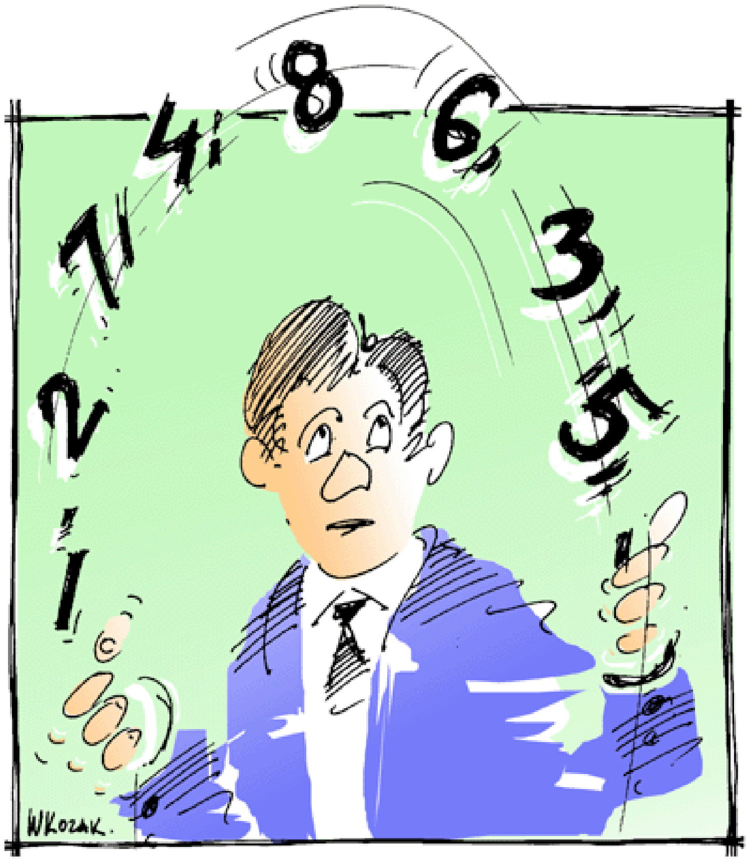
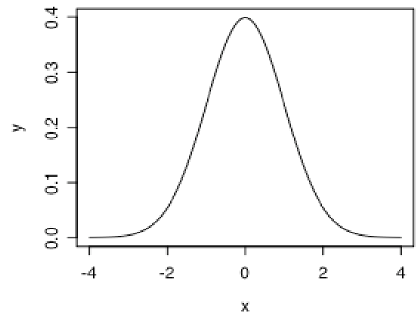
Chapter 5

Computer-Intensive Statistical Methods

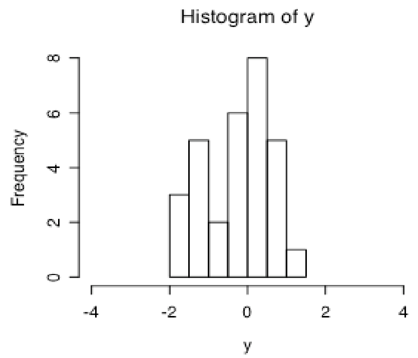


Last Week....



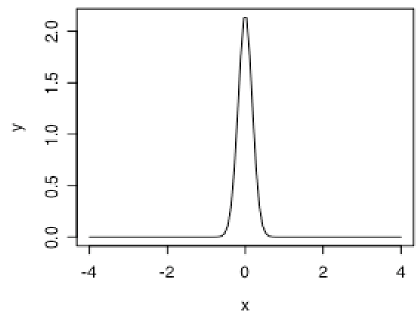
Assume Gaussian

Experiment
(take sample)



