

What is (are?) statistics?

Karl W. Broman

8 January 1998

Statistics as objects:

numbers used to describe/summarize data

Statistics as a discipline:

answering questions using data:

- exploratory data analysis
- inductive inference using probability
- quantifying uncertainty

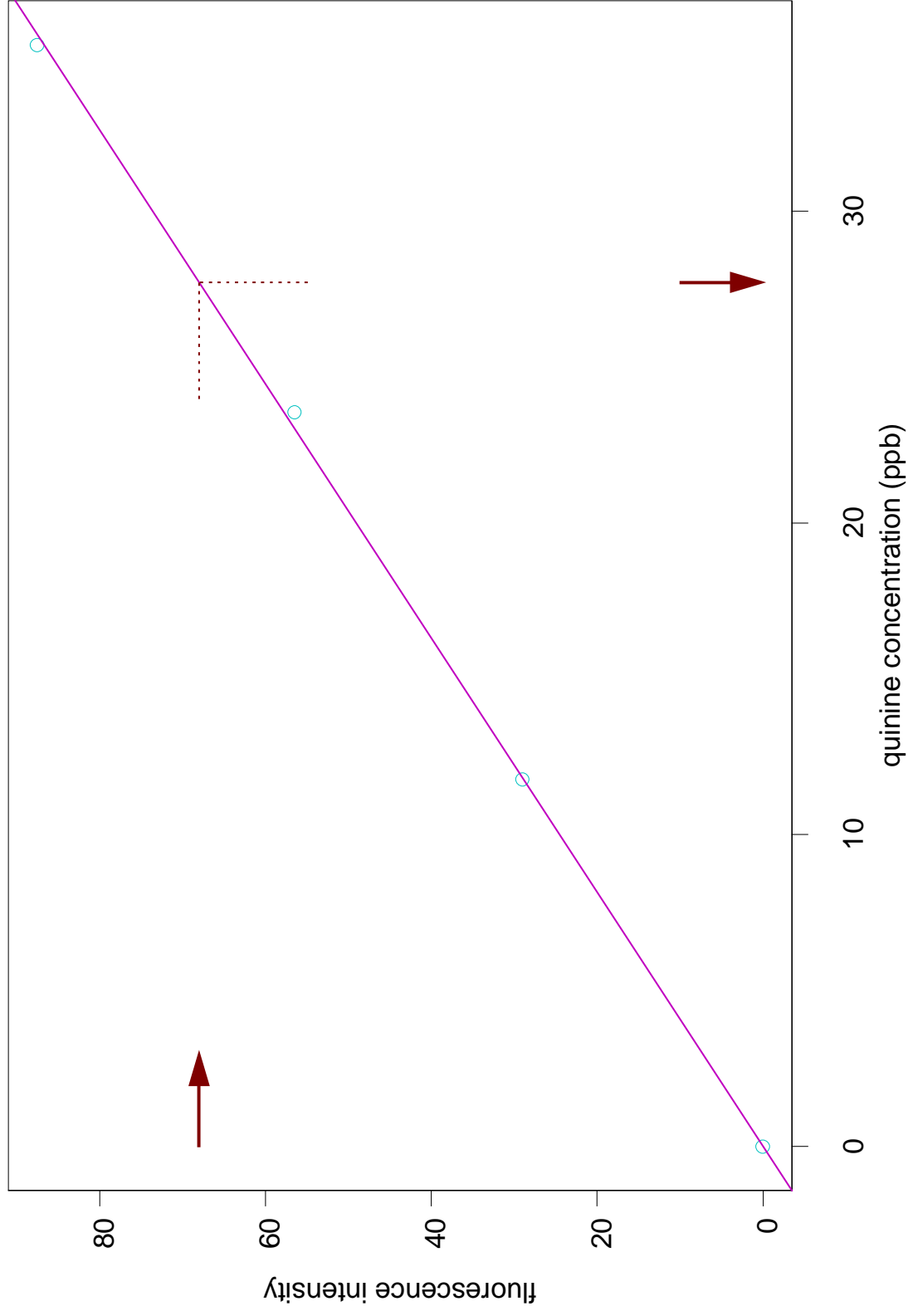
How I became interested in the subject:

Quantitative analytical chemistry:

determination of quinine concentration in tonic water

Experiment:

- prepare a set of solutions with known quinine conc.
- measure the fluorescence of each sample
- do the same for the unknown sample
- fit a line to the data on the standards



Questions:

