions
- Note that Joe's progress is more linear than Neugie's, based on the higher Pearson Correlation Coefficient

QUADRATIC (POLYNOMIAL DEGREE 2) REGRESSION

- You may recall from high school the quadratic equation:
  - Y=ax^2+bx+c
    - Coefficient a (in orange)
    - Coefficient b (in blue)
    - Constant c (in green)
- Y=.084(13^2)+1.08(13)+1.13
  - For both Neugie and Joe, the Quadratic regression correlates better
  - This implies that their growth is still accelerating

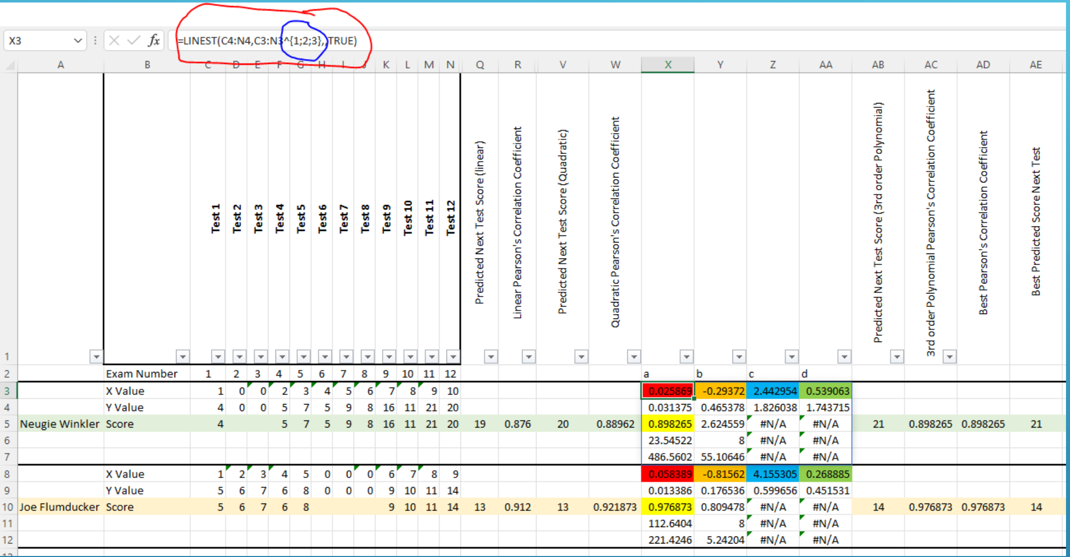


S3

=LINEST(C4:N4,C3:N3^{1;2},,TRUE)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Q	R	S	T	U	V	W
1			Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12	Predicted Next Test Score (linear)	Linear Pearson's Correlation Coefficient				Predicted Next Test Score (Quadratic)	Quadratic Pearson's Correlation Coefficient
2		Exam Number	1	2	3	4	5	6	7	8	9	10	11	12			a	b	c		
3		X Value	1	0	0	2	3	4	5	6	7	8	9	10			0.084004	1.093068	1.128319		
4		Y Value	4	0	0	5	7	5	9	8	16	11	21	20			0.080361	0.794165	1.561978		
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	19	0.876	0.88962	2.577449	#N/A	20	0.88962
6																	36.26824	9	#N/A		
7																	481.8775	59.78919	#N/A		
8		X Value	1	2	3	4	5	0	0	0	6	7	8	9			-0.05728	1.804508	0.926829		
9		Y Value	5	6	7	6	8	0	0	0	9	10	11	14			0.053082	0.45561	0.737497		
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	13	0.912	0.921873	1.402727	#N/A	13	0.921873
11																	53.0985	9	#N/A		
12																	208.9579	17.7088	#N/A		

POLYNOMIAL REGRESSION (3^{RD} ORDER)



