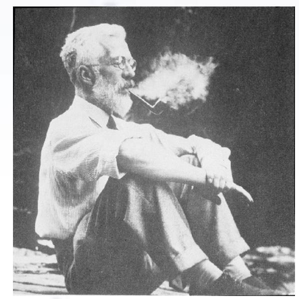
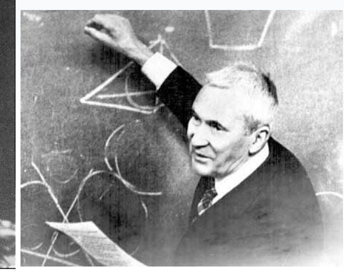
Applied Statistics

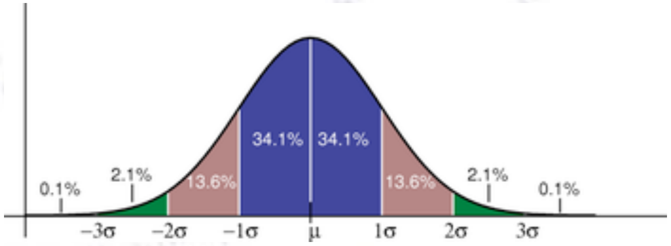
Error propagation





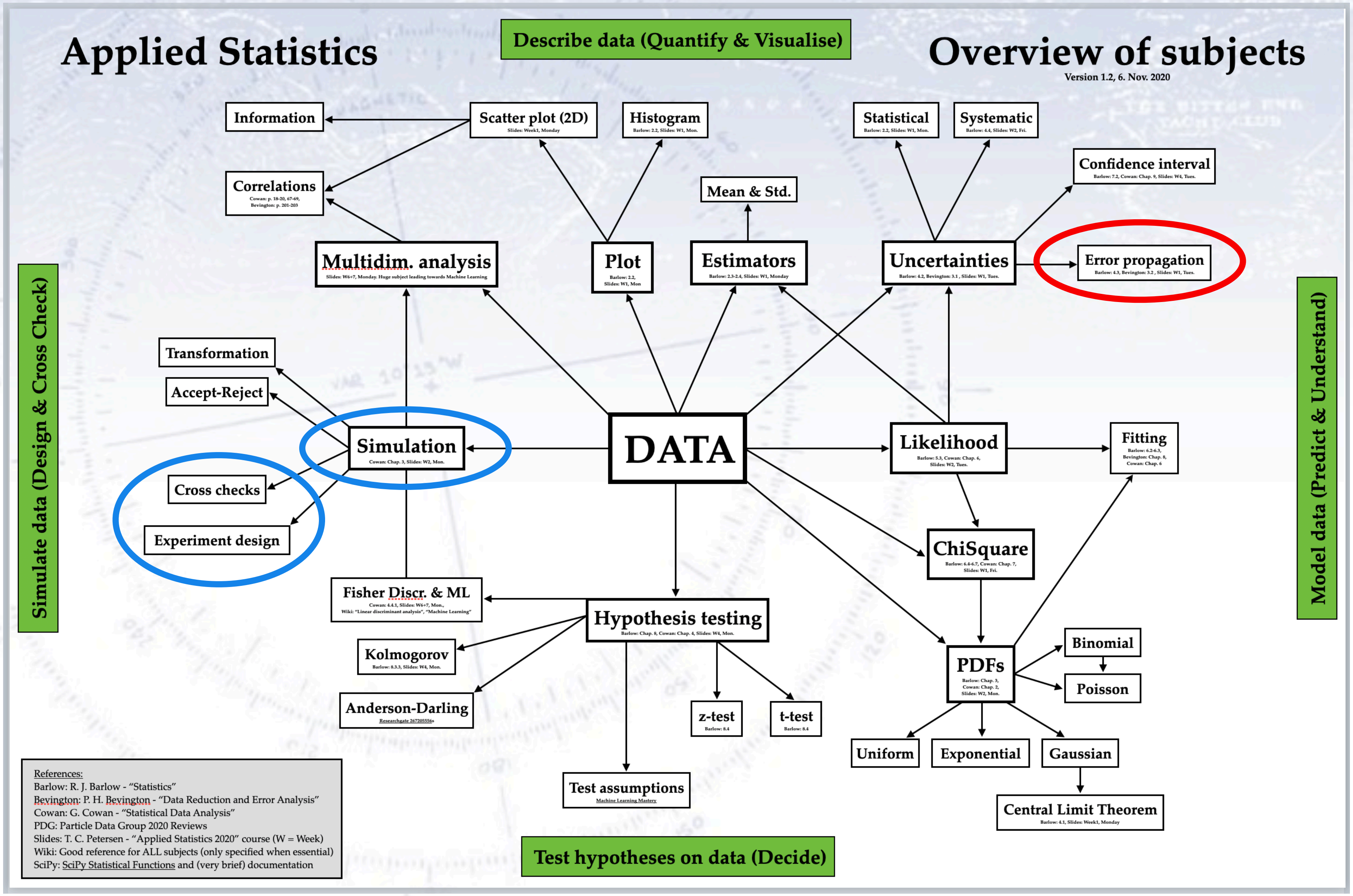



Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Error propagation



Error propagation

Imagine that y is a function of x_i

$$y(x_i)$$

and that we wish to find the error on y from the errors on x_i .