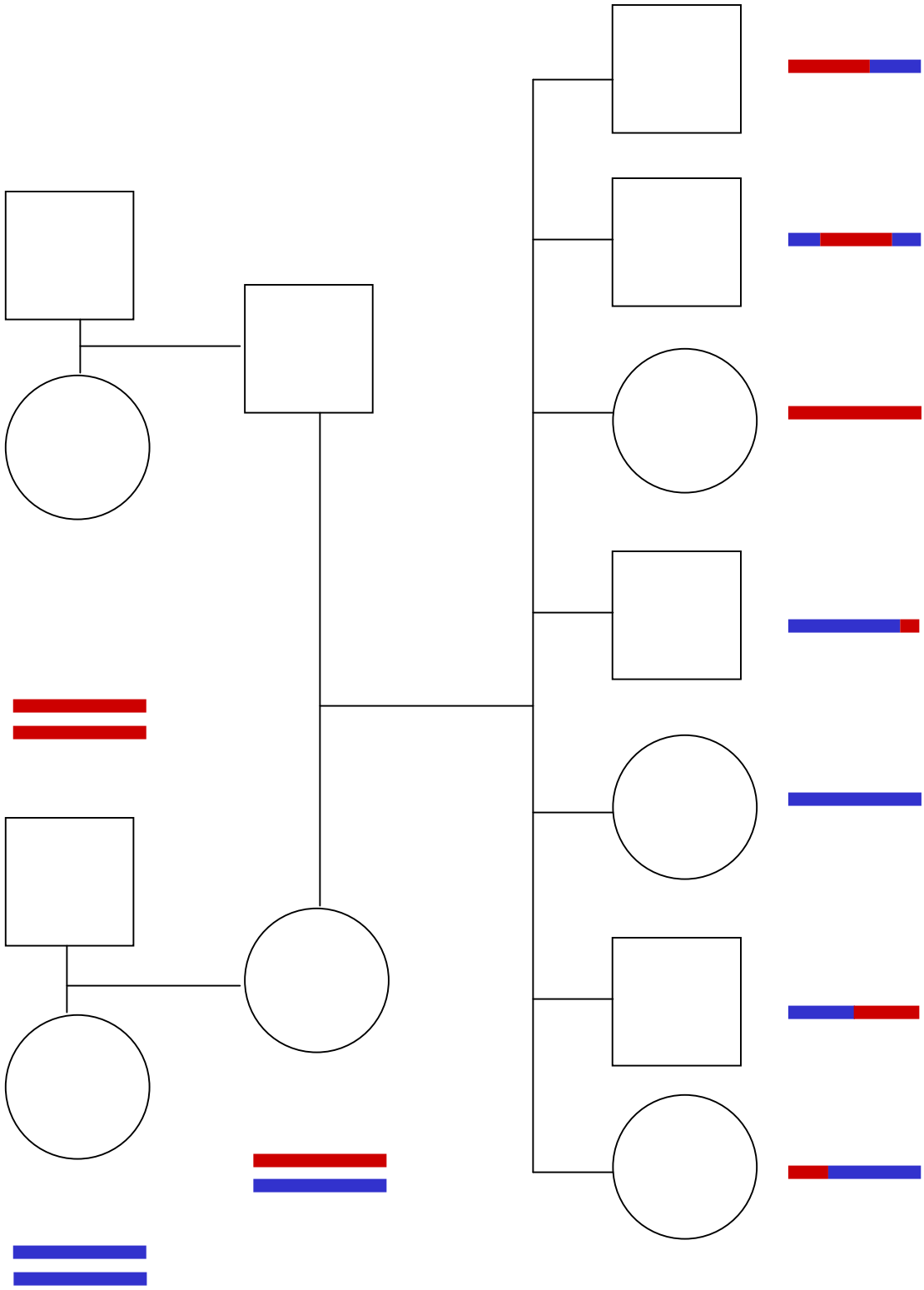
- (1) estimated concentration in sample?
- (2) estimated uncertainty?

27.7 +/- 0.1 ppb?
27.7 +/- 1.0 ppb?
27.7 +/- 10.0 ppb?

Dilution error and measurement error cause
variability in the fitted line which thus causes
variability in the estimated quinine concentration.

Exploratory data analysis

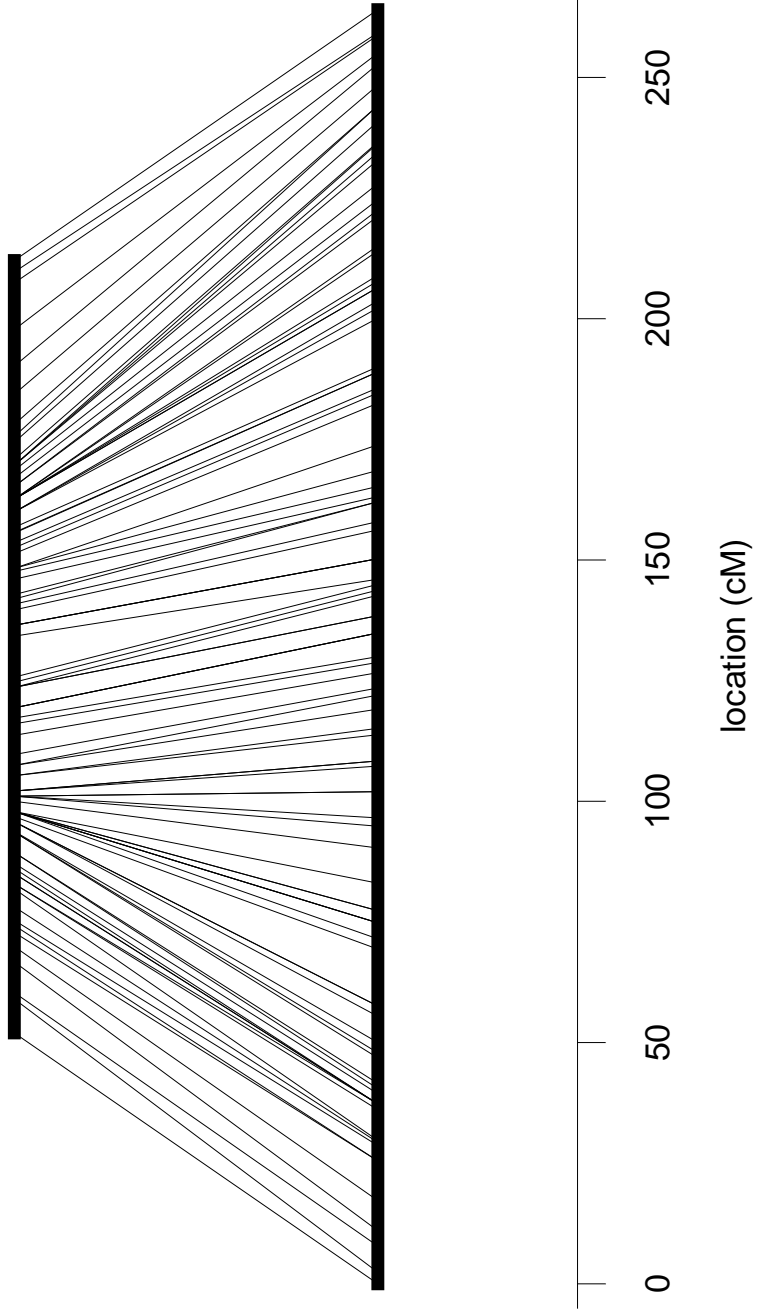
- **detective work**
find new relationships in data
- **visual design**
find better ways to display information



