Simulation-Based Statistics: Introduction to Resampling

Google

Tim Hesterberg, *Google Inc.*
July 2018

Outline

Google

- Case Study / Basics
  - 1-sample bootstrap
  - Idea behind bootstrap
  - 2-sample bootstrap
  - Permutation test
- Accuracy
- Bootstrap Regression
- Bootstrap Sampling Methods
- Permutation Tests

Meta goals:

Understand basic procedures

Useful for communication

Why Resample

Google

- Easier to understand
  - Communicate to clients
- Easier
  - Same procedure for many statistics
  - Don’t need to derive formulas
- More accurate
  - Depending on procedure

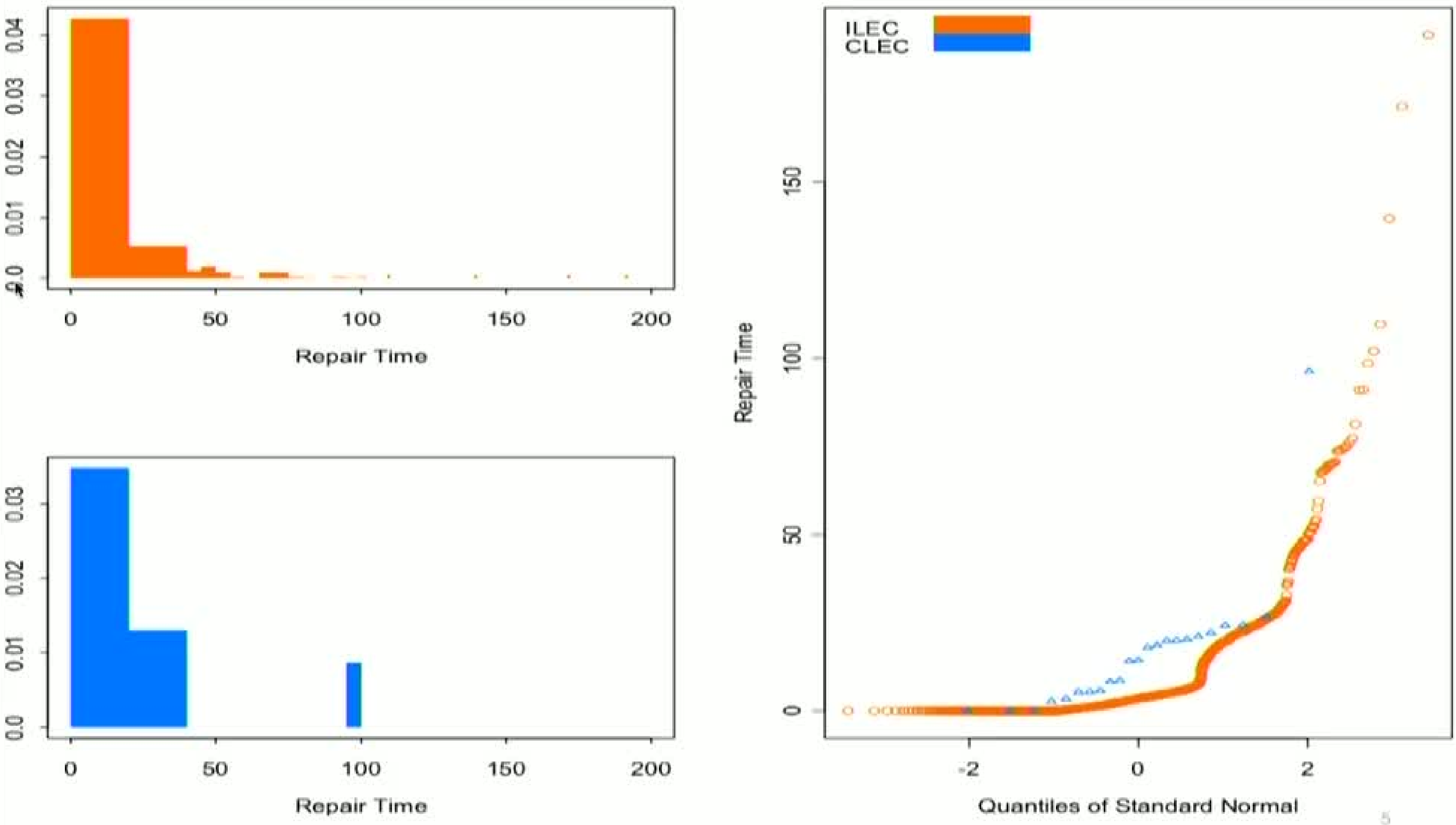
Skewed Data- Verizon

Google

	Number of Observations	Average Repair Time
ILEC (Verizon)	1664	8.4
CLEC (other carrier)	23	16.5

Is the difference statistically significant?

Example Data



Bootstrap Procedure

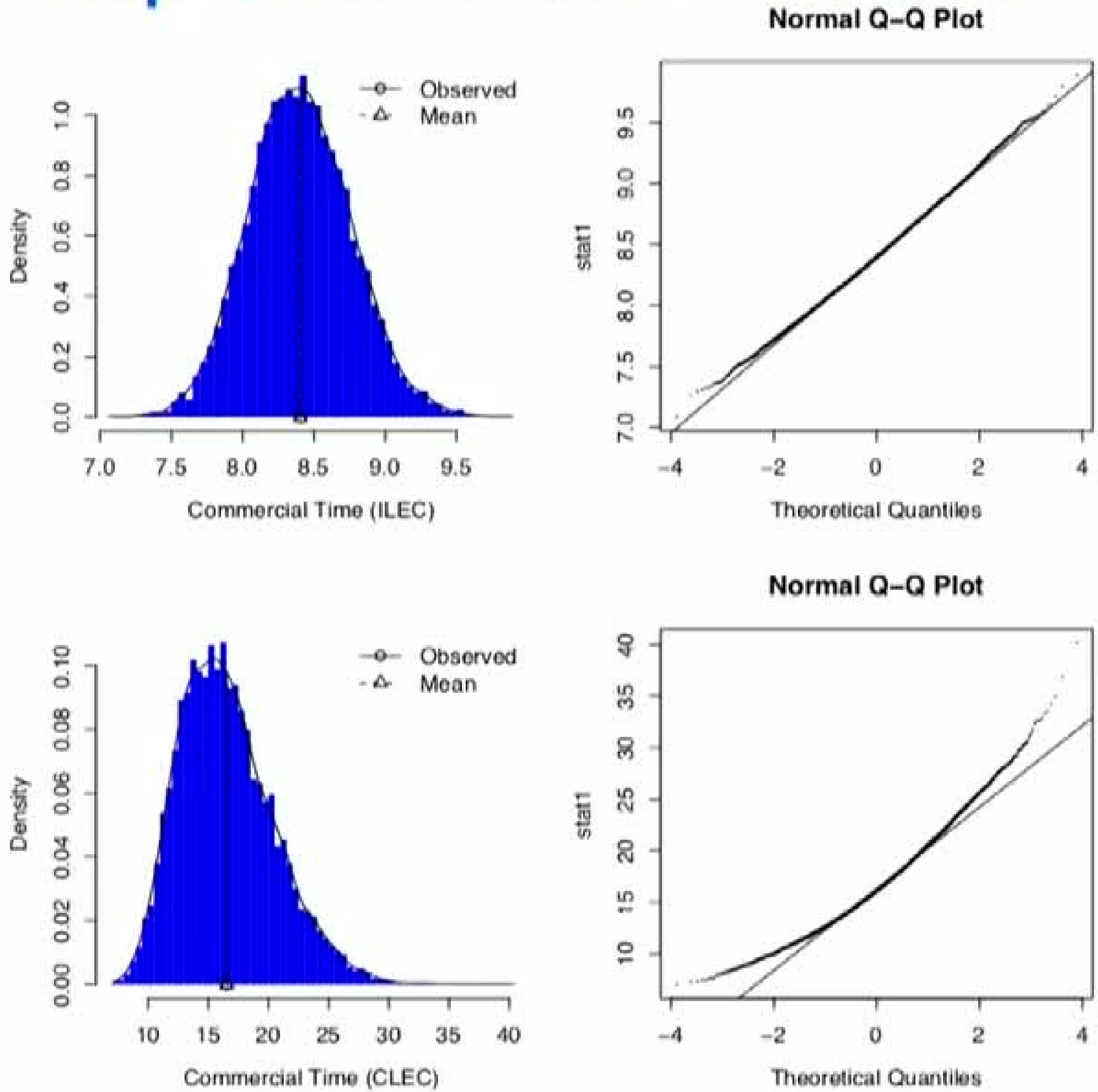
Google

Repeat 10000 times

- Draw a sample of size *n* with replacement from the original data ("bootstrap sample", or "resample")
- Calculate the sample mean for the resample

The 10000 bootstrap sample means comprise the "bootstrap distribution".

Bootstrap Distns for Verizon



Bootstrap Standard Error

Bootstrap standard error =

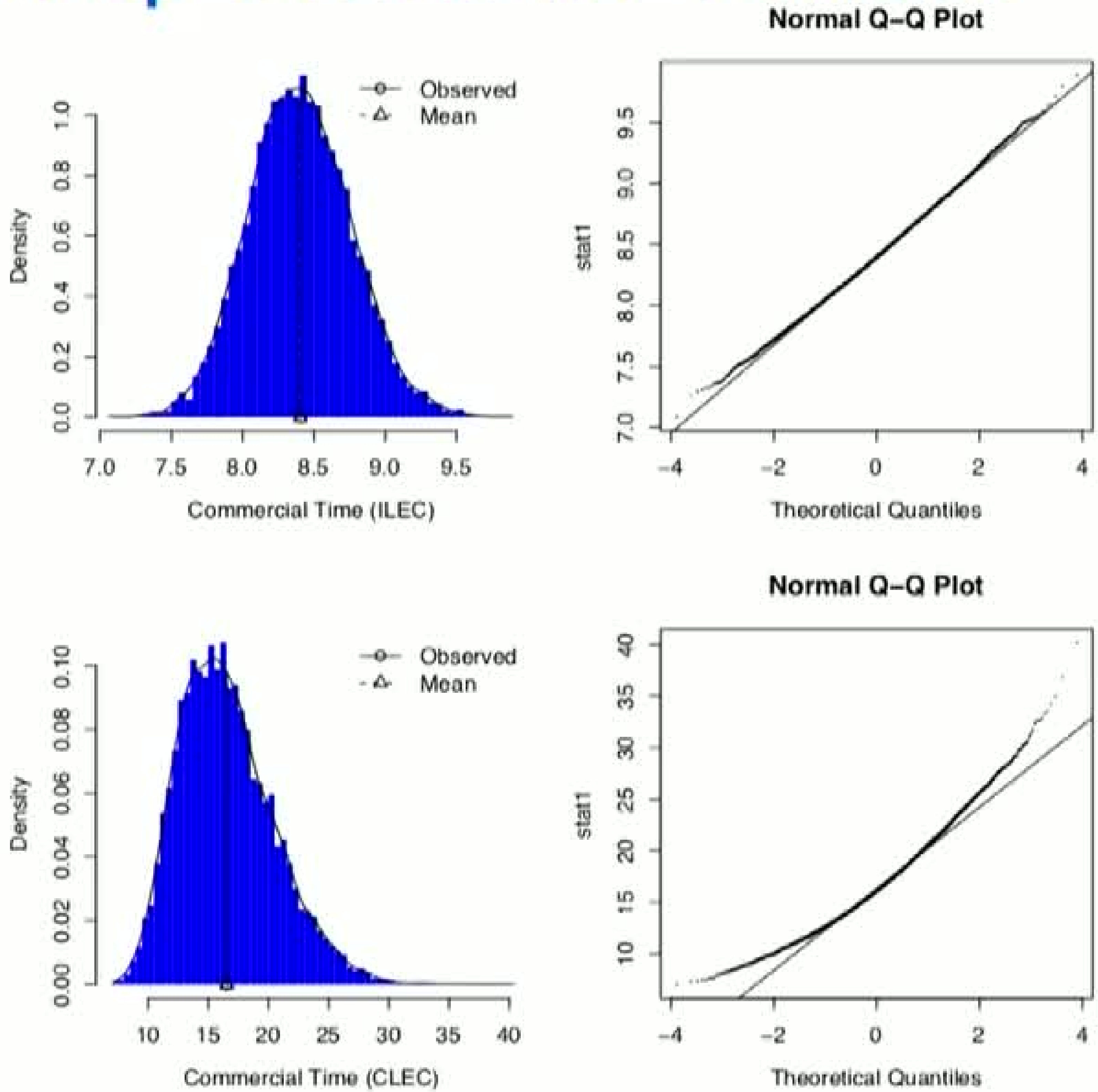
Standard deviation of the sampling distribution

Bootstrap bias =

Mean of the sampling distribution – Observed

```
Summary Statistics:
     Observed        SE     Mean       Bias
mean 16.50913 3.961816 16.53088  0.0217463
mean  8.41161 0.357599  8.40410 -0.0075031
```

