Outline

- Some parametric statistics and tests
- Some computer-intensive versions (bootstrap and randomization) and some novel computer-intensive tests
- Confidence intervals

t tests

One sample t test of whether a sample is drawn from a population with a given mean

$$H_0 : \mu = c$$

Two sample t test of whether the difference between population means is a constant

$$H_0 : \mu_1 - \mu_2 = c$$

Matched pairs t test is also for two groups, but the test is on the mean difference of matched pairs

$$H_0 : \frac{1}{n} \sum_{i}^{n} (x_1 - x_2) = c$$

One sample t test

Mean IQ in the class of 8 students is 135 with standard deviation 20. Test the hypothesis that this sample is drawn from a population of people with mean IQ = 100.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{135 - 100}{20 / \sqrt{8}} = 4.95$$

Reject the null hypothesis?

Two sample t test

Recall F = SIZE / (NUM-PROC * RUN-TIME) is a "parallelization factor" for the KOSO/KOSO* experiment. Test hypothesis that KOSO = KOSO*

$$H_0 : \phi_{koso} - \phi_{koso*} = 0$$

$$t = \frac{\bar{x}_{koso} - \bar{x}_{koso*} - 0}{\hat{\sigma}_{\bar{x}_{koso} - \bar{x}_{koso*}}} = \frac{.74 - .82}{.027} = -2.92$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

Testing whether KOSO = KOSO* in CLASP

- What is the p value (use CLASP CDF functions)
- One-tailed or two-tailed test?
- Reject or fail to reject null hypothesis?

Matched pairs t test

A	B
10	11
0	3
60	65
27	31

Mean(A) = 24.25, Mean(B) = 27.5

Mean difference: (10 - 11) = -1

(0 - 3) = -3

(60 - 65) = -5

(27 - 31) = -4

Mean difference = – 13 / 4 = – 3.25

Test whether the difference of the means is zero, or whether the mean difference is?

Matched pairs t test

A	B
10	11
0	3
60	65
27	31

Treated as unrelated samples, the variance in the row variable swamps any difference in the column variable (t = -.17, p=.87). But if the numbers in each row are matched then the mean difference between As and Bs is significant (t = -3.81, p = .03)

Assumptions of the t test

- Samples are drawn from a normal population
- If this is false but the sample size isn't very small and the p value isn't marginal, don't worry
- Alternatively, you could run t tests as computer-intensive tests and not worry about any assumptions.

Computer-intensive tests of means

- When the parameters of the population are known, the sampling distribution of a statistic can be estimated by Monte Carlo simulation
- Otherwise, use the bootstrap or randomization

How’s your IQ? We've done it with the Z test and Monte Carlo sampling, now let's try the bootstrap

- Pretend the population IQ is unknown but our null hypothesis is that its mean is 100.
- The mean IQ in this class, 23 students, is 130.
- Should we reject the null hypothesis that this class is no different in IQ from the population?

The Logic

- Our result: $R = 108$
- Assume Ho: $\Pi = 100$, this class is a random sample drawn from the population of people with mean IQ 100
- If the result is very unlikely under Ho, if $Pr(R=108 \mid \Pi = 100) \leq \alpha$, then we are inclined to reject Ho.
- Pick a value of α (say, .01) and calculate the conditional probability $p = Pr(R=108 \mid \Pi = 100)$
- Our residual uncertainty that Ho might be right is less than or equal to α

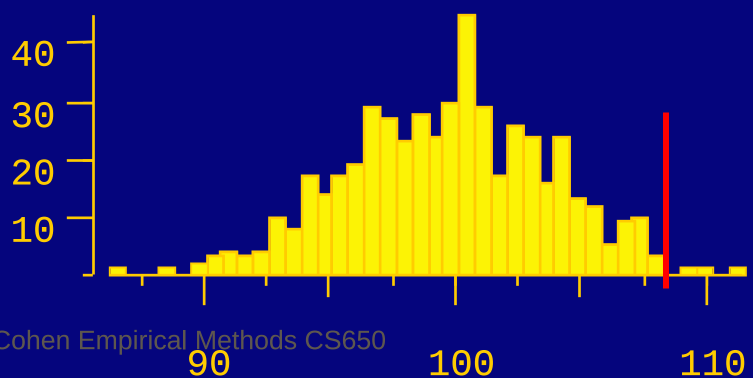
Calculating p = Pr(R=108 | Π = 100)