Assume ~ Population
Compute std. deviation, s

Central Limit Theorem



Assume Gaussian
Compute std. error
$(s/\sqrt{n})$
Make inferences

Motivations and Goals

You might me interested in things that

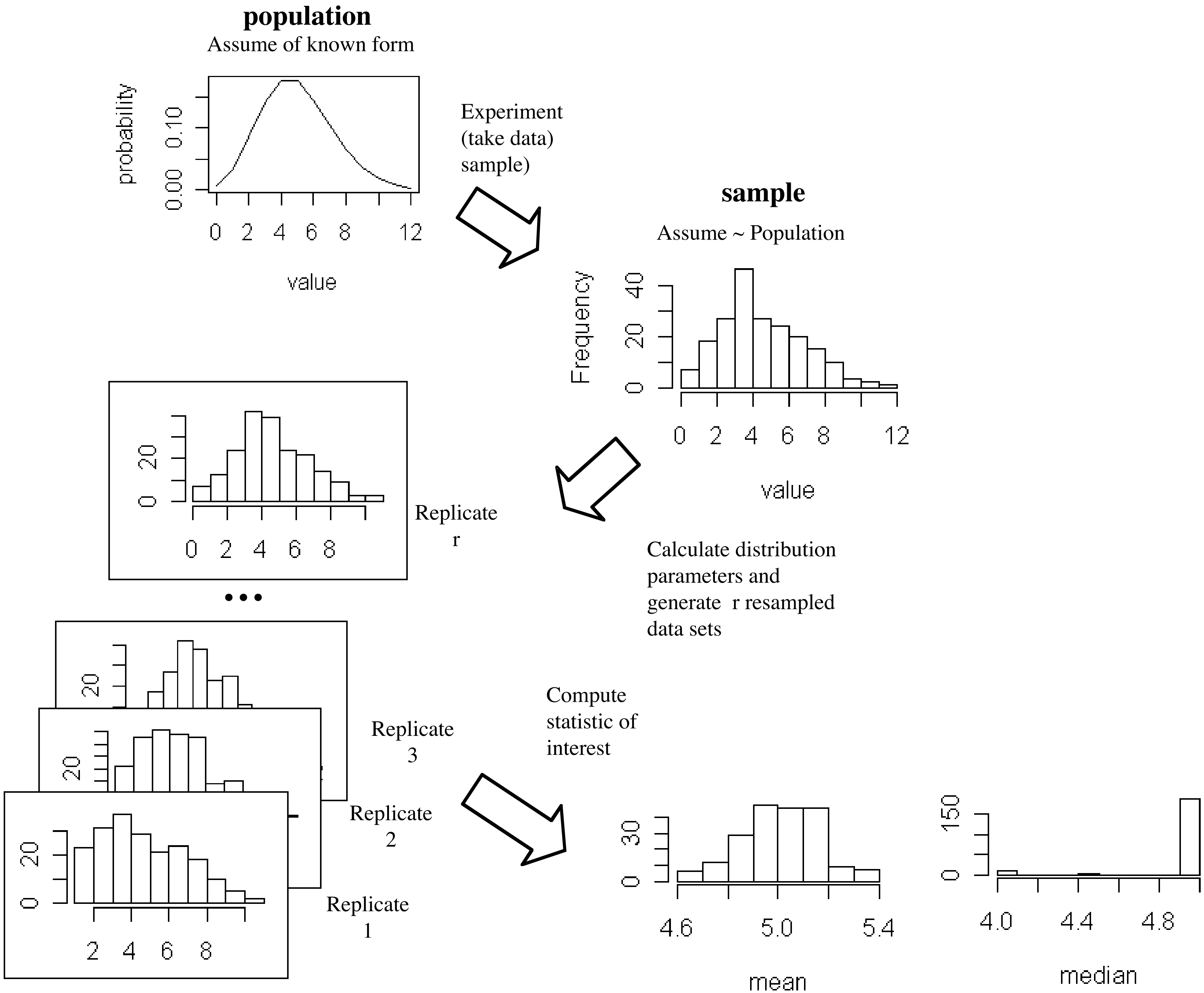
- aren’t normal
- aren’t mathematically computable in convenient form
- come from data with unknown distribution
- statistics of interest is not mean
- to test hypotheses only on samples!

Construct sampling distribution empirically with computers!

Monte Carlo Test

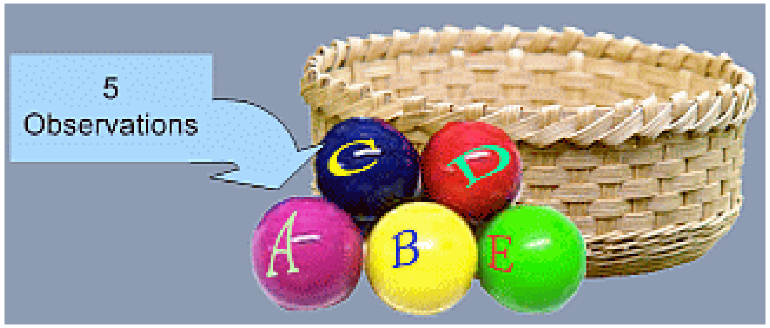
- The distribution of the population can be described parametrically e.g Binomial, Poisson
- The statistic of interest can be anything