Bootstrap Distns for Verizon



Bootstrap Standard Error

Google

Bootstrap standard error =

Standard deviation of the sampling distribution

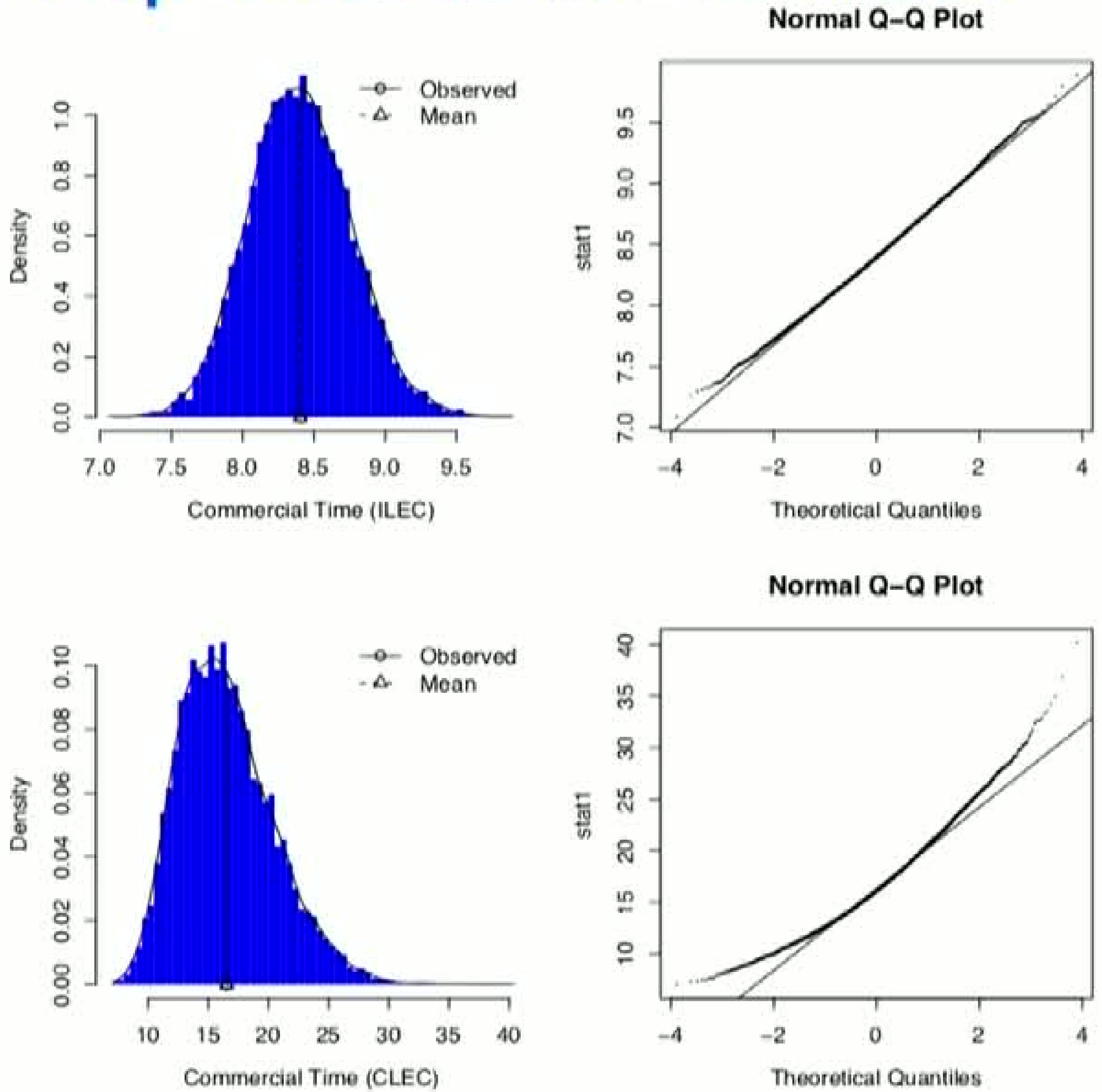
Bootstrap bias =

Mean of the sampling distribution – Observed

```
Summary Statistics:
     Observed        SE      Mean        Bias
mean 16.50913 3.961816 16.53088  0.0217463
mean  8.41161 0.357599  8.40410 -0.0075031
```

Bootstrap Distns for Verizon

Google



Bootstrap Standard Error

Bootstrap standard error =

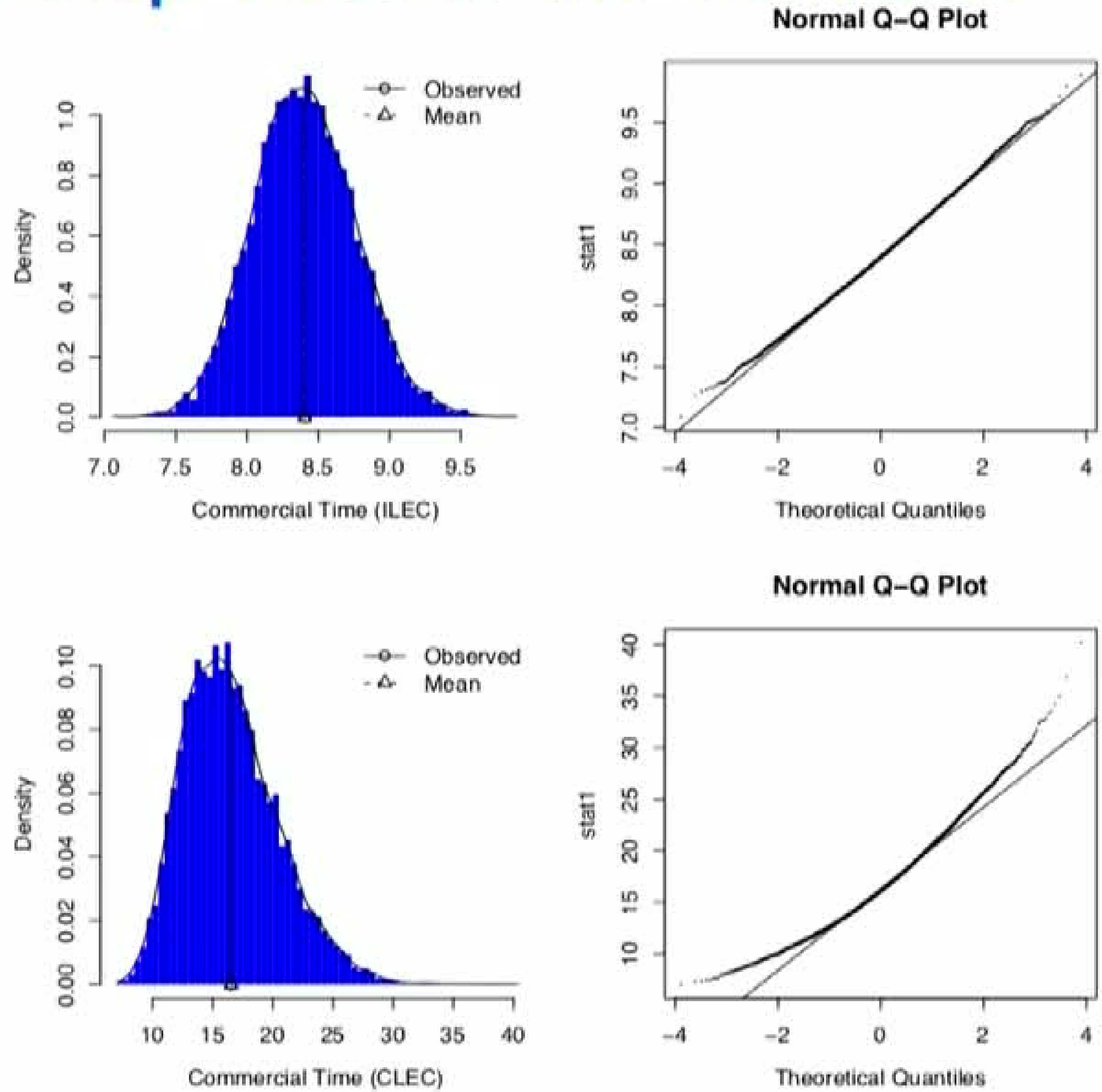
Standard deviation of the sampling distribution

Bootstrap bias =

Mean of the sampling distribution – Observed

```
Summary Statistics:
     Observed         SE      Mean      Bias
mean 16.50913 3.961816 16.53088  0.0217463
mean  8.41161 0.357599  8.40410 -0.0075031
```

Bootstrap Distns for Verizon



Bootstrap Standard Error

Google

Bootstrap standard error =

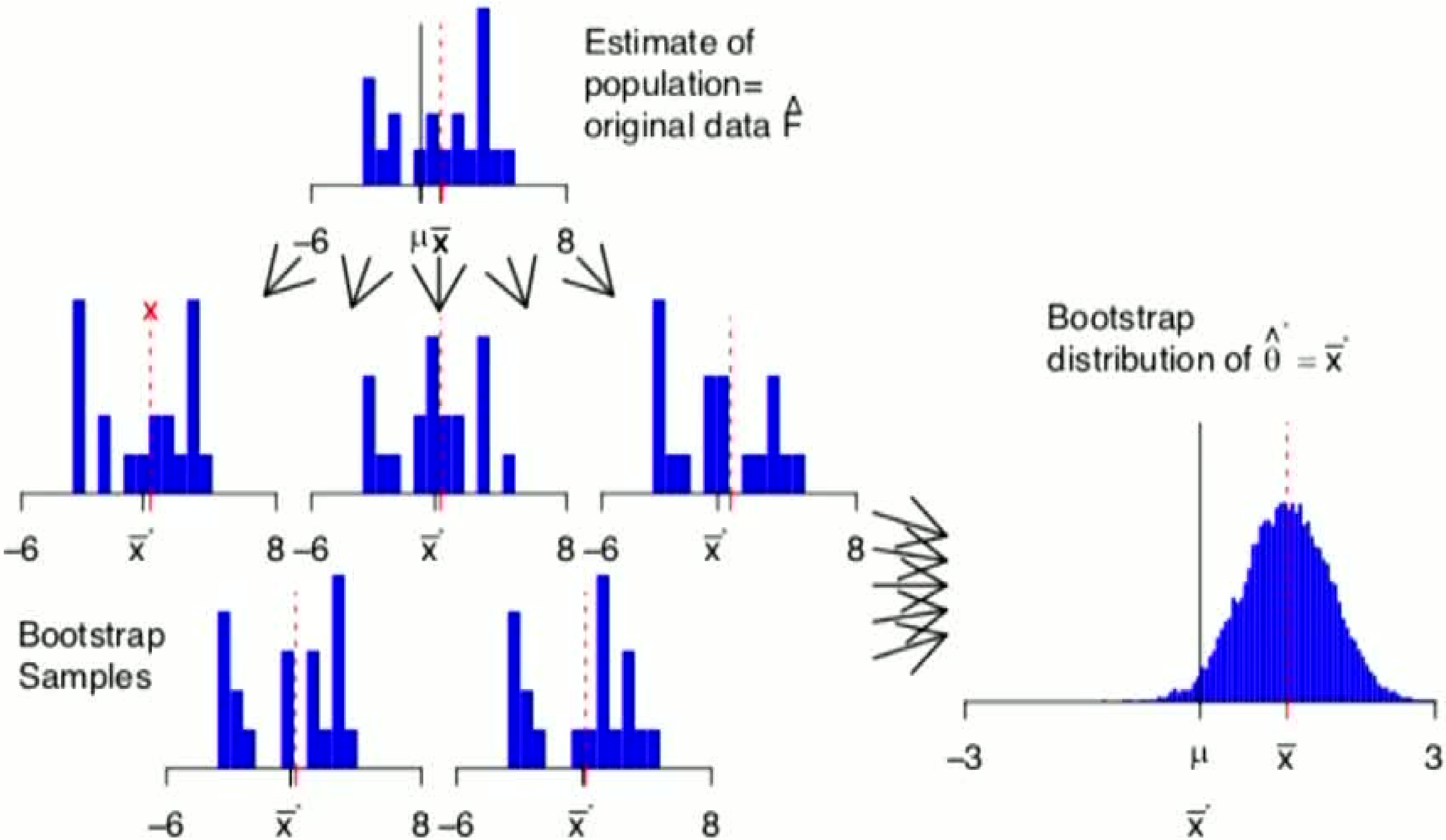
Standard deviation of the sampling distribution

Bootstrap bias =

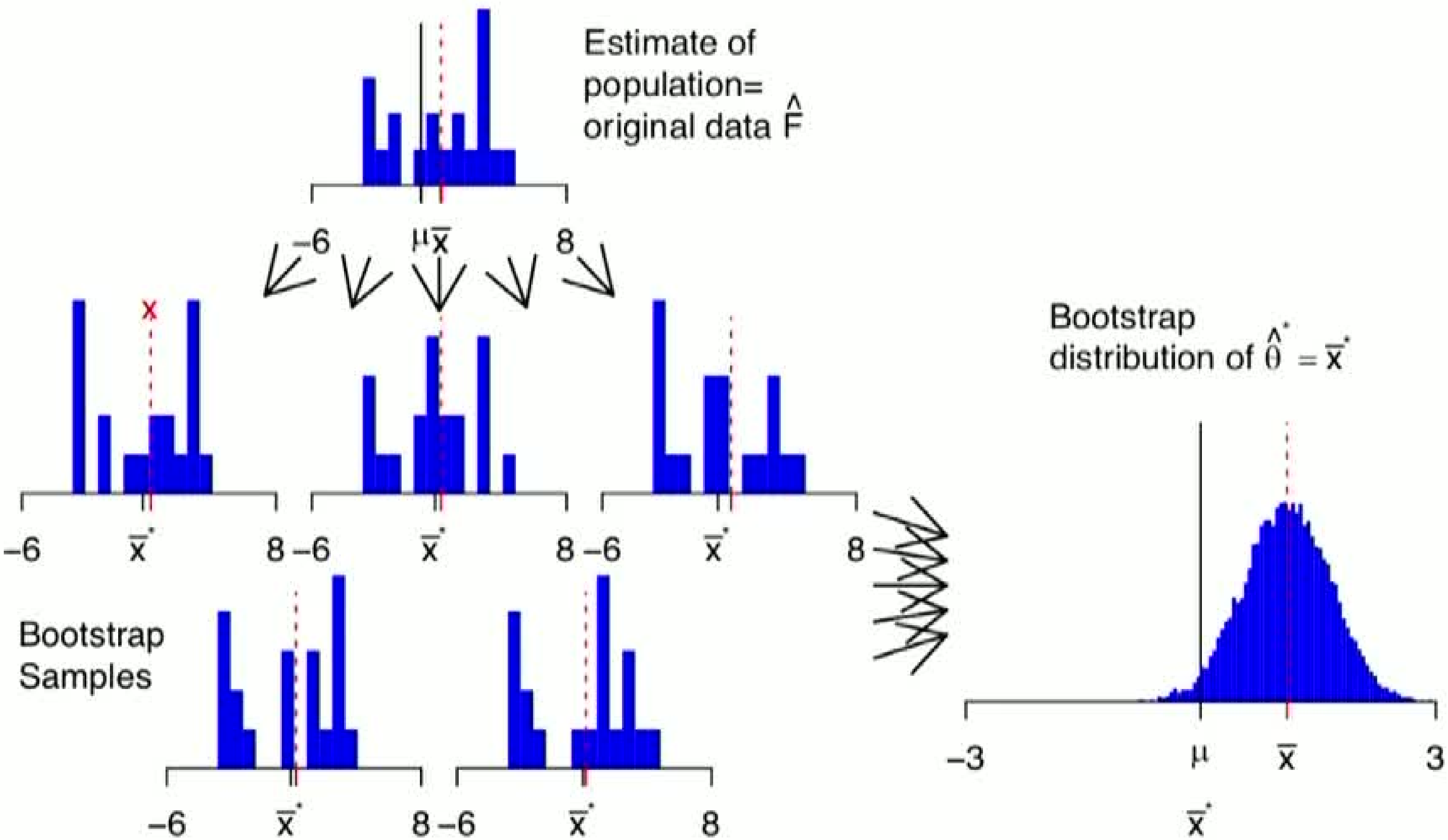
Mean of the sampling distribution – Observed

```
Summary Statistics:
     Observed          SE      Mean          Bias
mean 16.50913 3.961816 16.53088  0.0217463
mean  8.41161 0.357599  8.40410 -0.0075031
```

Bootstrap world



Bootstrap world



What to substitute?

