Summarizing Variation Matrix Algebra & Mx

Michael C Neale PhD
Virginia Institute for Psychiatric and Behavioral Genetics
Virginia Commonwealth University

19th International workshop on Methodology Twin and Family Studies

Overview

- Mean/Variance/Covariance
  - Calculating
  - Estimating by ML

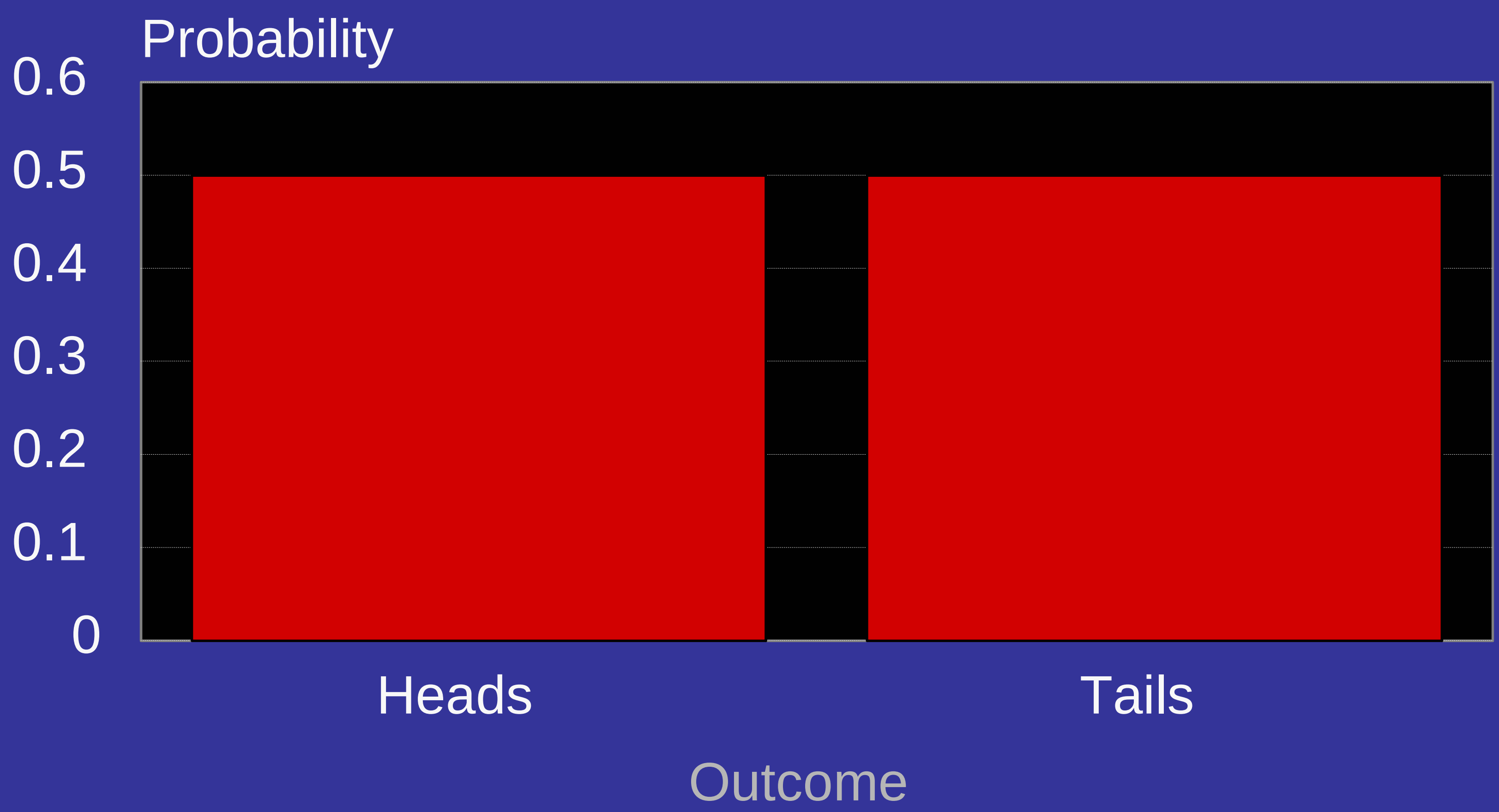
- Matrix Algebra
- Normal Likelihood Theory
- Mx script language

Computing Mean

- Formula $\sum(x_i)/N$
- Can compute with
  - Pencil
  - Calculator
  - SAS
  - SPSS
  - Mx

