# Variance and Co-variance

- Variance of *X+Y* depends on how *X* and *Y* co-vary:

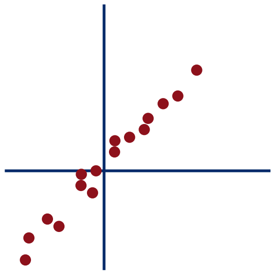
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}(X,Y) \equiv \left\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \right\rangle$$

$$\begin{aligned}
\mathrm{Var}(X+Y) &\equiv \left\langle [X + Y - \langle X + Y \rangle]^2 \right\rangle \\
&= \left\langle [X + Y - \langle X \rangle - \langle Y \rangle]^2 \right\rangle \\
&= \left\langle [(X - \langle X \rangle) + (Y - \langle Y \rangle)]^2 \right\rangle \\
&= \left\langle (X - \langle X \rangle)^2 + (Y - \langle Y \rangle)^2 + 2(X - \langle X \rangle)(Y - \langle Y \rangle) \right\rangle \\
&= \left\langle (X - \langle X \rangle)^2 \right\rangle + \left\langle (Y - \langle Y \rangle)^2 \right\rangle + 2\left\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \right\rangle \\
&= \quad \mathrm{Var}(X) \quad + \quad \mathrm{Var}(Y) \quad + \quad 2\,\mathrm{Cov}(X,Y)
\end{aligned}$$

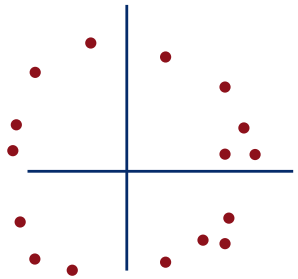
# Co-variance vs Independence

- **Cov > 0**

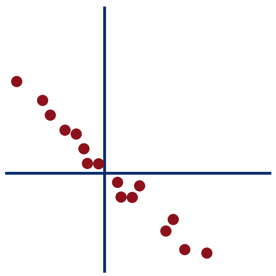


- **Cov = ?**

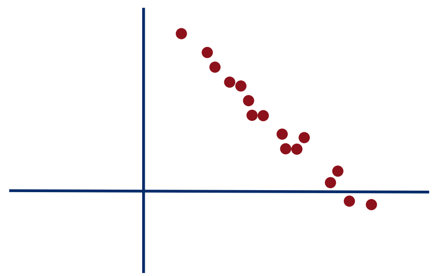
**Independent?**



- **Cov < 0**



- **Cov = ?**

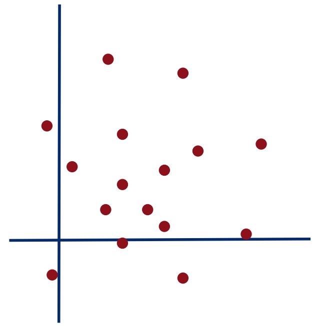


# Practice !

$$X = 1 \pm 1$$

$$Y = 1 \pm 1$$

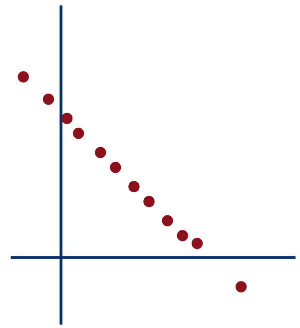
$$X + Y = ? \pm ?$$



$$X = 1 \pm 1$$

$$Y = 2 - X$$

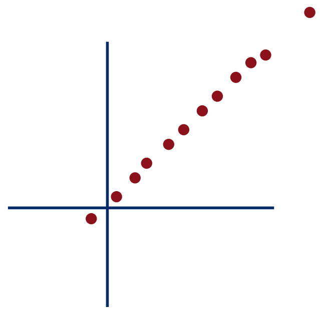
$$X + Y = ? \pm ?$$



$$X = 1 \pm 1$$

$$Y = X$$

$$X + Y = ? \pm ?$$



$$X = 1 \pm 1 \qquad Y = 2 \pm 1 \qquad \mathrm{Cov}[X,Y] = 0 \qquad a = 2 \qquad b = 1$$

$$Z = aX + bY \qquad \langle Z \rangle = ? \qquad \mathrm{Var}(Z) = ?$$

# Linear Transformations

- Scale and add any number of random variables:

$$\left\langle \sum_i a_i X_i \right\rangle = \sum_i a_i \langle X_i \rangle \qquad \mathrm{Var}\left[\sum_i a_i X_i\right] = \sum_{i,j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