$$\sigma(x_i) = 0.8$$
$$\sigma(y(x_i)) = ?$$

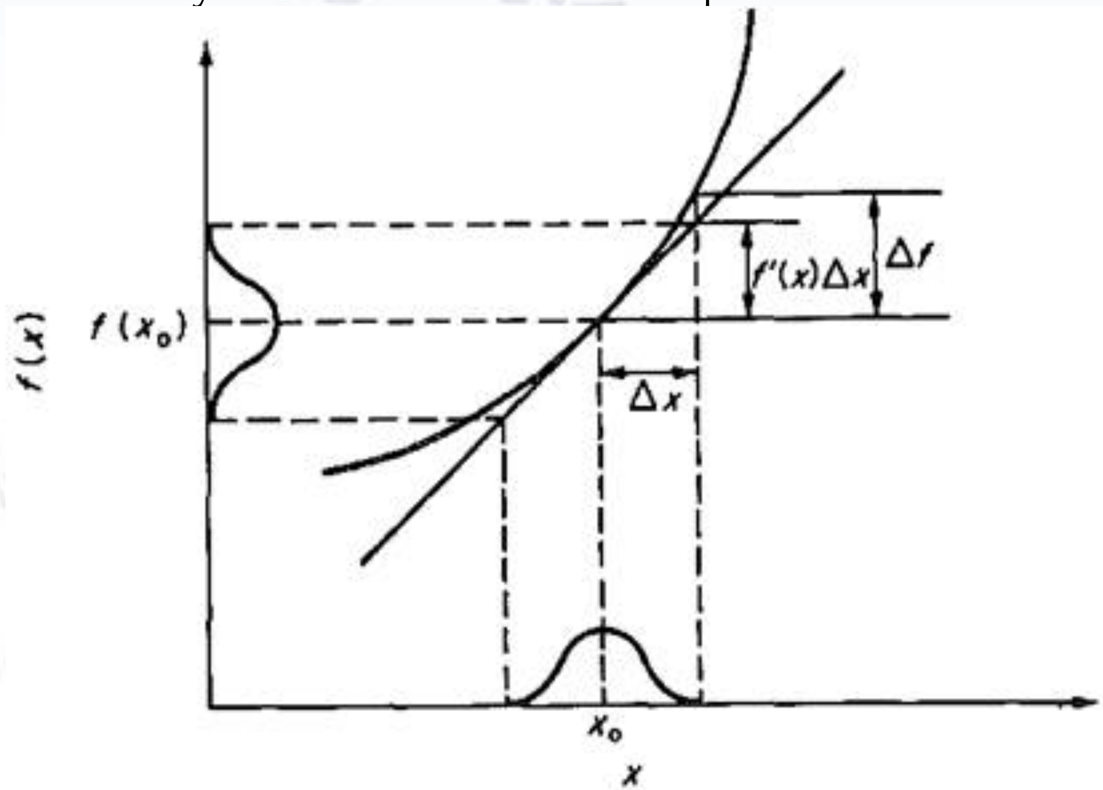
Error propagation

Imagine that y is a function of x_i

$$y(x_i)$$

and that we wish to find the error on y from the errors on x_i .