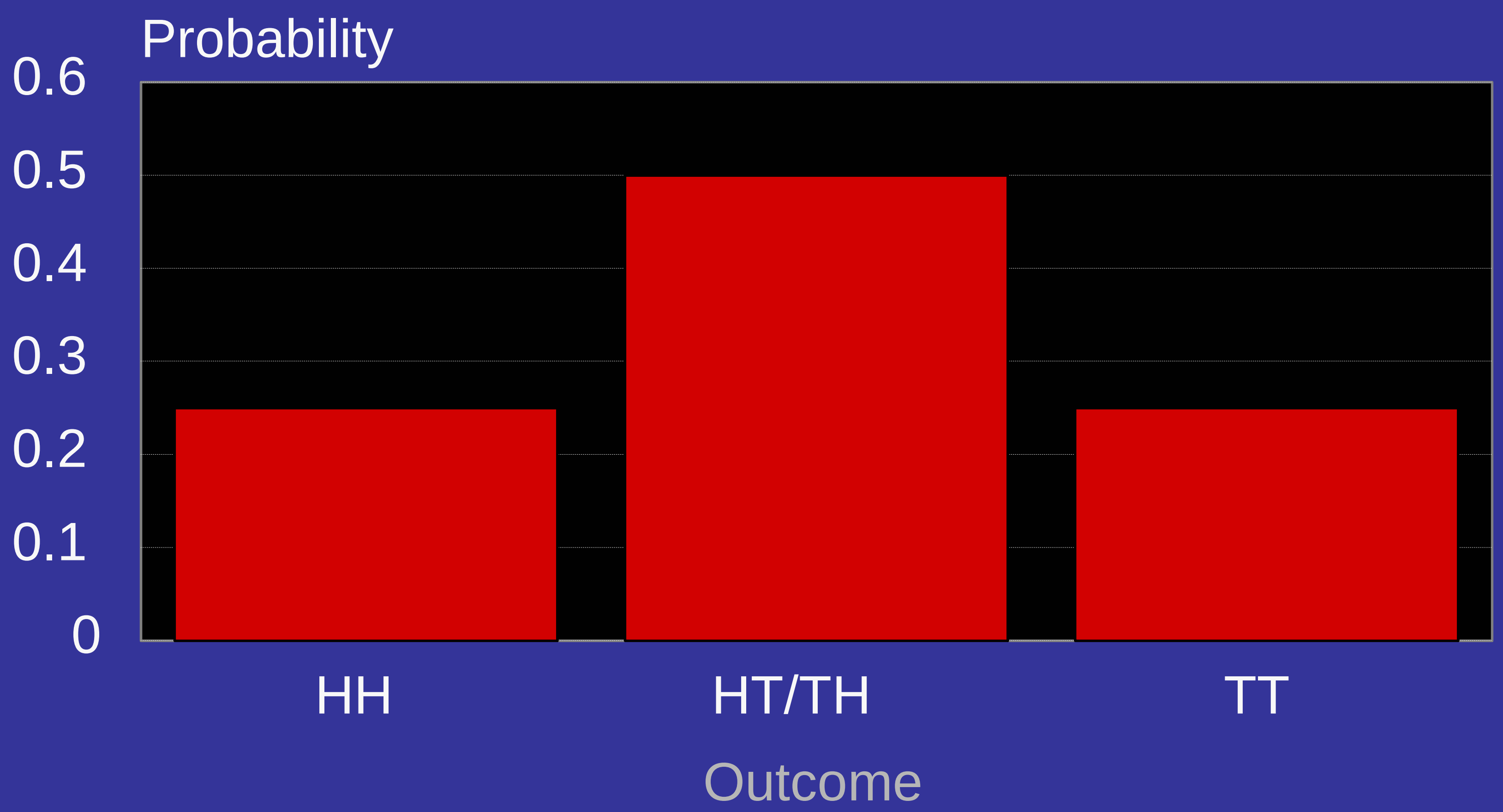
One Coin toss

2 outcomes



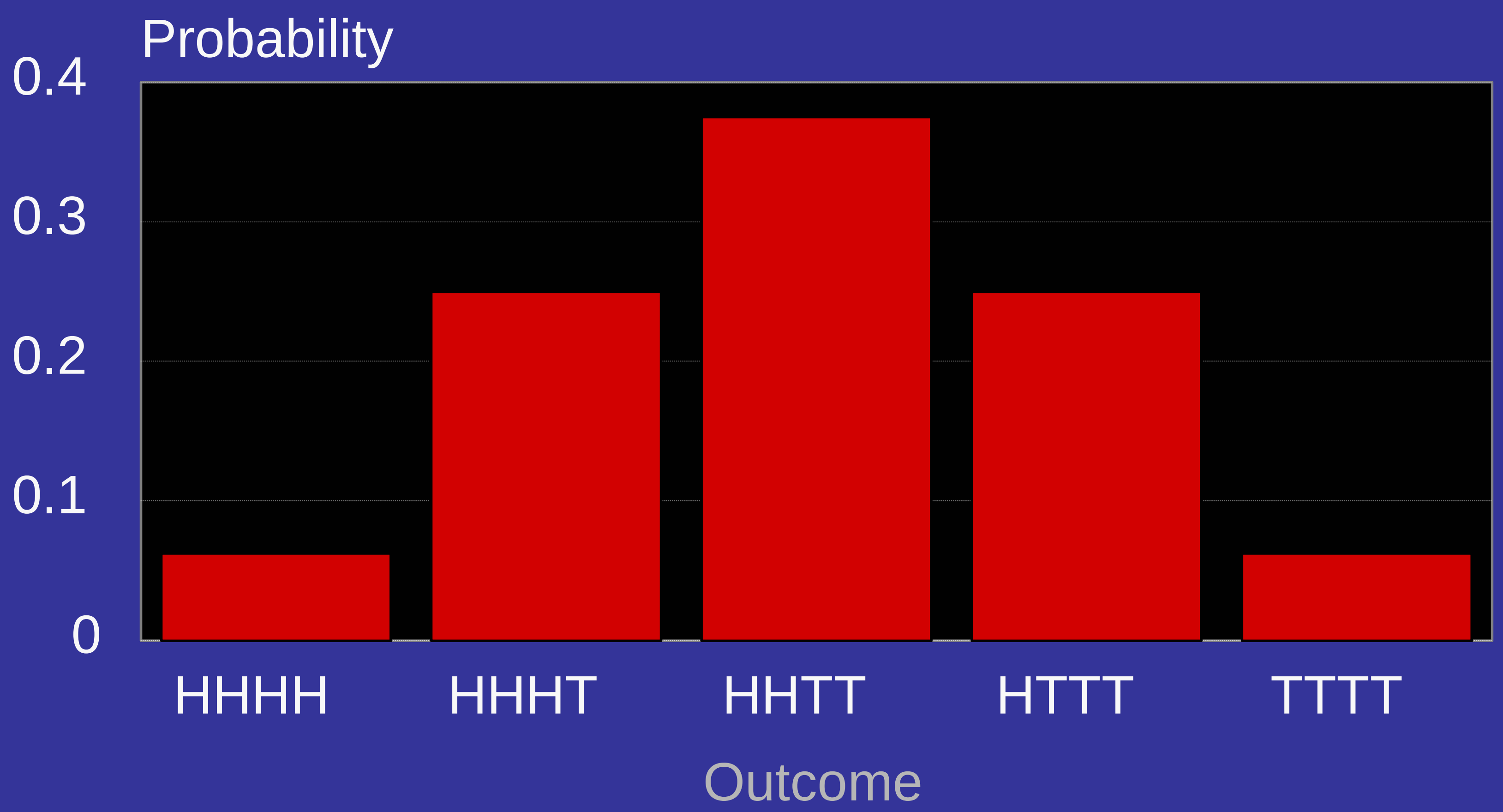
Two Coin toss

3 outcomes



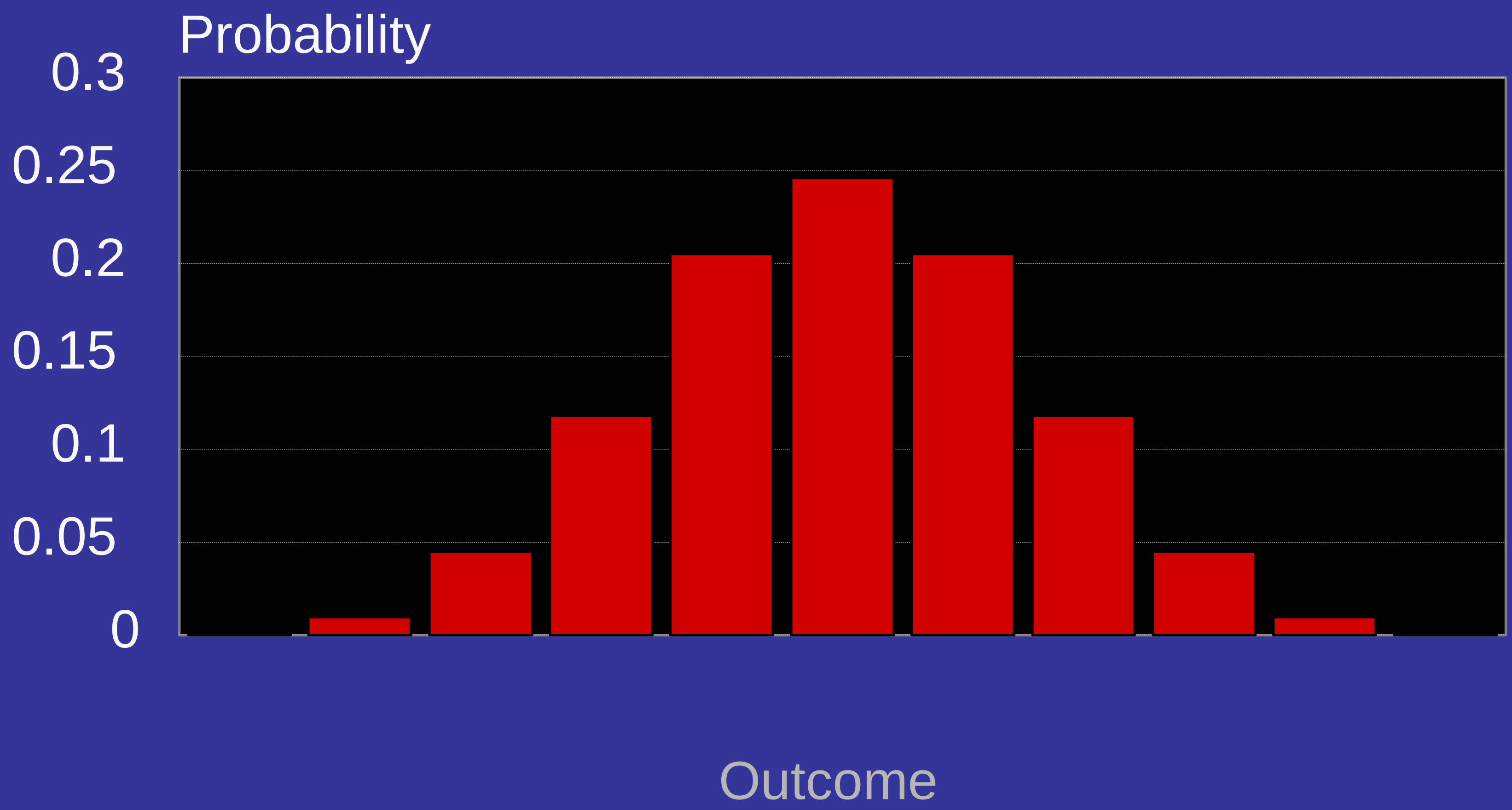
Four Coin toss

5 outcomes



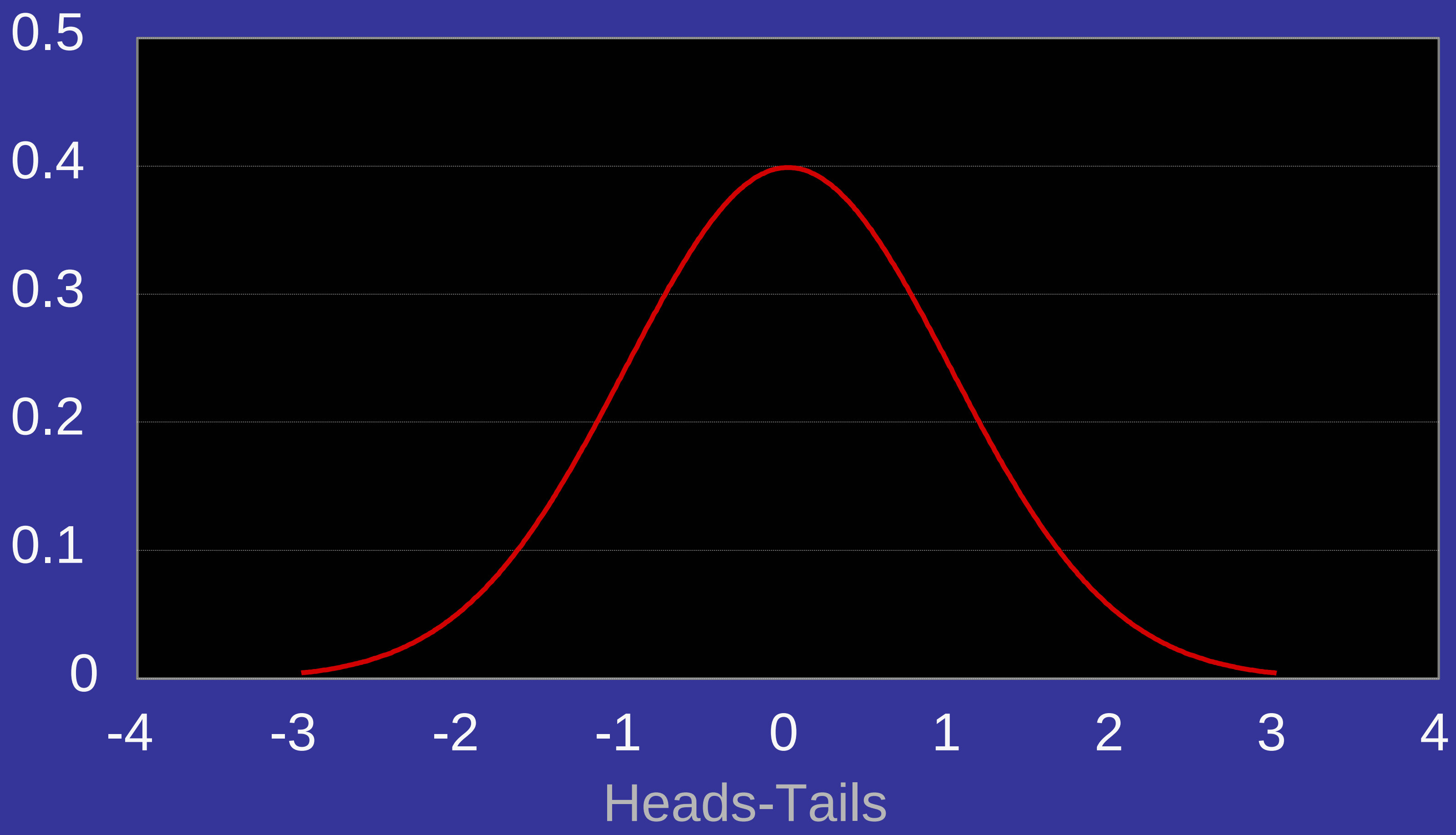
Ten Coin toss