Or, in terms of the (symmetric) **Co-variance Matrix**:

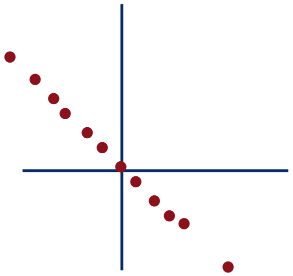
$$\mathrm{Var}\left[\sum_i a_i X_i\right] = \sum_{i,j} a_i C_{ij} a_j$$

$$\mathrm{Var}\left[\begin{pmatrix} a_i & \ldots & a_N \end{pmatrix} \begin{pmatrix} X_1 \\ \ldots \\ X_N \end{pmatrix}\right] = \begin{pmatrix} a_i & \ldots & a_N \end{pmatrix} \begin{pmatrix} C_{11} & \ldots & C_{1N} \\ \ldots & \ldots & \ldots \\ C_{N1} & \ldots & C_{NN} \end{pmatrix} \begin{pmatrix} a_1 \\ \ldots \\ a_N \end{pmatrix}$$

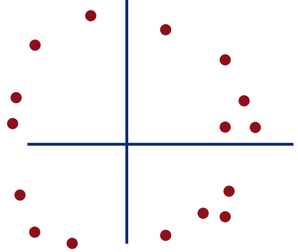
# Correlation Coefficient *R( X, Y )*

- **Correlation coefficient**:

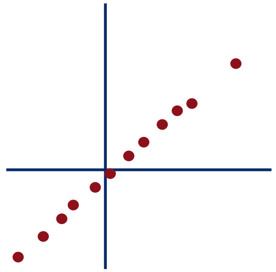
$$R(X,Y) \equiv \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$



