Making the grade

ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

Predicting students' grades

- Term 1 grade
- Term 2 grade
- Mother's education level
- Number of absences
- Final grade

	A	B	C	D	E
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences
2	6	5	6	4	6
3	6	5	5	1	4
4	10	7	8	1	10
5	15	15	14	4	2
6	10	6	10	3	4
7	15	15	15	4	10
8	11	12	12	2	0
9	6	6	5	4	6
10	19	16	18	3	0
11	15	14	15	3	0
12	9	10	8	4	0
13	12	10	12	2	4
14	14	14	14	4	2
15	11	10	10	4	2
16	16	14	16	2	0
17	14	14	14	4	4
18	14	13	14	4	6

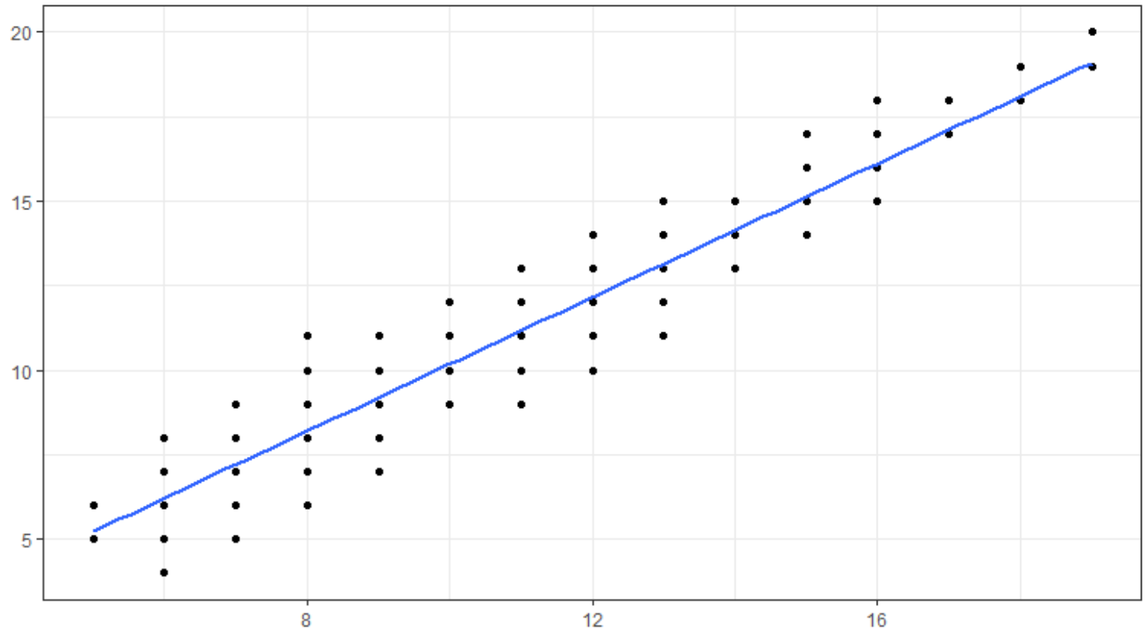
Weighted averages vs. linear models

- Weighted average
  - Interim measures
  - Adjustable weights

	A	B	C	D	E	F	G	H
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Weight: Period 1	Weight: Period 2	Weighted Average
2	6	5	6	4	6	30	70	5.7
3	6	5	5	1	4			5
4	10	7	8	1	10			7.7
5	15	15	14	4	2			14.3
6	10	6	10	3	4			8.8
7	15	15	15	4	10			15
8	11	12	12	2	0			12
9	6	6	5	4	6			5.3
10	19	16	18	3	0			17.4
11	15	14	15	3	0			14.7
12	9	10	8	4	0			8.6
13	12	10	12	2	4			11.4
14	14	14	14	4	2			14
15	11	10	10	4	2			10
16	16	14	16	2	0			15.4
17	14	14	14	4	4			14