9 outcomes



Fort Knox Toss

Infinite outcomes



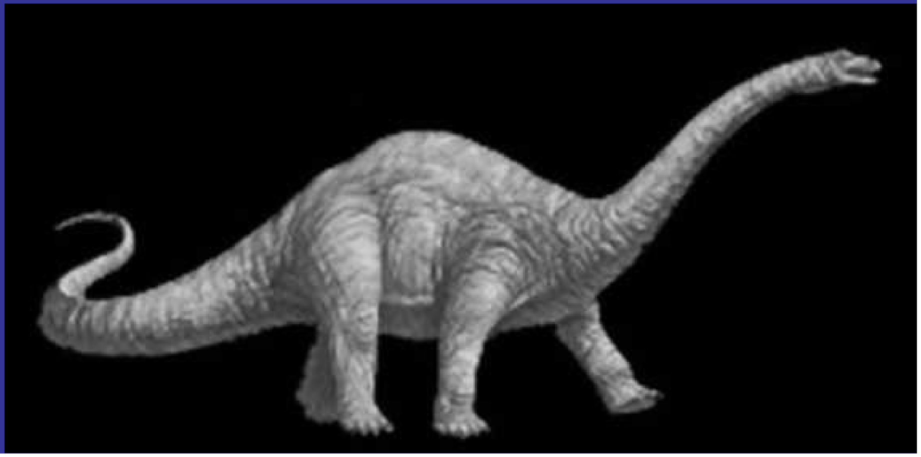
De Moivre 1733 Gauss 1827

Dinosaur (of a) Joke

- Elk:
  The Theory by A. Elk brackets Miss brackets. My theory is along the following lines.

- Host:
  Oh God.

- Elk:
  All brontosauruses are thin at one end, much MUCH thicker in the middle, and then thin again at the far end.



Pascal's Triangle

Frequency	Probability
1	1/1
1 1	1/2
1 2 1	1/4
1 3 3 1	1/8
1 4 6 4 1	1/16
1 5 10 10 5 1	1/32
1 6 15 20 15 6 1	1/64
1 7 21 35 35 21 7 1	1/128