*R = -1*



*R = 0*



*R = +1*

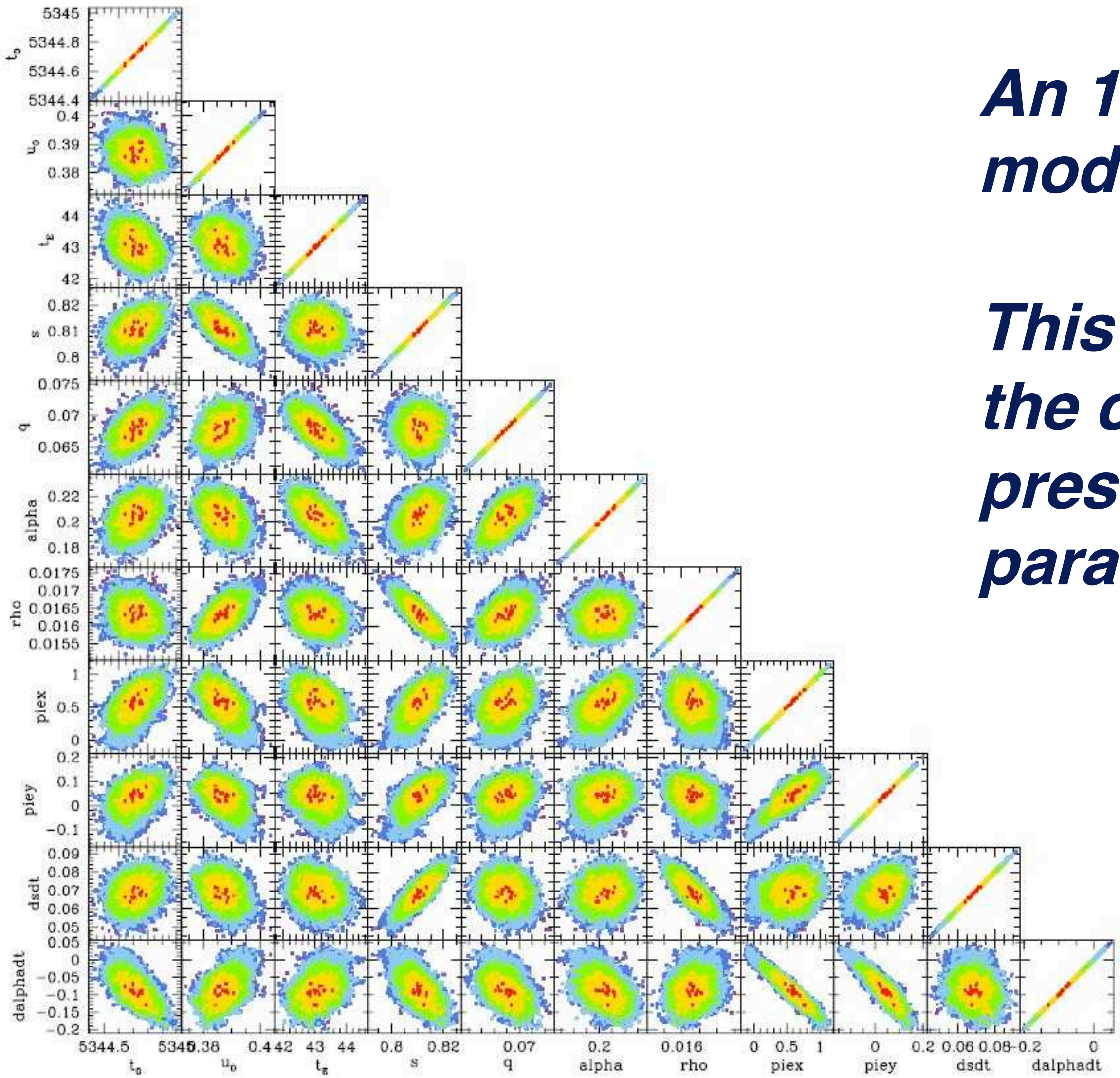
- **Correlation matrix**:

$$R_{ij} \equiv \frac{\mathrm{Cov}(X_i,X_j)}{\sigma(X_i)\,\sigma(X_j)} = \begin{bmatrix} 1 & . & . & . \\ . & 1 & . & . \\ . & . & 1 & . \\ . & . & . & 1 \end{bmatrix}$$

- **Variance:**

$$\mathrm{Var}\left[\sum_i a_i X_i\right] = \sum_i \sum_j a_i a_j \sigma(X_i)\sigma\left(X_j\right) R_{ij}$$

# Example:  Correlation Matrix



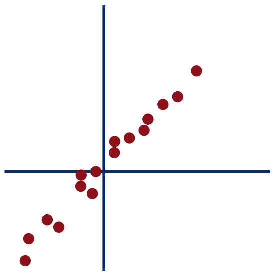

***An 11-parameter model fitted to data.***

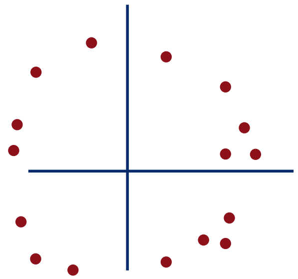
***This matrix shows the correlations present among the 11 parameters.***