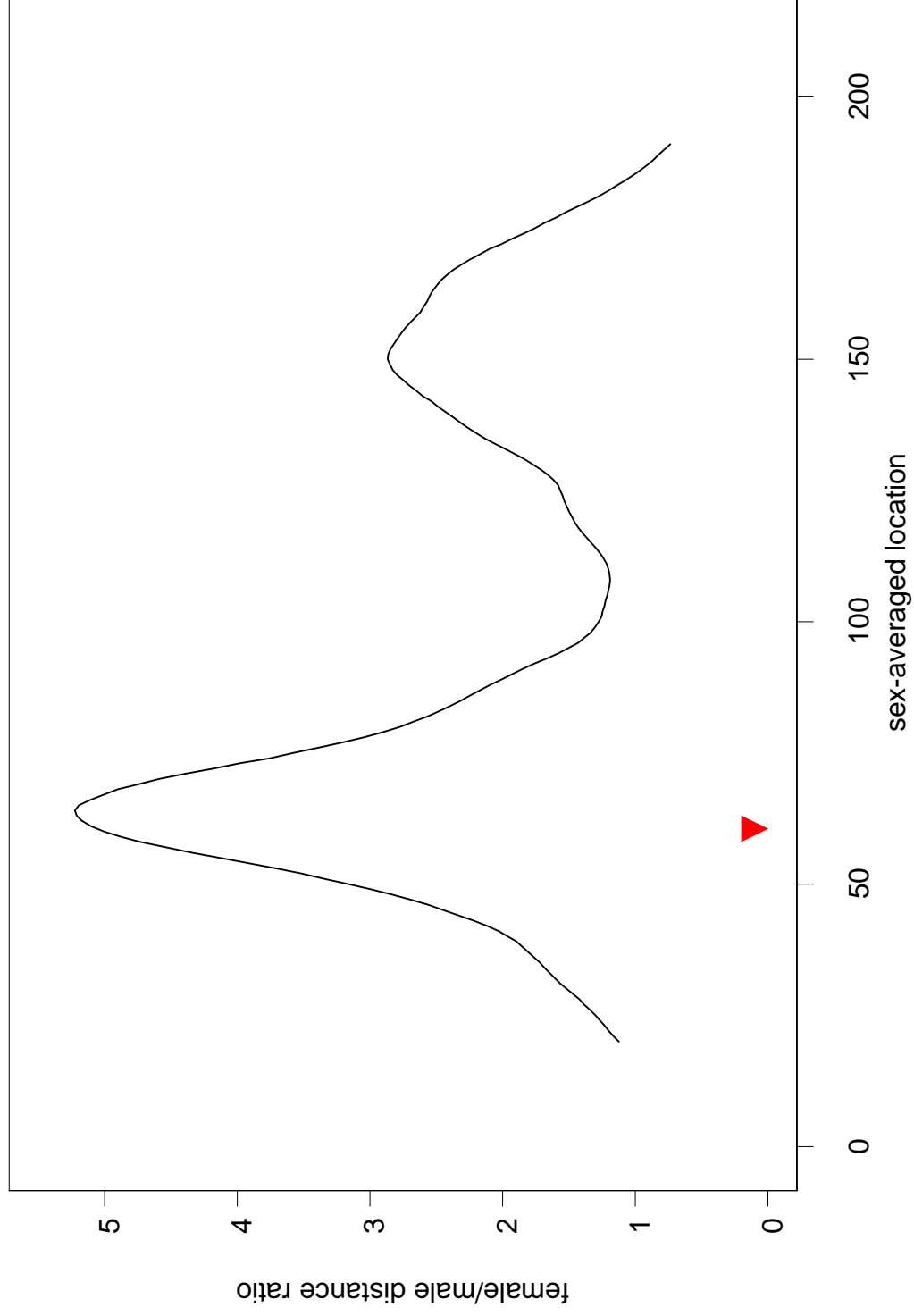
Chromosome 6 Family 884



Chromosome 4



chromosome 4



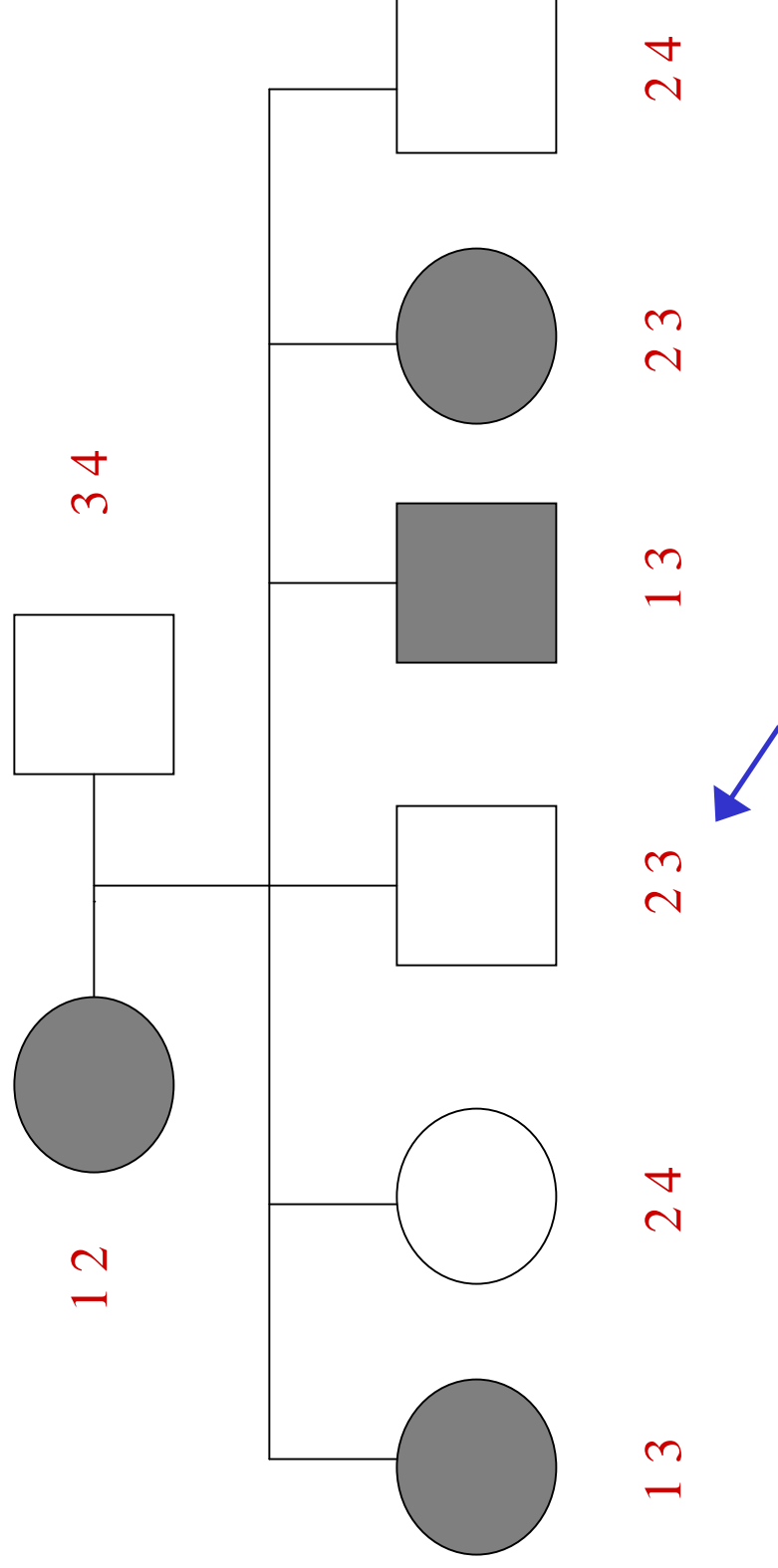
Statistical inference

- making general statements using limited data
- quantifying the uncertainty in such statements

Tools

mathematical models of random variation; probability

