Where are we in our discussion of error analysis?

Let's revisit:

CM3215

MichiganTech

Fundamentals of Chemical Engineering Laboratory

Statistics Lecture 4:

Error Propagation

Professor Faith Morrison

Department of Chemical Engineering
Michigan Technological University

1. Quick start—Replicate error
2. Reading Error
3. Calibration Error
4. **Error Propagation**

1

From Lecture 4—Error Propagation:

Summary: Error Analysis with Real Numbers

- To understand the accuracy of our numbers, we need to determine a ***confidence interval***.

$\bar{x} \pm 2e_s$ with 95.0% confidence

For replicate data with $n < 7$, replace "2" with $t_{0.025,n-1}$

- The **Standard error e_s** for a measured quantity is the largest of:
  - e_s determined by *replicates* $e_s = s/\sqrt{n}$ or
  - e_s by estimate of *reading error* $e_s = e_R/\sqrt{3}$ or
  - e_s by estimate of *calibration error* $e_s = \text{max error}/2$

- Standard error e_f for derived quantities (arrived at from equations), is obtained at through **error propagation,** which is a combination of *variances*.

2

From Lecture 4—Error Propagation:

Error Propagation

We use an analysis based on the Taylor series expansion of a nonlinear function.

Taylor series:

(higher order terms)

$$f(x_1, x_2, x_3) = f^0 + \frac{\partial f}{\partial x_1} x_1 + \frac{\partial f}{\partial x_2} x_2 + \frac{\partial f}{\partial x_3} x_3 + h.o.t.$$

A calculation of the function $f(x_1, x_2, x_3)$ from uncertain values of x_1, x_2, x_3 is a random variable of mean $\bar{f}$ and variance σ_f^2:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 \sigma_{x_3}^2 + \text{Covariance terms, if } x_i \text{ are correlated}$$

3

From Lecture 4—Error Propagation:

Error Propagation

We use an analysis based on the Taylor series expansion of a nonlinear function.

Taylor series:

(higher order terms)

$$f(x_1, x_2, x_3) = f^0 + \frac{\partial f}{\partial x_1} x_1 + \frac{\partial f}{\partial x_2} x_2 + \frac{\partial f}{\partial x_3} x_3 + h.o.t.$$

A calculation of the function $f(x_1, x_2, x_3)$ from uncertain values of x_1, x_2, x_3 is a random variable of mean $\bar{f}$ and variance σ_f^2:

neglect

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 \sigma_{x_3}^2 + \text{Covariance terms, if } x_i \text{ are correlated}$$

4

Note: covariance terms are not always zero or small; but they often are. For now, this is fine.

From Lecture 4—Error Propagation:

Michigan Technological University
Department of Chemical Engineering

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

This worksheet guides the user through the determination of the standard error e_{sf} of a quantity $f(x_1, x_2, x_3, x_4, x_5)$ that is calculated from measured quantities x_1, x_2, x_3, x_4 and x_5. The x_i are subject to random errors. The replicate standard deviations σ_i or the reading errors e_{R_i} for each variable x_i must be determined first, and these uncertainties are propagated to determine e_{sf} using the relationship given below. Note: if standard error e_{x_i} estimates via both replicates and reading error are available, use the larger of the two.

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f:	Representative value of f: (include units)	95% C.I. of f: $(f \pm 2e_{sf})$ (include units)

Measured quantities, x_i			$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{\sigma_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_s	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_i	Symbol	Representative value			
x_1					
x_2					
x_3					
x_4					
x_5					

$$e_{sf}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{sf}^2 =$

$e_{sf} =$ units

Standard error of calculated quantity f

Note: For some quantities, you will look up the uncertainty; for example the volume of a volumetric flask may be given as $100.00 \pm 0.04ml$. In these circumstances it is reasonable to assume that the reported uncertainty is $\pm 1.96e_s$. For example, if volume is given as $100.00 \pm 0.04ml$, then $1.96e_s \approx 2e_s = 0.04$. Reference: page 564 of Fritz and Schenk, *Quantitative Analytical Chemistry*, Allyn and Bacon, Boston, 1987.

Handy worksheet for error propagation

www.chem.mtu.edu/~fmorriso/cm3215/ErrorPropagationWorksheet.pdf

5

From Lecture 4—Error Propagation:

Example 1: What is the uncertainty (95% confidence interval) in $\rho_{bluefluid}$ as determined in the lab?