Plug-in principle

- Underlying distribution is unknown
- Substitute your best guess

What to substitute?

- Empirical distribution – ordinary bootstrap
- Smoothed distribution – smoothed bootstrap
- Parametric distribution – parametric bootstrap
- Satisfy assumptions, e.g. null hypothesis

Fundamental Bootstrap Principle

Google

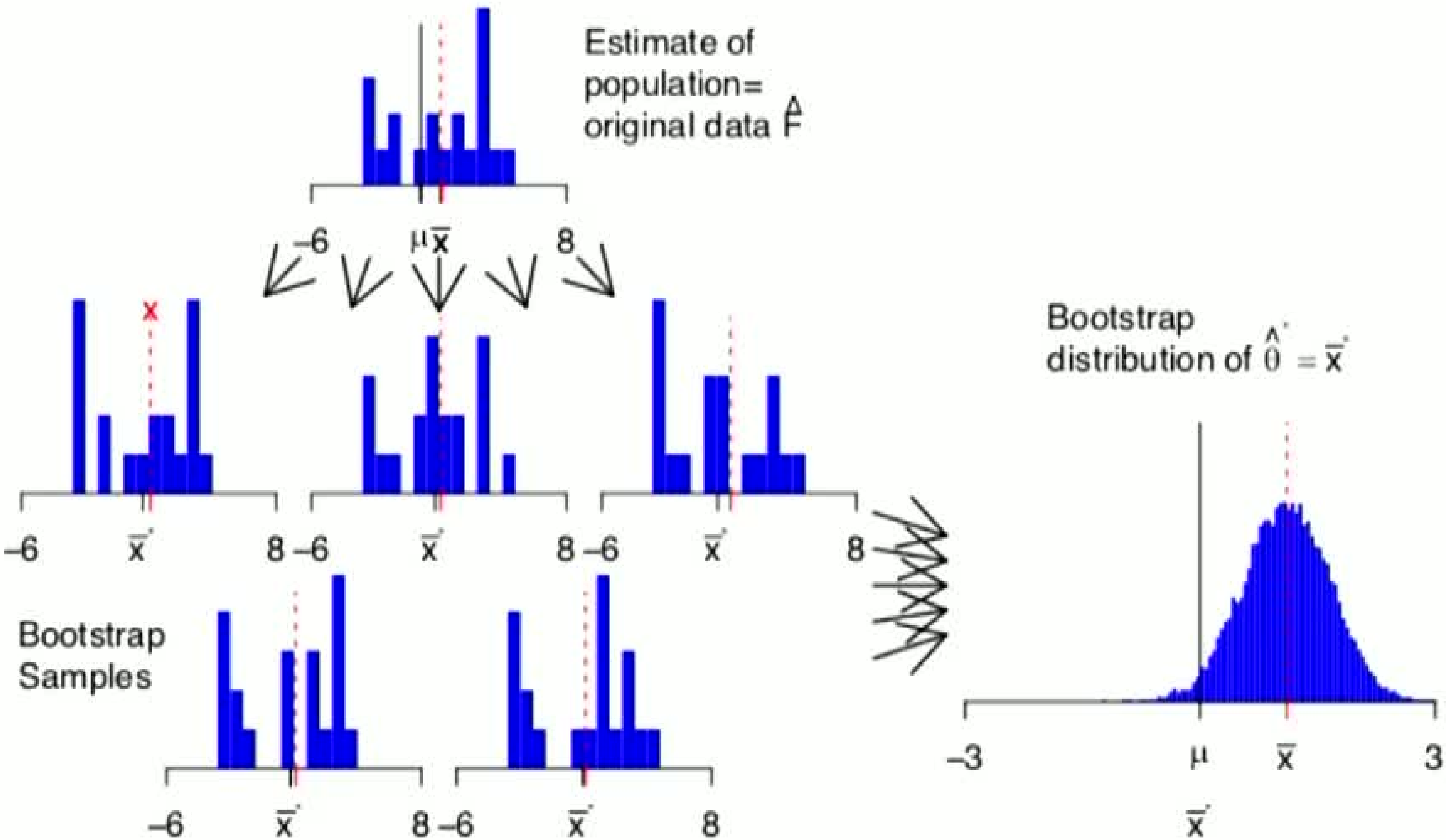
Plug-in principle

- Underlying distribution is unknown
- Substitute your best guess

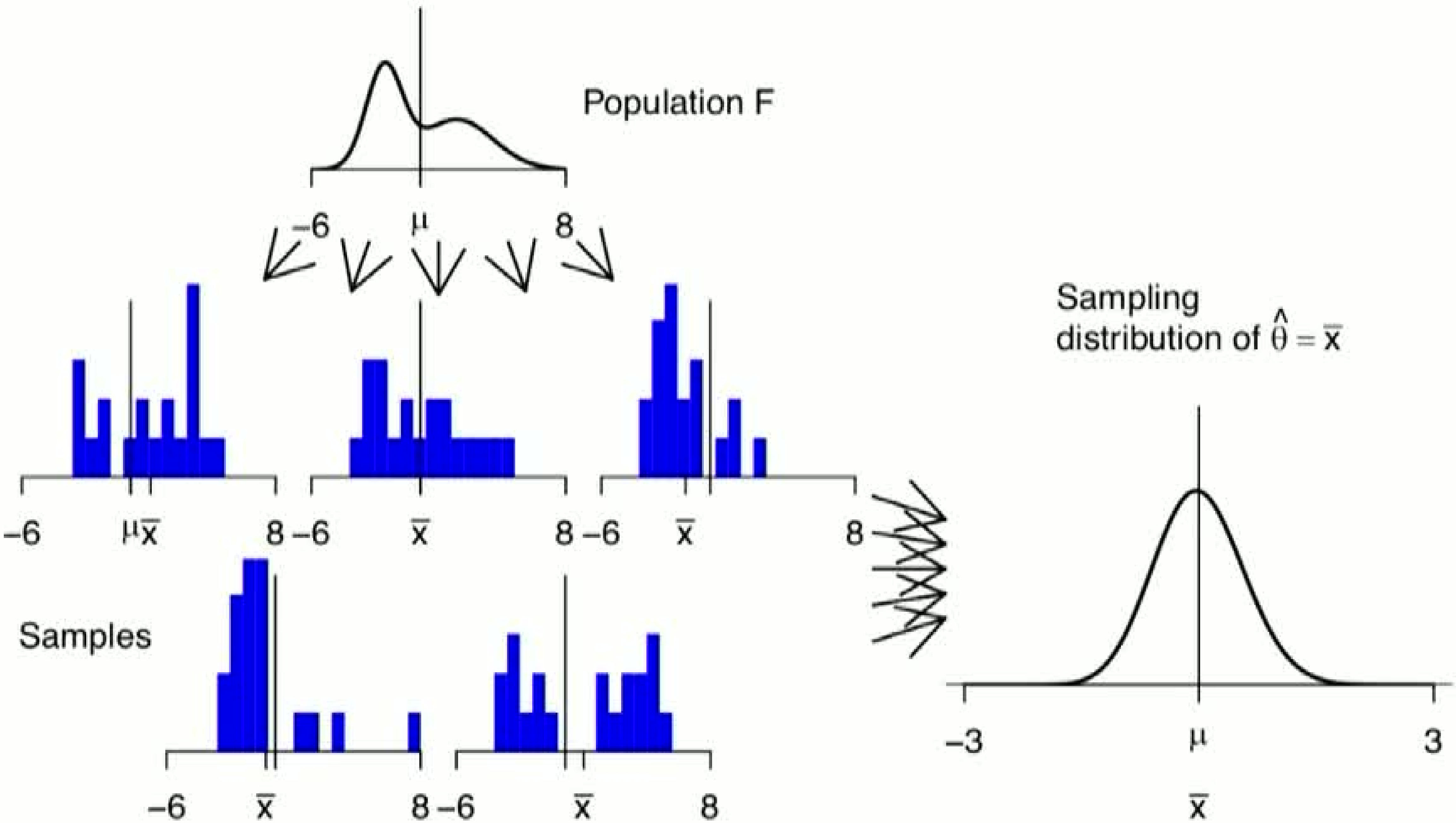
Fundamental Bootstrap Principle

- This substitution works ☺
- Not always ☹
  - Bootstrap distribution centered at statistic, not parameter
  - Too narrow for small *n*

Bootstrap world

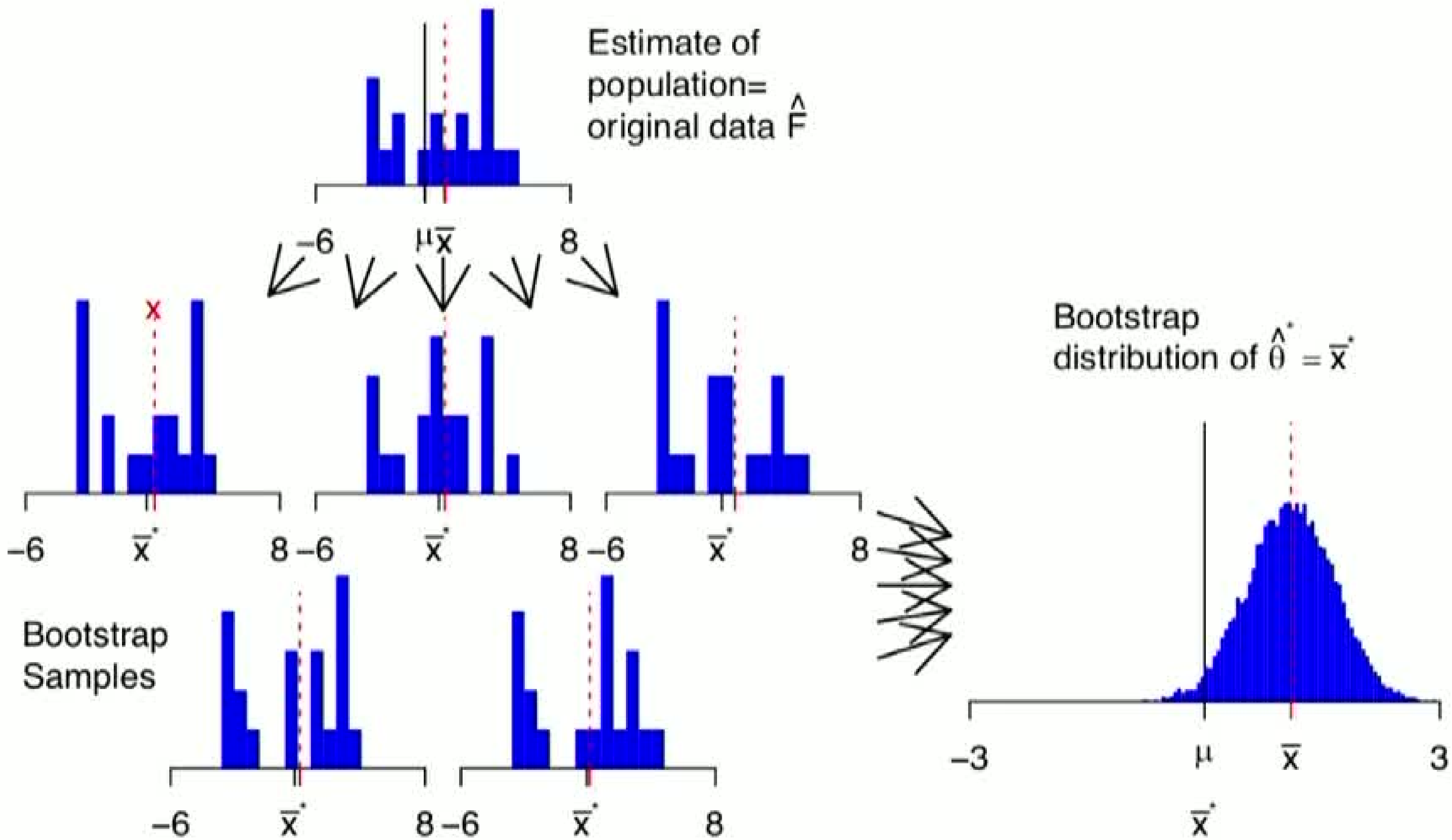


Ideal world



Population F
−6
μ
8
Sampling
distribution of $\hat{\theta} = \bar{x}$
−6
μ $\bar{x}$
8 −6
$\bar{x}$
8 −6
$\bar{x}$
8
Samples
−6
$\bar{x}$
8 −6
$\bar{x}$
8
−3
μ
3
$\bar{x}$

Bootstrap world



What to substitute?