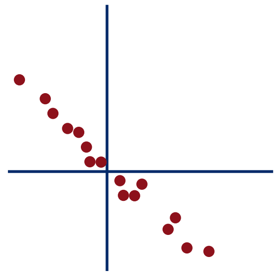
# Review: Correlation vs Independence

- **R > 0**
- **R = 0**
- ***R < 0***

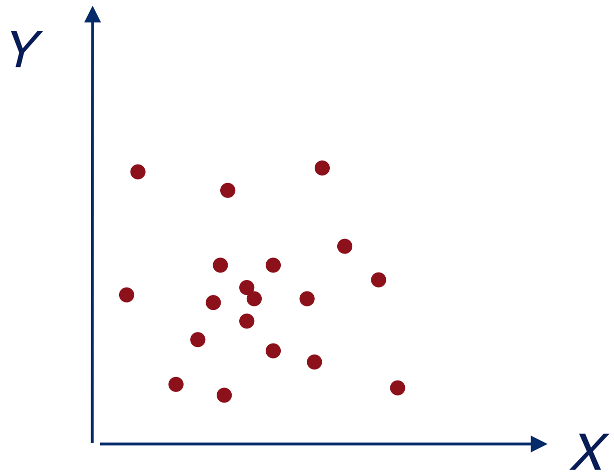
**Dependent :**





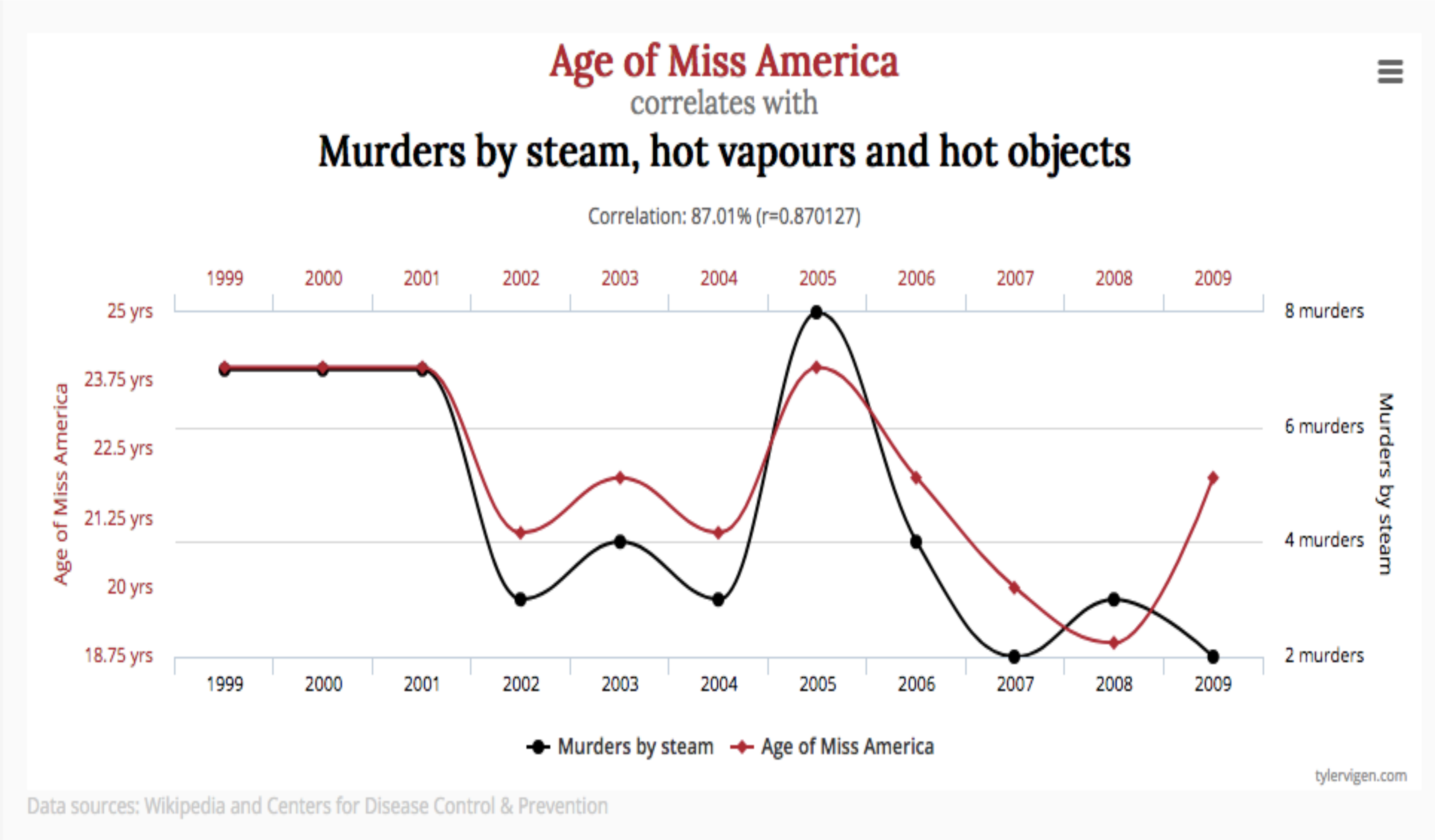


**Independent :**




$$f(X,Y) = f(X)\,f(Y)$$

# Example of a Spurious Correlation



# 1: Beware Spurious Correlations

- Two variables may appear to be strongly correlated.
- But, can be spurious if you look at many variables, to find the strongest correlations, then pretend you only looked at those.