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f: $\rho_{BF} = \frac{M_F - M_E}{V_{pyc}}$	Representative value of f: (include units) $1.739\ g/ml$	95% C.I. of f: $(f \pm 2e_{sf})$ (include units) $1.739 \pm 0.007\ g/ml$

Measured quantities, x_i			$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_i	Symbol	Representative value			
x_1	M_F	$30.800\ g$	$1/V_{pyc}$	$5.8 \times 10^{-5} g$	$3.3 \times 10^{-11}\ g^2/ml^2$
x_2	M_E	$13.410\ g$	$-1/V_{pyc}$	$5.8 \times 10^{-5} g$	$3.3 \times 10^{-11}\ g^2/ml^2$
x_3	V_{pyc}	$10.00\ ml$	$-(M_F - M_E)/V_{pyc}^2$	$0.02\ ml$	$1.21 \times 10^{-5}\ g^2/ml^2$
x_4					
x_5					

Review

$$e_{sf}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{sf}^2 = 1.21 \times 10^{-5}\ g^2/ml^2$

$e_{sf} = 0.0035$ units g/ml

Standard error of calculated quantity, f

6

From Lecture 4—Error Propagation:

Example 1: What is the uncertainty (95% confidence interval) in $\rho_{bluefluid}$ as determined in the lab?

Excel is an excellent tool for error propagation

Review

Error propagation Worksheet

$f(x_1,x_2,x_3)$			f	ρ_{BF}	1.739	g/ml	$2e_s$	0.007	g/ml
	x_i	value		df/dx_i	$(df/dx_i)^2$	e_{xi}	e_{xi}^2	$(df/dx_i)^2 e_{xi}^2$	
x_1	M_F	30.800	g	0.10	0.010	5.8E-05	3.3E-09	3.33E-11	g^2/ml^2
x_2	M_E	13.410	g	-0.10	0.010	5.8E-05	3.3E-09	3.33E-11	g^2/ml^2
x_3	V_{pyc}	10.000	ml	-0.174	0.0302	0.02	4.0E-04	1.210E-05	g^2/ml^2
							e_s^2	1.21E-05	g^2/ml^2
							e_s	0.0035	g/ml

7

From Lecture 4—Error Propagation:

Review

CM3215 MichiganTech

Fundamentals of Chemical Engineering Laboratory

Error Analysis for Laboratory Data

1. Quick start—Replicate error
2. Reading Error
3. Calibration Error
4. Error Propagation

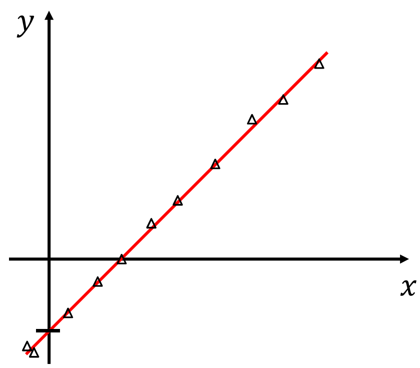
Summary: Error Analysis with Real Numbers

- To understand the accuracy of our numbers, we need to determine a ***confidence interval.***
  $\bar{x} \pm 2e_s$ with 95.0% confidence
  For replicate data with $n < 7$, replace "2" with $t_{0.025,n-1}$
- The **Standard error e_s** for a measured quantity is the largest of:
  e_s determined by *replicates* $e_s = s/\sqrt{n}$ or
  e_s by estimate of *reading error* $e_s = e_R/\sqrt{3}$ or
  e_s by estimate of *calibration error* $e_s = \text{max error}/2$
- Standard error e_f for derived quantities (arrived at from equations), is obtained through **error propagation,** which is a combination of *variances*.
- Replication always improves the **estimation of the mean.** The answer from replicates is more reliable than single values.
- The **prediction interval of the next value of x** should encompass 95% of all measured values. 95% PI: $\bar{x} \pm 2s$ or $\bar{x} \pm t_{0.025,n-1}s$ if $n < 7$
- The weighting values $\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$ indicate the **impact** of individual errors on the final value.
- **Estimates** for e_s (particularly those obtained through e_R) may need to be re-evaluated, if unreasonably narrow confidence intervals are identified.

8

Now, how do we determine ***uncertainty*** from numbers that we obtain as parameters in a curve-fit?

i	x_i	y_i
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n



$$y = mx + b$$

9

CM3215

Fundamentals of Chemical Engineering Laboratory

Uncertainty in Least Squares Curve Fitting: Excel's LINEST

Professor Faith Morrison

Department of Chemical Engineering
Michigan Technological University

1. Quick start—Replicate error
2. Reading Error
3. Calibration Error
4. Error Propagation
5. **Least Squares Curve Fitting**

Reference:
- www.chem.mtu.edu/~fmorriso/cm3215/UncertaintySlopeInterceptOfLeastSquaresFit.pdf

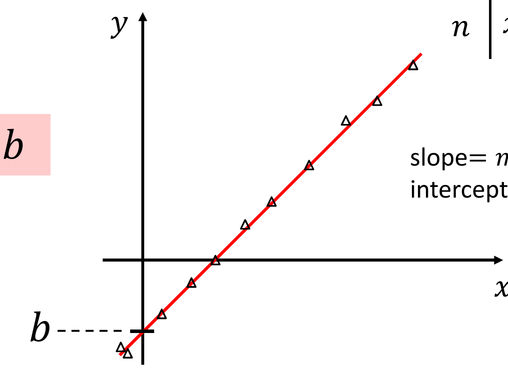
10

Ordinary, *Least Squares*, Linear Regression

Question: For a dataset of n data pairs x_i, y_i that is expected to show a linear relationship between y and x, what are the parameters m and b of the equation for the line?

