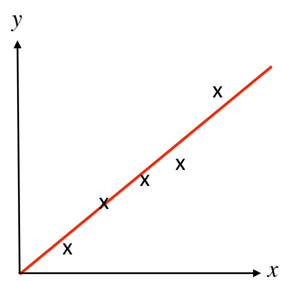
Curve fitting

- If the quantity of interest is a slope or an intercept
  - How do we obtain these values?
  - What is the associated error?
- Want to find the "best fit" of data to be represented by a straight line model
  - Least squares us a common criteria for best fit



$$F = \sum_i \left(y_i - y_{model}\right)^2 = \sum_i \left[y_i - \left(mx_i + b\right)\right]^2$$

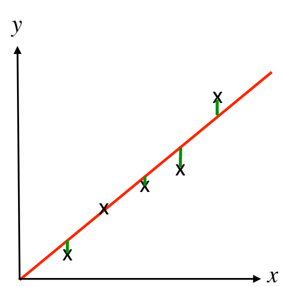
We want to find *m* and *b* that will give the smallest *F* value, e.g., find the *least* value of the *squared* errors

Curve fitting

- Condition for minimum: dF/dm = 0 and dF/db = 0

$$m = \frac{n\sum_i (x_i y_i) - \left(\sum_i x_i\right)\left(\sum_i y_i\right)}{n\left(\sum_i x_i^2\right) - \left(\sum_i x_i\right)^2}$$

$$b = \frac{\left(\sum_i x_i^2\right)\left(\sum_i y_i\right) - \left(\sum_i x_i\right)\left(\sum_i x_i y_i\right)}{n\left(\sum_i x_i^2\right) - \left(\sum_i x_i\right)^2}$$



Your calculator calculates this for you but what is the standard deviation in m and b?

Define: $\delta y_i = y_i - (mx_i + b)$

Standard deviation in y is:

$$S_y = \sqrt{\frac{\sum_i \left(\delta y_i\right)^2}{n-2}} \quad \text{and} \quad \lambda_y = \frac{t.S_y}{\sqrt{n}}$$

P = 5 and DF = n-2

Curve fitting

- Standard deviations in *m* and *b* are:

$$S_m = S_y \sqrt{\frac{n}{n\left(\sum_i x_i^2\right) - \left(\sum_i x_i\right)^2}}$$

$$S_b = S_y \sqrt{\frac{\left(\sum_i x_i^2\right)}{n\left(\sum_i x_i^2\right) - \left(\sum_i x_i\right)^2}}$$

and

$$\lambda_m = \frac{t.S_m}{\sqrt{n}}$$

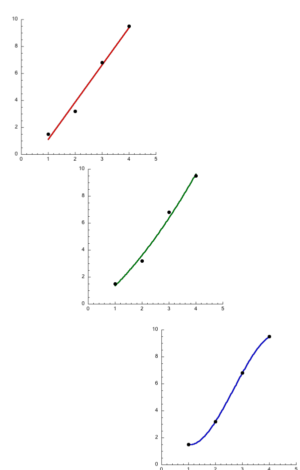
$$\lambda_b = \frac{t.S_b}{\sqrt{n}}$$

with $P = 5$ and DF = n-2

Fit validity

- Every DF of a system represents a measurement or data point
- Which fits are we allowed to perform?

DF (data points)	Model	Formula
1	None	None
2	Linear	$mx + b$
3	Parabola	$ax^2 + bx + c$
4	Cubic	$ax^3 + bx^2 + cx + d$
5	Quartic	$ax^4 + bx^3 + cx^2 + dx + e$
6	5th order	$ax^5 + bx^4 + cx^3 + dx^2 + ex + d$

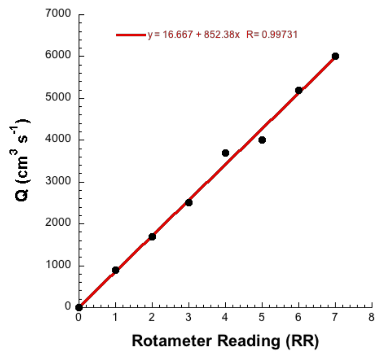


No quartic!

Example: rotameter calibration

- Measure Q with bucket/stopwatch; read RR; vary Q; repeat

RR (read)	Volume (L)	Time (s)	Q (cm³/s)
0	0	10	0
1	9	10	900
2	17	10	1700
3	25	10	2500
4	37	10	3700
5	40	10	4000
6	52	10	5200
7	60	10	6000



- Applying formula (or using software) with n=8 yields:

Q = 850.RR + 17 (best fit)

Example: rotameter calibration

- How much confidence should be place in Q?

$$S_Q = \sqrt{\frac{\sum_i (\delta Q_i)^2}{n-2}}$$

$\delta Q_i = Q_i - Q_{model} = Q_i - (850.RR_i + 17)$

$\delta Q_1 = 0 - 17 = -17$
$\delta Q_2 = 900 - (850 + 17) = 33$
$\delta Q_3 = 1700 - (1700 + 17) = -17$ ...

