Week 3

Note Title 1/23/2007

Data rejection

- Is it okay to throw out data?
- Consider a data set

$$3.8, 3.5, 3.9, 3.4, 1.8$$

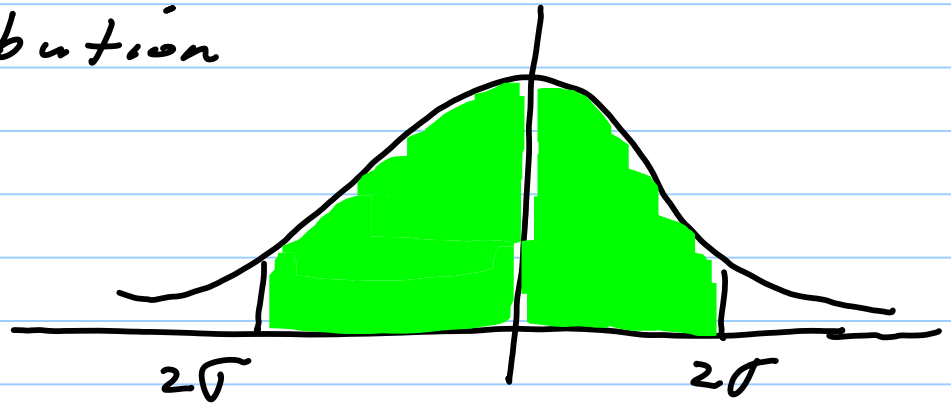
mean

$$\bar{x} = 3.4$$

Standard deviation

$$\sigma_x = .8$$

If we assume a Gaussian Distribution



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

$$= 1 - .95$$

$$= .05$$

Chauvenets Criterion

N measurements

$$x_1, \ldots x_N$$

Consider one measurement x_{sus}

$$t_{sus} = \frac{|x_{sus} - \bar{x}|}{\sigma_x}$$

Define n = (expected number as deviant as x_{sus}

So

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

If $n < .5$ reject

From our data set

$$t_s = \frac{1.8 - 3.4}{.8} = 2$$

Since

$$\text{Prob}(\text{outside } 2\sigma) = .05$$

$$n = 6 \times .05 = .3$$ reject!

New $\bar{x} = \dfrac{3.8 + 3.5 + 3.9 + 3.9 + 3.4}{5}$

$= 3.64$ Compares with 3.4

Is this process a good idea?

- Can you identify a reason
- reject data at your peril
- never a second time

Problem of Combining Measurements

Suppose you have two measurements of a lifetime

$$\text{Meas. } A = 10 \pm .5 \text{ sec.}$$

$$\text{Meas } B = 12 \pm 1 \text{ sec.}$$

Is it meaningful to combine these results?

Probability of value x_A

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

and for value x_B

$$\text{Prob}(x_B) = \frac{1}{\sigma_B} e^{-(x_B - \chi)^2 / 2\sigma_B^2}$$

Probability of x_A and x_B

$$\text{Prob}_x(x_A, x_B) = \text{Prob}_x(x_A)\,\text{Prob}_{x_B}(x_B)$$

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

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

Find minimum

$$\frac{d\chi^2}{d\bar{X}} = 0 = 2\frac{x_A - \bar{X}}{\sigma_A^2} + 2\frac{x_B - \bar{X}}{\sigma_B^2}$$