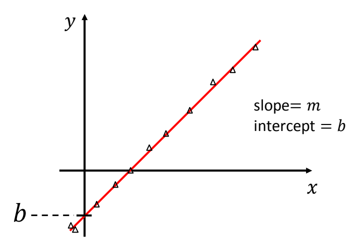
i	x_i	y_i
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

$$y = mx + b$$



11

Ordinary, *Least Squares*, Linear Regression



Solution:

- Assume you know the x_i with certainty (*"ordinary"* least squares)
- Guess a line, $\hat{y} = mx + b$
- Create a measure of the error between the *guess* and the *data* **(error measure should always be positive, so square it)**
- Add these individual error measures to calculate a *sum of squared errors*, SS_E
- Use *calculus* to find the values of m and b that result in the **least** sum of squared error.

i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	$\hat{y}_n$

$$\hat{y}_i = mx_i + b$$

$$SS_E \equiv \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

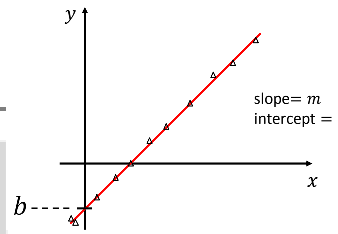
12

Ordinary, *Least Squares*, Linear Regression

Result:

$$\hat{m} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{b} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	$\hat{y}_n$

$$y = \hat{m}x + \hat{b}$$

In Excel:
$\hat{m}$ = SLOPE(y-range, x-range)
$\hat{b}$ =INTERCEPT(y-range,x-range)

Least squares slope $= \hat{m}$
Least squares intercept $= \hat{b}$

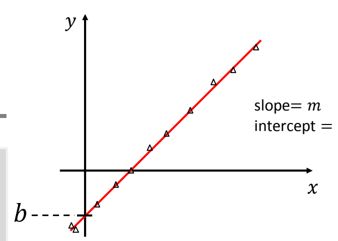
13

Ordinary, *Least Squares*, Linear Regression

Result:

$$\hat{m} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{b} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$



But, what are the error limits on $\hat{m}$ and $\hat{b}$?

$$y = \hat{m}x + \hat{b}$$

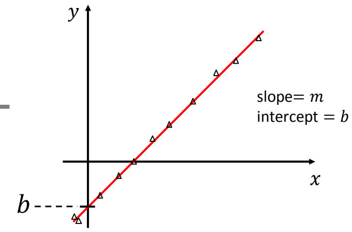
Least squares slope $= \hat{m}$
Least squares intercept $= \hat{b}$

14

Ordinary, *Least Squares*, Linear Regression

$$y = \hat{m}x + \hat{b}$$

$$\text{slope} = \hat{m} \pm ?$$
$$\text{Intercept} = \hat{b} \pm ?$$



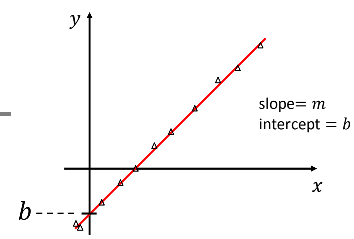
But, what are the error limits on $\hat{m}$ and $\hat{b}$?

15

Ordinary, *Least Squares*, Linear Regression

$$y = \hat{m}x + \hat{b}$$

$$\text{slope} = \hat{m} \pm ?$$
$$\text{Intercept} = \hat{b} \pm ?$$



But, what are the error limits on $\hat{m}$ and $\hat{b}$?

Answer:

$$\text{slope} = \hat{m} \ \pm 2e_s$$
$$\text{Intercept} = \hat{b} \ \pm 2e_s$$

But what is e_s?

16

(Later we will correct the "2" for small n)

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Michigan Tech
Michigan Technological University
Department of Chemical Engineering

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

This worksheet guides the user through the determination of the standard error e_{sf} of a quantity $f(x_1, x_2, x_3, x_4, x_5)$ that is calculated from measured quantities x_1, x_2, x_3, x_4 and x_5. The x_i are subject to random errors. The replicate standard deviations σ_i or the reading errors e_{R_i} for each variable x_i must be determined first, and these uncertainties are propagated to determine e_{sf} using the relationship given below. Note: If standard error e_{x_i} estimates via both replicates and reading error are available, use the larger of the two.

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f:	Representative value of f: (include units)	95% C.I. of f: $(f \pm 2e_{sf})$ (include units)

Measured quantities, x_i			$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{\sigma_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_s	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_i	Symbol	Representative value			
x_1					
x_2					
x_3					
x_4					
x_5					

$$e_{sf}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{sf}^2 =$

$e_{sf} =$ units

Standard error of calculated quantity f

Note: For some quantities, you will look up the uncertainty; for example the volume of a volumetric flask may be given as $100.00 \pm 0.04ml$. In these circumstances it is reasonable to assume that the reported uncertainty is $\pm 1.96 e_s$. For example, if volume is given as $100.00 \pm 0.04ml$, then $1.96e_s \approx 2e_s = 0.04$. Reference: page 564 of Fritz and Schenk, *Quantitative Analytical Chemistry*, Allyn and Bacon, Boston, 1987.