Monte Carlo Test

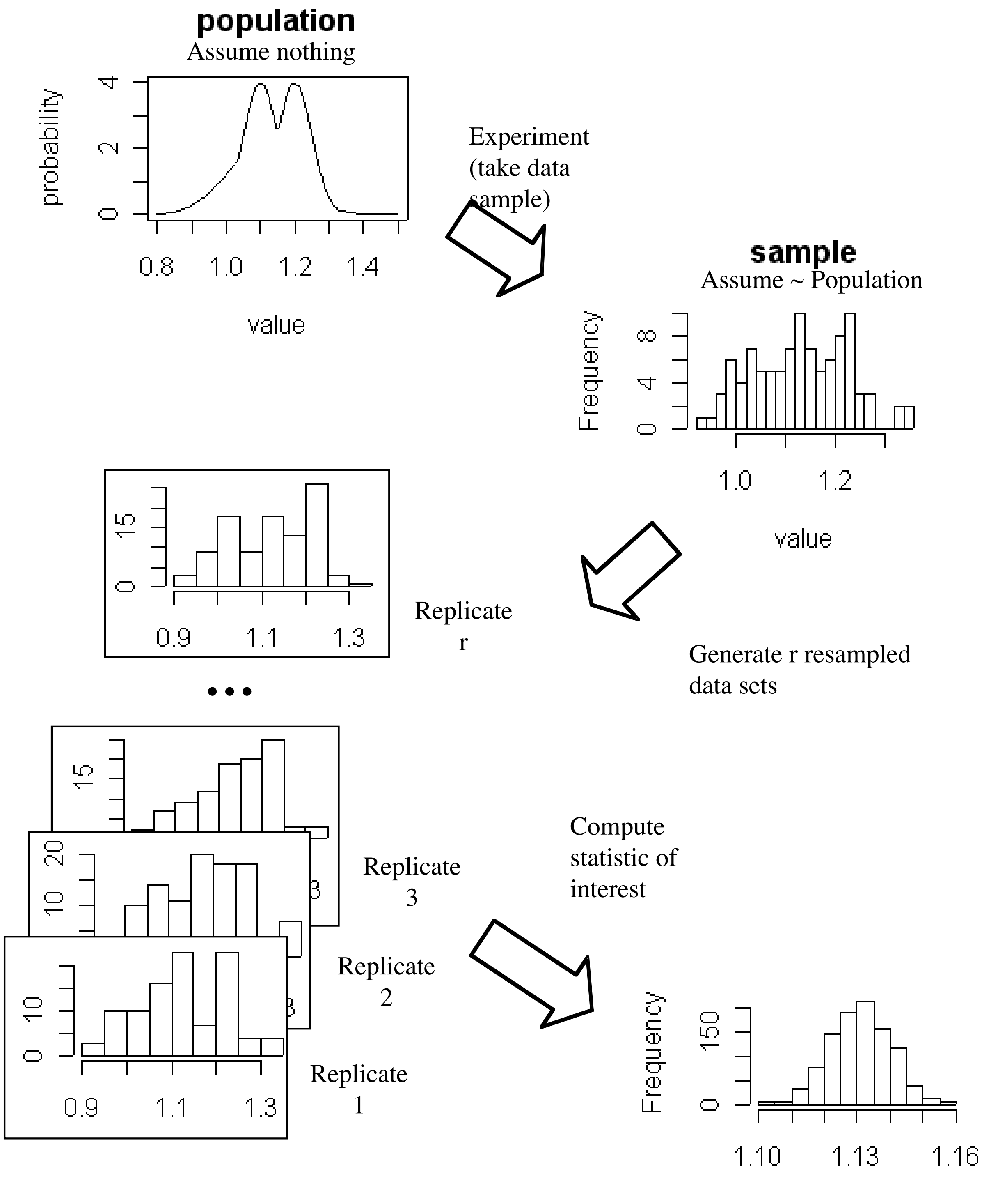


Bootstrap Method

- A special case of Monte Carlo test
- The distribution need not be described parametrically
- The statistics of interest can be anything
- Assumption: the data be your best estimate of the distribution of population
- Draw samples with **replacement** from your data many times