Solve for $\bar{X}$

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

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

Weighted average becomes

$$\chi_{wAv} = \frac{w_A x_A + w_B x_B}{w_A + w_B}$$

For our example $x_A = 10 \pm .5$ $x_B = 12 \pm 1$

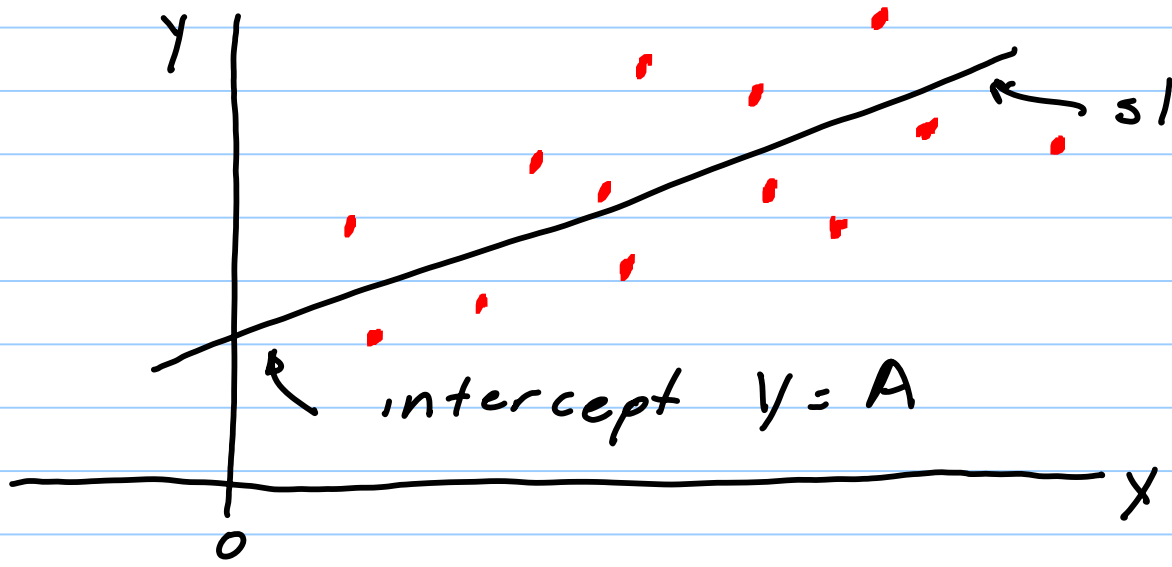
Simple average = 11

Weighted average = 10.7

Karl Friedich Gauss

Least Squares 1795

Least Squares Fitting (Regression)



$$Y = A + BX$$

Chacterise line by A & B

If we knew the constants A & B

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

$$\text{Prob}_{A,B}(y_i) \propto \frac{1}{\sigma_y} e^{-(y_i - A - Bx_i)^2 / 2\sigma_y^2}$$

For complete set of measurements

$$\text{Prob}_{A,B}(y_1 \ldots y_N) = \text{Prob}_{A,B}(y_1) \cdots \text{Prob}_{A,B}(y_N)$$

$$\propto \frac{1}{\sigma_y^N} e^{-\chi^2/2}$$

where exponent

$$\chi^2 = \sum_{i=1}^{N} \frac{(Y_i - A - BX_i)^2}{\sigma_y^2}$$

Minimum value χ^2

$$\frac{\partial \chi^2}{\partial A} = \frac{-2}{\sigma_y^2} \sum_{i=1}^{N} (Y_i - A - BX_i) = 0$$

and

$$\frac{\partial \chi^2}{\partial B} = \frac{-2}{\sigma_y^2} \sum_{i=1}^{N} (Y_i - A - BX_i) = 0$$

Two equations

$$AN + B\Sigma x_i = \Sigma y_i$$

$$A\Sigma x_i + B\Sigma x_i^2 = \Sigma x_i y_i$$

Solving

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

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

$$\Delta = N\Sigma x^2 - (\Sigma x)^2$$

Uncertainty in Measurements of y

- good guess

$$\sigma_y = \sqrt{\frac{1}{N}\sum (y_i - A - Bx_i)^2}$$

- better answer

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

Covariance.