17

Error limits on $\hat{m}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f: $\hat{m} = \frac{n\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$	Representative value of f: (include units) $\hat{m} =$	95% C.I. of f: $(f \pm 2e_{sf})$ (include units) $\hat{m} = \pm(2)(s_m)$

Measured quantities, x_i			$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_i	Symbol	Representative value			
x_1	y_1		$\frac{\partial \hat{m}}{\partial y_1}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_1}\right)^2 s_{y,x}^2$
x_2	y_2		$\frac{\partial \hat{m}}{\partial y_2}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_2}\right)^2 s_{y,x}^2$
x_3	y_3		$\frac{\partial \hat{m}}{\partial y_3}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_3}\right)^2 s_{y,x}^2$
x_4	y_4		$\frac{\partial \hat{m}}{\partial y_4}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_4}\right)^2 s_{y,x}^2$
x_5	$\vdots$		$\vdots$	$\vdots$	$\vdots$

$$e_{sf}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{sf}^2 = s_m^2$

$e_{sf} = s_m$ units

Error limits on $\hat{m}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f:	Representative value of f: (include units)	95% C.I. of f: $(f \pm 2e_{sf})$ (include units)
	$\hat{m} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$	$\hat{m} =$	$\hat{m} = \pm(2)(s_m)$

Only the y_i are variables; we assumed we knew the x_i with certainty

x_i	Symbol		$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_1	y_1		$\frac{\partial \hat{m}}{\partial y_1}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_1}\right)^2 s_{y,x}^2$
x_2	y_2		$\frac{\partial \hat{m}}{\partial y_2}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_2}\right)^2 s_{y,x}^2$
x_3	y_3		$\frac{\partial \hat{m}}{\partial y_3}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_3}\right)^2 s_{y,x}^2$
x_4	y_4		$\frac{\partial \hat{m}}{\partial y_4}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_4}\right)^2 s_{y,x}^2$
x_5	$\vdots$		$\vdots$	$\vdots$	$\vdots$

$$e_{s_f}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{s_f}^2 = s_m^2$

$e_{s_f} = s_m$ units

Error limits on $\hat{m}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

Assume that the variances of the y_i are the same for all y_i

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f:	Re… (in…	95% C.I. of f: $(f \pm 2e_{sf})$ (include units)
	$\hat{m} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$		$\hat{m} = \pm(2)(s_m)$

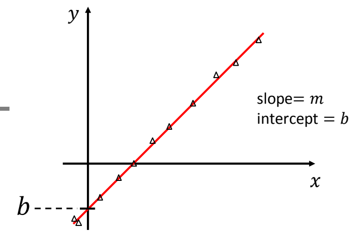
Measured quantities, x_i					
x_i	Symbol	Representative value	$\frac{\partial f}{\partial x_i}$	$\frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_1	y_1		$\frac{\partial \hat{m}}{\partial y_1}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_1}\right)^2 s_{y,x}^2$
x_2	y_2		$\frac{\partial \hat{m}}{\partial y_2}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_2}\right)^2 s_{y,x}^2$
x_3	y_3		$\frac{\partial \hat{m}}{\partial y_3}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_3}\right)^2 s_{y,x}^2$
x_4	y_4		$\frac{\partial \hat{m}}{\partial y_4}$	$s_{y,x}$	$\left(\frac{\partial \hat{m}}{\partial y_4}\right)^2 s_{y,x}^2$
x_5	$\vdots$		$\vdots$	$\vdots$	$\vdots$

$$e_{s_f}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{s_f}^2 = s_m^2$

$e_{s_f} = s_m$ units

Ordinary, *Least Squares*, Linear Regression



$s_{y,x}^2$

The variance of y, given x

$$s_{y,x}^2 \equiv \left(\frac{1}{n-2}\right)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

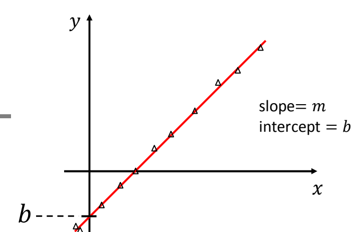
(from the *definition* of variance)

The variance of the mean value of y at a given x

In Excel:
$s_{y,x}$ = STEYX(y-range, x-range)

21

Ordinary, *Least Squares*, Linear Regression



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

What are the error limits on $\hat{m}$?