$$\sigma(x_i) = 0.8$$
$$\sigma(y(x_i)) = ?$$



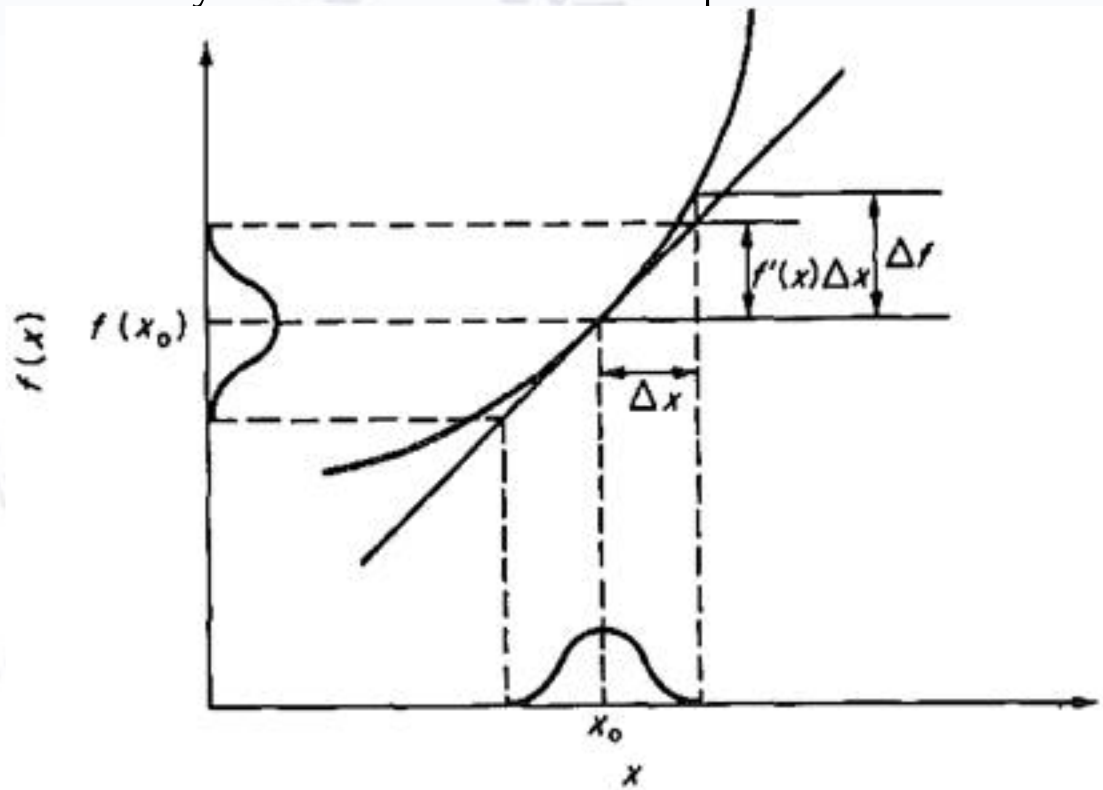
Error propagation

Imagine that y is a function of x_i

$$y(x_i)$$

and that we wish to find the error on y from the errors on x_i .

$$\sigma(x_i) = 0.8$$
$$\sigma(y(x_i)) =$$
$$\frac{\partial y}{\partial x_i} \times 0.8$$



Error propagation

<u>Note, the approximation here:</u>

The derivative of y - dy/dx_i - should be relatively constant.

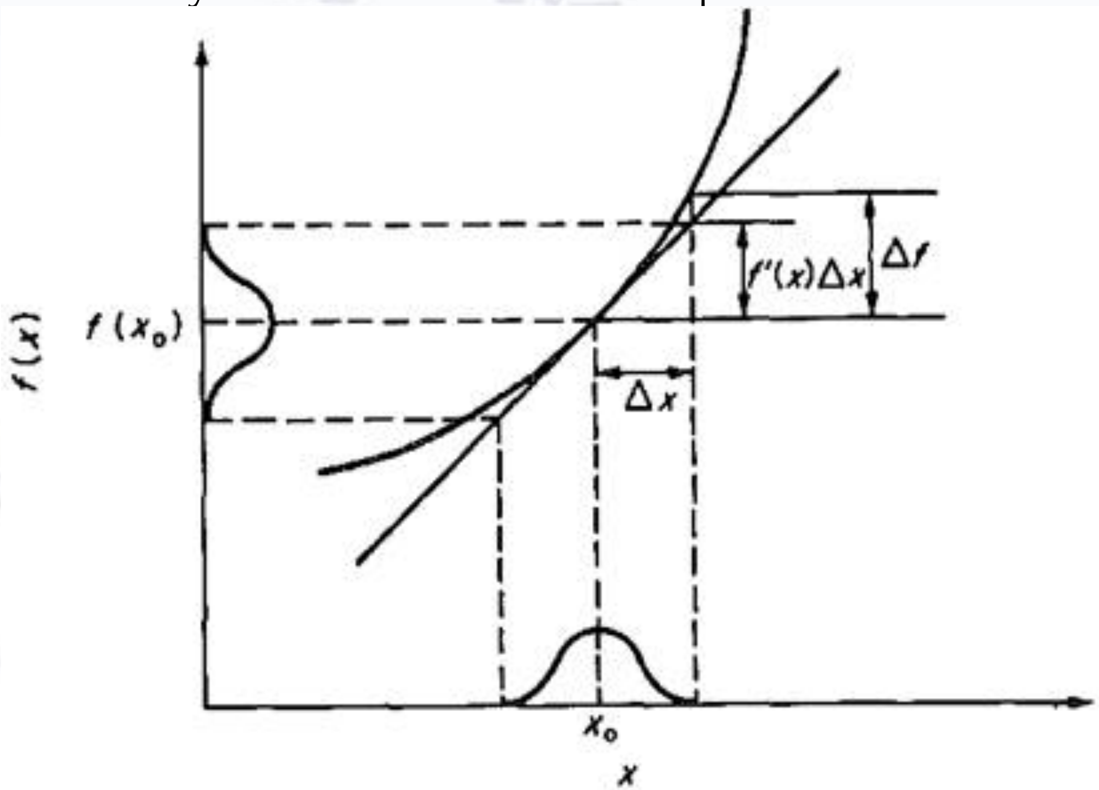
If not, the error propagation formula breaks down.

and that we wish to find the error on y from the errors on x_i .

$$\sigma(x_i) = 0.8$$

$$\sigma(y(x_i)) =$$

$$\frac{\partial y}{\partial x_i} \times 0.8$$



General formula

(i.e. can always be used!)