$$S_Q^2 = \frac{\sum_i (\delta Q_i)^2}{8-2} = 27585$$

and $S_Q = 166$ cm³/s

From the critical values of *t* table for P = 5 and and DF = n-2 = 8-2 = 6, we have *t* = 2.45

$$\lambda_Q = \frac{t.S_Q}{\sqrt{n}} = \frac{2.45 \times 166}{\sqrt{8}} = 144 \text{ cm}^3\text{s}^{-1}$$

Thus: $Q = [850.RR + 17] \pm 144$ cm³s⁻¹ (95% confidence, best fit)

So, if RR is set at 3.5 $Q = 2992 \pm 144$ cm³s⁻¹ $\approx 3000 \pm 140$ cm³s⁻¹

What if the expression is not linear?

- Reduce to linear model
  - Hydrodynamic power law: $C_D = k.\mathrm{Re}^n$ => $\ln C_D = \ln k + n \ln \mathrm{Re}$
  - Langmuir adsorption: $\theta = \dfrac{k.c}{1+k.c}$ => $\dfrac{1}{\theta} = 1 + \left(\dfrac{1}{k}\right).\dfrac{1}{c}$
  - Arrhenius kinetics $k = A.e^{-\frac{E_a}{R.T}}$ => $\ln k = \ln A - \left(\dfrac{E_a}{R}\right).\dfrac{1}{T}$

Correlation coefficient

- The correlation coefficient r^2 describes how much of the variation in y is described by the model ($y_{model} = m x_i + b$)
- In other words, r^2 is the fraction of the variation in y that can be described by the variation in x

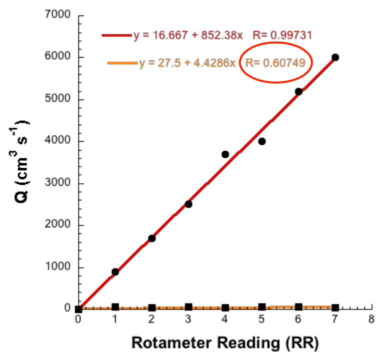
$$r^2 = \frac{\left[n\sum_i (x_i y_i) - \left(\sum_i x_i\right)\left(\sum_i y_i\right)\right]^2}{\left[n\left(\sum_i x_i^2\right) - \left(\sum_i x_i\right)^2\right]\left[n\left(\sum_i y_i^2\right) - \left(\sum_i y_i\right)^2\right]}$$

recall that $m = \dfrac{n\sum_i (x_i y_i) - \left(\sum_i x_i\right)\left(\sum_i y_i\right)}{n\left(\sum_i x_i^2\right) - \left(\sum_i x_i\right)^2}$

Correlation coefficient

- The correlation coefficient r^2 describes how much of the variation in y is described by the model ($y_{model} = m x_i + b$)
- In other words, r^2 is the fraction of the variation in y that can be described by the variation in x

$$r^2 = \frac{\left[n \sum_i (x_i y_i) - \left(\sum_i x_i \right) \left(\sum_i y_i \right) \right]^2}{\left[n \left(\sum_i x_i^2 \right) - \left(\sum_i x_i \right)^2 \right] \left[n \left(\sum_i y_i^2 \right) - \left(\sum_i y_i \right)^2 \right]}$$



- If $r^2 = 1$, *all* of the variation in y is described by the model

However, a nearly horizontal data set will have low r^2 since there is little variation in y to be described by the model

Propagation of errors

- Goal: to calculate the mean and standard deviation (or confidence interval) in a variable A that depends on measured quantities

$$A = f(x,y,z)$$

x_i	y_i	z_i
x_1	y_1	z_1
x_2	y_2	z_2
.	.	.
.	.	.
.	.	.
x_n	y_n	z_n

$$\bar{x} = \frac{\sum_i x_i}{n} \quad \bar{y} = \frac{\sum_i y_i}{n} \quad \bar{z} = \frac{\sum_i z_i}{n}$$

$$s_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} \quad s_y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n-1}} \quad s_z = \sqrt{\frac{\sum_i (z_i - \bar{z})^2}{n-1}}$$

Possible approach

$A_1 = f(x_1,y_1,z_1)$
$A_2 = f(x_2,y_2,z_2)$
.
.
.
$A_n = f(x_n,y_n,z_n)$

$$\bar{A} = \frac{\sum_i A_i}{n} \quad s_A = \sqrt{\frac{\sum_i (A_i - \bar{A})^2}{n-1}}$$

OK for *estimate* but propagation of error gives much more info

Propagation of errors

- Better approach is to calculate $\overline{A}$ from the means of measured variables

$$\overline{A} = f(\overline{x},\overline{y},\overline{z})$$

- And to calculate S_A (or λ_A) from the knowledge of S_x, S_y and S_z (or λ_x, λ_y and λ_z)