Answer:

$$\text{slope} = \hat{m} \pm 2s_m$$

$$s_m^2 = \frac{s_{y,x}^2}{SS_{xx}}$$

(final result of the algebra indicated on previous slide)

for $n-2 \leq 6$:
slope $= \hat{m} \pm t_{0.025,n-2}s_m$

In Excel:

- $s_m^2 = \dfrac{(\text{STEYX(y–range, x–range)})^2}{(\text{DEVSQ(x–range)})}$, or
- use LINEST

22

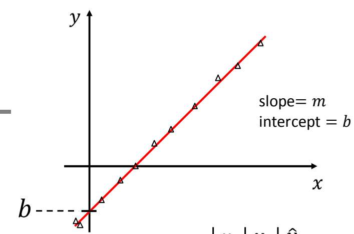
Ordinary, *Least Squares*, Linear Regression

What are the error limits on $\hat{b}$?

Answer:

$$\text{intercept} = \hat{b} \ \pm 2e_s$$

?



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

23

Error limits on $\hat{b}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f: $\hat{b} = \frac{(\sum_{i=1}^n x_i)^2(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i y_i)(\sum_{i=1}^n x_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$	Representative value of f: (include units) $\hat{b} =$	95% C.I. of f: $(f \pm 2e_{sf})$ (include units) $\hat{b} = \pm(2)(s_b)$

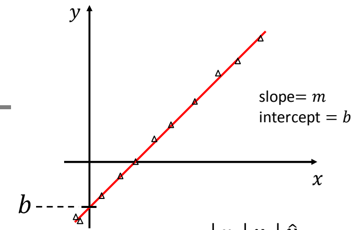
Measured quantities, x_i			$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_i	Symbol	Representative value			
x_1	y_1		$\frac{\partial \hat{b}}{\partial y_1}$	$s_{y,x}$	$\left(\frac{\partial \hat{b}}{\partial y_1}\right)^2 s_{y,x}^2$
x_2	y_2		$\frac{\partial \hat{b}}{\partial y_2}$	$s_{y,x}$	$\left(\frac{\partial \hat{b}}{\partial y_2}\right)^2 s_{y,x}^2$
x_3	y_3		$\frac{\partial \hat{b}}{\partial y_3}$	$s_{y,x}$	$\left(\frac{\partial \hat{b}}{\partial y_3}\right)^2 s_{y,x}^2$
x_4	y_4		$\frac{\partial \hat{b}}{\partial y_4}$	$s_{y,x}$	$\left(\frac{\partial \hat{b}}{\partial y_4}\right)^2 s_{y,x}^2$
x_5	⋮		⋮	⋮	⋮

$$e_{s_f}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{s_f}^2 = s_b^2$

$e_{s_f} = s_b$ units

Ordinary, *Least Squares*, Linear Regression



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

What are the error limits on $\hat{b}$?

Answer:

$$\text{intercept} = \hat{b} \pm 2s_b$$

$$s_b^2 = s_{y,x}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SS_{xx}}\right)$$

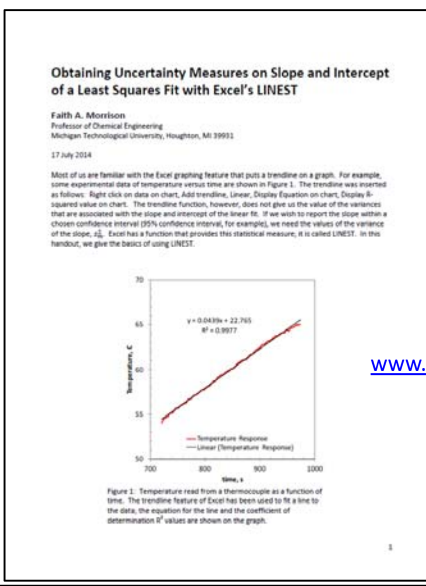
(final result of the algebra indicated on previous slide)

for $n - 2 \leq 6$:
$\text{intercept} = \hat{b} \pm t_{0.025,n-2}s_b$

In Excel:

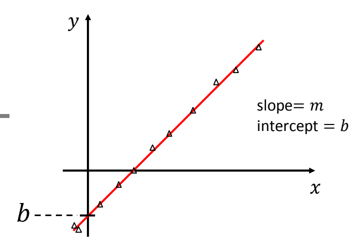
- Calculate s_b^2 from STEYX(y–range, x–range) and DEVSQ(x–range) and the formula above, or
- use LINEST

Ordinary, *Least Squares*, Linear Regression


Obtaining Uncertainty Measures on Slope and Intercept of a Least Squares Fit with Excel's LINEST

Faith A. Morrison
Professor of Chemical Engineering
Michigan Technological University, Houghton, MI 39931

17 July 2014

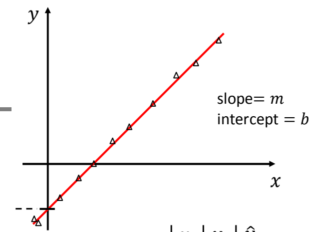


For instructions on how to use Microsoft Excel's LINEST function, see the handout on the web:

www.chem.mtu.edu/~fmorriso/cm3215/UncertaintySlopeInterceptOfLeastSquaresFit.pdf

(the appendix has some derivations, if you're interested)

Ordinary, *Least Squares*, Linear Regression



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

What are the error limits on a value of y obtained from the equation $y = \hat{m}x + \hat{b}$?

at x_p,

$$y_p = \left(\hat{m}x_p + \hat{b}\right) \pm 2e_s$$

?