Monte Carlo

Infinite population

Many samples and values of R*

Empirical MC sampling distribution of R*



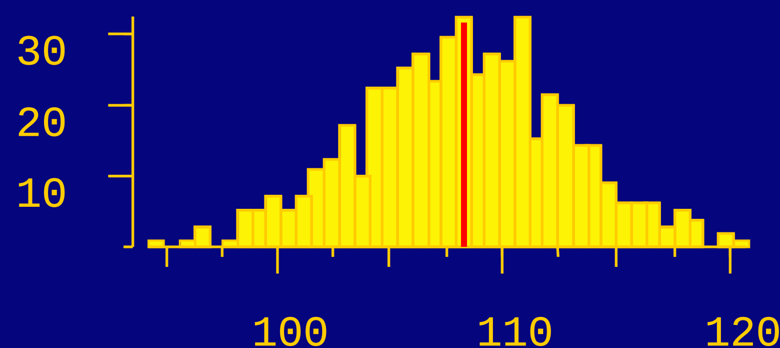
Bootstrap

Sample

Resample with replacement

Many samples and values of R*

Empirical bootstrap sampling distribution of R*

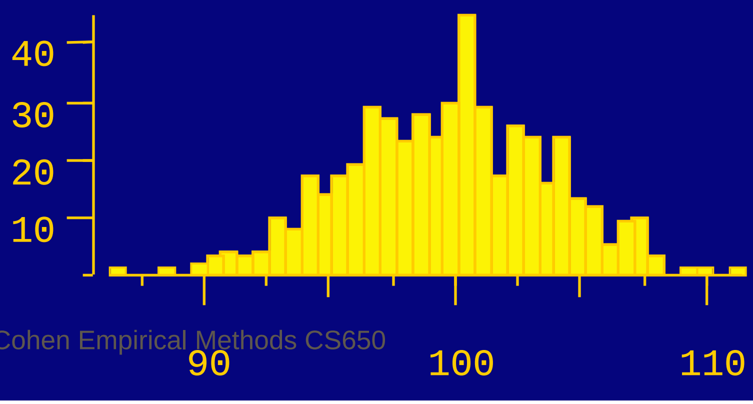


Wait…there's a problem:

Monte Carlo

This is the sampling distribution of R under the null hypothesis that Ho: Π = 100.

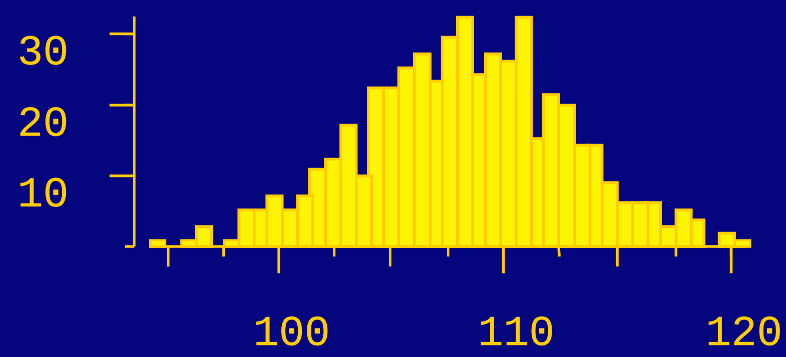
Ho was enforced by sampling from a distribution with Π = 100.



Bootstrap

This is not the sampling distribution of R under Ho: Π = 100.