Weighted averages vs. linear models

- Linear model
  - Lines of "best fit"



Prediction methods

- Weighted averages
  - Simple to compute
  - Good predictors for interim measures
  - Bad at generalizing
- Linear models
  - Easily generalizable
  - Reasonable predictions from novel data
  - May oversimplify relationships

Weighted averages

- `AVERAGE.WEIGHTED(values, weights)`
  - `values` : data to average
  - `weights` : proportions

	A	B	C	D	E	F	G	H
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Weight: Period 1	Weight: Period 2	Weighted Average
2	6	5	6	4	6	50		
3	6	5	5	1	4			
4	10	7	8	1	10			
5	15	15	14	4	2			
6	10	6	10	3	4			
7	15	15	15	4	10			
8	11	12	12	2	0			
9	6	6	5	4	6			
10	19	16	18	3	0			
11	15	14	15	3	0			
12	9	10	8	4	0			
13	12	10	12	2	4			
14	14	14	14	4	2			
15	11	10	10	4	2			
16	16	14	16	2	0			
17	14	14	14	4	4			

Weighted averages

- `= 100 - F2`

	A	B	C	D	E	F	G	H
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Weight: Period 1	Weight: Period 2	Weighted Average
2	6	5	6	4	6	50	50	
3	6	5	5	1	4			
4	10	7	8	1	10			
5	15	15	14	4	2			
6	10	6	10	3	4			
7	15	15	15	4	10			
8	11	12	12	2	0			
9	6	6	5	4	6			
10	19	16	18	3	0			
11	15	14	15	3	0			
12	9	10	8	4	0			
13	12	10	12	2	4			
14	14	14	14	4	2			
15	11	10	10	4	2			
16	16	14	16	2	0			
17	14	14	14	4	4			

Weighted averages

- `AVERAGE.WEIGHTED(B2:C2, $F$2:$G$2)`

	A	B	C	D	E	F	G	H	I
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences		Weight: Period 1	Weight: Period 2	Weighted Average
2	6	5	6	4	6		30	70	5.7
3	6	5	5	1	4				5
4	10	7	8	1	10				7.7
5	15	15	14	4	2				14.3
6	10	6	10	3	4				8.8
7	15	15	15	4	10				15
8	11	12	12	2	0				12
9	6	6	5	4	6				5.3
10	19	16	18	3	0				17.4
11	15	14	15	3	0				
12	9	10	8	4	0				
13	12	10	12	2	4				
14	14	14	14	4	2				
15	11	10	10	4	2				
16	16	14	16	2	0				
17	14	14	14	4	4				
18	14	13	14	4	6				
19	10	8	10	3	4				
20	5	6	5	3	16				
21	10	8	10	4	4				
22	15	13	14	4	0				
23	15	12	15	4	0				
24	16	15	15	4	2				
25	12	13	13	2	0				
26	8	10	9	2	2				
27	8	6	9	2	14				

Weighted averages

- `AVERAGE.WEIGHTED(B2:C2, F2:G2)`

	A	B	C	D	E	F	G	H
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Weight: Period 1	Weight: Period 2	Weighted Average
2	6	5	6	4	6	50	50	5.5
3	6	5	5	1	4			5
4	10	7	8	1	10			7.5
5	15	15	14	4	2			14.5
6	10	6	10	3	4			8
7	15	15	15	4	10			15
8	11	12	12	2	0			12
9	6	6	5	4	6			5.5
10	19	16	18	3	0			17
11	15	14	15	3	0			14.5
12	9	10	8	4	0			9
13	12	10	12	2	4			11
14	14	14	14	4	2			14
15	11	10	10	4	2			10
16	16	14	16	2	0			15
17	14	14	14	4	4			14