27

Error limits on $y = \hat{m}x + \hat{b}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f:	Representative value of f: (include units)	95% C.I. of f: $(f \pm 2e_{sf})$ (include units)
	$y_p = \hat{m}x_p + \hat{b}$	$y_p =$	$y_p = \pm(2)(s_{y_p})$

Measured quantities, x_i			$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_i	Symbol	Representative value			
x_1	$\hat{m}$		$\frac{\partial y_p}{\partial \hat{m}} = x_p$	s_m	$(x_p s_m)^2$
x_2	x_p		$\frac{\partial y_p}{\partial x_p} = \hat{m}$	0	0
x_3	$\hat{b}$		$\frac{\partial y_p}{\partial \hat{b}} = 1$	s_b	s_b^2
x_4					
x_5					

$$e_{s_f}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{s_f}^2 = s_{y_p}^2$

units

$e_{s_f} = s_{y_p}$

Error limits on $y = \hat{m}x + \hat{b}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f: $y_p = \hat{m}x_p + \hat{b}$	Representative value of f: (include units) $y_p =$	95% C.I. of f: $(f \pm 2e_{sf})$ (include units) $y_p = \pm(2)(s_{y_p})$

x_i	Measured quantities, x_i: Symbol	Representative value	$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_1	$\hat{m}$		$\frac{\partial y_p}{\partial \hat{m}} = x_p$	s_m	$(x_p s_m)^2$
x_2	x_p		$\frac{\partial y_p}{\partial x_p} = \hat{m}$	0	0
x_3	$\hat{b}$		$\frac{\partial y_p}{\partial \hat{b}} = 1$	s_b	s_b^2
x_4					
x_5					

But, $\hat{m}$ and $\hat{b}$ are not **independent** (both are calculated from the y_i).

$$e_{s_f}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{s_f}^2 = s_{y_p}^2$

$e_{s_f} = s_{y_p}$ units

Error limits on $y = \hat{m}x + \hat{b}$

Ordinary, *Least Squares*, Linear Regression

Answer: $e_s = e_{sf}$

Error Propagation Worksheet
CM3215 Fundamentals of Chemical Engineering Lab
Prof. Faith Morrison

$f(x_1, x_2, x_3, x_4, x_5)$:	Formula for f: $y_p = \hat{m}x_p + \hat{b}$	Representative value of f: (include units) $y_p =$	95% C.I. of f: $(f \pm 2e_{sf})$ (include units) $y_p = \pm(2)(s_{y_p})$

x_i	Measured quantities, x_i: Symbol	Representative value	$\frac{\partial f}{\partial x_i}$	$e_{x_i} = \frac{s_i}{\sqrt{N}}$ or $\frac{e_{R_i}}{\sqrt{3}}$ or e_{s_i}	$\left(\frac{\partial f}{\partial x_i}\right)^2 e_{x_i}^2$
x_1	$\hat{m}$		$\frac{\partial y_p}{\partial \hat{m}} = x_p$	s_m	$(x_p s_m)^2$
x_2	x_p		$\frac{\partial y_p}{\partial x_p} = \hat{m}$	0	0
x_3	$\hat{b}$		$\frac{\partial y_p}{\partial \hat{b}} = 1$	s_b	s_b^2
x_4					
x_5					

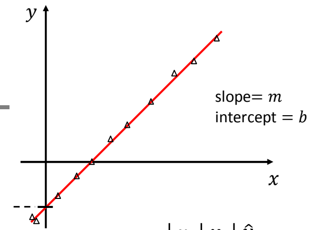
$$+2\left(\frac{\partial y_p}{\partial \hat{m}}\right)\left(\frac{\partial y_p}{\partial \hat{b}}\right)\mathrm{Cov}(\hat{m}, \hat{b})$$

$$e_{s_f}^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 e_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 e_{x_2}^2 + \left(\frac{\partial f}{\partial x_3}\right)^2 e_{x_3}^2 + \left(\frac{\partial f}{\partial x_4}\right)^2 e_{x_4}^2 + \left(\frac{\partial f}{\partial x_5}\right)^2 e_{x_5}^2$$

$e_{s_f}^2 = s_{y_p}^2$

$e_{s_f} = s_{y_p}$ units

Ordinary, *Least Squares*, Linear Regression



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

What are the error limits on a value of y obtained from the equation $y = \hat{m}x + \hat{b}$?

Answer:

at x_p, $y_p = \left(\hat{m}x_p + \hat{b}\right) \pm 2s_{y_p}$

$$s_{y_p}^2 = s_{y,x}^2\left(\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}\right)$$

(final result of the algebra indicated on previous slide see Appendix B of the handout.)

for $n - 2 \leq 6$,
replace "2" with $t_{0.025,n-2}$

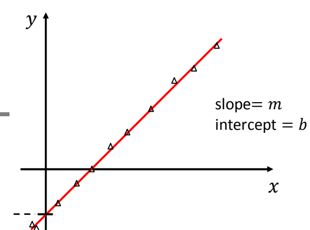
Use this for error limits on the fit (95% CI).