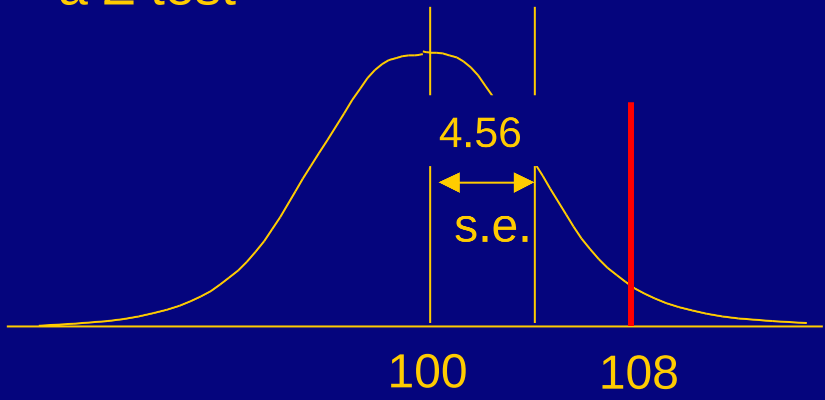
It was obtained by resampling from a sample, no null hypothesis was enforced.



Turning a bootstrap sampling distribution into a null hypothesis bootstrap sampling distribution

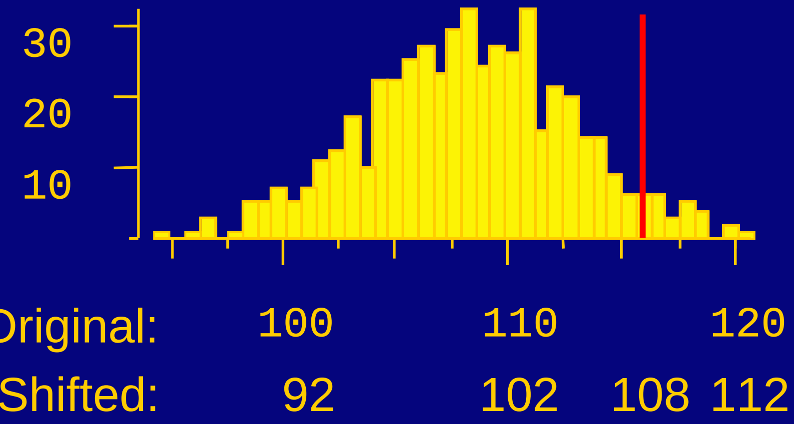
Normal approximation method

Assume the Ho distribution is normal with the Ho mean and a standard error equal to the standard deviation of the bootstrap distribution, then run a Z test



Shift method

Assume the Ho distribution has the same shape and shift the bootstrap distribution until its mean coincides with the Ho mean.



Bootstrap code

```
Sample:  110 104 102 100 131 104 96 115 147 125 70
         101 93 75 117 136 129 127 113 79 125 122 72
```

```
defun bootstrap (sample statistic k)
  (let* ((n (length sample))
         (s (make-array n :initial-contents sample))
         (s* (make-array (length sample)))
         (dist nil))
    (dotimes (replications k)
      (dotimes (i n)
        (setf (aref s* i)(aref s (random n))))
      (push (funcall statistic s*) dist))
    (values dist)))
```

Bootstrap for two-sample tests

- Method 1: Resample S^*_1 and S^*_2 from S_1 and S_2 separately, recalculate the test statistic, collect a sampling distribution of pseudostatistics, apply the shift method or normal approoaximation method to get an Ho distribution