# 2 : Correlation is not Causation

- Correlation of 2 variables does not mean that one causes the other. Both could be side effects of something else.

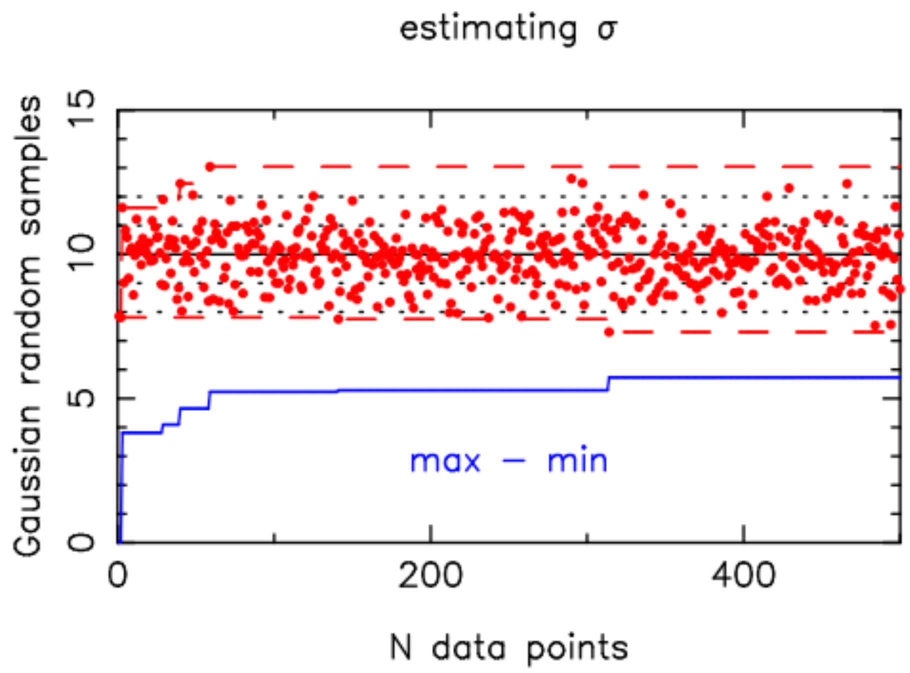
# Misleading Significance Claims

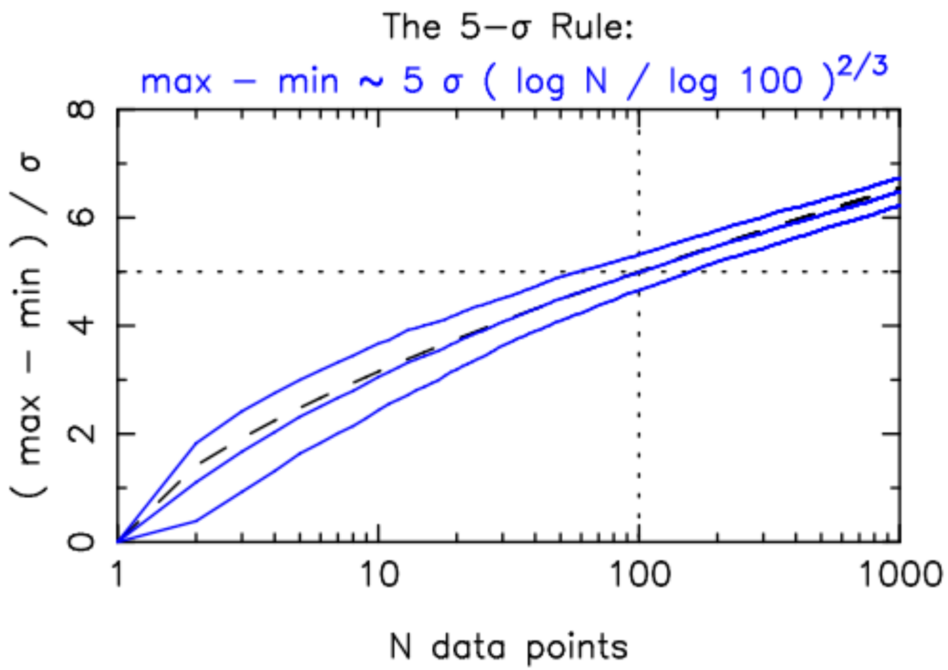
If we look at 100 points, we typically find 2 that are 5-sigma apart.