In Excel:

- $s_{y,x}$ = STEYX(y–range,x–range)
- SS_{xx} = DEVSQ(x–range)
- $\bar{x}$ = AVERAGE(x–range)

31

Ordinary, *Least Squares*, Linear Regression



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

What are the error limits on a predicted next value of y obtained from the equation $y = \hat{m}x + \hat{b}$?

Answer:

at x_p, we predict a new measurement of y will fall in the prediction interval:

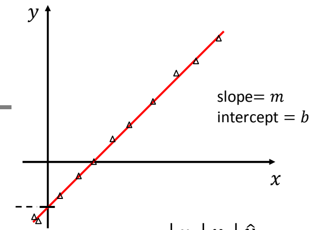
$$y_{\hat{p}} = \left(\hat{m}x_p + \hat{b}\right) \pm 2e_s$$

?

32

Ordinary, *Least Squares*, Linear Regression

What are the error limits on a predicted <u>next value</u> of y obtained from the equation $y = \hat{m}x + \hat{b}$?



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

Answer:

at x_p, we predict a <u>new measurement</u> of y will fall in the prediction interval:

$$y_{\hat{p}} = \left(\hat{m}x_p + \hat{b}\right) \pm 2e_s$$

?

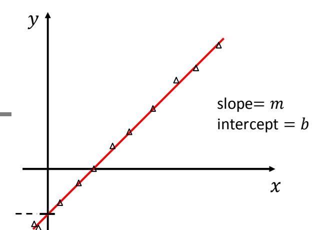
The new measurement $y_{\hat{p}}$ will have the same scatter as the source measurements and is less certain than the prediction of the mean value y_p at x_p.

Solve with same approach as we have been using: write the equation to calculate the quantity, then propagate the error.

(See Appendix B of the handout.)

Ordinary, *Least Squares*, Linear Regression

What are the error limits on a predicted <u>next value</u> of y obtained from the equation $y = \hat{m}x + \hat{b}$?



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

Answer:

at x_p, we predict a <u>new measurement</u> of y will fall in the prediction interval:

$$y_{\hat{p}} = \left(\hat{m}x_p + \hat{b}\right) \pm 2s_{y_{\hat{p}}}$$

$$s_{y_{\hat{p}}}^2 = s_{y,x}^2 \left(1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}\right)$$

(See Appendix B of the handout.)

for $n - 2 \leq 6$,
replace "2" with $t_{0.025,n-2}$

Use this for predicting next likely y (95% PI).

Ordinary, *Least Squares*, Linear Regression

Confidence interval for the fit:

$$y_p = (\hat{m}x_p + \hat{b}) \pm 2s_{y_p}$$

$$s_{y_p}^2 = s_{y,x}^2\left(\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}\right)$$

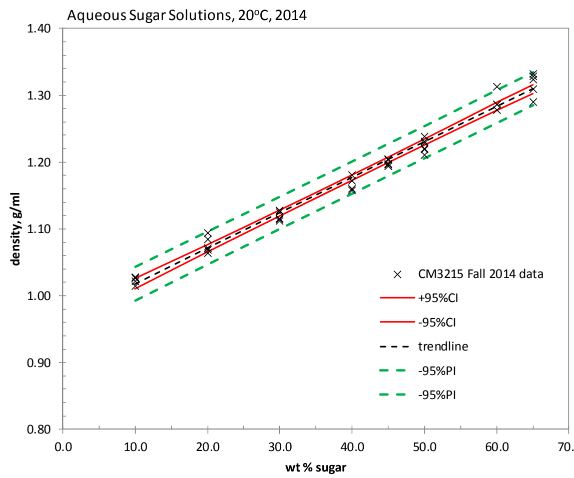
(for large n, the mean of y at each point is well predicted (CI is narrow))

Prediction interval:

$$y_{\hat{p}} = (\hat{m}x_p + \hat{b}) \pm 2s_{y_{\hat{p}}}$$

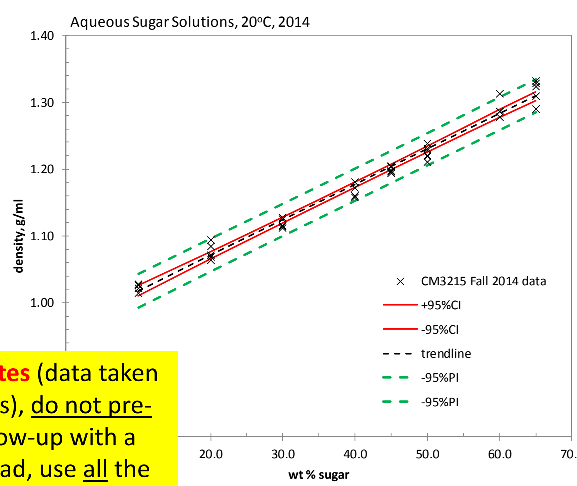
$$s_{y_{\hat{p}}}^2 = s_{y,x}^2\left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}\right)$$

(Notice that $\approx 95\%$ of the data points fall within the PI; that's what it means to be a PI.)