Google

Plug-in principle

- Underlying distribution is unknown
- Substitute your best guess

What to substitute?

- Empirical distribution – ordinary bootstrap
- Smoothed distribution – smoothed bootstrap
- Parametric distribution – parametric bootstrap
- Satisfy assumptions, e.g. null hypothesis

Fundamental Bootstrap Principle

Google

Plug-in principle

- Underlying distribution is unknown
- Substitute your best guess

Fundamental Bootstrap Principle

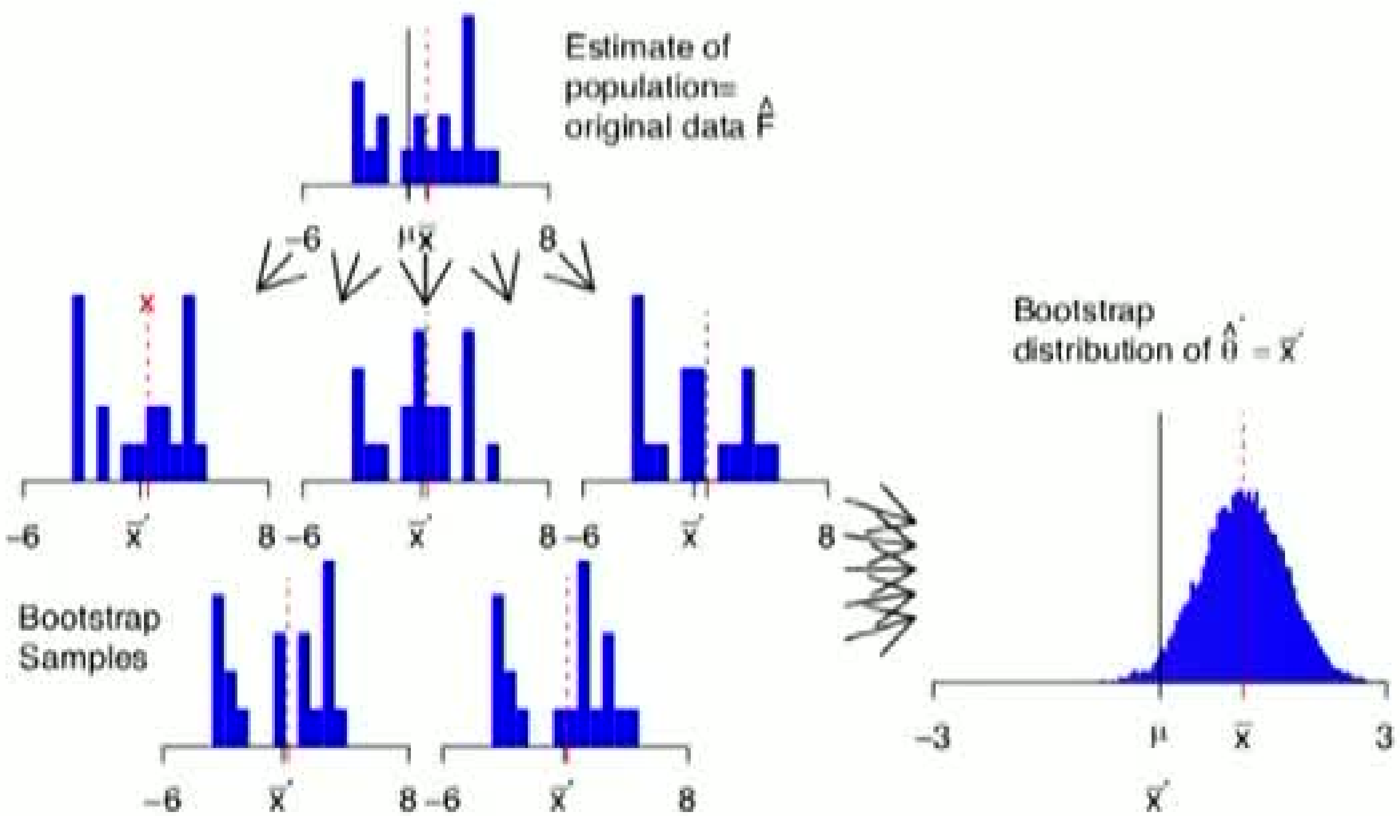
- This substitution works ☺
- Not always ☹
  - Bootstrap distribution centered at statistic, not parameter
  - Too narrow for small *n*

Bootstrap Implementation

Google

Nonparametric bootstrap

- Exact: n^n samples
- Monte Carlo, typically 10000 samples



Two-sample Bootstrap

Google

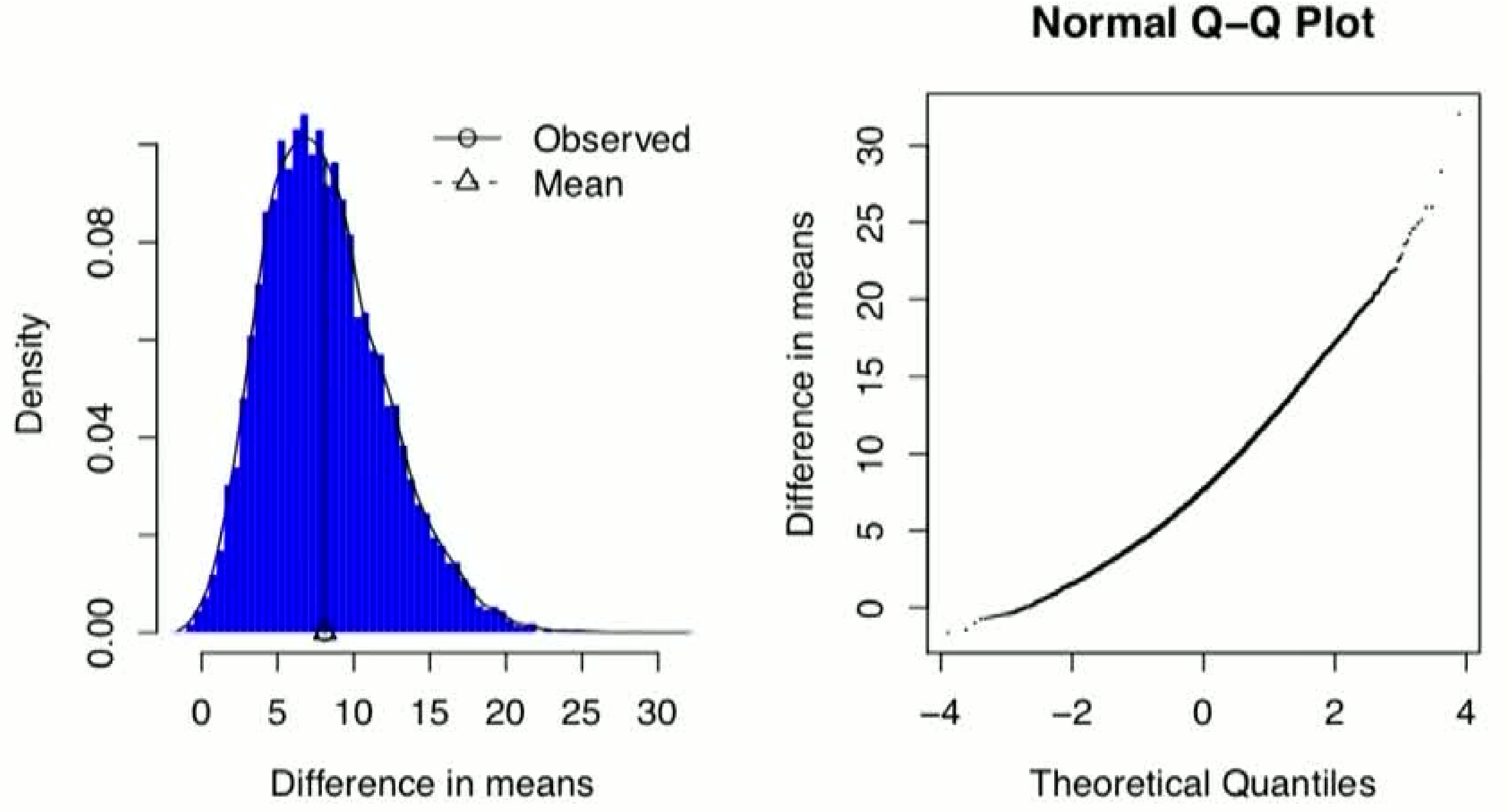
Given independent SRSs from two populations:

Repeat 10000 times

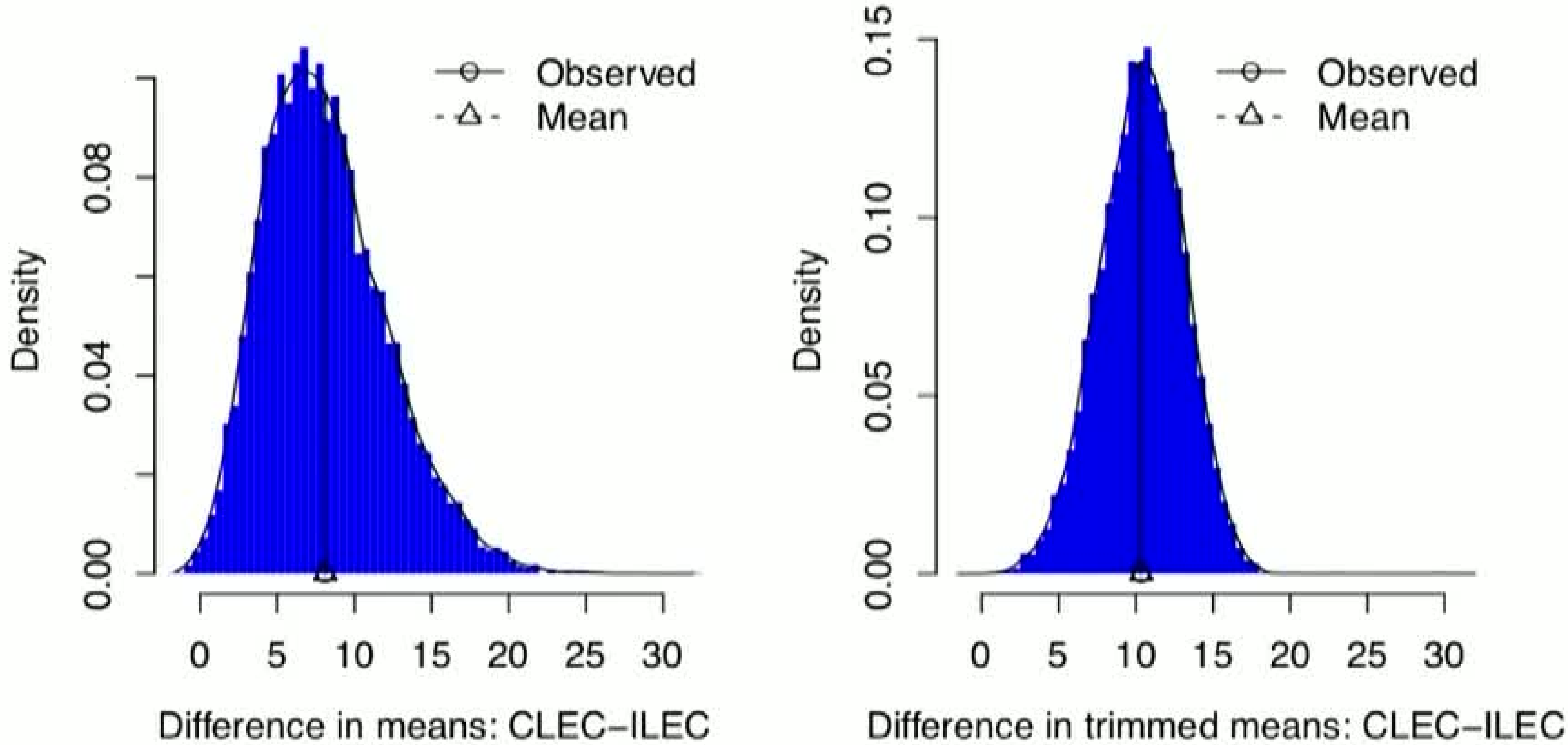
- Draw sample size n_1 from sample 1
- Draw sample size n_2 from sample 2, independently
- Compute statistic, e.g. difference in means

The 10000 bootstrap statistics comprise the bootstrap distribution.