Error propagation

Imagine that y is a function of x_i, and that we wish to find the error on y from the errors on x_i. Making a Taylor expansion of the function y gives:

$$y(\bar{x}) \simeq y(\bar{\mu}) + \sum_{i}^{n} \frac{\partial y}{\partial x_i}(x_i - \mu_i)$$

In order to get the uncertainty of y as a function of the variables x_i we calculate:

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 = E[x^2] - E^2[x]$$

$$E[y(\bar{x})] \simeq y(\bar{\mu})$$

$$E[y^2(\bar{x})] \simeq y^2(\bar{\mu}) + \sum_{i,j}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right] V_{ij}$$

Error propagation formula

Subtracting the two formulae, we obtain:

$$\sigma_y^2 = \sum_i^n \sum_j^n \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right] V_{ij}$$

Error propagation formula

Subtracting the two formulae, we obtain:

$$\sigma_y^2 = \sum_i^n \sum_j^n \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right] V_{ij}$$

If there are no correlations, only the diagonal (individual errors) enter:

$$\sigma_y^2 = \sum_i^n \left[\frac{\partial y}{\partial x_i} \right]_{\bar{x}}^2 \sigma_i^2$$

Error propagation formula

The "simple" (i.e. uncorrelated) formula is thus:

$$\sigma_y = \sqrt{\left[\frac{\partial y}{\partial x_1}\right]^2_{\bar{x}_1} \sigma_1^2 + \left[\frac{\partial y}{\partial x_2}\right]^2_{\bar{x}_1} \sigma_2^2 + \ldots}$$

Error propagation formula

The "simple" (i.e. uncorrelated) formula is thus:

$$\sigma_y = \sqrt{\left[\frac{\partial y}{\partial x_1}\right]^2_{\bar{x}_1} \sigma_1^2 + \left[\frac{\partial y}{\partial x_2}\right]^2_{\bar{x}_1} \sigma_2^2 + \ldots}$$

Note, that **each term** represents the **individual contributions** of x_i to the uncertainty on y.

Thus, the uncertainty on y from e.g. x_1 is:

$$\sigma_y = \sqrt{\left[\frac{\partial y}{\partial x_1}\right]^2_{\bar{x}_1} \sigma_1^2} = \frac{\partial y}{\partial x_1}\sigma_1$$

Specific formula

(i.e. for special simple cases!)

Specific error propagation formula

Addition

$$y = x_1 + x_2$$

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + 2V_{x_1,x_2}$$

$$y = ax_1 + bx_2$$

$$\sigma_y^2 = a^2\sigma_{x_1}^2 + b^2\sigma_{x_2}^2 + 2abV_{x_1,x_2}$$

"When adding numbers, their errors add in quadrature"

Specific error propagation formula

Multiplication

$$y = x_1 x_2$$

$$\sigma_y^2 = (x_2\sigma_{x_1})^2 + (x_1\sigma_{x_2})^2 + 2x_1x_2V_{x_1,x_2}$$

Dividing by x^2 to get relative terms, we obtain:

$$\frac{\sigma_y^2}{y^2} = \frac{\sigma_{x_1}^2}{x_1^2} + \frac{\sigma_{x_2}^2}{x_2^2} + 2\frac{V_{x_1,x_2}}{x_1x_2}$$

"When multiplying numbers, their RELATIVE errors add in quadrature"

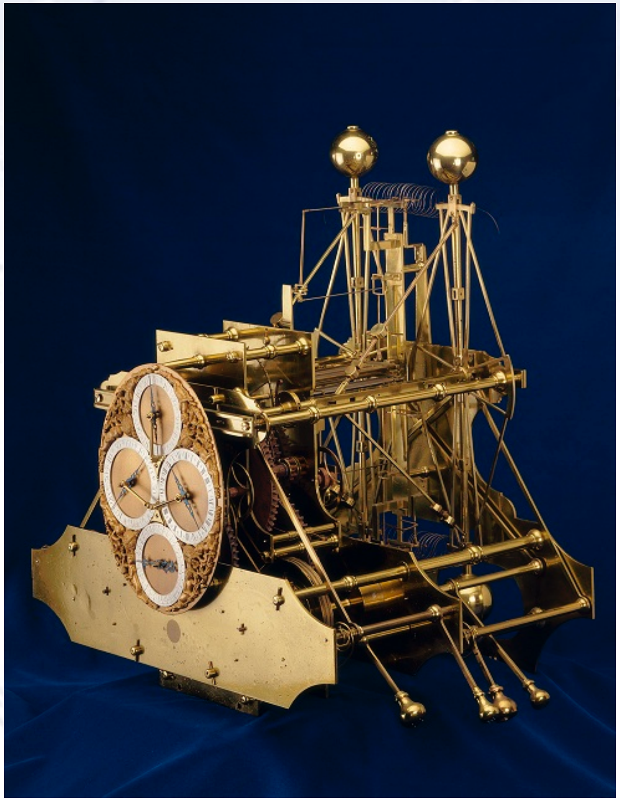
Error propagation at work...

John Harrison (24 March 1693 – 24 March 1776)

British clockmaker extraordinaire

“Won” the Longitude Act prize (3 sec/day).