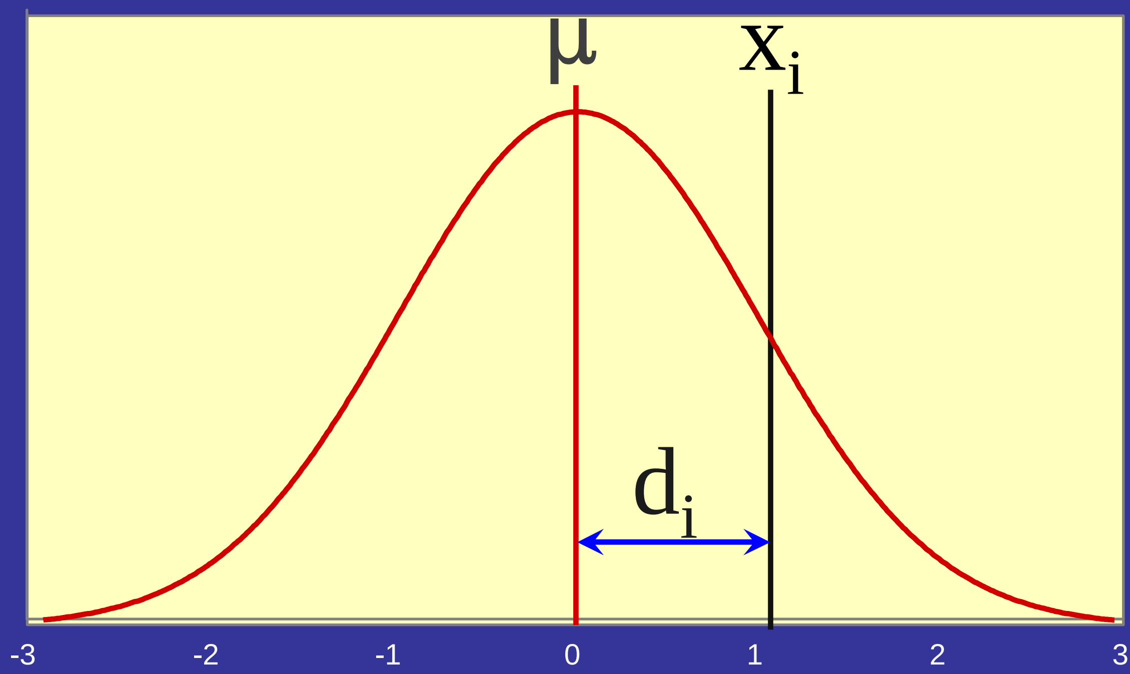
Pascal's friend Chevalier de Mere 1654; Huygens 1657; Cardan 1501-1576

Variance

- Measure of Spread
- Easily calculated
- Individual differences

Average squared deviation

Normal distribution



$$\text{Variance} = \Sigma\, d_i^2/N$$

Measuring Variation

Weighs & Means

- Absolute differences?
- Squared differences?
- Absolute cubed?
- Squared squared?