Example: linkage between gene for simple recessive trait and a genetic marker



θ = recombination fraction between marker
and gene

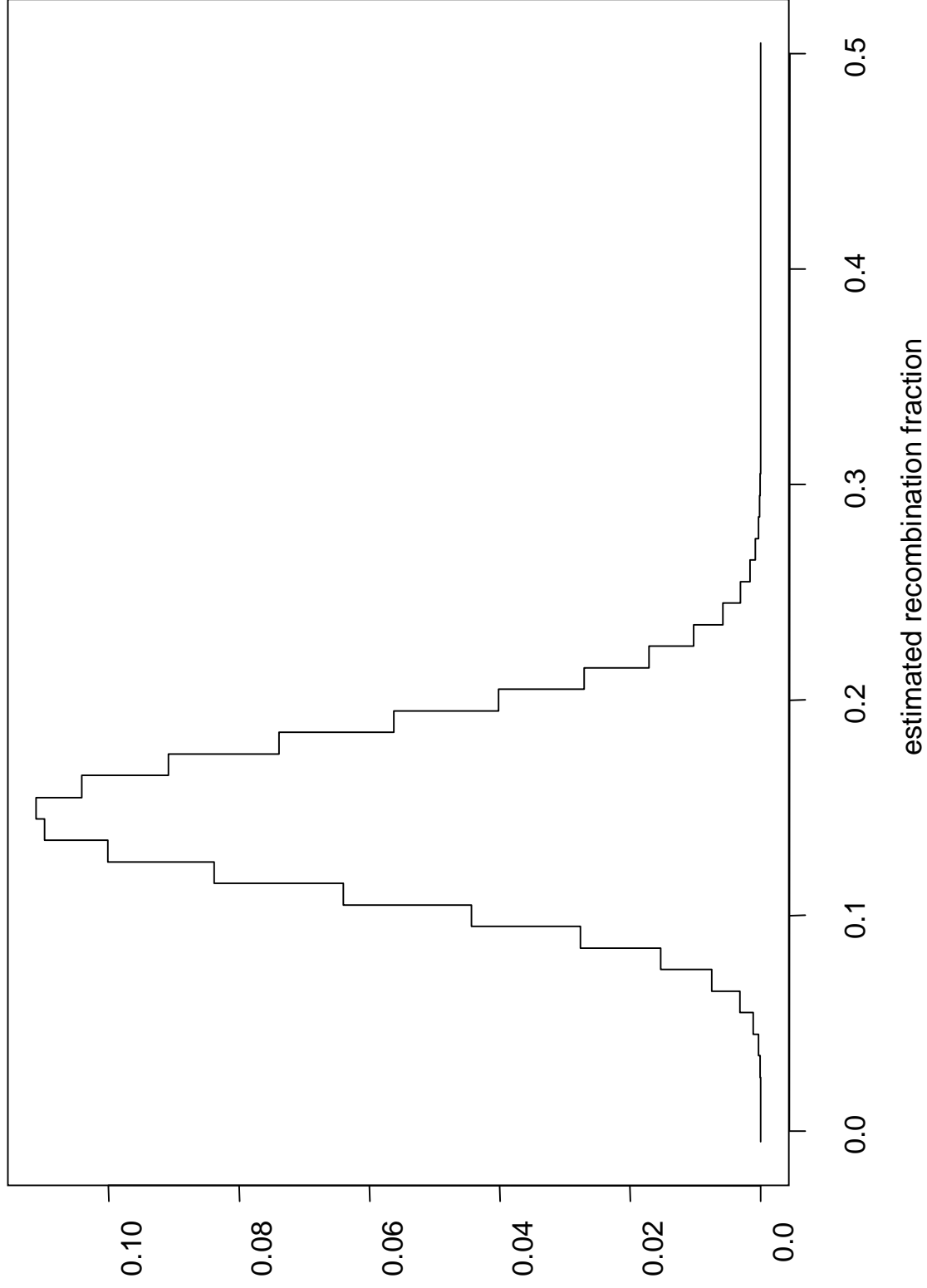
$$L(\theta) = \Pr(\text{data}|\theta) \propto \theta^k (1-\theta)^{n-k} + \theta^{n-k} (1-\theta)^k$$