35 © Faith A. Morrison, Michigan Tech U.

Ordinary, *Least Squares*, Linear Regression



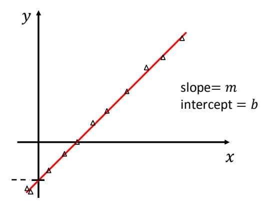
Note: if your data are **replicates** (data taken repeatedly at chosen x values), do not pre-average the y-data and follow-up with a least-squares curve fit. Instead, use all the replicates as individual values, and let LINEST find the least squared error among all points.

36 © Faith A. Morrison, Michigan Tech U.

Summary:

Uncertainty Ordinary, *Least Squares*, Linear Regression

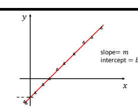
- The **Ordinary** ***Least Squares*** **Linear Regression** method provides the equations needed to obtain model parameters ***slope*** and ***intercept***.
- The equations for the parameters may be used with error propagation to obtain the variances associated with the parameters
- 95% confidence intervals on the parameters are constructed with $\pm 2e_s$ for large n
- For $n-2 \leq 6$, the 95% CI is constructed as $\pm\, t_{0.025,n-2}\, e_s$
- We can construct 95% CI on the mean value of y at a chosen x. These CI are used for **error range** on the fit.
- We can construct 95% prediction intervals (PI) on a next value of y at a chosen x. These are used for bracketing likely observed next values of y.



i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$

37

i	x_i	y_i	$\hat{y}_i$
1	x_1	y_1	$\hat{y}_1$
2	x_2	y_2	$\hat{y}_2$
⋮	⋮	⋮	⋮
n	x_n	y_n	$\hat{y}_n$



Excel Summary:

Uncertainty Ordinary, *Least Squares*, Linear Regression

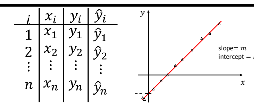
- $\bar{x}$ = AVERAGE(range)
- s^2 = VAR.S(range)
- s = STDEV.S(range)
- n = COUNT(range)
- SS_{xx} = DEVSQ(x-range)
- $\hat{m}$ = SLOPE(y-range, x-range)
- $\hat{b}$ =INTERCEPT(y-range,x-range)
- $s_{y,x}$ = STEYX(y-range, x-range)
- LINEST (see handout)
- LOGEST (look it up)

- $s_m^2 = \dfrac{s_{y,x}^2}{SS_{xx}}$
- $s_b^2 = s_{y,x}^2\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{SS_{xx}}\right)$

Use for CI error bars on y-values
- $s_{y_p}^2 = s_{y,x}^2\left(\dfrac{1}{n} + \dfrac{(x_p-\bar{x})^2}{SS_{xx}}\right)$

Use for PI on next value of y
- $s_{y_{\hat{p}}}^2 = s_{y,x}^2\left(1 + \dfrac{1}{n} + \dfrac{(x_p-\bar{x})^2}{SS_{xx}}\right)$

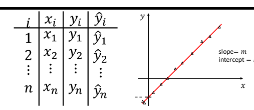
38



Excel Handy List:

Uncertainty Ordinary, *Least Squares*, Linear Regression

- $\hat{y}(x_p)$ = TREND(known-y's, known-x's,x_p) for y and x related by $y = mx + b$
- $\hat{y}(x_p)$ = GROWTH(known-y's, known-x's, x_p) for y and x related by $y = ae^{bx}$



One final piece of advice:

Uncertainty Ordinary, *Least Squares*, Linear Regression

Often, you can **transform** your data to make it linear, allowing you to use linear regression. For example, if you know the y-data vary as the square root of the x-data, then

$$y \text{ versus } \sqrt{x}$$

will be linear. If data plotted with log-log scaling (using scatterplot) look quadratic, then

$$\log y \text{ versus } \log x$$

will be quadratic, and we can use trendline to obtain a fit:

$$\log y = a(\log x)^2 + b(\log x) + c$$

Transforming data can greatly broaden our ability to fit empirical models to data.

CM3215 Fundamentals of Chemical Engineering Laboratory — Statistics *Quick Start*: Random Error and Replicates

CM3215 Fundamentals of Chemical Engineering Laboratory — Statistics Lecture 2: Reading Error

CM3215 Fundamentals of Chemical Engineering Laboratory — Statistics Lecture 3: Calibration Error

CM3215 Fundamentals of Chemical Engineering Laboratory — Statistics Lecture 4: Error Propagation

CM3215 Fundamentals of Chemical Engineering Laboratory — Uncertainty in Least Squares Curve Fitting: Excel's LINEST

Engineering Error Analysis: 5 Practical Lessons

MichiganTech

Professor Faith Morrison

Department of Chemical Engineering
Michigan Technological University

Done!

41