Measuring Variation

Ways & Means

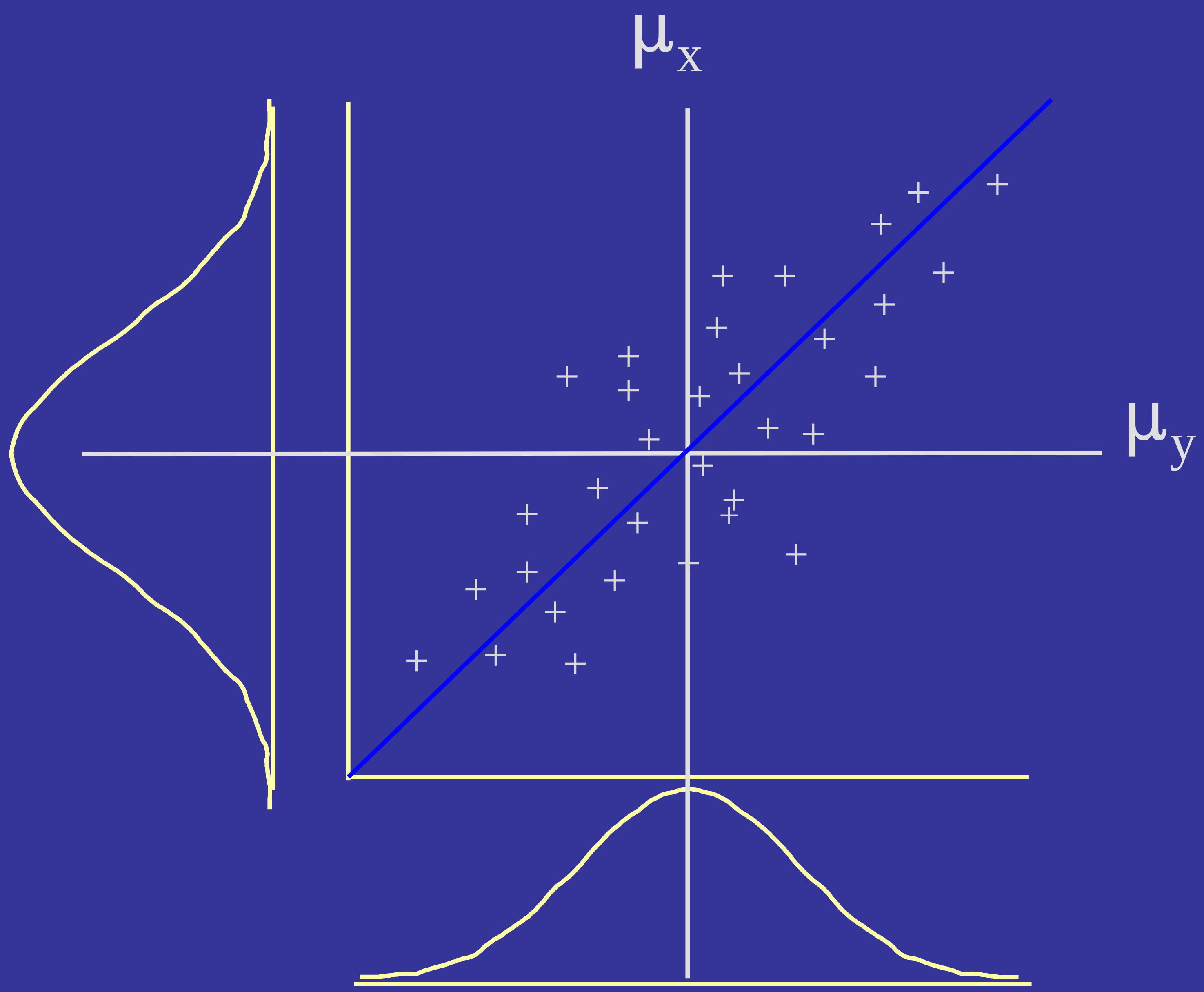
- Squared differences

Fisher (1922) Squared has minimum variance under normal distribution

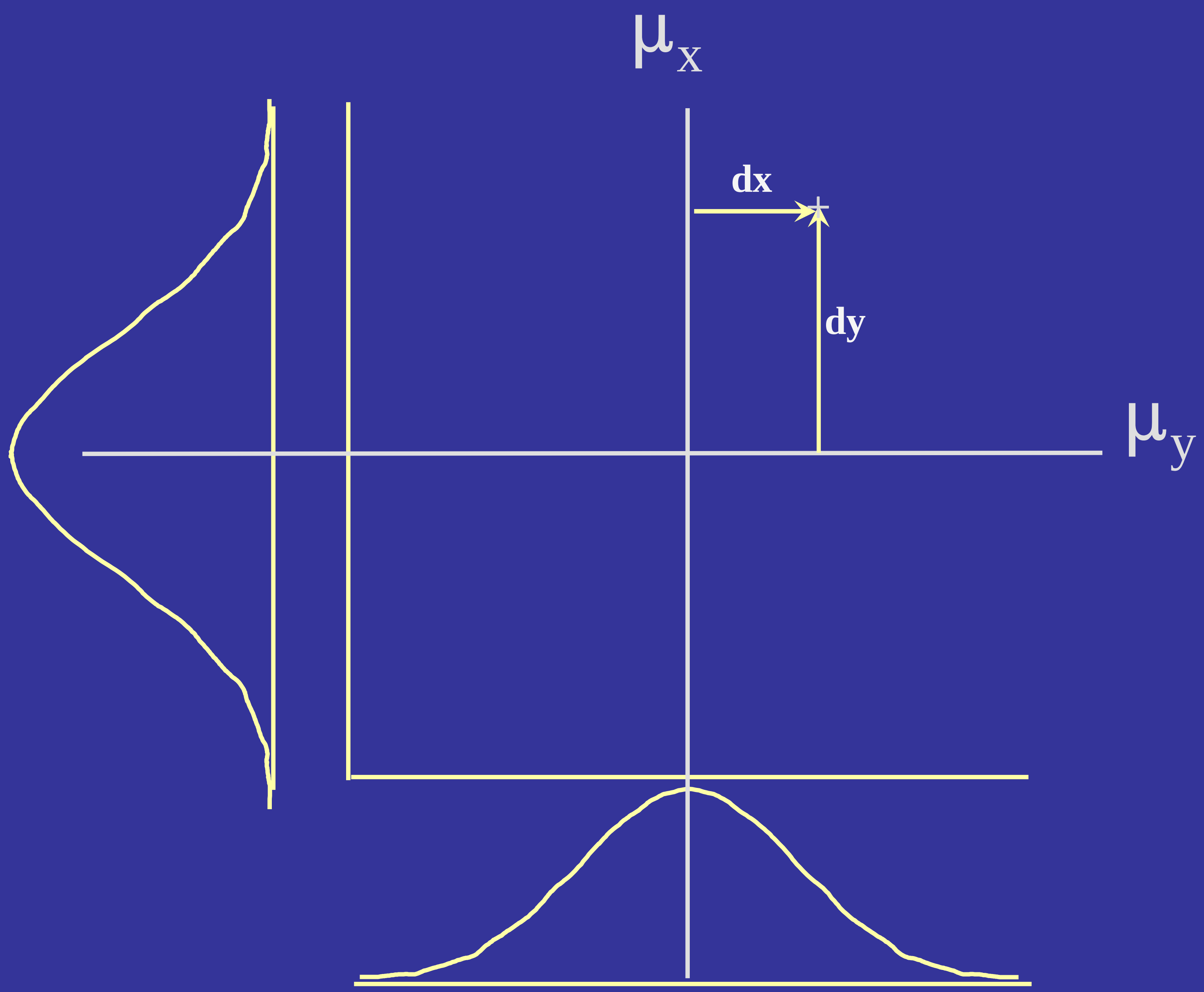
Covariance

- Measure of association between two variables
- Closely related to variance
- Useful to partition variance