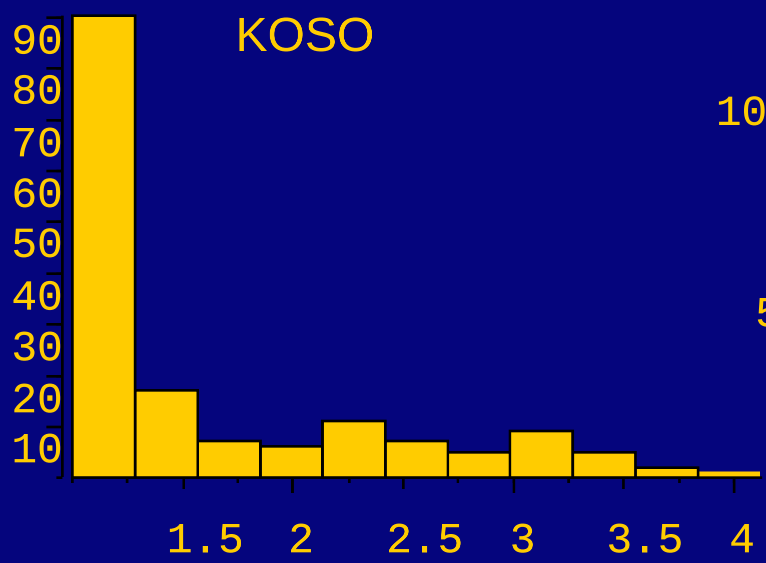
- Method 2: Shuffle the elements of the samples together into **S,** resample S^*_1 and S^*_2 from **S**, collect a sampling distribution of pseudostatistics. This is a null hypothesis distribution!

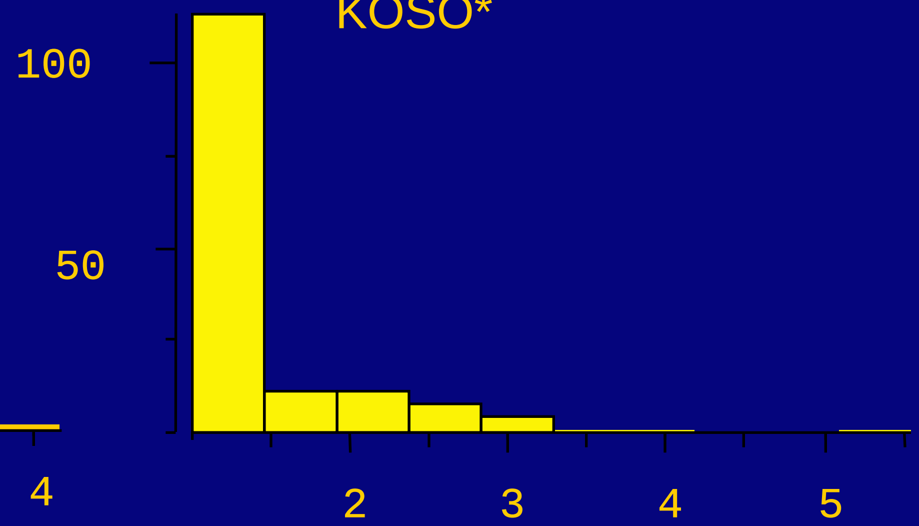
Are KOSO runtimes more variable than KOSO* runtimes? Use the interquartile range.

First transform runtime into multiples of optimal runtime:

$$r = 1 / (size / (num\text{-}proc\ run\text{-}time))$$

IQR(KOSO) = 1.09   IQR(KOSO*) = .39 . A significant difference?





Testing for a difference in interquartile ranges Method 1. The Logic

- Resample with replacement from KOSO sample to k and from KOSO* sample to k*
- Calculate the interquartile ranges of k and k*
- Collect the difference IQR(k) – IQR(k*)
- Repeat

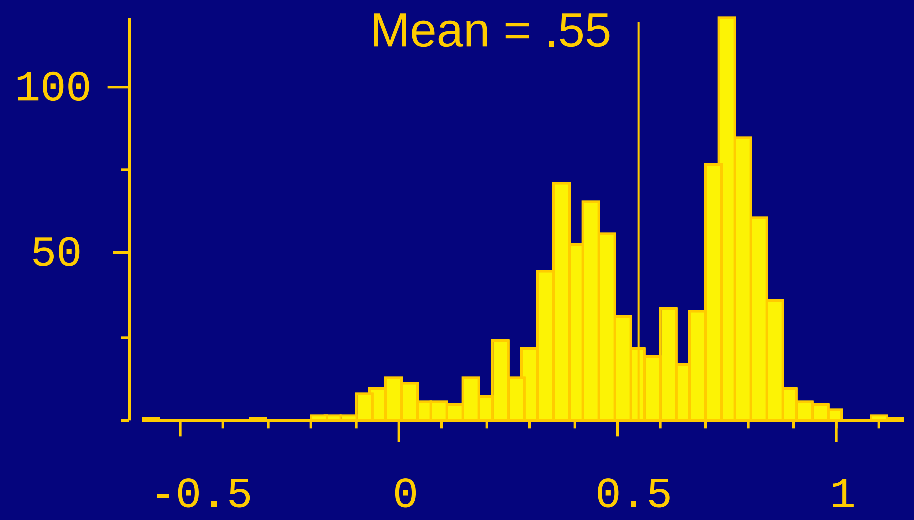
- The resulting distribution is then shifted to have mean zero, enforcing Ho: IQR(KOSO) = IQR(KOSO*)