Bootstrap Distns for Verizon



Trimmed Means



	Observed	SE	Mean	Bias
Difference in means	8.097	3.979	8.113	0.01594
Difference in trimmed means	10.336	2.728	10.35	0.02298

Confidence Intervals

Google

Quick & Dirty:

Bootstrap Percentile Interval =
Middle 95% of the bootstrap distribution
(1.63, 17.00)

t interval (with bootstrap SE) =
Observed $\pm$ 2 SE = (-0.16, 16.35)

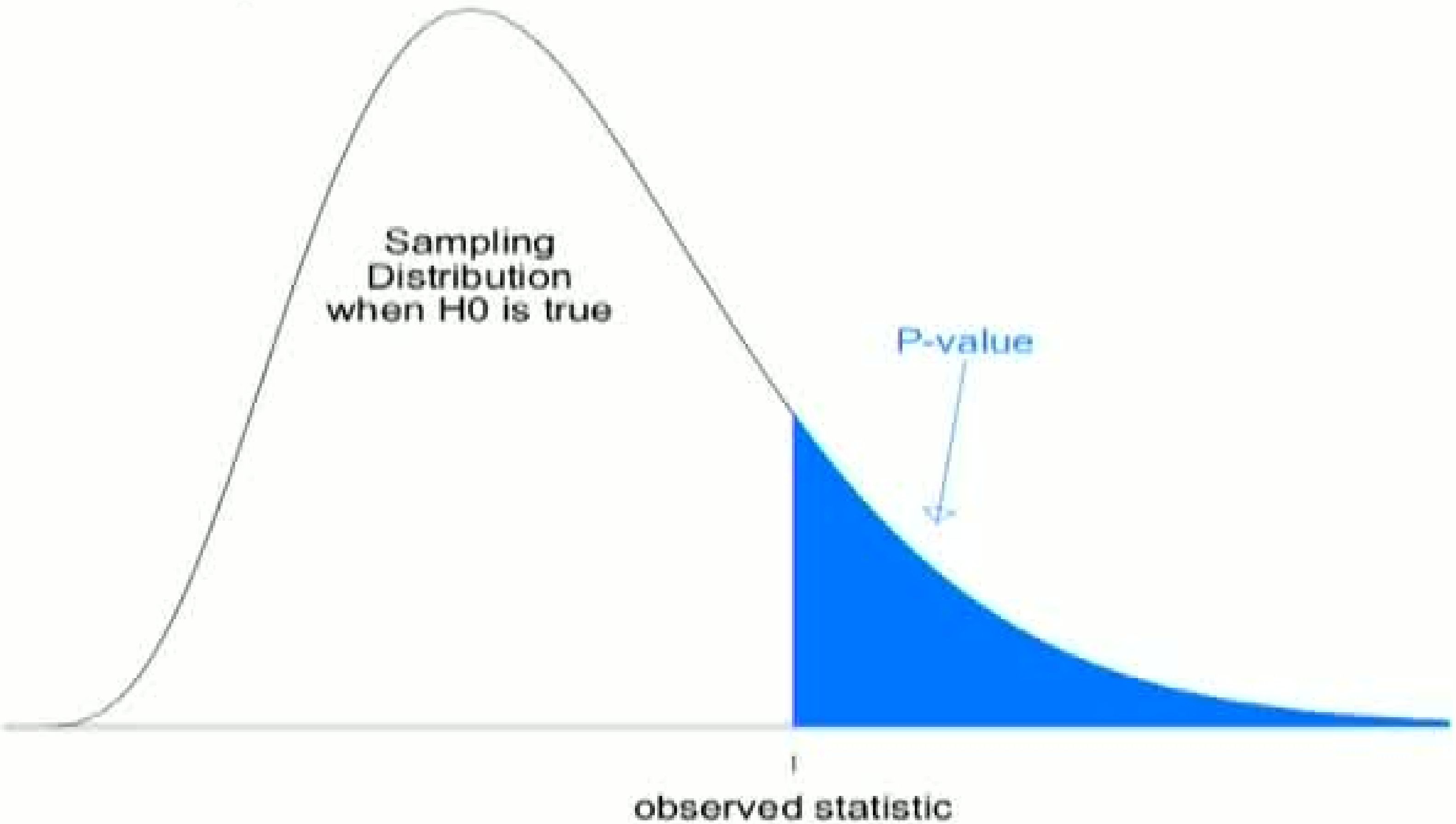
Later: better interval – bootstrap *t*

Google

Resampling for Hypothesis Tests

Sample in a manner consistent with H_0

P-value = P_0(random value exceeds observed)



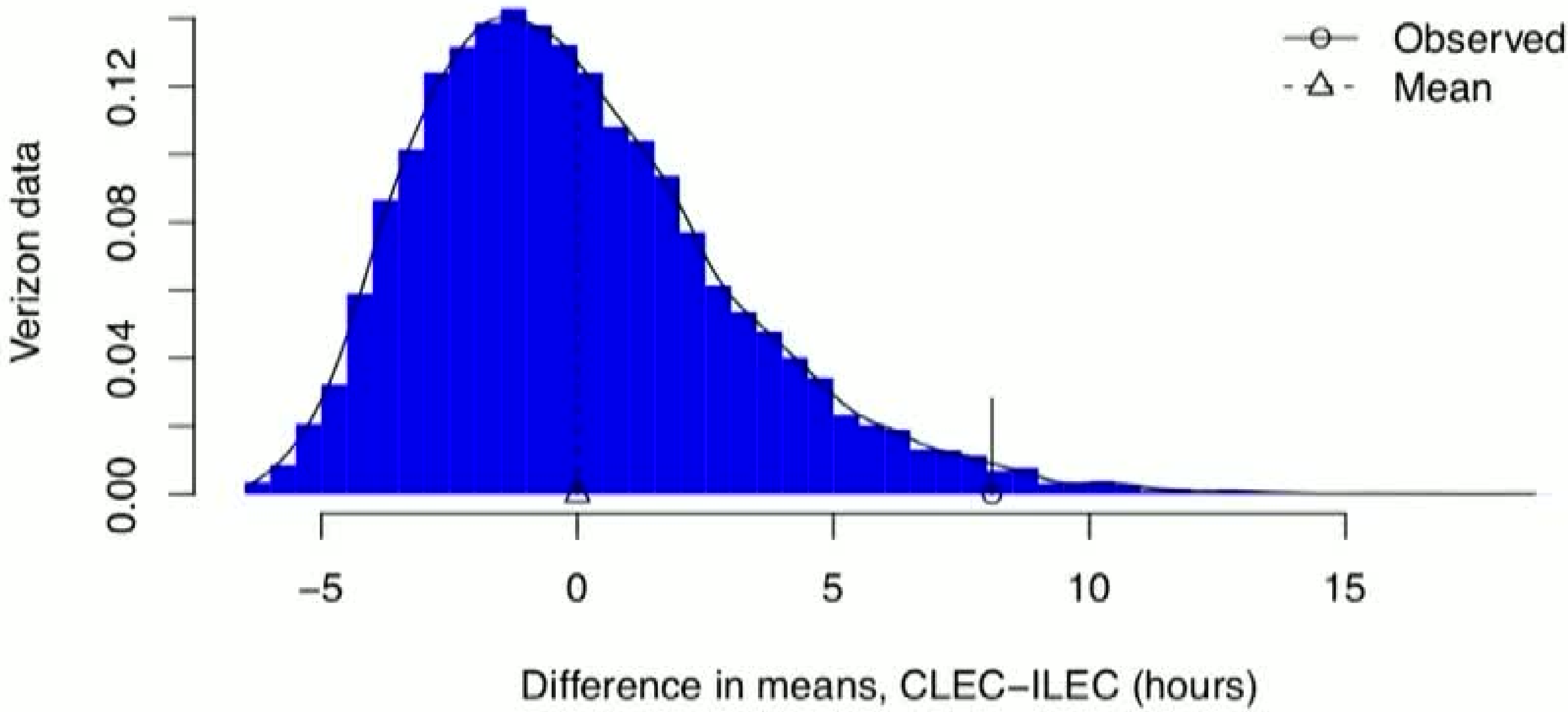
Permutation Test for 2-samples

H_0: no real difference between groups; observations could come from one group as well as the other

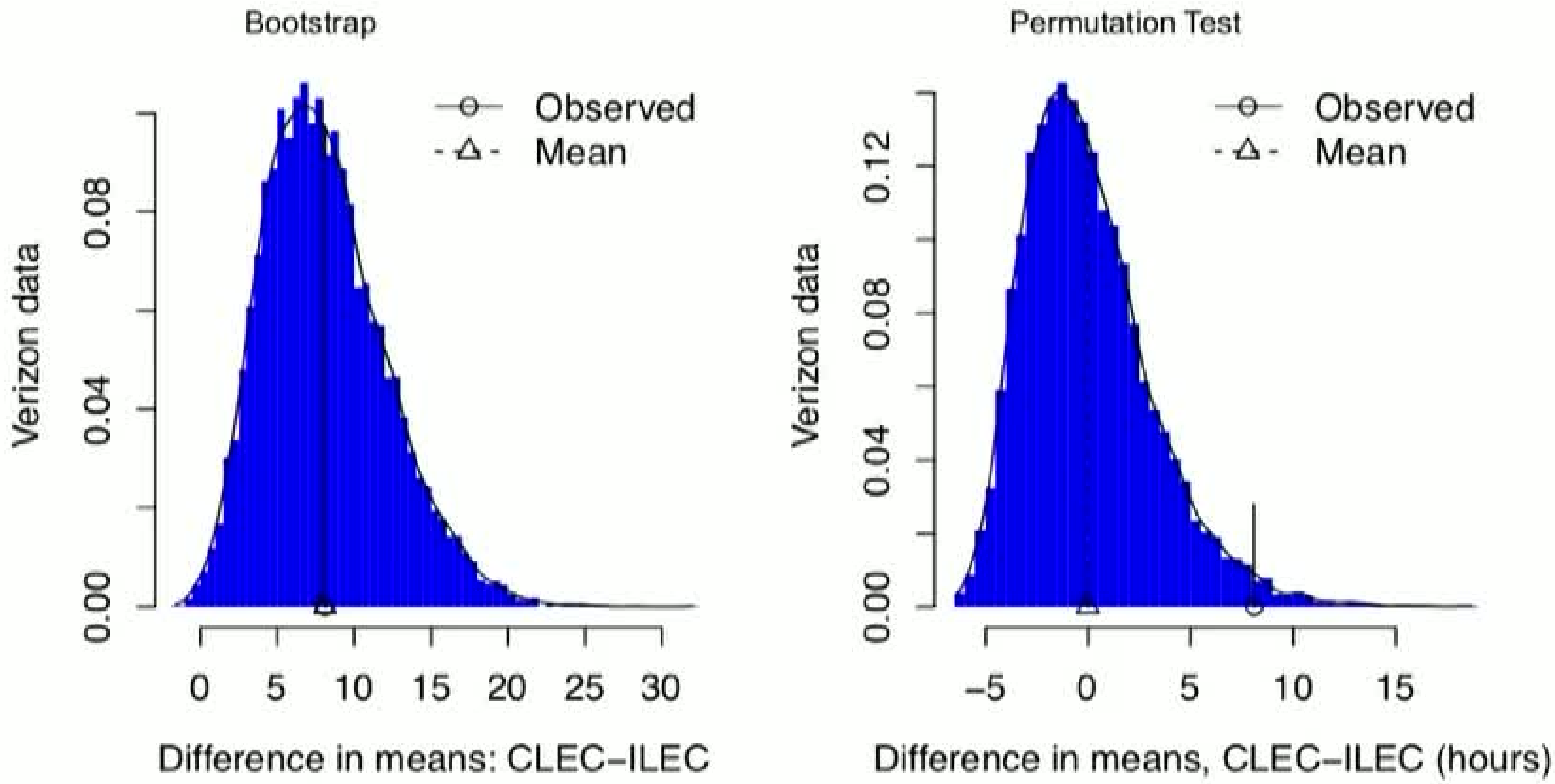
Resample: randomly choose n_1 observations for group 1, rest for group 2.

Equivalent to permuting all n, first n_1 into group 1.

Verizon permutation test



Bootstrap & Permutation



Permutation Test P-value

Google

Results

```
Replications: 9999
Two samples, sample sizes are 23 1664

Summary Statistics for the difference between samples 1 and 2:
                 Observed          Mean Alternative PValue
mean: CLEC-ILEC  8.09752 -0.0006686392     greater 0.0171
```

Comment: Work directly with difference in means; don't need *t*-statistic. More natural.

Assumptions

Permutation Test:

- Same distribution for two populations
  - When H_0 is true
  - Population variances must be the same; sample variances may differ
- Does not require normality
- Does not require that data be a random sample from a larger population

Permutation vs Pooled Bootstrap

Google

Pooled bootstrap test

- Pool all *n* observations
- Choose n_1 *with* replacement for group 1
- Choose n_2 *with* replacement for group 2

Permutation test is better

- Same number of outliers as observed data

Easy to Understand

Google

Concrete Analogs to Abstract Concepts:

- Sampling variability
- Standard Error
- Bias
- Confidence Interval
- *P*-value for a significance test
- Central Limit Theorem (normality)

Outline

- Case Study, Basics
- Accuracy
  - Outrageous example
  - Pictures
    - Large *n*
    - Small *n*
    - Sample Median
    - Skewness
  - How many resamples
  - Coverage
    - Small *n*
    - Skewness

Meta goals:

Understand when bootstrap works or not.

How accurate is the bootstrap?

How accurate are formula methods?

SE of Variance Estimates

Google

- CLEC data – skewed
- Verizon CLEC data: n=23, s^2=380
- Classical
  - chi-square CI: (227, 762)
  - SE = 114 (Assume normality)
- Bootstrap
  - Percentile CI: (59, 931)
  - SE = 267 (Don't assume normality)
- George Box: *To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!*

Two Sources of Variation in Bootstrap Distribution

Google

Original sample is chosen randomly from the population

Bootstrap samples are chosen randomly from the original sample: "*Monte Carlo implementation variability*"

Permutation vs Pooled Bootstrap

Pooled bootstrap test

- Pool all *n* observations
- Choose n_1 *with* replacement for group 1
- Choose n_2 *with* replacement for group 2

Permutation test is better

- Same number of outliers as observed data

Permutation vs Pooled Bootstrap

Google

Pooled bootstrap test

- Pool all *n* observations
- Choose n_1 *with* replacement for group 1
- Choose n_2 *with* replacement for group 2

Permutation test is better

- Same number of outliers as observed data

Permutation Test P-value

Google

Results

```
Replications: 9999
Two samples, sample sizes are 23 1664

Summary Statistics for the difference between samples 1 and 2:
                  Observed          Mean Alternative PValue
mean: CLEC-ILEC  8.09752 -0.0006686392     greater 0.0171
```

Comment: Work directly with difference in means; don't need *t*-statistic. More natural.