Deviations in two dimensions



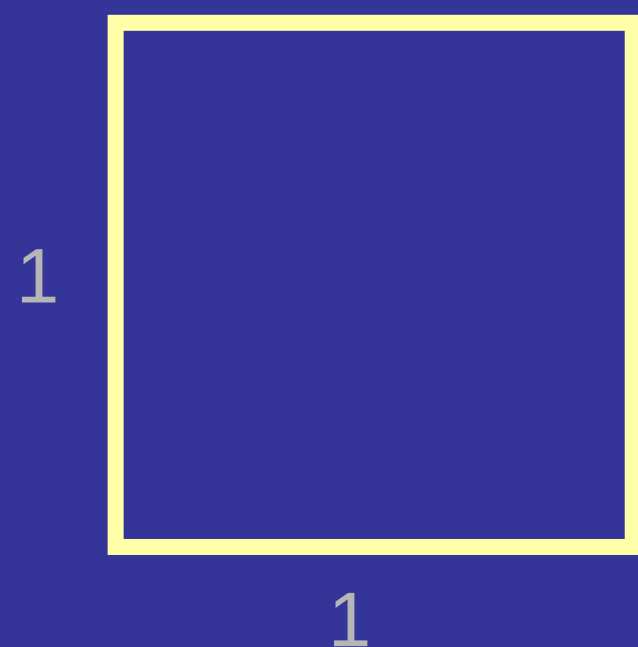
Deviations in two dimensions



Measuring Covariation

Concept: Area of a rectangle

- A square, perimeter 4
- Area 1



Measuring Covariation

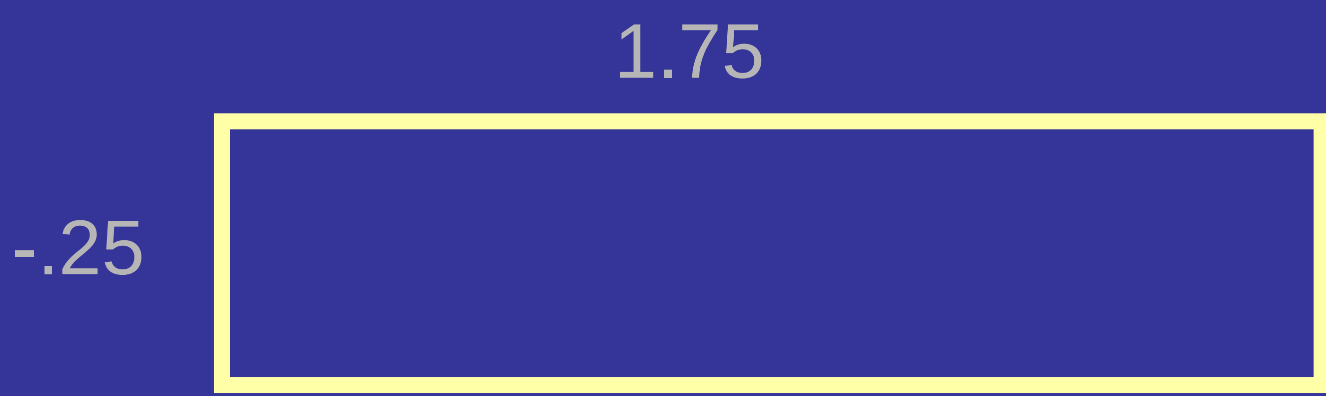
Concept: Area of a rectangle

- A skinny rectangle, perimeter 4
- Area .25*1.75 = .4385

Measuring Covariation

Concept: Area of a rectangle

- Points can contribute negatively
- Area -.25*1.75 = -.4385



Measuring Covariation

Covariance Formula: Average cross-product of deviations from mean

$$\sigma_{xy} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}$$

Correlation

- Standardized covariance
- Lies between -1 and 1

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 * \sigma_y^2}}$$

Summary

Formulae for sample statistics; i=1…N observations

$$\mu = (\Sigma x_i)/N$$

$$\sigma_x^2 = \Sigma (x_i - \mu_x)^2 / (N)$$

$$\sigma_{xy} = \Sigma (x_i - \mu_x)(y_i - \mu_y) / (N)$$

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \ \sigma_x^2}}$$

Variance covariance matrix

Several variables

$$\begin{bmatrix} Var(X) & Cov(X,Y) & Cov(X,Z) \\ Cov(X,Y) & Var(Y) & Cov(Y,Z) \\ Cov(X,Z) & Cov(Y,Z) & Var(Z) \end{bmatrix}$$

Variance covariance matrix

Univariate Twin Data

$$\begin{bmatrix} \text{Var(Twin1)} & \text{Cov(Twin1,Twin2)} \\ \text{Cov(Twin2,Twin1)} & \text{Var(Twin2)} \end{bmatrix}$$

Only suitable for complete data
Good conceptual perspective

Conclusion

- Means and covariances
- Basic input statistics for “Traditional SEM”
- Easy to compute
- Can use raw data instead