$\hat{\theta}$ = value of θ maximizing $L(\theta) \approx k / n$

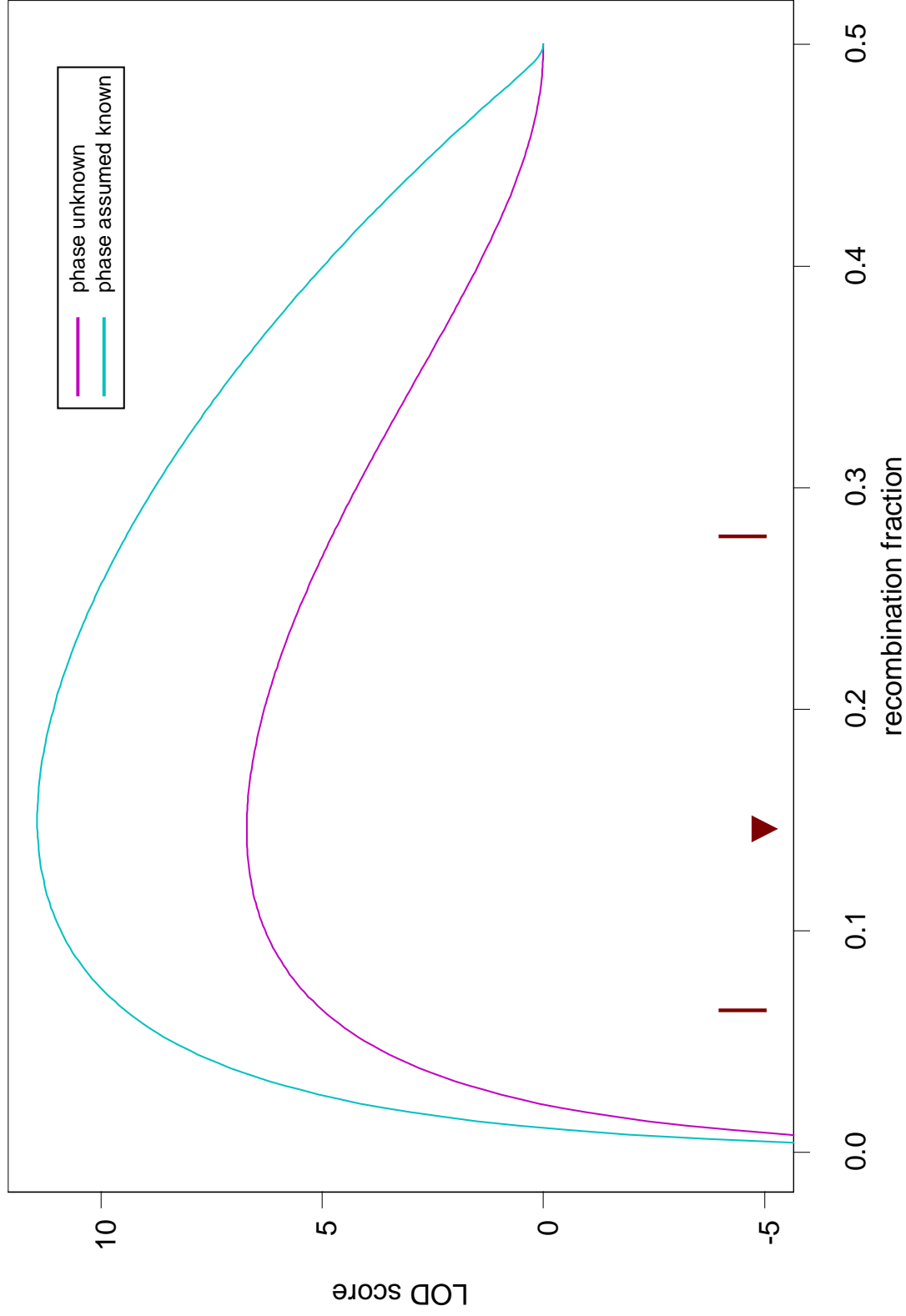
$$\text{LOD}(\theta) = \log_{10}[L(\theta) / L(\frac{1}{2})]$$