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

Advanced prediction strategies

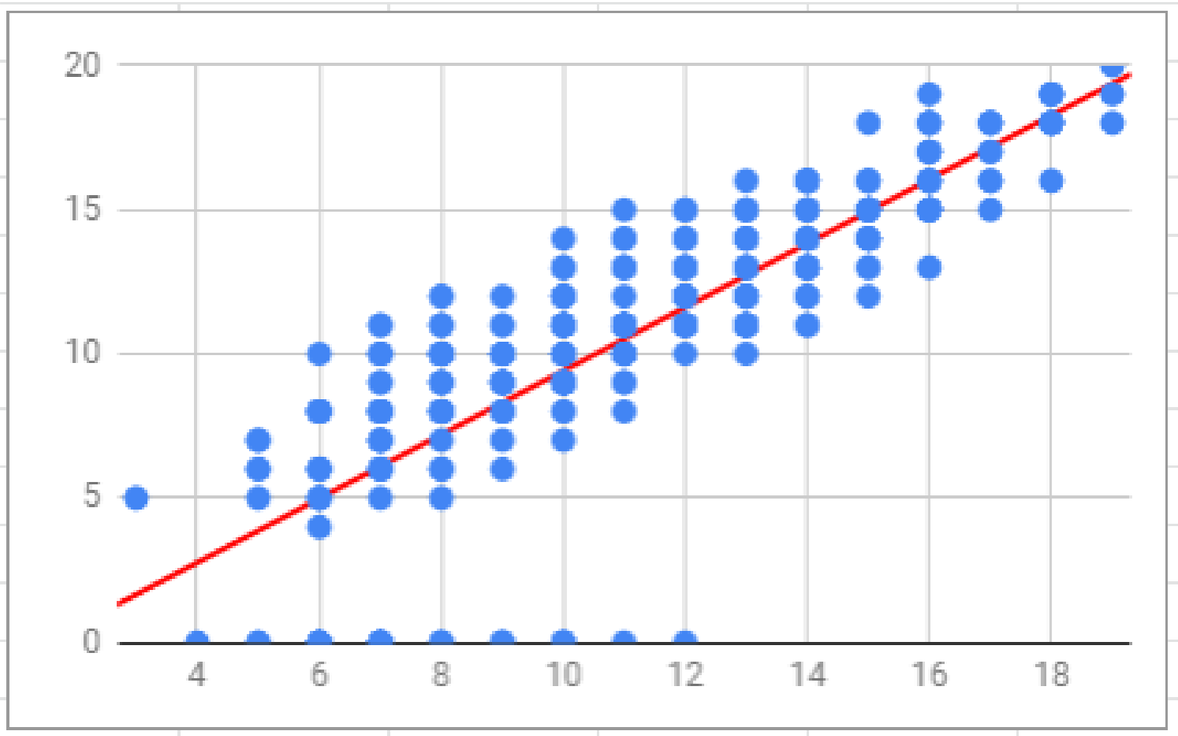
ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

datacamp

Linear models

- Predict unobserved values
- Predict using other measures
- Independent variable
  - Causes some change
- Dependent variable
  - Outcomes



Forecasting

- `FORECAST()` function
  - `x` : value to use to predict
  - `data_y` : actual outcomes
  - `data_x` : predictors

	A	B	C	D	E	F
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Predicted Grade: P1
2	6	5	6	4	6	5.21991593
3	6	5	5	1	4	
4	10	7	8	1	10	
5	15	15	14	4	2	
6	10	6	10	3	4	
7	15	15	15	4	10	
8	11	12	12	2	0	
9	6	6	5	4	6	
10	19	16	18	3	0	
11	15	14	15	3	0	
12	9	10	8	4	0	
13	12	10	12	2	4	
14	14	14	14	4	2	
15	11	10	10	4	2	
16	16	14	16	2	0	
17	14	14	14	4	4	
18	14	13	14	4	6	