Permutation vs *t* test

T tests

```
Pooled-variance t-test
   t = -2.6125, df = 1685, p-value = 0.0045

Non-pooled-variance t-test
   t = -1.9834, df = 22.3463548265907, p-value = 0.0299
```

Permutation test

```
Number of Replications: 499999

Summary Statistics:
    Observed      Mean    SE alternative p.value
Var   -8.098 -0.001288 3.105        less 0.01825
```

Assumptions

Permutation Test:

- Same distribution for two populations
  - When H_0 is true
  - Population variances must be the same; sample variances may differ
- Does not require normality
- Does not require that data be a random sample from a larger population

Two Sources of Variation in Bootstrap Distribution

Original sample is chosen randomly from the population

Bootstrap samples are chosen randomly from the original sample: "*Monte Carlo implementation variability*"

Two Sources of Variation in Bootstrap Distribution

Google

Original sample is chosen randomly from the population

Bootstrap samples are chosen randomly from the original sample: "*Monte Carlo implementation variability*"

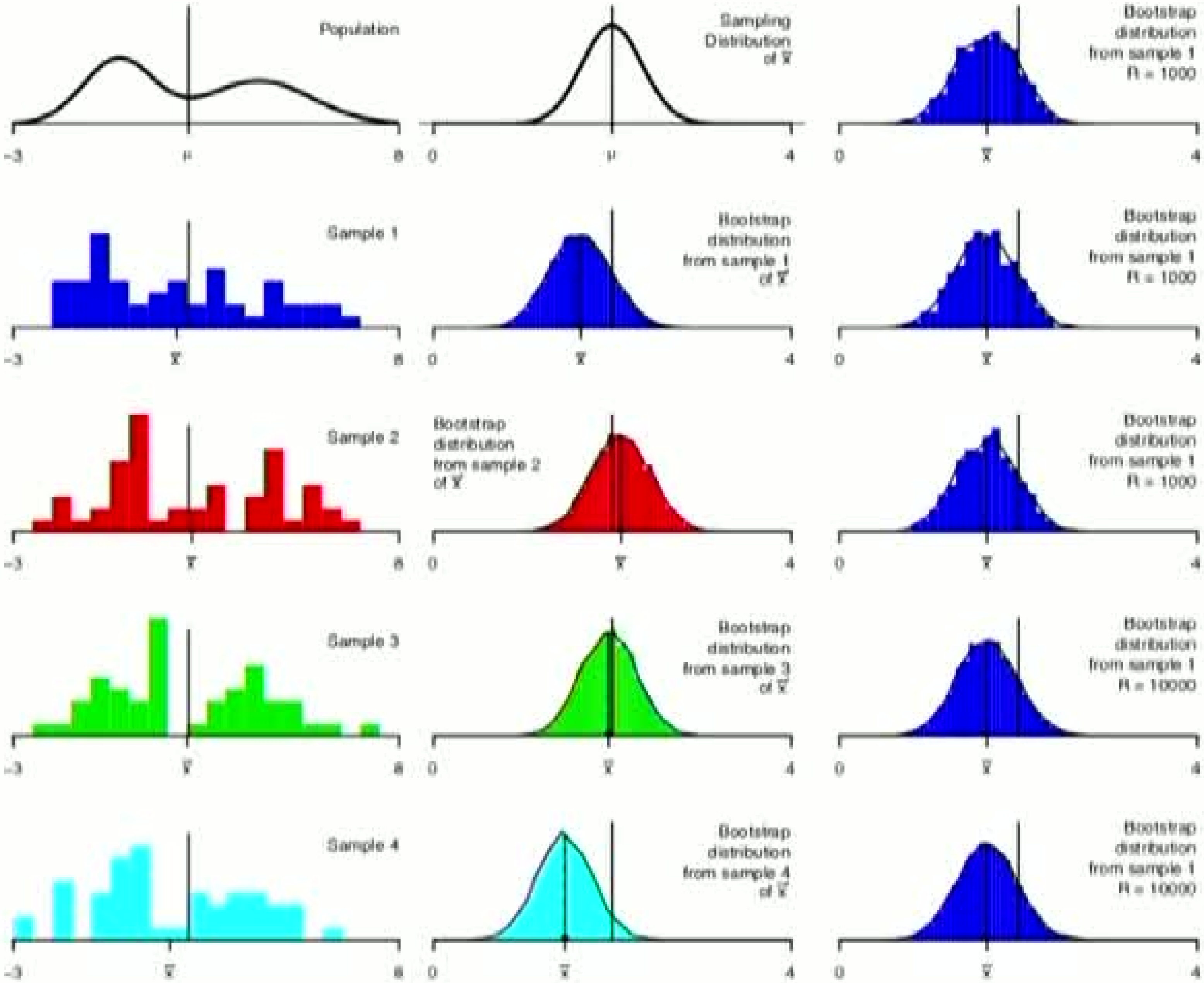
Large samples

Bootstrap distributions are centered close to statistic values

Shape and spread of bootstrap distribution vary *a bit*, due to shape and spread of original sample

With $R = 10^4$, MC variability is small

Large samples



Large samples

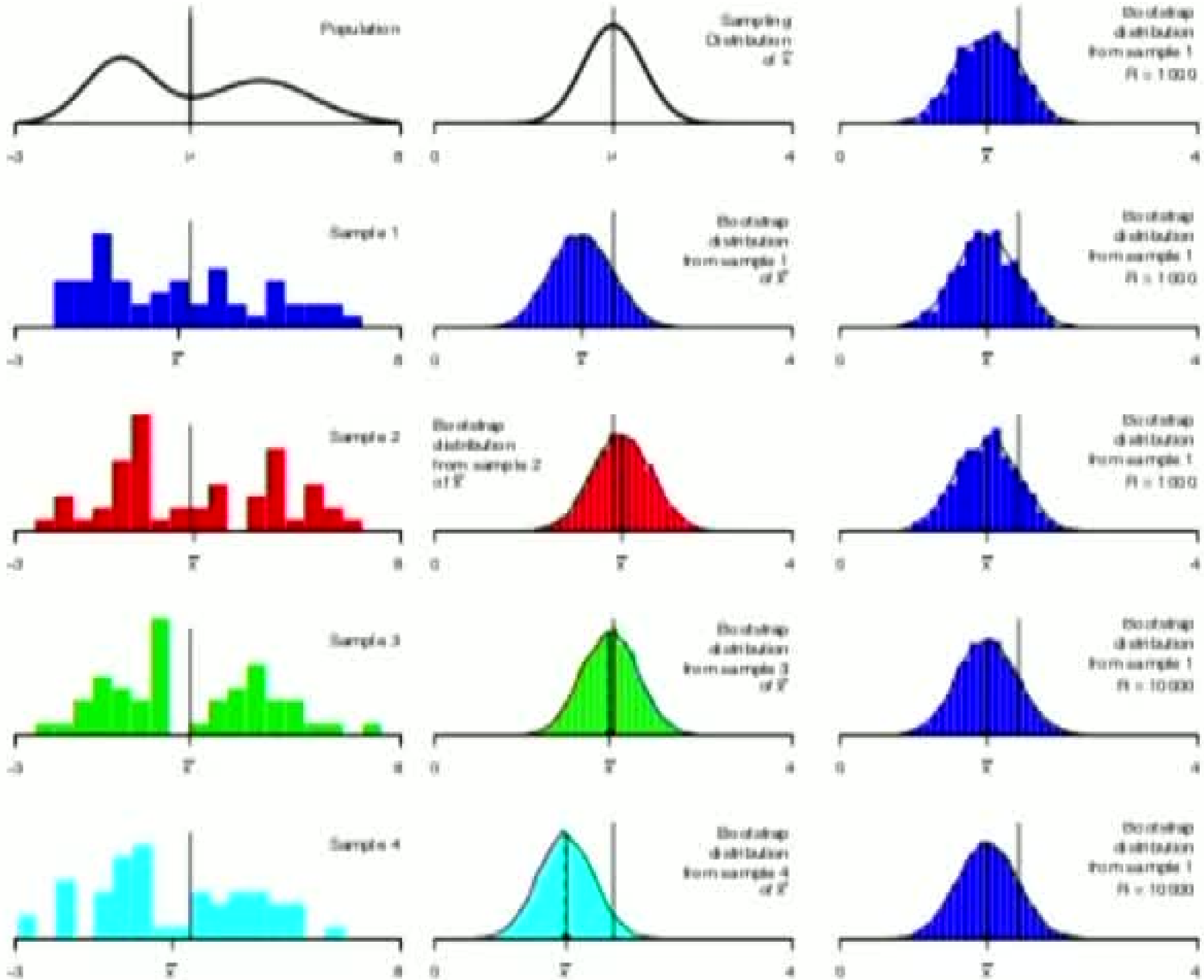
Google

Bootstrap distributions are centered close to statistic values

Shape and spread of bootstrap distribution vary *a bit*, due to shape and spread of original sample

With $R = 10^4$, MC variability is small

Large samples



Small samples

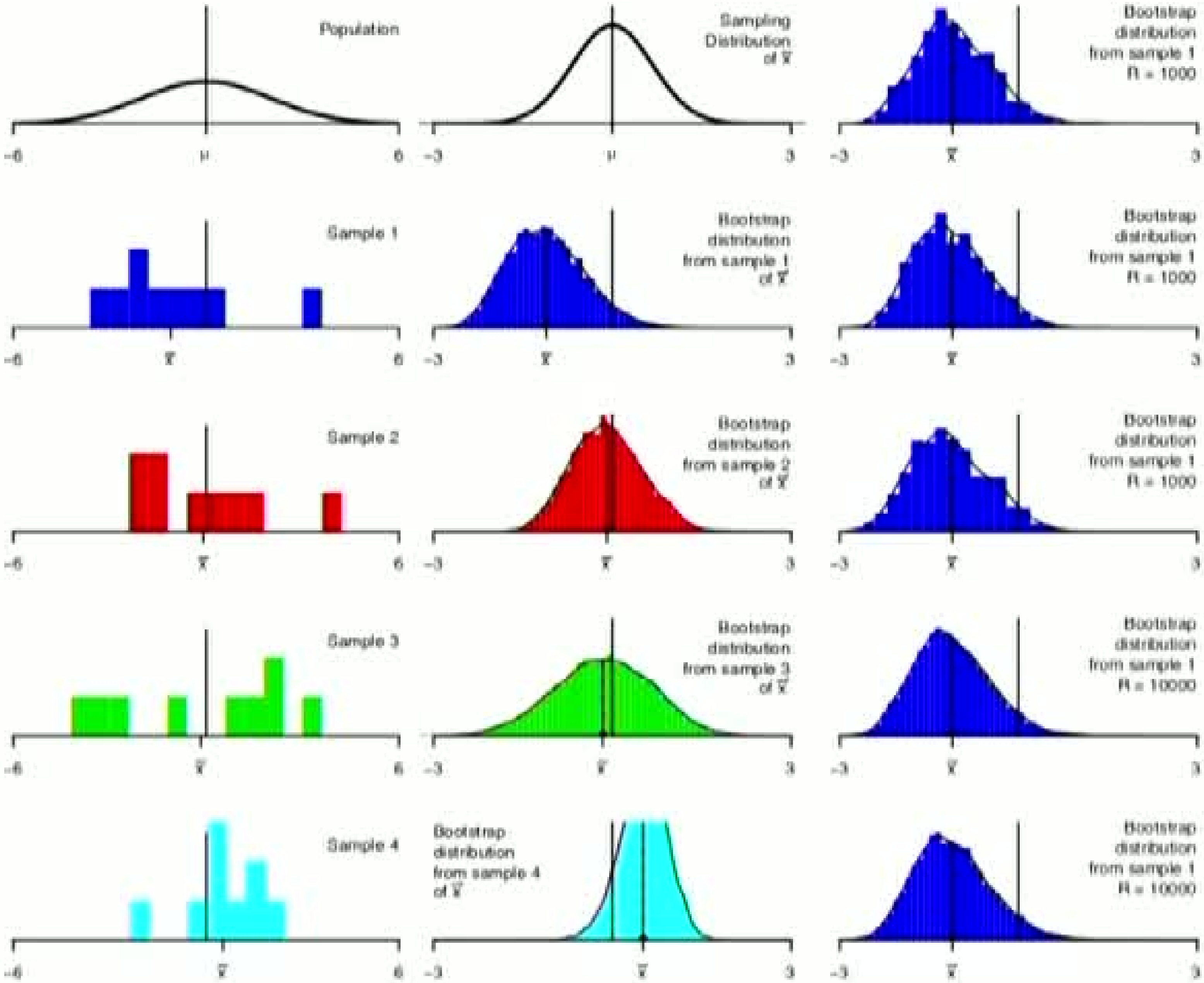
Bootstrap distributions are centered close to statistic values

Shape and spread of bootstrap distn vary *substantially*, due to shape and spread of original sample

Bootstrap distributions too narrow

With $R = 10^4$, MC variability is small

Small samples



Sample Median

Google

Bootstrap distribution poor estimate of sampling distribution (except for large *n*)

Depends heavily on middle values of empirical distribution

Empirical distn is discrete (underlying distn continuous?).