Code for the two-sample test

```
(defun IQR-diff (sample1 sample2)
  (- (interquartile-range sample1)
     (interquartile-range sample2)))

(defun two-sample-bootstrap (sample1 sample2 statistic k)
  (let* ((n1  (length sample1))
         (n2  (length sample2))
         (s1* (make-array n1))
         (s2* (make-array n2))
         (dist nil))
    (dotimes (i k)
      (dotimes (j n1)
        (setf (aref s1* j)(aref sample1 (random n1))))
      (dotimes (j n2)
        (setf (aref s2* j)(aref sample2 (random n2))))
      (push (funcall statistic s1* s2*) dist))
    (values dist)))
```

Empirical sampling distribution of differences of interquartile ranges

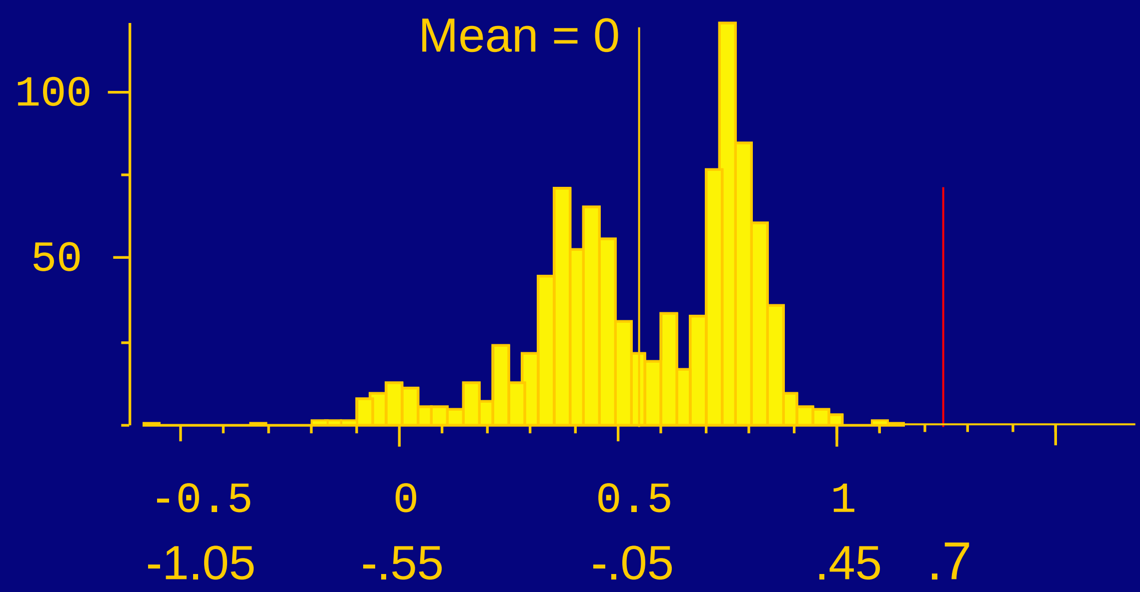


IQR(KOSO) = 1.09   IQR(KOSO*) = .39 .

A difference of 0.7. Is it a significant difference?

Hint: What's the null hypothesis?

Empirical sampling distribution of differences of interquartile ranges



To test Ho: IQR(KOSO – IQR(KOSO*) = 0, shift the distribution so its mean is zero by subtracting .55 from each value

IQR(KOSO) = 1.09 IQR(KOSO*) = .39 . Is 0.7 a significant difference?