X3

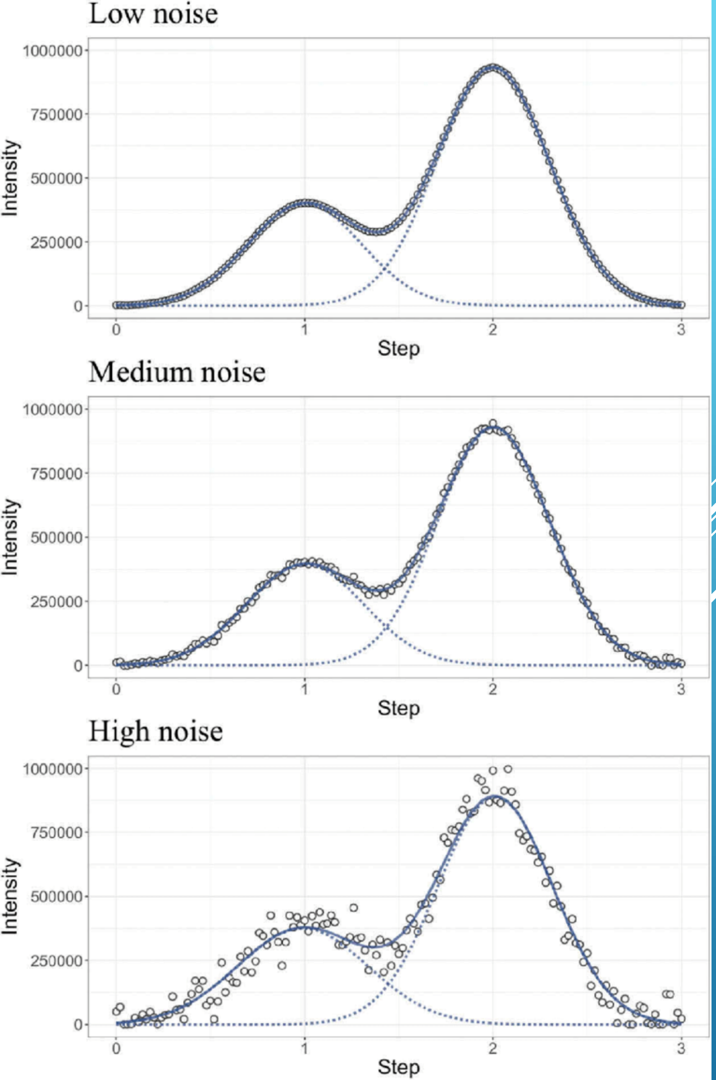
=LINEST(C4:N4,C3:N3^{1;2;3},,TRUE)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Q	R	V	W	X	Y	Z	AA	AB	AC	AD	AE
1			**Test 1**	**Test 2**	**Test 3**	**Test 4**	**Test 5**	**Test 6**	**Test 7**	**Test 8**	**Test 9**	**Test 10**	**Test 11**	**Test 12**	Predicted Next Test Score (linear)	Linear Pearson's Correlation Coefficient	Predicted Next Test Score (Quadratic)	Quadratic Pearson's Correlation Coefficient					Predicted Next Test Score (3rd order Polynomial)	3rd order Polynomial Pearson's Correlation Coefficient	Best Pearson's Correlation Coefficient	Best Predicted Score Next Test
2		Exam Number	1	2	3	4	5	6	7	8	9	10	11	12					a	b	c	d				
3		X Value	1	0	0	2	3	4	5	6	7	8	9	10					0.025869	-0.29372	2.442954	0.539063				
4		Y Value	4	0	0	5	7	5	9	8	16	11	21	20					0.031375	0.465378	1.826038	1.743715				
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	19	0.876	20	0.88962	0.898265	2.624559	#N/A	#N/A	21	0.898265	0.898265	21
6																			23.54522	8	#N/A	#N/A				
7																			486.5602	55.10646	#N/A	#N/A				
8		X Value	1	2	3	4	5	0	0	0	6	7	8	9					0.058389	-0.81562	4.155305	0.268885				
9		Y Value	5	6	7	6	8	0	0	0	9	10	11	14					0.013386	0.176536	0.599656	0.451531				
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	13	0.912	13	0.921873	0.976873	0.809478	#N/A	#N/A	14	0.976873	0.976873	14
11																			112.6404	8	#N/A	#N/A				
12																			221.4246	5.24204	#N/A	#N/A				