Remedy: smoothed bootstrap

With $R = 10^4$, MC variability is small

Skewed Population

Mean-variance relationship; spread of bootstrap distribution depends on statistic

Need to adjust confidence intervals

- Bootstrap percentile is better than t
- But still not good enough

Remedy: bootstrap-*t*

$$t = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{s_{\hat{\theta}^*}} = \frac{\overline{x}^* - \overline{x}}{s^*/\sqrt{n}}$$

Skewed Population

Google

Mean-variance relationship; spread of bootstrap distribution depends on statistic

Need to adjust confidence intervals

- Bootstrap percentile is better than t
- But still not good enough

Remedy: bootstrap-*t*

$$t = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{s_{\hat{\theta}^*}} = \frac{\overline{x}^* - \overline{x}}{s^*/\sqrt{n}}$$

Skewed Population

Google

Mean-variance relationship; spread of bootstrap distribution depends on statistic

Need to adjust confidence intervals

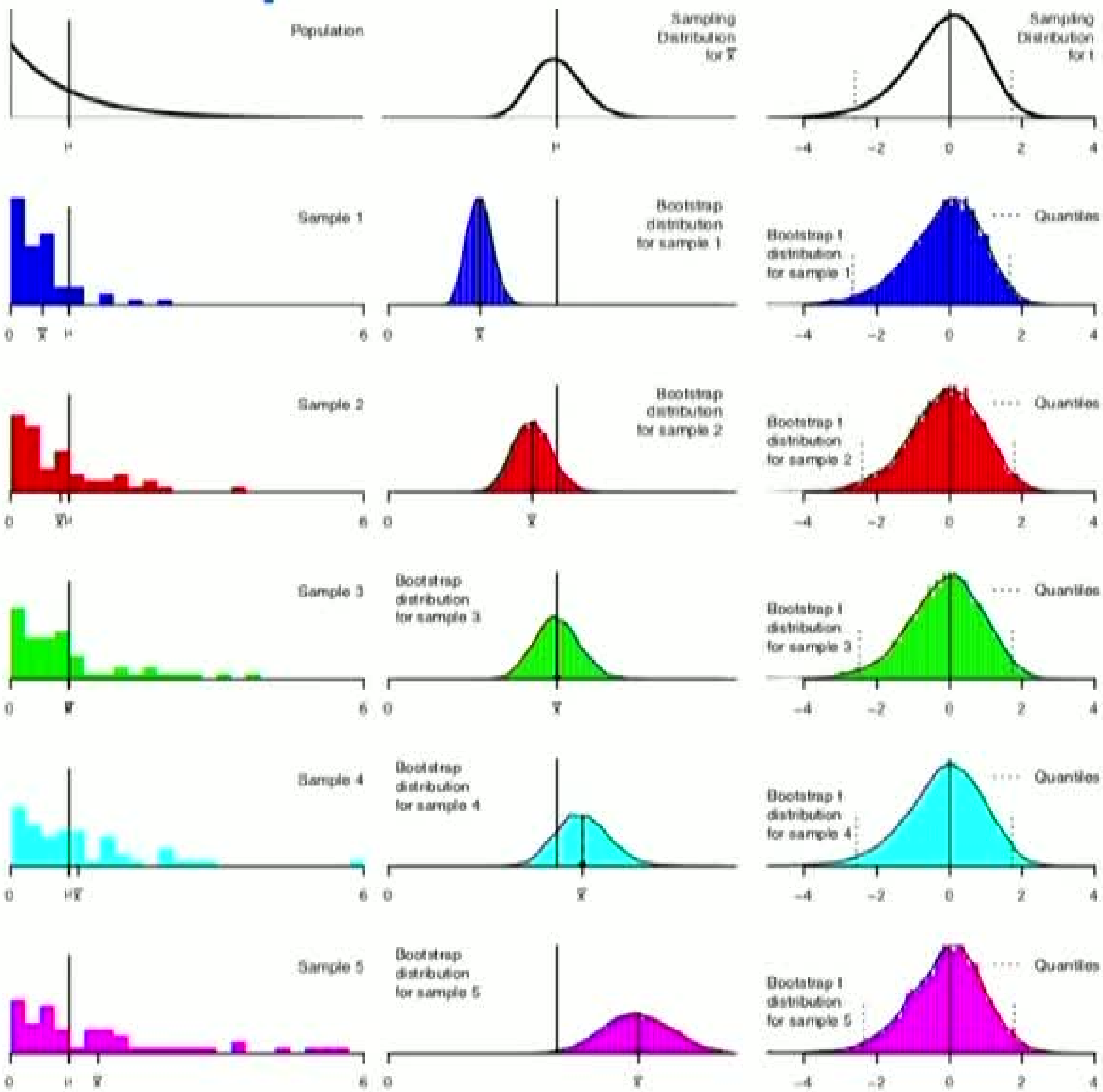
- Bootstrap percentile is better than t
- But still not good enough

Remedy: bootstrap-*t*

$$t = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{s_{\hat{\theta}^*}} = \frac{\overline{x}^* - \overline{x}}{s^*/\sqrt{n}}$$

Skewed Population



Skewed Population

Google

Mean-variance relationship; spread of bootstrap distribution depends on statistic

Need to adjust confidence intervals

- Bootstrap percentile is better than t
- But still not good enough

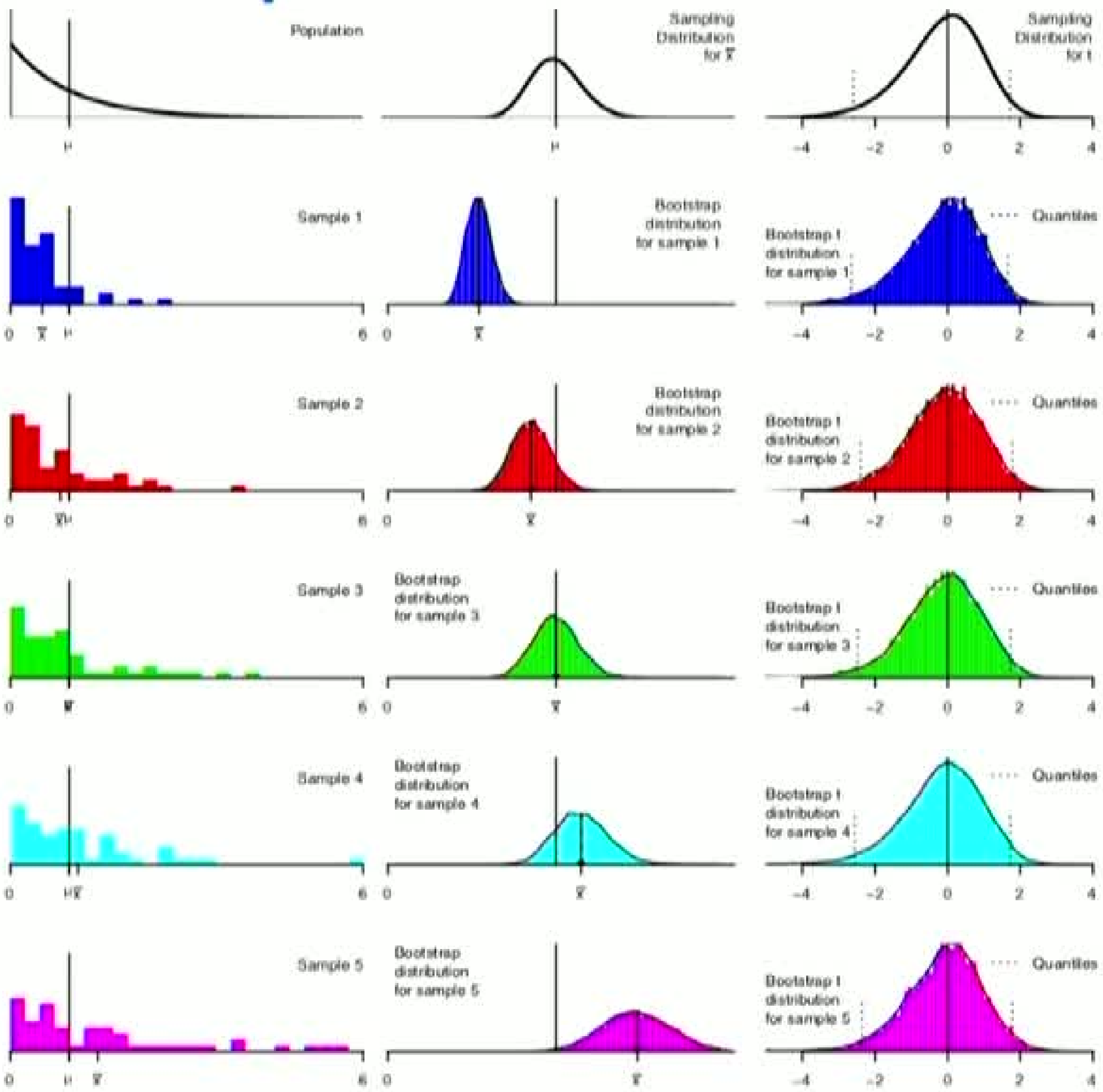
Remedy: bootstrap-*t*

$$t = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{s_{\hat{\theta}^*}} = \frac{\overline{x}^* - \overline{x}}{s^*/\sqrt{n}}$$

Skewed Population

Google



Skewed Population

Google

Mean-variance relationship; spread of bootstrap distribution depends on statistic

Need to adjust confidence intervals

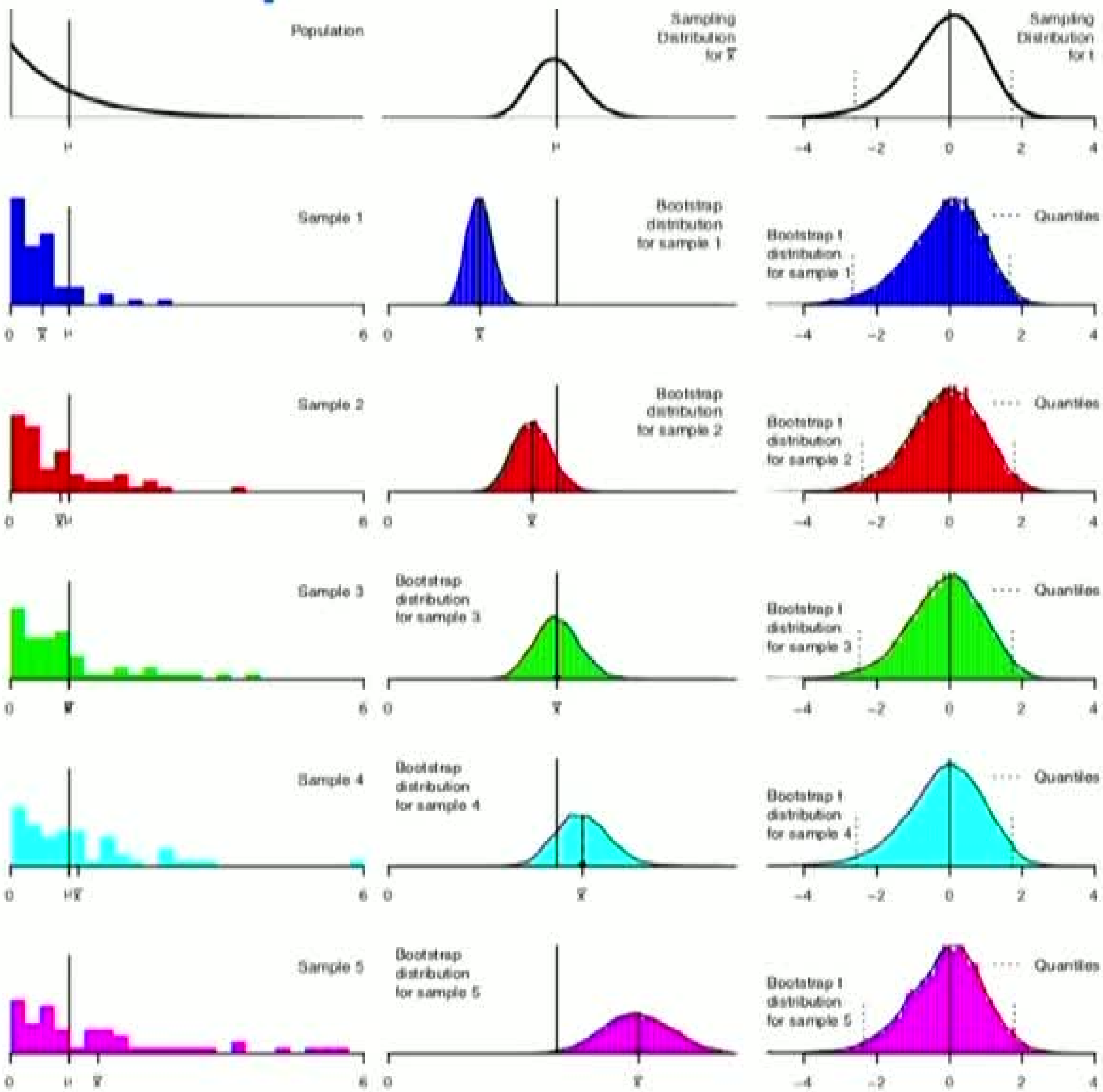
- Bootstrap percentile is better than t
- But still not good enough

Remedy: bootstrap-*t*

$$t = \frac{\hat\theta - \theta}{s_{\hat\theta}} = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

$$t^* = \frac{\hat\theta^* - \hat\theta}{s_{\hat\theta^*}} = \frac{\overline{x}^* - \overline{x}}{s^*/\sqrt{n}}$$

Skewed Population



How many bootstrap samples?

Sample of size R from theoretical n^n distribution

Accuracy improves at rate $1/\sqrt{R}$

How many

- Quick and dirty: 1000
- Better: 10000

Why more?