sampling distribution

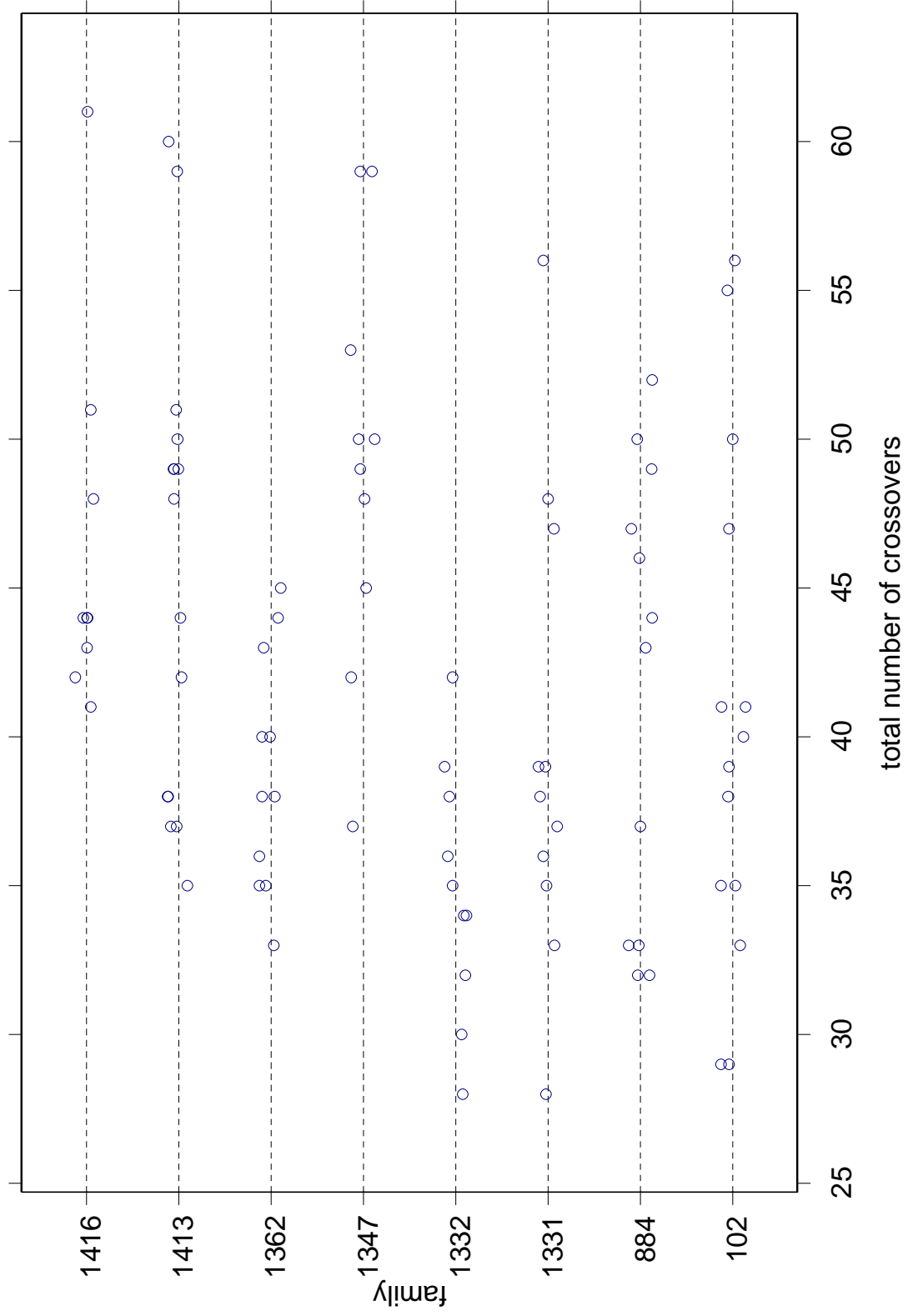
20 families with 5 progeny each; assume $\theta = 0.15$