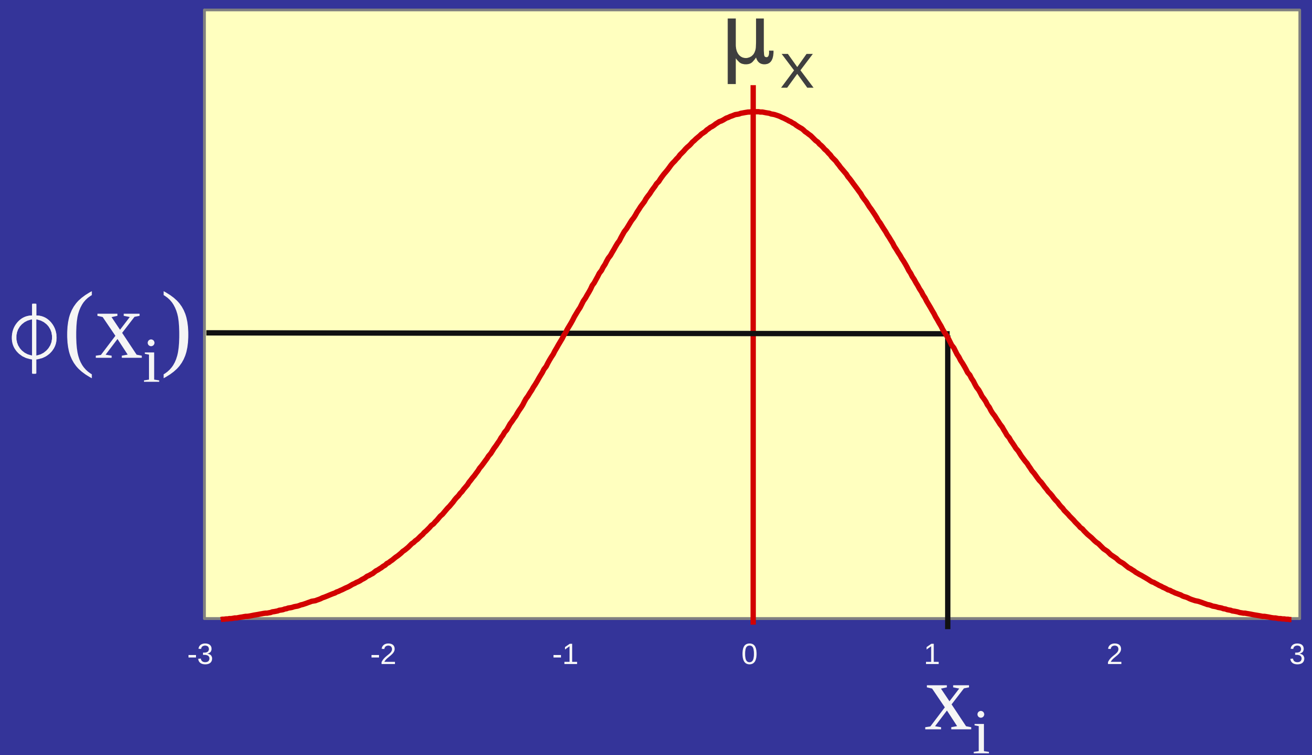
Likelihood computation

Calculate height of curve

- Univariate - height of normal pdf
  - $\phi(x) =$
  - $(2\Pi\sigma^2)^{-.5}\, e^{-.5((x_i - \mu)\hat{}2)/\sigma\hat{}2)}$
- Multivariate - height of multinormal pdf
  - $|2\Pi\Sigma|^{-n/2}\, e^{-.5((\mathbf{x_i} - \mu)\Sigma^{-1}(\mathbf{x_i} - \mu)')}$

Height of normal curve

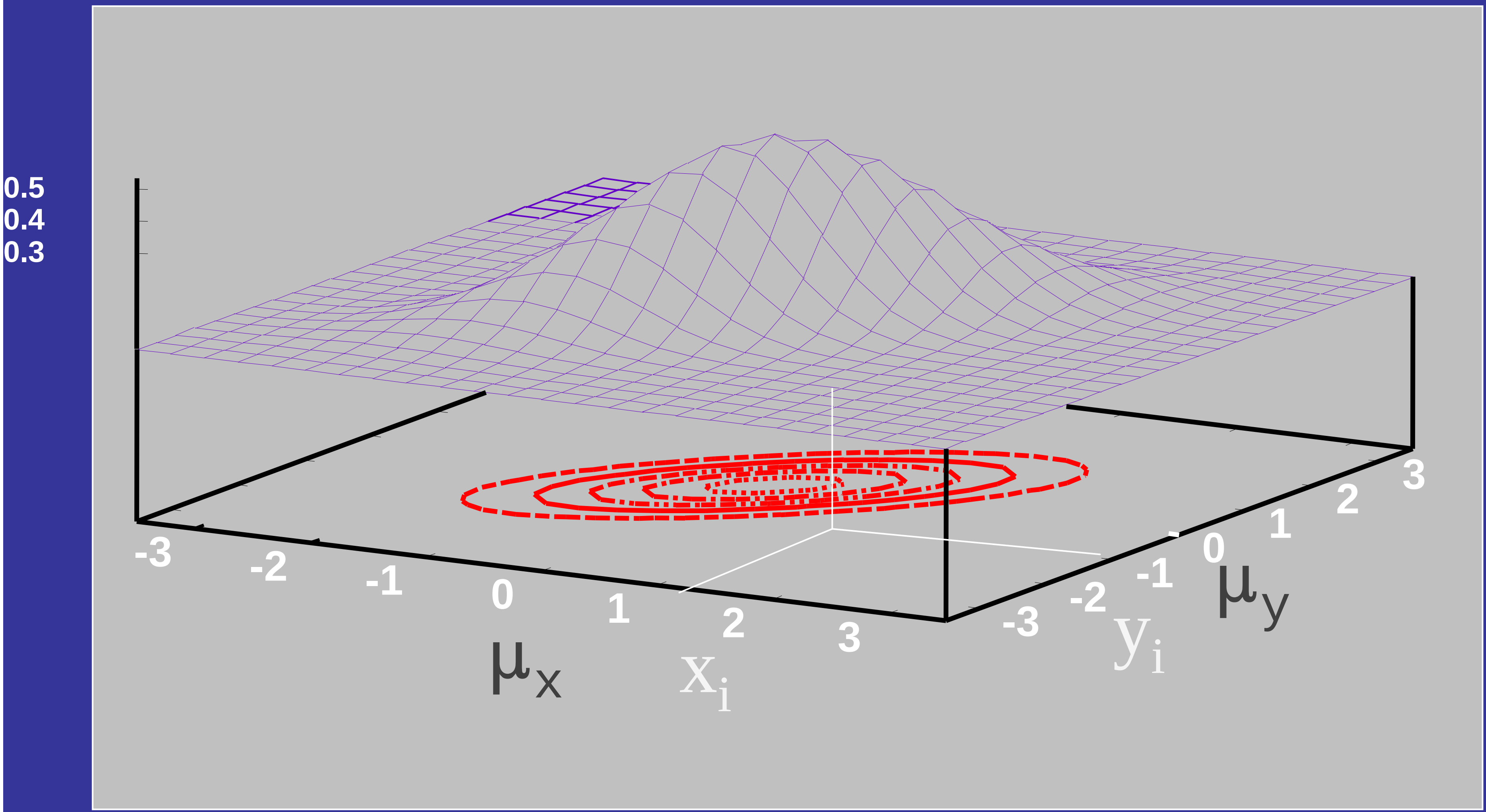
Probability density function



$\phi(x_i)$ is the likelihood of data point x_i for particular mean & variance estimates

Height of bivariate normal curve

An unlikely pair of (x,y) values



Exercises: Compute Normal PDF

- Get used to Mx script language
- Use matrix algebra
- Taste of likelihood theory