If we pull out those 2

(and omit the others)

we can't honestly claim to have a 5-sigma result.



# Review: Algebra of Random Variables

$$\langle a \rangle = a \qquad \mathrm{Var}[a] = 0$$

$$\langle a X \rangle = a \langle X \rangle \qquad \mathrm{Var}[a X] = a^2 \, \mathrm{Var}[X]$$

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle \qquad \mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X, Y]$$

Co - variance :

$$\mathrm{Cov}[X, Y] \equiv \Big\langle \big( X - \langle X \rangle \big)\big( Y - \langle Y \rangle \big) \Big\rangle \qquad \mathrm{Var}[X] \equiv \mathrm{Cov}[X, X]$$

Linear transformations :

$$\left\langle \sum_i a_i X_i \right\rangle = \sum_i a_i \langle X_i \rangle \qquad \mathrm{Var}\left[ \sum_i a_i X_i \right] = \sum_i \sum_j a_i \, a_j \, \sigma_i \, \sigma_j \, R_{ij}$$

Correlation Matrix :

$$R_{ij} \equiv \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_\mathrm{i} \, \sigma_\mathrm{j}} \qquad \sigma_\mathrm{i} \equiv \sigma(X_i)$$

# Practice the “fuzzy” algebra of random variables

$$6\,(1 \pm 1) \quad =$$

$$(1 \pm 1) \quad + \quad (2 \pm 2) \quad =$$

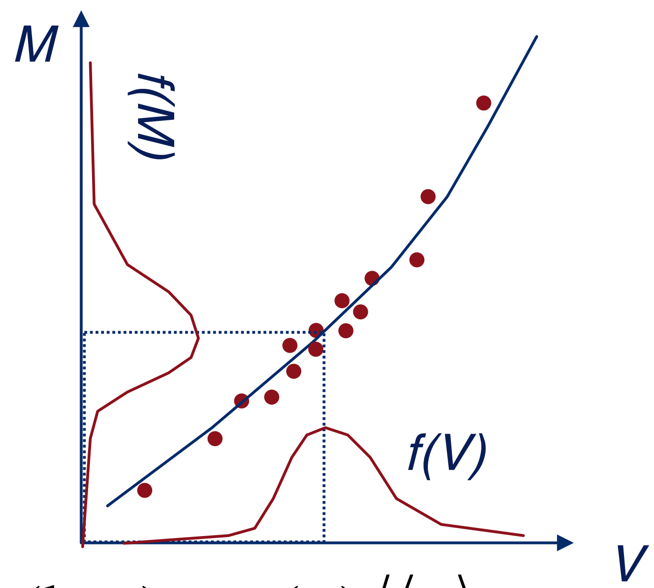
$$(1 \pm 2) \quad - \quad (2 \pm 2) \quad =$$

*Practice until this becomes automatic …*

# Functions of Random Variables

Often what we can measure is not what we are most interested in!

Example: mass of binary-star system:




$$M = \frac{V^2 a}{G} = \frac{V^3 P}{2\pi G}$$

We want *M,* but can only measure *V* and *P*.

*P* = accurate, but *V* usually less certain.

What is the uncertainty in *M* ?