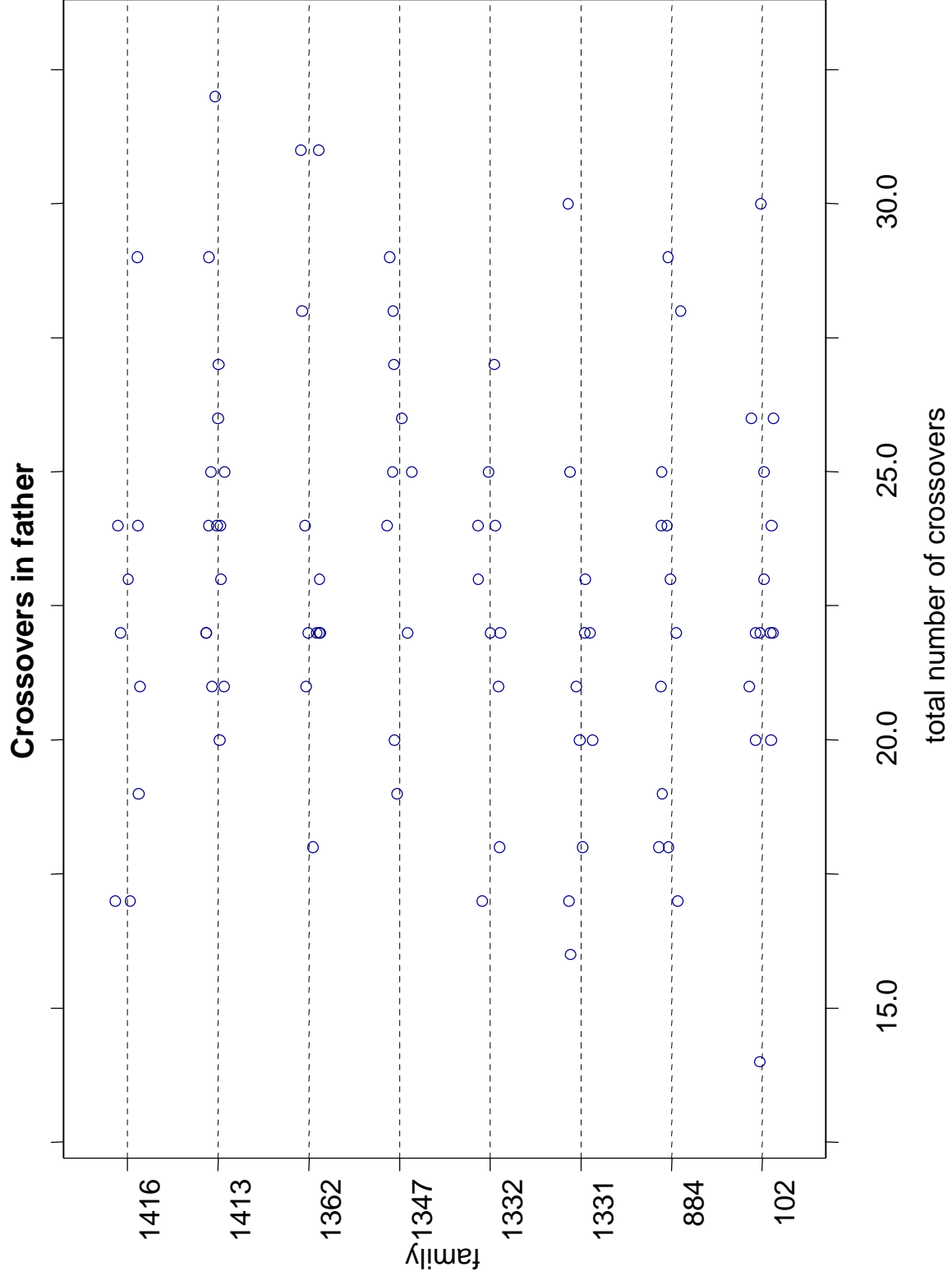


20 families, each with 5 progeny

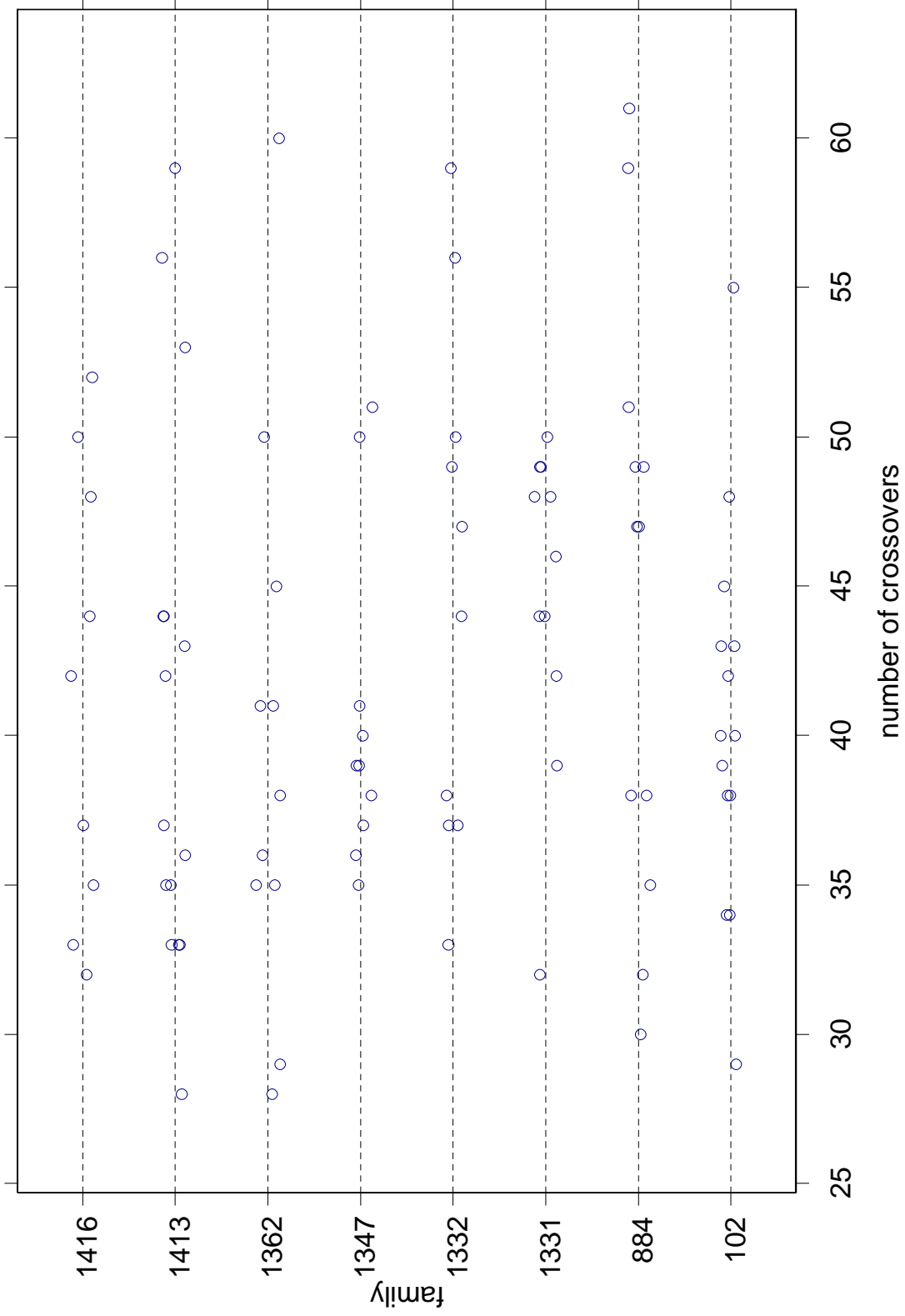


Crossovers in mother

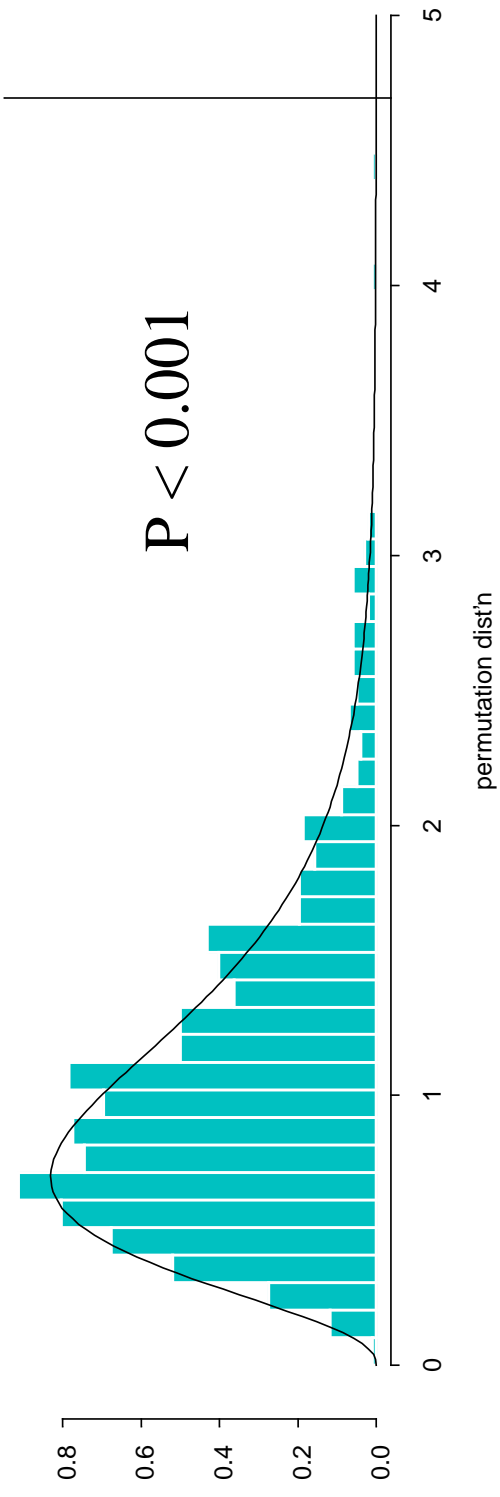




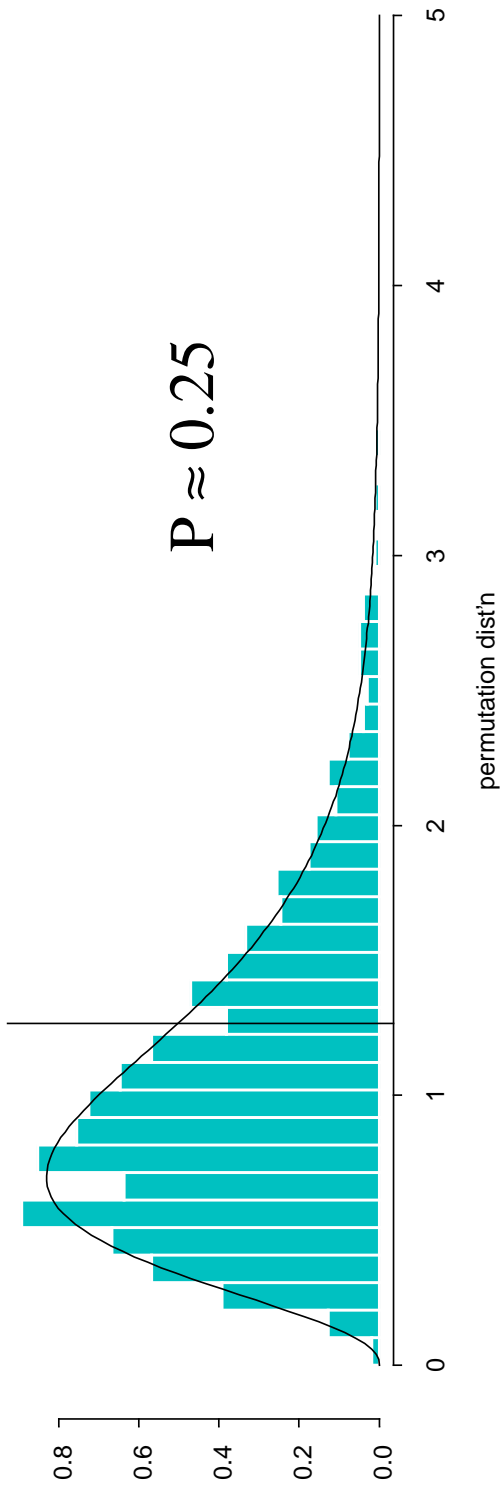
Crossovers in mother after random permutation



mothers



fathers



Summary

- exploratory data analysis
 - find new ways to look at data
 - follow up strange observations
- statistical inference
 - probability model for data
 - sampling distribution of estimator
 - confidence interval, hypothesis test