Harrison's first sea clock (H1)

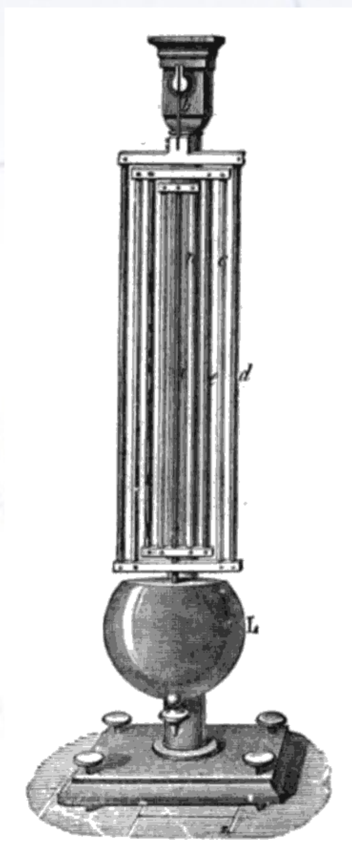


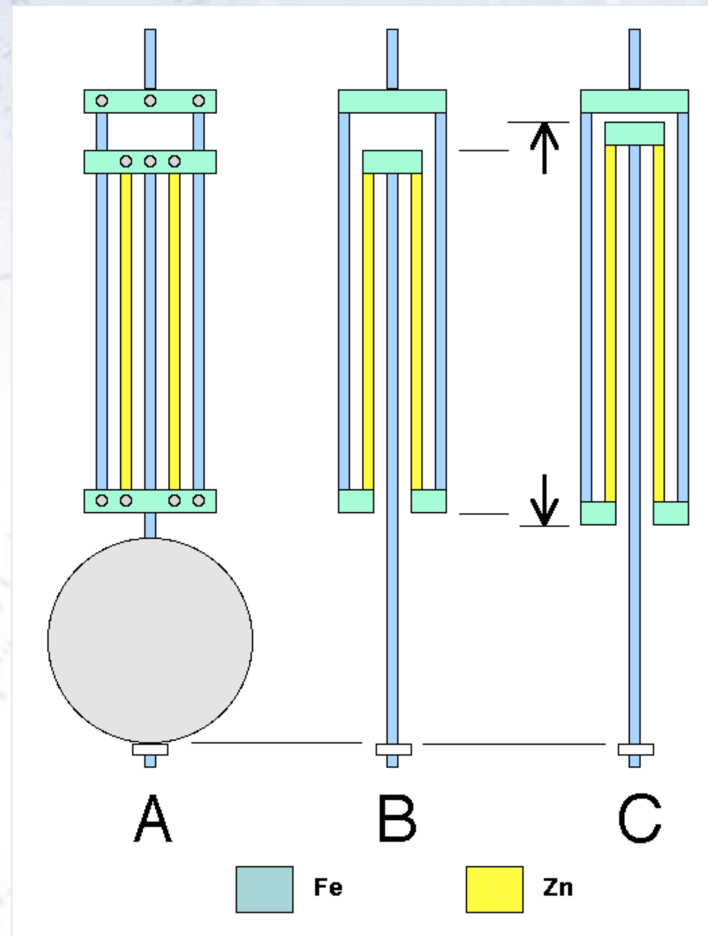
Harrison build H1-H5.

K1 (Copy of H4) was used by James Cook.

Error propagation at work...

Harrison's Gridiron pendulum is designed to cancel the change in length (in fact moment of inertia) with temperature.





Coefficient of thermal expansion:
Iron = 11.8×10^{-6}/C° Zinc = 30.2×10^{-6}/C°

Error propagation at more work...

Analysis of tiny differences in Uranus' orbit from Newtonian prediction led to the prediction and discovery of Neptune!

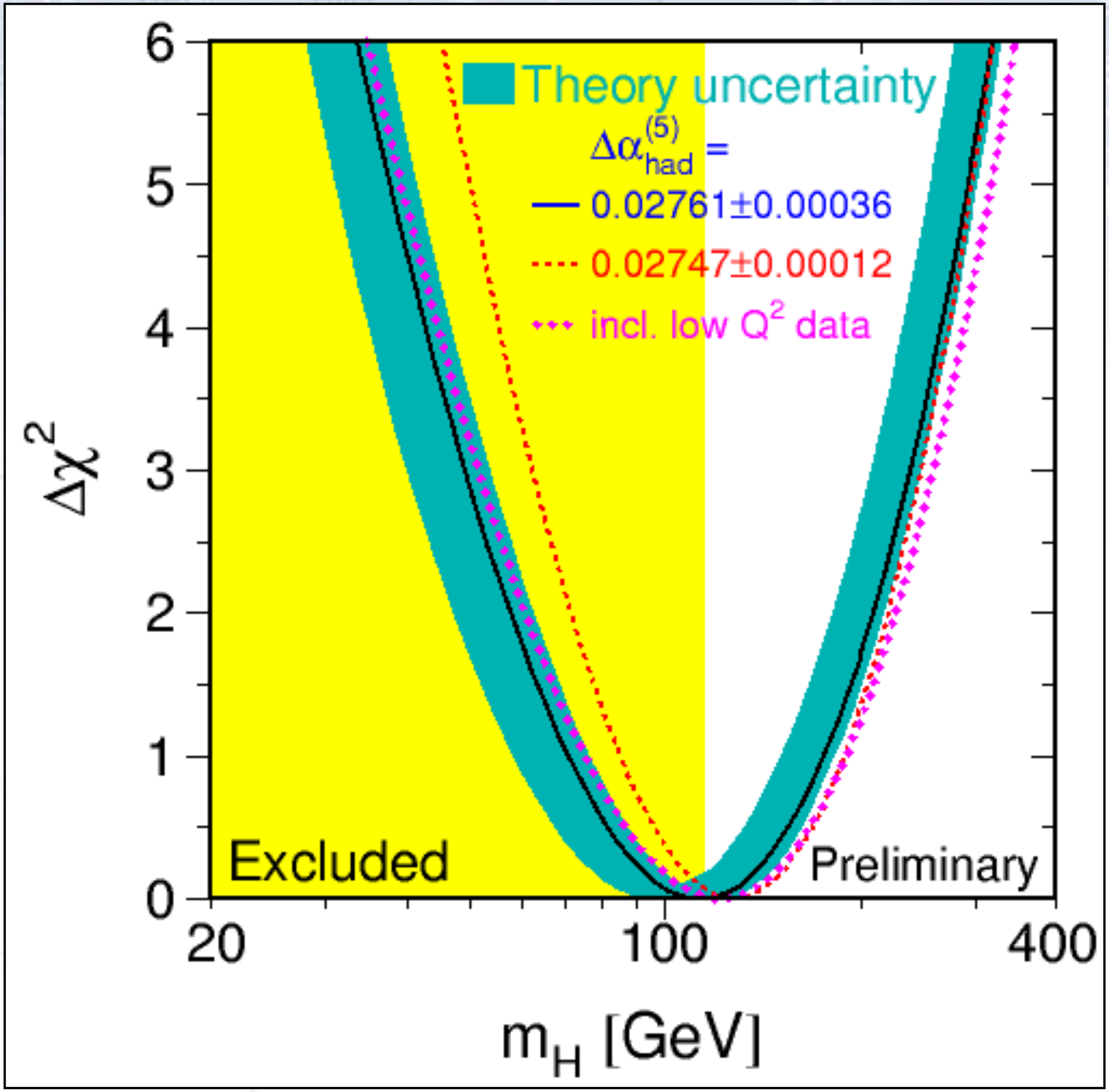
Continuing with Mercury...