For power-laws: $\ln M = 3\ln V + \ln P + \text{const.} \quad \sigma(\ln x) \approx \sigma(x) / \langle x \rangle$

$$\left( \frac{\sigma_M}{\langle M \rangle} \right)^2 \approx \left( 3 \frac{\sigma_V}{\langle V \rangle} \right)^2 + \left( \frac{\sigma_P}{\langle P \rangle} \right)^2$$

(valid for ***small*** and ***independent*** errors in *V* and *P*).

***How do error bars propagate through non-linear functions?***

# Functions of a Random Variable

$$Y = y(X) \qquad \frac{dY}{dX} = y'(X)$$

**Conserve probability**:

$$d(\text{Prob}) = f(Y)\,|\text{d}Y| = f(X)\,|\text{d}X|$$

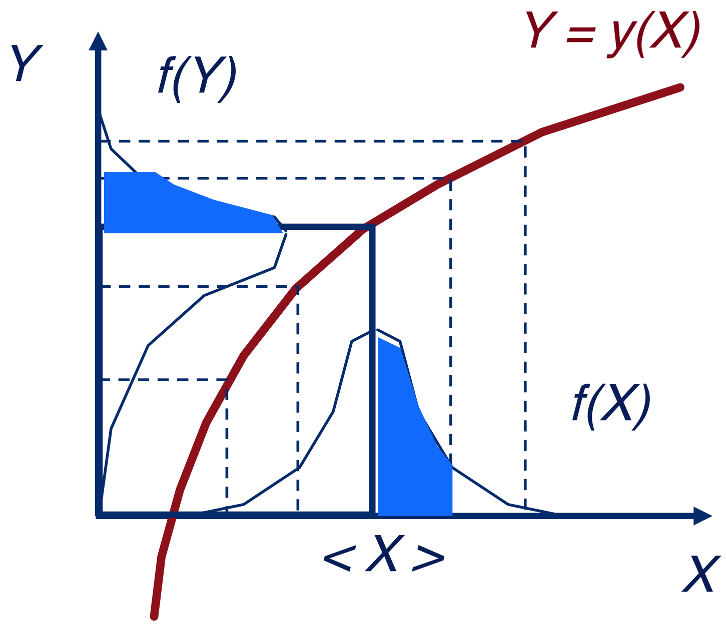
$$f(Y) = f(X)\left|\frac{dX}{dY}\right| = \frac{f(X)}{|y'(X)|}$$

mean value (biased)

$$\langle Y \rangle = y(\langle X \rangle) + \frac{1}{2}\,y''(\langle X \rangle)\,\sigma_X^2 + \quad \ldots$$

standard deviation (stretched)

$$\sigma_Y = \sigma_X \left|\frac{dy}{dx}\right|_{X=\langle X \rangle} + \quad \ldots$$




Negative curvature:

Long tail for $Y < y(<X>)$

*Bias:* $<Y> < y(<X>)$.

Median is not biased:

Med$(Y) = y($Med$(X))$

# Examples of Non-linear Transformations

Spectral Energy Distributions: per unit **wavelength** (erg cm$^{-2}$s$^{-1}$ A$^{-1}$), or per unit **frequency** (erg cm$^{-2}$ s$^{-1}$ Hz$^{-1}$)

$$f_\nu(\lambda)\,|\mathrm{d}\nu| = f_\lambda(\lambda)\,|\mathrm{d}\lambda|$$

$$\nu = \frac{c}{\lambda} \qquad \mathrm{d}\nu = -\frac{c}{\lambda^2}\mathrm{d}\lambda \quad \Rightarrow f_\nu(\lambda) = \left|\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}\right| f_\lambda(\lambda) = \frac{\lambda^2}{c} f_\lambda(\lambda)$$

Converting a **flux** to a **magnitude**:

- Measure Flux: Gaussian distribution: $f \sim G(<f>, \sigma_f^2)$
- Nonlinear transformation induces a bias:

$$m = m_0 - 2.5 \log f$$

$$\langle m \rangle = m_0 - 2.5 \log\langle f \rangle + a\,\sigma_m^2$$

- PROBLEM: evaluate $a$, $\sigma_m$ in terms of $<f>$, $\sigma_f$.