- You may recall from high school the quadratic equation:
  - Y=ax^3+bx^2+cx+d
    - Coefficient a (in red)
    - Coefficient b (in orange)
    - Coefficient c (in blue)
- Y=.026(13^3)-.29(13^2)+2.44(13)+.54
  - Joe's correlation is very strong ~ .97

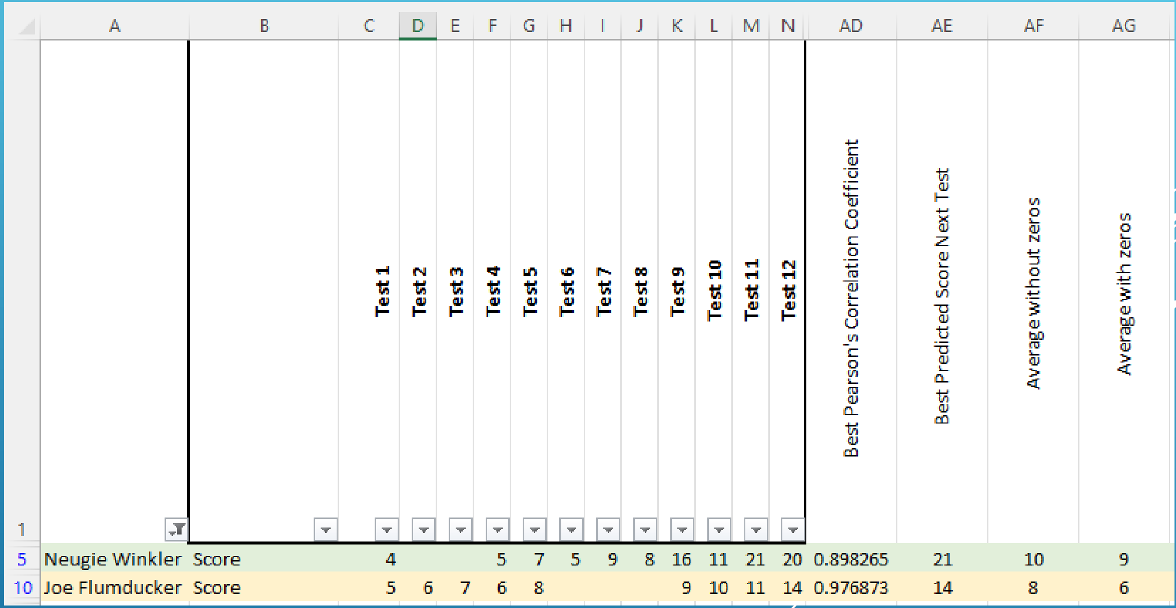
HOW FAR COULD YOU GO?

- At some point you're just fitting noise



SUMMARY

- Don’t make up data in its absence
- Consider equal increment scales like 1-5, rather than A-F
- You’re as good as your next test
- Simple regressions, like averaging, arrest noise but they also lose the signal
- The Excel LINEST function offers a defensible way to deal with incomplete data



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	AD	AE	AF	AG
1			Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Test 11	Test 12	Best Pearson's Correlation Coefficient	Best Predicted Score Next Test	Average without zeros	Average with zeros
5	Neugie Winkler	Score	4			5	7	5	9	8	16	11	21	20	0.898265	21	10	9
10	Joe Flumducker	Score	5	6	7	6	8				9	10	11	14	0.976873	14	8	6

You have been an awesome audience. Thank you!
CHANGE MY MIND
imgflip.com