Bootstrap Method



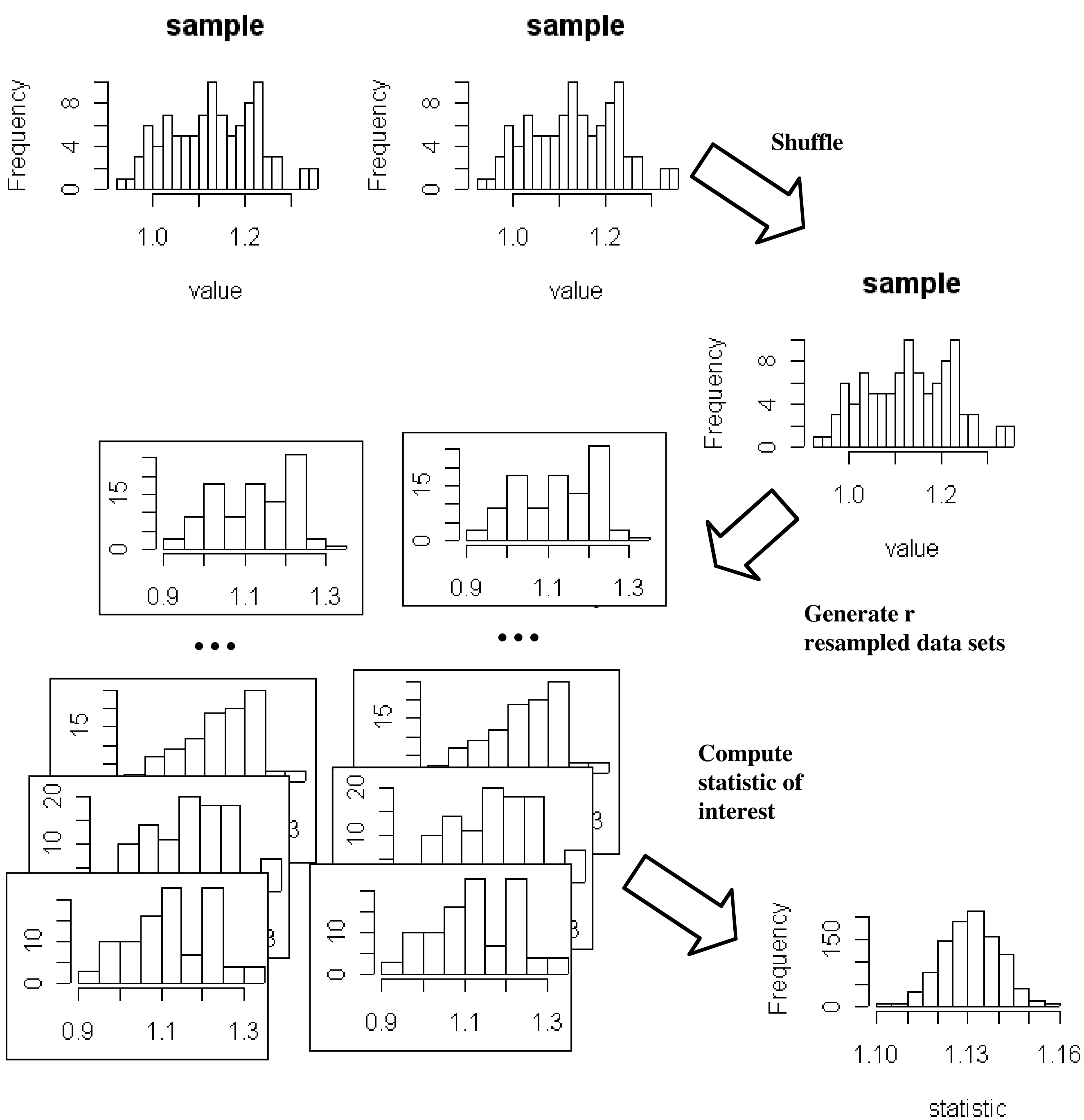
Bootstrap Sampling Distribution

- Shift Method:
  - bootstrap and null hypothesis sampling distributions are different but they have the same shape
- Normal Approximation Method
  - null hypothesis sampling distribution is normal
  - bootstrap sampling distribution - to estimate the variance of the normal distribution
  - Z-test to estimate the p value

Randomization Test

- Resampling without replacement
- To test hypothesis about samples without drawing any conclusion about population
- Mainly to test if the samples are unrelated

Randomization Test



Nonparametric Test

- Cohen's view - rather pessimistic!
- Number and nature of the parameters is flexible and not fixed in advance - distribution free
- Less power than the appropriate parametric tests, but are more robust
- Sampling distributions are usually exact/analytical

Wilcoxon Signed-Rank Test

- compute absolute difference |Xa—Xb| for each pair;
- omit cases where |Xa—Xb|=0;
- rank the remaining absolute differences, from smallest to largest, employing tied ranks where appropriate;
- "+" sign when Xa—Xb>0 and "-" sign when Xa—Xb<0
- sum the signed ranks T+, T- and choose the minimum

$$\mu = n(n+1)/4$$

$$\sigma^2 = n(n+1)(2n+1)/4$$

$$z = (T - \mu)/\sigma$$

Take Home Message