Testing for a difference in interquartile ranges
Method 2. The Logic

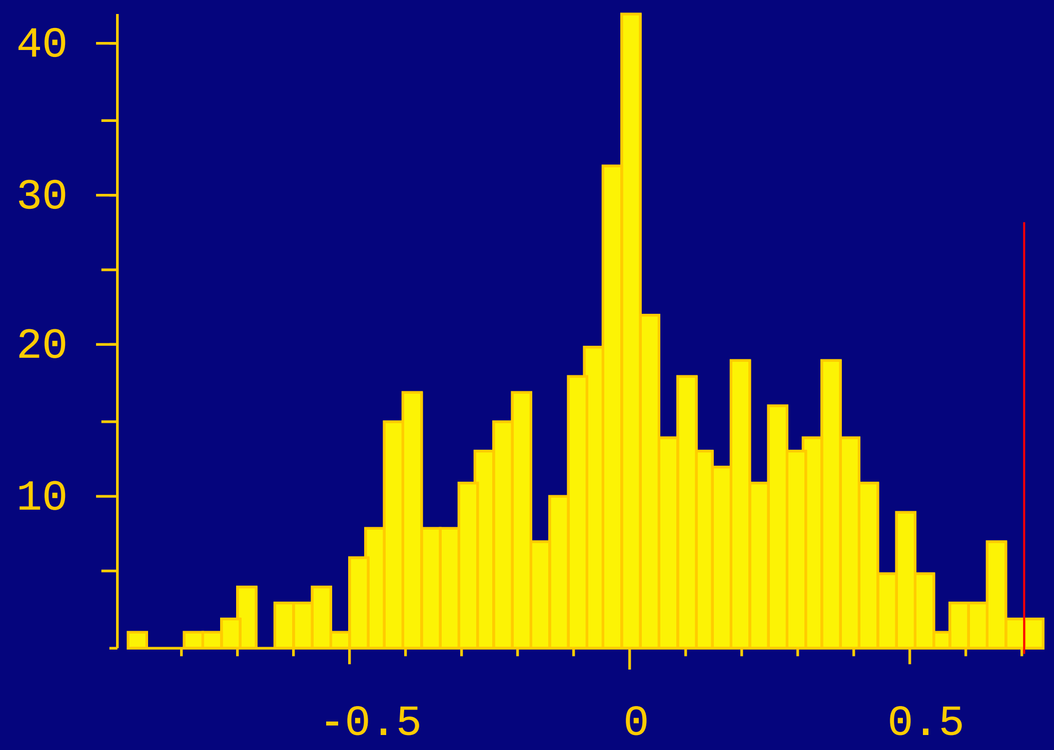
- Merge the KOSO and KOSO* samples into one sample **S** and shuffle it thoroughly.
- Resample with replacement from **S** to k and from **S** to k*
- Calculate the interquartile ranges of k and k*
- Collect the difference IQR(k) – IQR(k*)
- Repeat

- The merging and shuffling enforces Ho: IQR(KOSO) = IQR(KOSO*) so no shift is necessary

Code for the shuffled two-sample bootstrap

```
(defun two-sample-bootstrap (sample1 sample2 statistic k)
  (let* ((n1  (length sample1))
         (n2  (length sample2))
         (n (+ n1 n2))
         (s   (make-array n :initial-contents
                          (append sample1 sample2)))
         (s1* (make-array n1))
         (s2* (make-array n2))
         (dist nil))
    (shuffle-vec s 1000)
    (dotimes (i k)
      (dotimes (j n1)
        (setf (aref s1* j)(aref s (random n))))
      (dotimes (j n2)
        (setf (aref s2* j)(aref s (random n))))
      (push (funcall statistic s1* s2*) dist))
    (values dist)))
```

Shuffled-bootstrap sampling distribution of the difference of interquartile ranges, KOSO and KOSO*



IQR(KOSO) = 1.09 IQR(KOSO*) = .39 . Is 0.7 a significant difference?