Table II. Contributions to the motion of the perihelia of Mercury and the earth.

Cause	m^{-1}	Motion of perihelion: Mercury	Motion of perihelion: Earth
Mercury	6 000 000 ±1 000 000	0″.025±0″.00	−13″.75±2″.3
Venus	408 000 ± 1 000	277.856±0.68	345.49±0.8
Earth	329 390 ± 300	90.038±0.08	
Mars	3 088 000 ± 3 000	2.536±0.00	97.69±0.1
Jupiter	1 047.39± 0.03	153.584±0.00	696.85±0.0
Saturn	3 499 ± 4	7.302±0.01	18.74±0.0
Uranus	22 800 ± 300	0.141±0.00	0.57±0.0
Neptune	19 500 ± 300	0.042±0.00	0.18±0.0
Solar oblateness		0.010±0.02	0.00±0.0
Moon			7.68±0.0
General precession (Julian century, 1850)		5025.645±0.50	5025.65±0.5
Sum		5557.18 ±0.85	6179.1 ±2.5
Observed motion		5599.74 ±0.41	6183.7 ±1.1
Difference		42.56 ±0.94	4.6 ±2.7
Relativity effect		43.03 ±0.03	3.8 ±0.0

Urbain Le Verrier (1811-1877)

Advanced example of error propagation (Higgs particle mass):



Reporting uncertainties

The systematic uncertainties of a measurement should be reported in a table, and if measurements are combined, the correlation needs consideration.

Selection criteria	Systematic uncertainties (%)
K/π PID	1.0
μ PID	0.6
Muon selection	0.6
Trigger	1.0
Yields of reference channel	0.4
Efficiency modeling	5.3

Reporting uncertainties

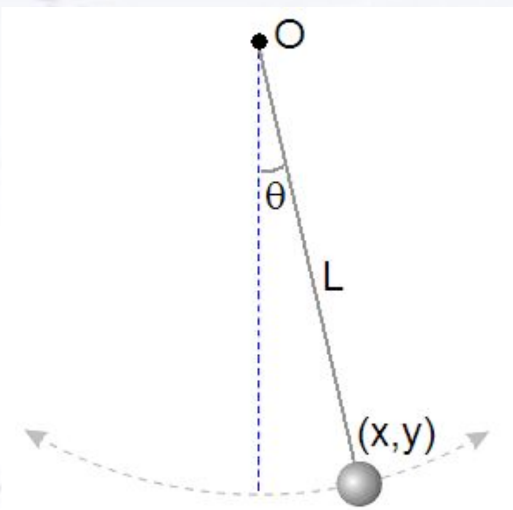
The systematic uncertainties of a measurement should be reported in a table, and if measurements are combined, the correlation needs consideration.

CDF II preliminary $L = 200\ pb^{-1}$

m_T Uncertainty [MeV]	Electrons	Muons	Common
Lepton Scale	30	17	17
Lepton Resolution	9	3	0
Recoil Scale	9	9	9
Recoil Resolution	7	7	7
$u_{\|\|}$ Efficiency	3	1	0
Lepton Removal	8	5	5
Backgrounds	8	9	0
$p_T(W)$	3	3	3
PDF	11	11	11
QED	11	12	11
Total Systematic	39	27	26
Statistical	48	54	0
Total	62	60	26

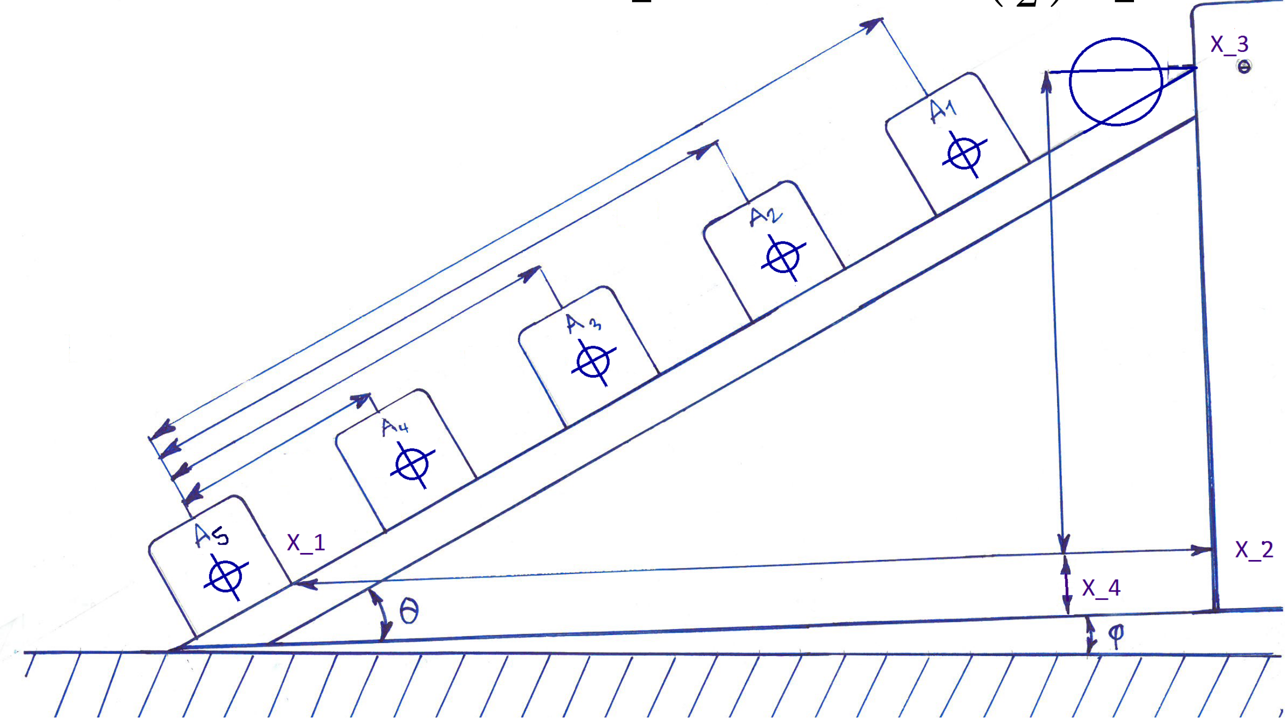
Applying error propagation

For the project, we'll be working with two measurements, which result from two formulae. Work out the error propagation formula for these two cases, and use these, when we discuss typical size of uncertainties.



$$g = L\left(\frac{2\pi}{T}\right)^2$$

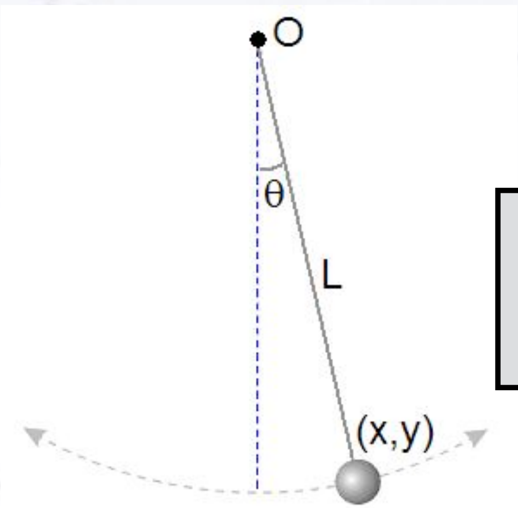
$$g = \frac{a}{\sin(\theta \pm \Delta\theta)}\left[1 + \frac{2}{5}\frac{R^2}{R^2 - \left(\frac{d}{2}\right)^2}\right]$$



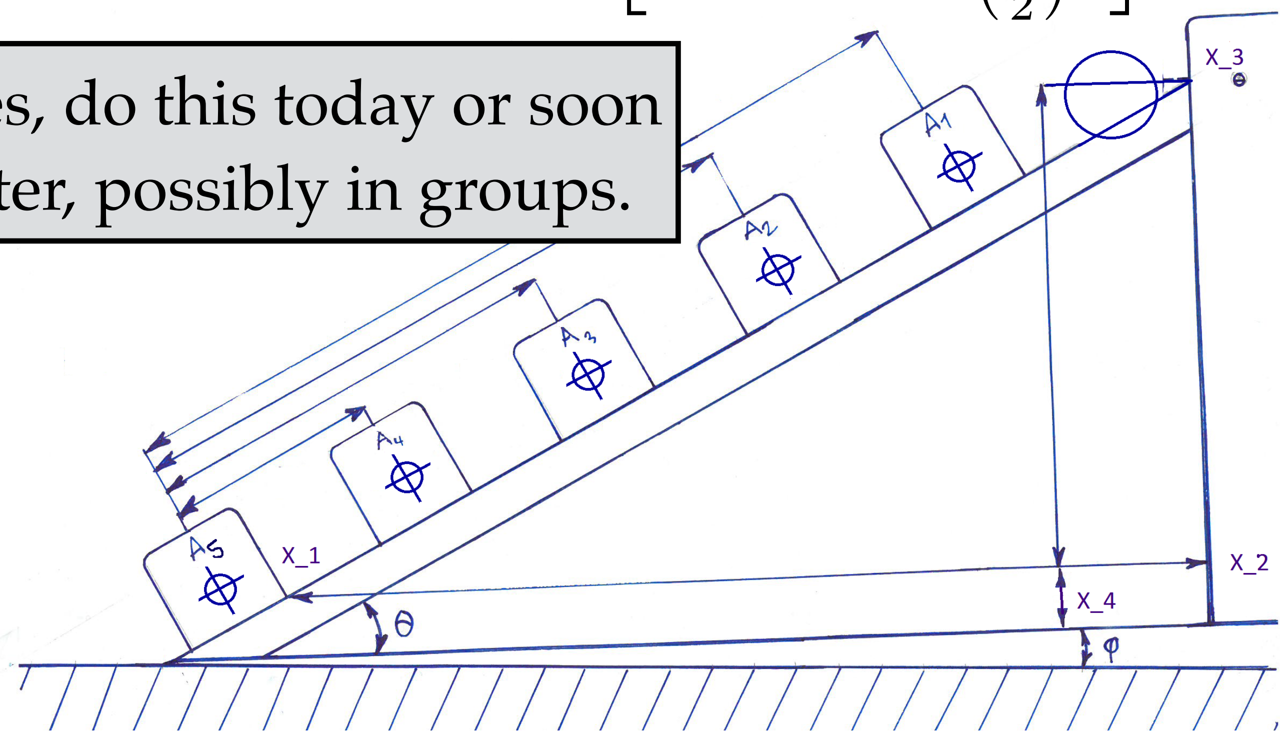
Applying error propagation

For the project, we'll be working with two measurements, which result from two formulae. Work out the error propagation formula for these two cases, and use these, when we discuss typical size of uncertainties.

$$g = \frac{a}{\sin(\theta \pm \Delta\theta)} \left[1 + \frac{2}{5} \frac{R^2}{R^2 - \left(\frac{d}{2}\right)^2} \right]$$



Yes, do this today or soon after, possibly in groups.



$$g = L \left(\frac{2\pi}{T} \right)^2$$

Simulating error propagation

Imagine that y is a very complicated function of x_i, perhaps not even parametric (i.e. not a function, but rather a model).

A simple method is to use simulation:

- Choose random values of x_i, corresponding to mean and Std of each x.
- Calculate $y(x_i)$ and record the resulting values.
- The standard deviation (and distribution) of y reflects the impact of x_i.

Note that the distribution of y may NOT be Gaussian, if the error propagation formula breaks down. It is then important to make this clear to the reader.

However, simulation exactly allows one to see to what degree the resulting distribution in y is Gaussian.

Errors on errors

The "uncertainty on the uncertainty" follows the approximate rule:

$$\sigma_\sigma = \frac{1}{\sqrt{2N-2}}$$

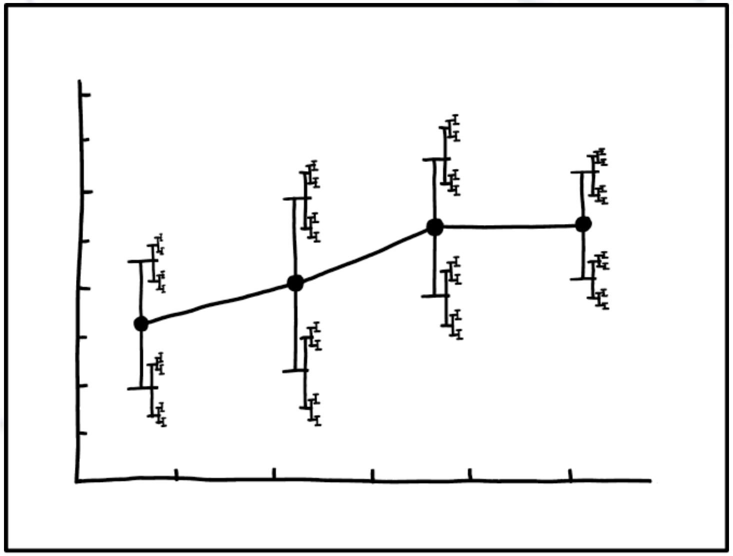
As we don't want to pursue an infinite line of uncertainties, we simply state the uncertainty, and only include one or two significant digits.

Errors on errors

The "uncertainty on the uncertainty" follows the approximate rule:

$$\sigma_\sigma = \frac{1}{\sqrt{2N-2}}$$

As we don't want to pursue an infinite line of uncertainties, we simply state the uncertainty, and only include one or two significant digits.



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.