$z_i = z(x_i, y_i)$ $\quad i = 1 \cdots N$

N data points $(x_1, y_1) \cdots (x_N, y_N)$

mean $\bar{z}$

standard deviation σ_z

We approximate

$z_i = z(x_i, y_i)$

$\approx z(\bar{x}, \bar{y}) + \frac{\partial z}{\partial x}(x_i - \bar{x})$

$+ \frac{\partial z}{\partial y}(y_i - \bar{y})$

$$\frac{\partial}{\partial x} = 2$$

$$g(\bar{x}, \bar{y}) = 2$$

Since $\sum_{i=1}^{n} (x_i - \bar{x}) = 0 = \sum_{i=1}^{n} (y_i - \bar{y}) = 0$

$$\left[(y_i - \bar{y}) \frac{\partial^2}{\partial x^2} + (x_i - \bar{x}) \frac{\partial^2}{\partial y} \right] g(\bar{x}, \bar{y}) + \frac{1}{n} \sum_{i=1}^{n} \frac{N}{1} =$$

$$\sum_{i=1}^{n} \frac{N}{1} = 2$$

$$\frac{N}{n} \geq 2$$

Hence proved

$$2\left[(y_i-\bar{y})\frac{\partial z}{\partial y}\right]$$

$$+(x_i-\bar{x})\frac{\partial z}{\partial x}\Big]\frac{N}{T}=\Delta_2$$

Then

$$(y_i-\bar{y})\frac{\partial z}{\partial y}+(x_i-\bar{x})\frac{\partial z}{\partial x}\approx z_i$$

Since

$$\frac{1}{2}\sum(z_i-\bar{z})^2=\frac{N}{T}=\Delta_2$$

$$\sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 + 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial y}\sigma_{xy}$$

Then

$$\sigma_{xy} = \frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y})$$

Define Covariance

$$\sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2 \frac{1}{N}\sum (x_i - \bar{x})^2 + \left(\frac{\partial z}{\partial y}\right)^2 \frac{1}{N}\sum (y_i - \bar{y})^2$$

$$+ 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial y}\frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y})$$

Correlation Coefficient

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$|r| \sim 1$ strong correlation

$|r| \sim 0$ weak correlation

—— Simulation ——

Poisson Distribution

- event distribution

ν number of counts in given interval

$\mu = \bar{\nu}$ mean number of counts in that interval

$$\text{Prob}(\nu \text{ counts in a definite interval}) = P_\mu(\nu)$$

$$= e^{-\mu} \frac{\mu^\nu}{\nu!}$$

$\bar{V} = \mu$ = average number of counts if we repeat experiment many times

μ = rate x time = RT

— Poisson Simulator —

Current Example from Physics

- Proton decay

Consider N protons

$$N = N_0 e^{-\lambda t}$$

where $\lambda = \frac{1}{\tau} \approx 10^{-33}/\text{year}$

$$e^{-\lambda t} \approx 1 - \lambda t$$

so

$$N \approx N_0 (1 - \lambda t)$$

Super Kamiokanda Neutrino Detector

Volume $\approx 7.5 \times 10^{33}$ protons

Then

$$N_0 - N = N_0 \lambda t = (7.5 \times 10^{33} \times 10^{-33})$$

$$= 7.5$$

40% of tank covered by detectors so we expect to see ~ 3 decays

Thus far no decays seen

Probability of no observations

$$P_{\mu}(\nu) = e^{-\mu} \frac{\mu^{\nu}}{\nu!}$$

$$P_{3}(0) = \frac{e^{-3} 3^{0}}{0!} \approx .05$$

Suggests that 10^{33} years is to short!

Properties of a Poisson Distribution

- Standard deviation

$$\sigma_\nu^2 = \overline{(\nu - \bar{\nu})^2}$$

or

$$\sigma_\nu^2 = \bar{\nu}_2 - (\bar{\nu})^2$$

and

$$\bar{\nu}^2 = \mu^2 + \mu$$

$$\sigma_\nu = \sqrt{\mu}$$

Thus if we make a measurement of a number of events in time t the mean count is then

$$\gamma \pm \sqrt{\gamma}$$