- Using a Taylor series development of Ai in the vicinity of zero, it can be shown that for independent variables x, y and z:

$$S_A^2 = \left(\frac{\partial A}{\partial x}\right)^2_{\overline{y},\overline{z}} S_x^2 + \left(\frac{\partial A}{\partial y}\right)^2_{\overline{x},\overline{z}} S_y^2 + \left(\frac{\partial A}{\partial z}\right)^2_{\overline{x},\overline{y}} S_z^2$$

and

$$\lambda_A^2 = \left(\frac{\partial A}{\partial x}\right)^2_{\overline{y},\overline{z}} \lambda_x^2 + \left(\frac{\partial A}{\partial y}\right)^2_{\overline{x},\overline{z}} \lambda_y^2 + \left(\frac{\partial A}{\partial z}\right)^2_{\overline{x},\overline{y}} \lambda_z^2$$

Example

- Propagation of error is more conservative *and* provides feedback on experimental design
- Consider pressure drop measurement

Approach 1

Run	P_{in}	P_{out}	ΔP
1	553	335	218
2	495	330	165
3	515	310	205
4	565	340	225
5	525	315	210
6	510	326	184

$$\overline{\Delta P} = 201$$
$$S_{\Delta P} = 22.6$$

Example

- Propagation of error is more conservative *and* provides feedback on experimental design
- Consider pressure drop measurement

Approach 2

Run	P_{in}	P_{out}
1	553	335
2	495	330
3	515	310
4	565	340
5	525	315
6	510	326

$\overline{P_{in}} = 527$, $S_{Pin} = 26.8$

$\overline{P_{out}} = 326$, $S_{Pout} = 11.6$

Calculate $\overline{\Delta P}$ from individual means:

$$\overline{\Delta P} = 527 - 326 = 201$$

Calculate variance from partial derivatives evaluated at means:

$$s_{\Delta P}^2 = \left(\frac{\partial(P_{in} - P_{out})}{\partial P_{in}}\right)^2_{\overline{P}_{out}} s_{P_{in}}^2 + \left(\frac{\partial(P_{in} - P_{out})}{\partial P_{out}}\right)^2_{\overline{P}_{in}} s_{P_{out}}^2$$

$$s_{\Delta P}^2 = 1.s_{P_{in}}^2 + 1.s_{P_{out}}^2 = 718.2 + 134.6$$

$$s_{\Delta P} = \sqrt{718.2 + 134.6} = 29.2$$

SD larger => More conservative
P_{in} is largest contributor to error!

Example 2

- Density of a rod determined by measuring length, diameter and mass

$$M = 5.0 \pm 0.1 \text{ kg (95\% conf, n = 6)}$$
$$L = 0.75 \pm 0.02 \text{ m (95\% conf, n = 5)}$$
$$D = 0.14 \pm 0.01 \text{ m (95\% conf, n = 6)}$$

Calculate $\overline{\rho}$ from sample means:

$$\overline{\rho} = \frac{M}{V} = \frac{M}{\left(\frac{\pi D^2}{4}\right)L} = \frac{4}{\pi}.\frac{M}{D^2 L} = \frac{4 \times 5}{\pi (0.14)^2 \times 0.75} = 433 \ \ kg/m^3$$

Calculate partial derivatives and evaluate at sample means:

$\frac{\partial \rho}{\partial M} = \frac{4}{\pi}.\frac{1}{D^2 L}$ at $\overline{D}$ and $\overline{L}$, we have: $\frac{\partial \rho}{\partial M} = \frac{4}{\pi}.\frac{1}{(0.14)^2 \times 0.75} = 86.6$

$\frac{\partial \rho}{\partial L} = -\frac{4}{\pi}.\frac{M}{D^2 L^2}$ at $\overline{D}$ and $\overline{L}$ and $\overline{M}$ we have: $\frac{\partial \rho}{\partial L} = -\frac{4}{\pi}.\frac{5}{(0.14)^2 (0.75)^2} = -577$

$\frac{\partial \rho}{\partial D} = -\frac{8}{\pi}.\frac{M}{D^3 L}$ at $\overline{D}$ and $\overline{L}$ and $\overline{M}$ we have: $\frac{\partial \rho}{\partial D} = -\frac{8}{\pi}.\frac{5}{(0.14)^3 (0.75)} = -6187$

Calculate confidence interval for density:

$$\lambda_\rho = \sqrt{\left(\frac{\partial \rho}{\partial M}\right)^2 \lambda_M^2 + \left(\frac{\partial \rho}{\partial L}\right)^2 \lambda_L^2 + \left(\frac{\partial \rho}{\partial D}\right)^2 \lambda_D^2} = \sqrt{(86.6)^2 (0.1)^2 + (-577)^2 (0.02)^2 + (-6187)^2 (0.01)^2} = \sqrt{74.99 + 133.17 + 3828} = 63.53 \ \ kg/m^3$$

$D = 433 \pm 63$ kg/m^3 (95% conf, propagated) D biggest contributor to error: use calipers!