Forecasting

- `FORECAST(C2, data_y, data_x)`

	A	B	C	D	E	F
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Predicted Grade: P1
2	6	5	6	4	6	5.21991593
3	6	5	5	1	4	
4	10	7	8	1	10	
5	15	15	14	4	2	
6	10	6	10	3	4	
7	15	15	15	4	10	
8	11	12	12	2	0	
9	6	6	5	4	6	
10	19	16	18	3	0	
11	15	14	15	3	0	
12	9	10	8	4	0	
13	12	10	12	2	4	
14	14	14	14	4	2	
15	11	10	10	4	2	
16	16	14	16	2	0	
17	14	14	14	4	4	
18	14	13	14	4	6	

Forecasting

- `FORECAST(C2, A2:A, data_x)`

	A	B	C	D	E	F
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Predicted Grade: P1
2	6	5	6	4	6	5.21991593
3	6	5	5	1	4	
4	10	7	8	1	10	
5	15	15	14	4	2	
6	10	6	10	3	4	
7	15	15	15	4	10	
8	11	12	12	2	0	
9	6	6	5	4	6	
10	19	16	18	3	0	
11	15	14	15	3	0	
12	9	10	8	4	0	
13	12	10	12	2	4	
14	14	14	14	4	2	
15	11	10	10	4	2	
16	16	14	16	2	0	
17	14	14	14	4	4	
18	14	13	14	4	6	

Forecasting

- `FORECAST(C2, A2:A, C2:C)`

	A	B	C	D	E	F
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Predicted Grade: P1
2	6	5	6	4	6	5.21991593
3	6	5	5	1	4	
4	10	7	8	1	10	
5	15	15	14	4	2	
6	10	6	10	3	4	
7	15	15	15	4	10	
8	11	12	12	2	0	
9	6	6	5	4	6	
10	19	16	18	3	0	
11	15	14	15	3	0	
12	9	10	8	4	0	
13	12	10	12	2	4	
14	14	14	14	4	2	
15	11	10	10	4	2	
16	16	14	16	2	0	
17	14	14	14	4	4	
18	14	13	14	4	6	

Forecasting

- `FORECAST(C2, A2:A, C2:C)`

	A	B	C	D	E	F
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences	Predicted Grade: P1
2	6	5	6	4	6	3.8784766
3	6	5	5	1	4	
4	10	7	8	1	10	
5	15	15	14	4	2	
6	10	6	10	3	4	
7	15	15	15	4	10	
8	11	12	12	2	0	
9	6	6	5	4	6	
10	19	16	18	3	0	
11	15	14	15	3	0	
12	9	10	8	4	0	
13	12	10	12	2	4	
14	14	14	14	4	2	
15	11	10	10	4	2	
16	16	14	16	2	0	

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

How clear is your crystal ball?

ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

datacamp

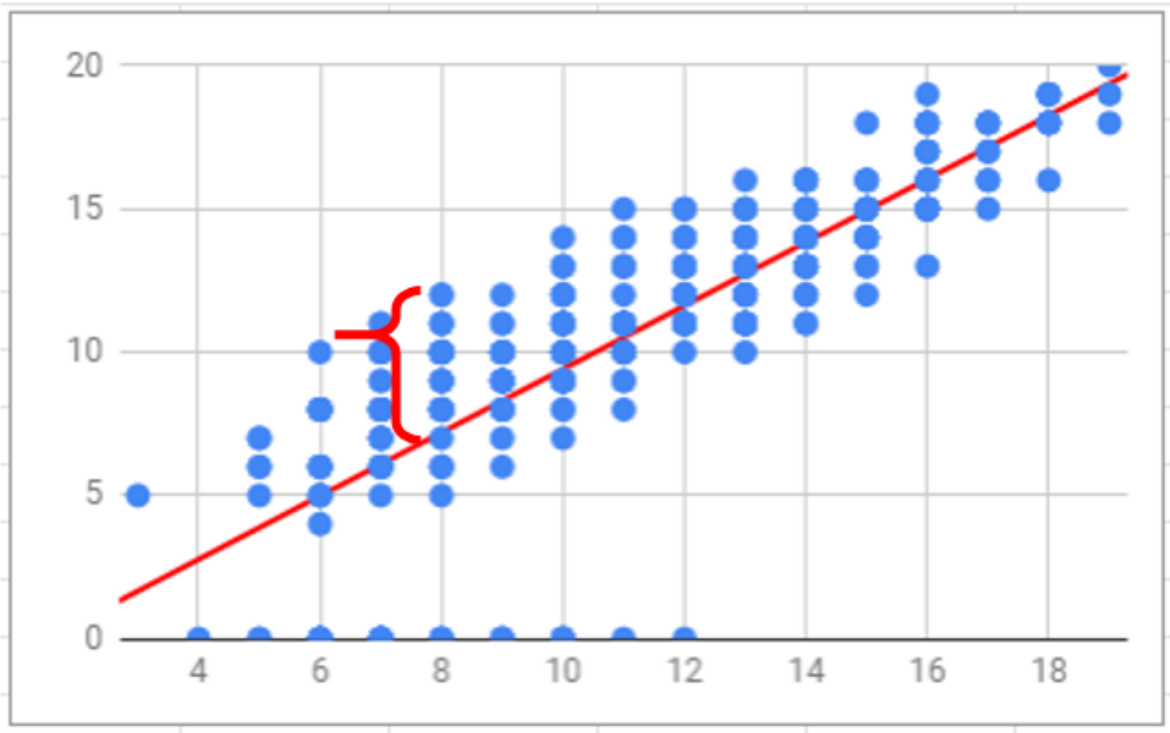
Predicted vs. actual

- Compare forecasts to observations
  - `prediction - observation`

	A	B	C	D	E	F	G	H	I	J	K
1	Final Grade	Period 1 Grade	Period 2 Grade	Mother's Education Level	Absences		Predicted Grade: Weighted Average	Predicted Grade: P1 Forecast	Predicted Grade: P2 Forecast		Absolute Deviation: Weighted Average
2	6	5	6	4	6		5.8	3.87847664	5.21991593		0.2
3	6	5	5	1	4		5	3.87847664	4.11780357		1
4	10	7	8	1	10		7.8	6.09098883	7.42414064		2.2
5	15	15	14	4	2		14.2	14.9410375	14.0368147		0.8
6	10	6	10	3	4		9.2	4.98473274	9.6283653		0.8
7	15	15	15	4	10		15	14.9410375	15.1389271		0
8	11	12	12	2	0		12	11.6222693	11.8325900		1
9	6	6	5	4	6		5.2	4.98473274	4.11780357		0.8
10	19	16	18	3	0		17.6	16.0472936	18.4452642		1.4
11	15	14	15	3	0		14.8	13.8347815	15.1389271		
12	9	10	8	4	0		8.4	9.40975711	7.42414064		
13	12	10	12	2	4		11.6	9.40975711	11.8325900		
14	14	14	14	4	2		14	13.8347815	14.0368147		
15	11	10	10	4	2		10	9.40975711	9.6283653		
16	16	14	16	2	0		15.6	13.8347815	16.2410394		
17	14	14	14	4	4		14	13.8347815	14.0368147		
18	14	13	14	4	6		13.8	12.7285254	14.0368147		

Absolute deviation

- `ABS()` function
  - `=ABS(prediction - observation)`



Mean absolute deviation

- `AVERAGE()` of absolute deviations
- Useful for comparing predictions
  - Lower deviation, better predictions

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

Statistical significance

ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

Statistical significance

- <5% probability due to chance
  - Not 0%
- `T.TEST()`
  - Likelihood that differences are significant

T-tests

- `T.TEST()`
  - `range1`
  - `range2`
  - `tails`
  - `type`

T-tests

- `tails`
  - `1` : Upper **or** lower distribution
  - `2` : Upper **and** lower distribution
- `type`
  - `1` : Paired-samples
  - `2` : Two-sample (equal variance)
  - `3` : Two-sample (unequal variance)

Confidence levels

- Significance
  - Confidence level
    - Probability due to chance
- `T.TEST()` <= 0.05

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS