Rejection of Data

3.8, 3.5, 3.9, 3.9, 3.4, 1.8

suspect

$$\bar{x} = \frac{1}{N}\sum x_i \qquad \sigma_x = \sqrt{\frac{1}{N-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$\bar{x} = 3.4\ \text{s}$$

$$\sigma = 0.8\ \text{s}$$

$$\bar{x} - x_{sus} = 3.4 - 1.8 = 1.6 = 2\sigma$$

Table A

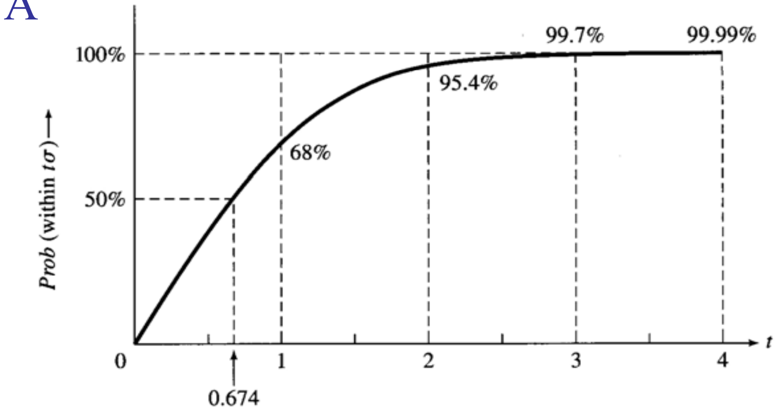
$$\text{Prob(outside } 2\sigma) = 1 - \text{Prob(within } 2\sigma) =$$

$$=1 - 0.95 = 0.05$$

$$n = (\text{expected number as deviant as } 1.8) =$$

$$= N \times \text{Prob(outside } 2\sigma) =$$

$$= 6 \times 0.05 = 0.3$$



erf(t) – error function

if $n < 0.5$ the measurement is “improbable” and can be rejected according to Chauvenet’s criterion

$$x_1, \ldots, x_N$$

$$t_{sus} = \frac{\left|x_{sus} - \bar{x}\right|}{\sigma}$$

$$n = N \times \text{Prob(outside } t_{sus}\sigma)$$

if $n < \frac{1}{2}$, then x_{sus} can be rejected

Chauvenet’s criterion

Example Problem

A student makes 5 measurements of the period of a pendulum and gets
T = 2.8, 2.5, 2.7, 2.7, 2.3 s.
Should any of these measurements be dropped?

Calculate the average

$$\bar{T} = \frac{2.8 + 2.5 + 2.7 + 2.7 + 2.3}{5} = 2.6\text{ s} \qquad \longleftarrow \qquad \bar{x} = \frac{1}{N}\sum x_i$$

Calculate the standard deviation

$$\sigma = \sqrt{\frac{1}{4}\left(0.2^2 + 0.1^2 + 0.1^2 + 0.1^2 + 0.3^2\right)} = 0.2\text{ s} \qquad \longleftarrow \qquad \sigma_x = \sqrt{\frac{1}{N-1}\sum\left(x_i - \bar{x}\right)^2}$$

The measurement furthest from the mean is 2.3 s
giving $t_{sus} = 0.3/0.2 = 1.5$ $\qquad \longleftarrow \qquad t_{sus} = \frac{\left|x_{sus} - \bar{x}\right|}{\sigma}$

Look up the probability to be further off, $P = 13.36\ \%$ ⟵ Table A

Multiply by the number of trials to get the expected
number of events that far off, $n = 5 \times 0.1336 \approx 0.67$

$0.67 \geq 0.5 \rightarrow$ Do not drop this measurement (or any other)

Weighted Averages

$$A: \quad x = x_A \pm \sigma_A$$
$$B: \quad x = x_B \pm \sigma_B$$

combining separate measurements: what is the best estimate for *x* ?

assume that measurements are governed by Gauss distribution with true value *X* $G_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}$

$$\text{Prob}_X(x_A) \propto \frac{1}{\sigma_A} e^{-(x_A-X)^2/2\sigma_A^2}$$

$$\text{Prob}_X(x_B) \propto \frac{1}{\sigma_B} e^{-(x_B-X)^2/2\sigma_B^2}$$

probability that A finds x_A

$$\text{Prob}_X(x_A x_B) = \text{Prob}_X(x_A) \cdot \text{Prob}_X(x_B)$$

probability that A finds x_A and B finds x_B

$$\propto \frac{1}{\sigma_A \sigma_B} e^{-\chi^2/2}$$

find maximum of probability

principle of maximum likelihood

the best estimate for *X* is that value for which *Prob*$_X$*(*x_A*,* x_B*)* is maximum

$$\chi^2 = \left(\frac{x_A - X}{\sigma_A}\right)^2 + \left(\frac{x_B - X}{\sigma_B}\right)^2$$

chi squared – “sum of squares”

find minimum of χ^2

method of least squares

$$\frac{d\chi^2}{dX} = 0 \quad \Rightarrow \quad -2\frac{x_A - X}{\sigma_A{}^2} - 2\frac{x_B - X}{\sigma_B{}^2} = 0$$

$$(\text{best estimate for } X) = \left(\frac{x_A}{\sigma_A^2} + \frac{x_B}{\sigma_B^2}\right) \Big/ \left(\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2}\right)$$

$$= \frac{w_A x_A + w_B x_B}{w_A + w_B} = x_{wav}$$

weighted average

$$w_A = \frac{1}{\sigma_A^2} \qquad w_B = \frac{1}{\sigma_B^2}$$

weights

Weighted Averages

$x_1, x_2, \ldots, x_N$ - measurements of a single quantity x with uncertainties $\sigma_1, \sigma_2, \ldots, \sigma_N$

$$x_1 \pm \sigma_1,\ x_2 \pm \sigma_2,\ \ldots,\ x_N \pm \sigma_N$$

$$x_{wav} = \frac{\sum w_i x_i}{\sum w_i}$$

← weighted average

$$w_i = \frac{1}{\sigma_i^2}$$

← weights

$$\sigma_{wav} = \frac{1}{\sqrt{\sum w_i}}$$

← uncertainty in x_{wav}

can be calculated
using error propagation

Example of Weighted Average

$R_1 = 11 \pm 1 \quad (\Omega)$
$R_2 = 12 \pm 1$
$R_3 = 10 \pm 3$

three measurements of a resistance
what is the best estimate for *R* ?

$\sigma_1 = 1 \qquad w_1 = 1$
$\sigma_2 = 1 \qquad w_2 = 1$
$\sigma_3 = 3 \qquad w_3 = \frac{1}{9}$

$$\longleftarrow \; w_i = \frac{1}{\sigma_i^2}$$

$$R_{WAV} = \frac{\Sigma w_i R_i}{\Sigma w_i} = \frac{(1 \times 11) + (1 \times 12) + (\frac{1}{9} \times 10)}{1 + 1 + \frac{1}{9}} = 11.42\ \Omega$$

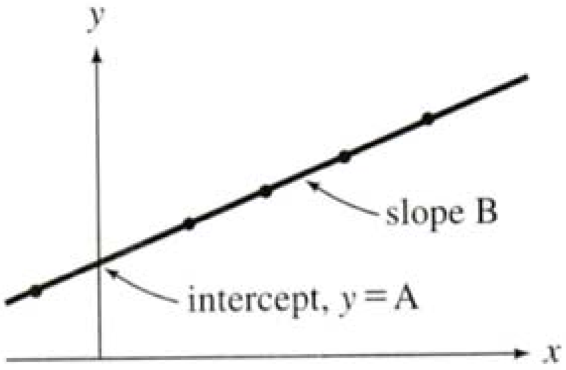
$$\sigma_{WAV} = \frac{1}{\sqrt{\Sigma w_i}} = \frac{1}{\sqrt{1 + 1 + \frac{1}{9}}} = 0.69$$

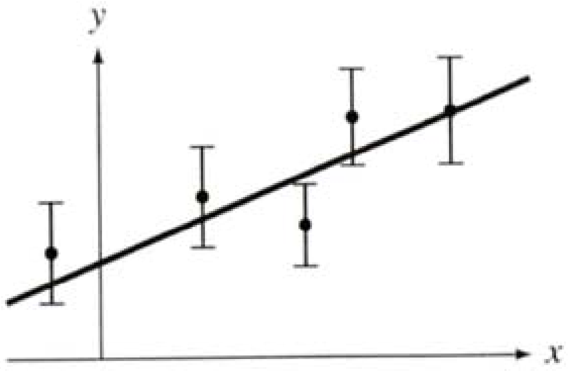
$$\underline{R = 11.4 \pm 0.7\ \Omega}$$

Least-Squares Fitting

consider two variables *x* and *y* that are connected by a linear relation

$$y = A + Bx$$





graphical method of finding the best straight line to fit a series of experimental points

$$\begin{matrix} x_1, x_2, \ldots, x_N \\ y_1, y_2, \ldots, y_N \end{matrix} \longrightarrow \text{find } A \text{ and } B$$

analytical method of finding the best straight line to fit a series of experimental points is called **linear regression** or **the least-squares fit for a line**

Calculation of the Constants *A* and *B*

$$(\text{true value for } y_i) = A + Bx_i$$

$$\text{Prob}_{A,B}(y_1) \propto \frac{1}{\sigma_y} e^{-(y_1 - A - Bx_1)^2/2\sigma_y^2}$$ ← probability of obtaining the observed value of y_1

$$\text{Prob}_{A,B}(y_1,\ldots,y_N) = \text{Prob}_{A,B}(y_1)\cdots\text{Prob}_{A,B}(y_N)$$ ← probability of obtaining the set y_1, … , y_N

$$\propto \frac{1}{\sigma_y^N} e^{-\chi^2/2}$$ ← find maximum of probability

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y^2}$$ ← chi squared – “sum of squares”
find minimum of χ^2
least squares fitting

$$\left|\begin{aligned} \frac{\partial\chi^2}{\partial A} &= \frac{-2}{\sigma_y^2}\sum_{i=1}^{N}(y_i - A - Bx_i) = 0 \\ \frac{\partial\chi^2}{\partial B} &= \frac{-2}{\sigma_y^2}\sum_{i=1}^{N}x_i(y_i - A - Bx_i) = 0 \end{aligned}\right.$$

$$\left|\begin{aligned} &\sum y_i - AN - B\sum x_i = 0 \\ &\sum x_i y_i - A\sum x_i - B\sum x_i^2 = 0 \end{aligned}\right.$$ →

$$\boxed{\begin{aligned} A &= \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta} \\ B &= \frac{N\sum xy - \sum x \sum y}{\Delta} \\ \Delta &= N\sum x^2 - \left(\sum x\right)^2 \end{aligned}}$$

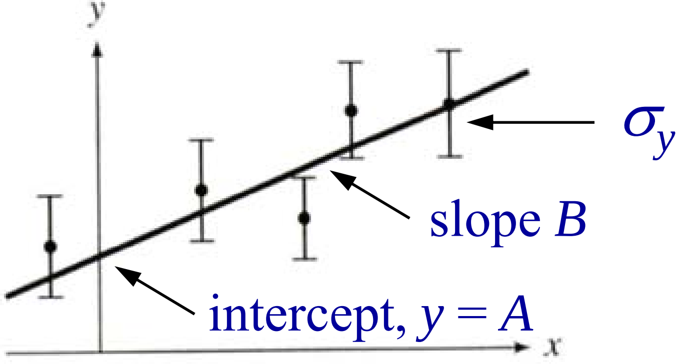
Uncertainties in *y*, *A*, and *B*

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

$$B = \frac{N \sum xy - \sum x \sum y}{\Delta}$$

$$\Delta = N \sum x^2 - \left(\sum x\right)^2$$

$y = A + Bx$



$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} \left(y_i - A - Bx_i\right)^2}$$

uncertainty in the measurement of *y*

$$\sigma_A = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

uncertainties in the constants *A* and *B*

given by error propagation in terms
of uncertainties in $y_1, \ldots, y_N$

Example of Calculation of the Constants *A* and *B*

$$\overset{y}{T} = A + B\,\overset{x}{P}$$

if volume of an ideal gas is kept constant,
its temperature is a linear function of its pressure

i	P_i	T_i
1	65	-20
2	75	17
3	85	42
4	95	94
5	105	127

absolute zero of temperature $A = ?$

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

$$B = \frac{N\sum xy - \sum x \sum y}{\Delta}$$

$$\Delta = N\sum x^2 - \left(\sum x\right)^2$$

$\Sigma P = 425$

$\Sigma P^2 = 37{,}125$

$\Sigma T = 260$

$\Sigma PT = 25{,}810$

$\Delta = N\,\Sigma P^2 - (\Sigma P)^2 = 5{,}000$

$$A = \frac{\Sigma P^2 \Sigma T - \Sigma P \Sigma PT}{\Delta} = -263.35$$

$$B = \frac{N \Sigma PT - \Sigma P \Sigma T}{\Delta} = 3.71$$

$$\sigma_y = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}\left(y_i - A - Bx_i\right)^2}$$

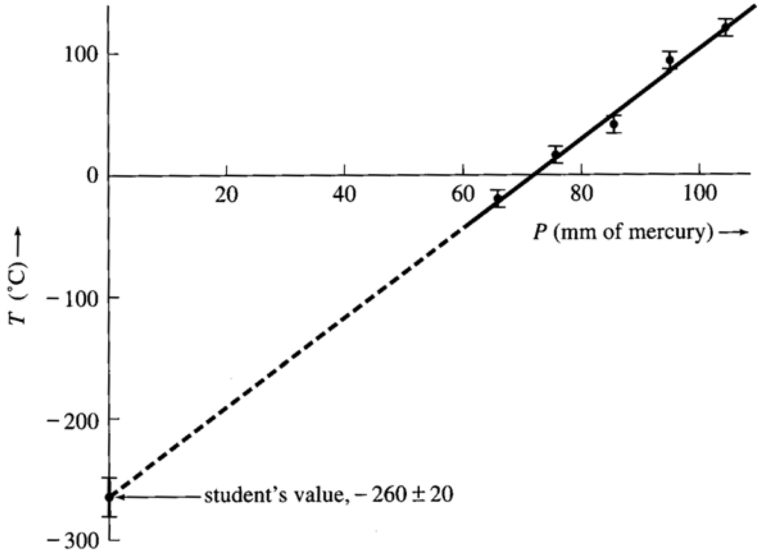
$$\sigma_A = \sigma_y\sqrt{\frac{\sum x^2}{\Delta}}$$

$$\sigma_T = \sqrt{\frac{1}{N-2}\,\Sigma(T_i - A - BP_i)^2} = 6.7$$

$$\sigma_A = \sigma_T\sqrt{\frac{\Sigma P^2}{\Delta}} = 18$$

$$A = -263 \pm 18^0 C$$

absolute zero of
temperature = – 273.15^0 C



Example Problem

Two students measure the radius of a planet and get final answers R_A=25,000±3,000 km and R_B=19,000±2,500 km.
(a) Assuming all errors are independent and random, what is the discrepancy and what is its uncertainty?
(b) Assuming all quantities are normally distributed as expected, what would be the probability that the two measurements would disagree by more than this?
Do you consider the discrepancy in the measurements significant (at the 5% level)?

(a) $$R_A - R_B = 25,000 - 19,000 = \underline{6,000 km}$$

$$\sigma_{R_A - R_B} = \sqrt{\sigma_{R_A}{}^2 + \sigma_{R_B}{}^2} = \sqrt{3,000^2 + 2,500^2} = 3,905 km \rightarrow \underline{4,000 km}$$

$$R_A - R_B = 6,000 \pm 4,000 km$$

(b) $$t = \frac{6,000}{4,000} = 1.5$$

Table A: Probability to be within 1.5σ is 86.64 % ≈ 87 %. Therefore, the probability that the two measurements would disagree by more than this is 100 – 87 = 13 %.
The discrepancy in the measurements is not significant (at the 5% level).

Example Problem

Two students measure the radius of a planet and get final answers
R_A=25,000±3,000 km and R_B=19,000±2,500 km.
The best estimate of the true radius of a planet is the weighted average. Find the best estimate of the true radius of a planet and the error in that estimate.

$$x_{wav} = \frac{w_A x_A + w_B x_B}{w_A + w_B} \qquad w_A = \frac{1}{\sigma_A^2} \qquad w_B = \frac{1}{\sigma_B^2} \qquad \sigma_{wav} = \frac{1}{\sqrt{w_A + w_B}}$$

$$R_{wav} = \frac{\frac{R_A}{\sigma_A{}^2} + \frac{R_B}{\sigma_B{}^2}}{\frac{1}{\sigma_A{}^2} + \frac{1}{\sigma_B{}^2}} = \frac{\frac{25,000}{3,000^2} + \frac{19,000}{2,500^2}}{\frac{1}{3,000^2} + \frac{1}{2,500^2}} = 21,459km \rightarrow \underline{21,500km}$$

$$\sigma_{wav} = \frac{1}{\sqrt{\frac{1}{\sigma_A{}^2} + \frac{1}{\sigma_B{}^2}}} = \frac{1}{\sqrt{\frac{1}{3,000^2} + \frac{1}{2,500^2}}} = 1,921km \rightarrow \underline{1,900km}$$

$$\underline{R_{wav} = 21,500 \pm 1,900km}$$