- Faster computers
- Treat data as fixed; small MC variation

Accuracy for small *n*

Percentile CI ~ $\overline{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{(n-1)/n}$

Too narrow!

bootstrap SE for mean = $\hat{\sigma}/\sqrt{n}$

where $\hat{\sigma}^2 = (1/n) \sum (x_i - \overline{x})^2$

Expanded Percentile Interval:

Adjusted percentiles ($\alpha'/2$, 1-$\alpha'/2$)

Pick α' to mimic usual *t* interval

Accuracy for skewness

Google

Much worse!

Not only problem for bootstrap – ordinary *t* intervals are poor.

Bootstrap *t* works well.

Common Statistical Practice

Google

Is n >= 30, and data not too skewed?

- Yes: use t intervals and tests
- No: use them anyway ☹

How big a problem is this?

Common Statistical Practice

Google

Is *n* >= 30, and data not too skewed?

- Yes: use *t* intervals and tests
- No: use them anyway ☹

How big a problem is this?

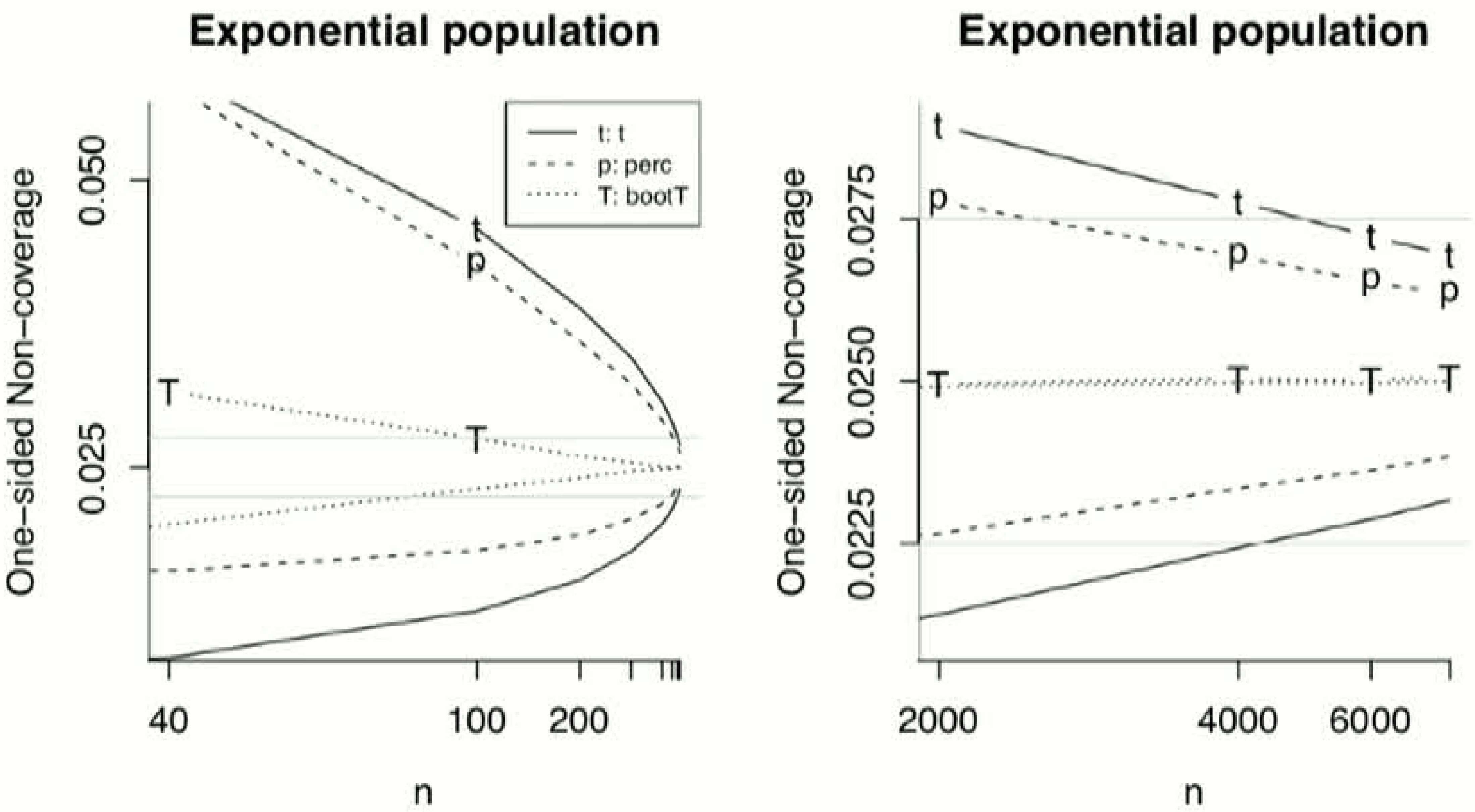
Common Statistical Practice

Is *n* >= 30, and data not too skewed?

- Yes: use *t* intervals and tests
- No: use them anyway ☹

How big a problem is this?

Need n > 5000 for t accuracy!



Why isn't this known?

Google

Inertia

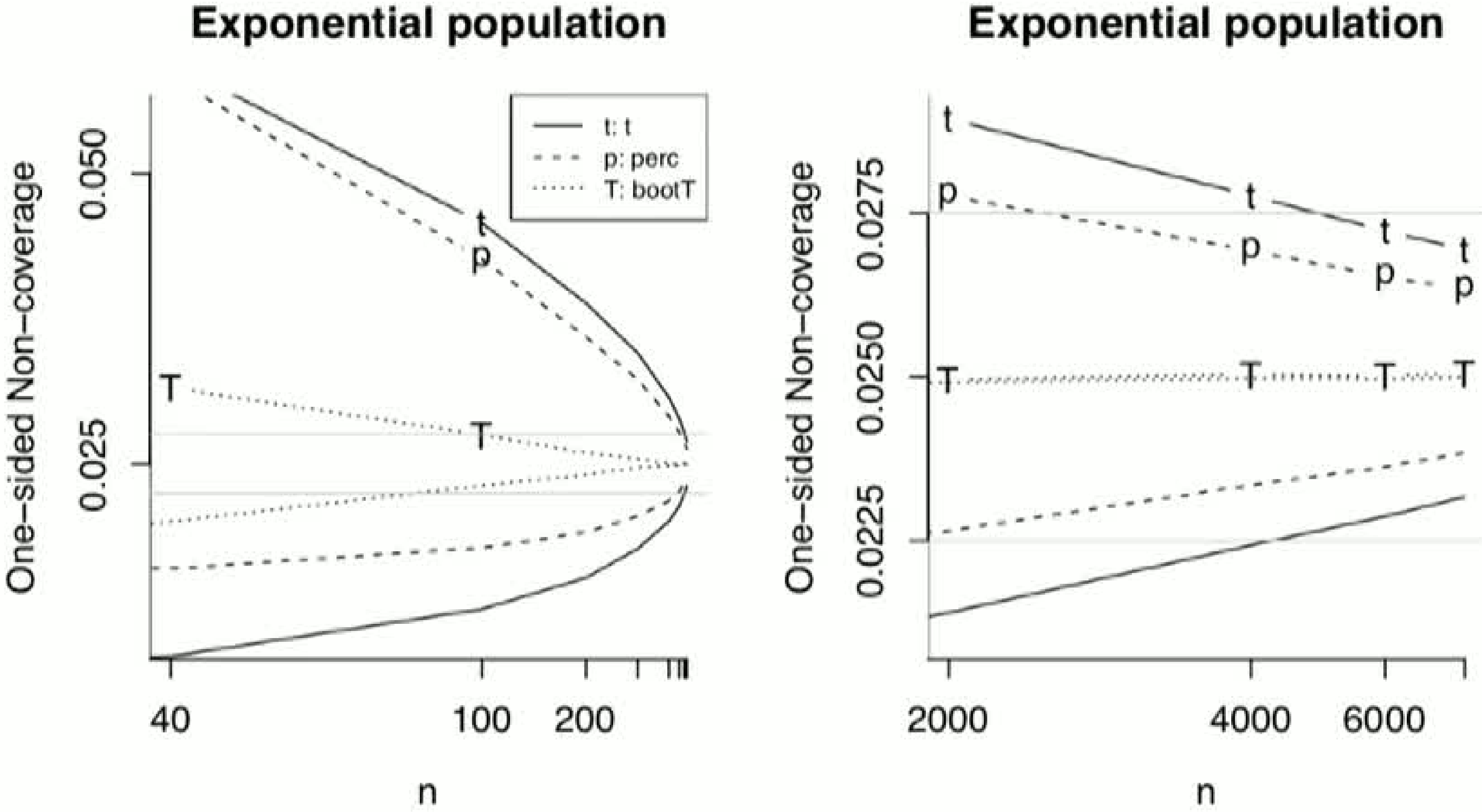
Need bootstrap diagnostics

Need 64,000 bootstrap samples

Need fast computers

- 20,000 hours for simulations in 1981

Need n > 5000 for t accuracy!



Why isn't this known?

Google

Inertia

Need bootstrap diagnostics

Need 64,000 bootstrap samples

Need fast computers

- 20,000 hours for simulations in 1981

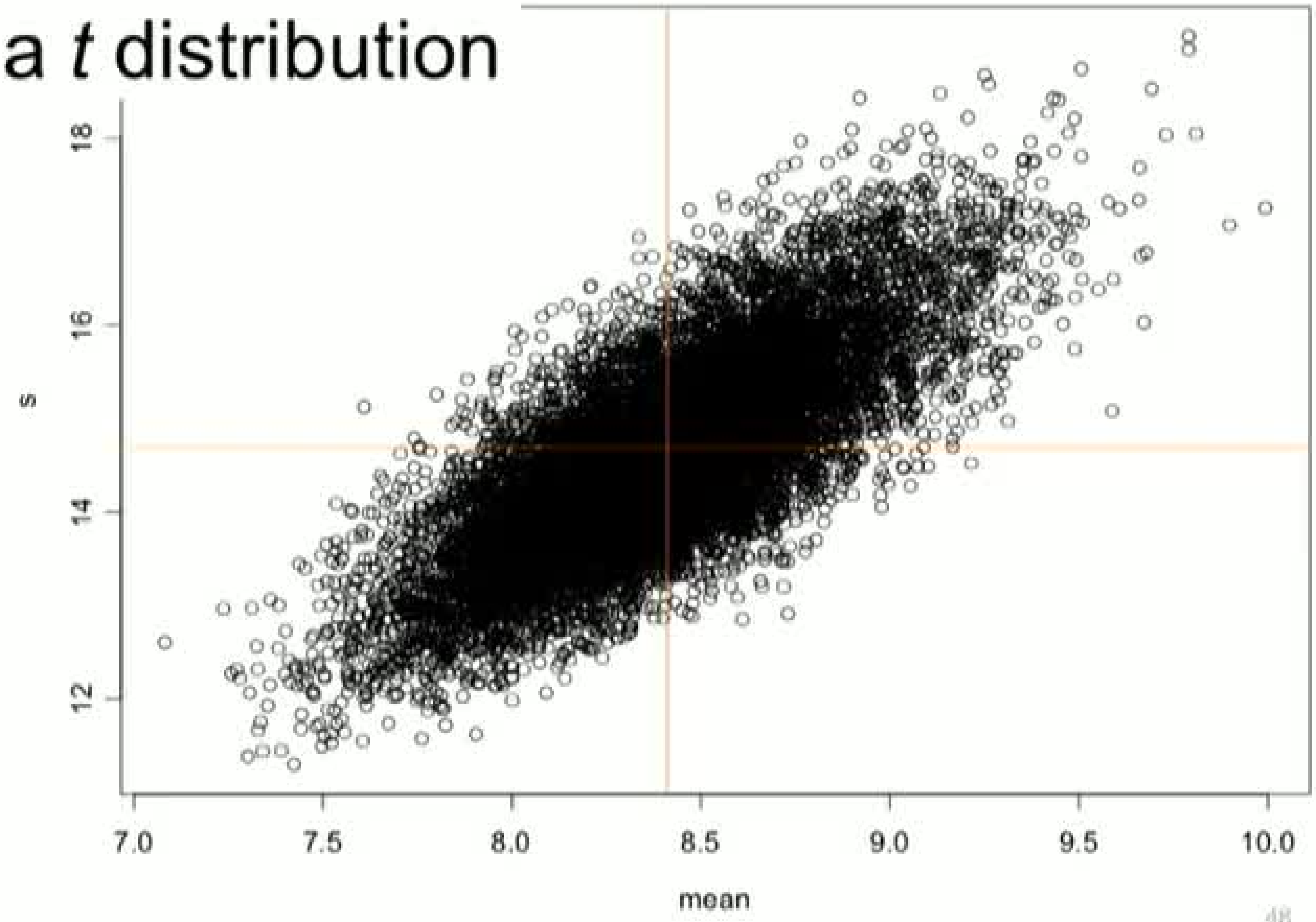
Sampling distribution of *t*-statistic

If population Normal,

- $\overline{x}$ and *s* are independent
- *t*-statistic has a *t* distribution

Not if skewed!

ILEC data (n=1664)



Bootstrap *t* interval

Google

Do not assume a *t*-statistic has a *t* distribution.

Instead, bootstrap to estimate quantiles, and solve for μ.

Interval: $$\left(\overline{x} - t^*_{.975}\frac{s}{\sqrt{n}}, \overline{x} - t^*_{.025}\frac{s}{\sqrt{n}}\right)$$

```
      bootstrapT tWithBootSE percentile
2.5%        7.77         7.71        7.73
97.5%       9.17         9.11        9.13
=xbar +  (-.65, .76) (-.7, .7) (-.68, .72)
=xbar + (-1.79, 2.11)s/sqrt(n), ...
```

More Accurate Intervals

Percentile and *t* interval:

- first-order correct
- Consistent, coverage error O(1/sqrt(*n*))

Bootstrap *t*, BCa, …

- second-order correct
- coverage error O(1/*n*)
- Handle bias, skewness, and transformations

Diagnosing Accuracy

Google

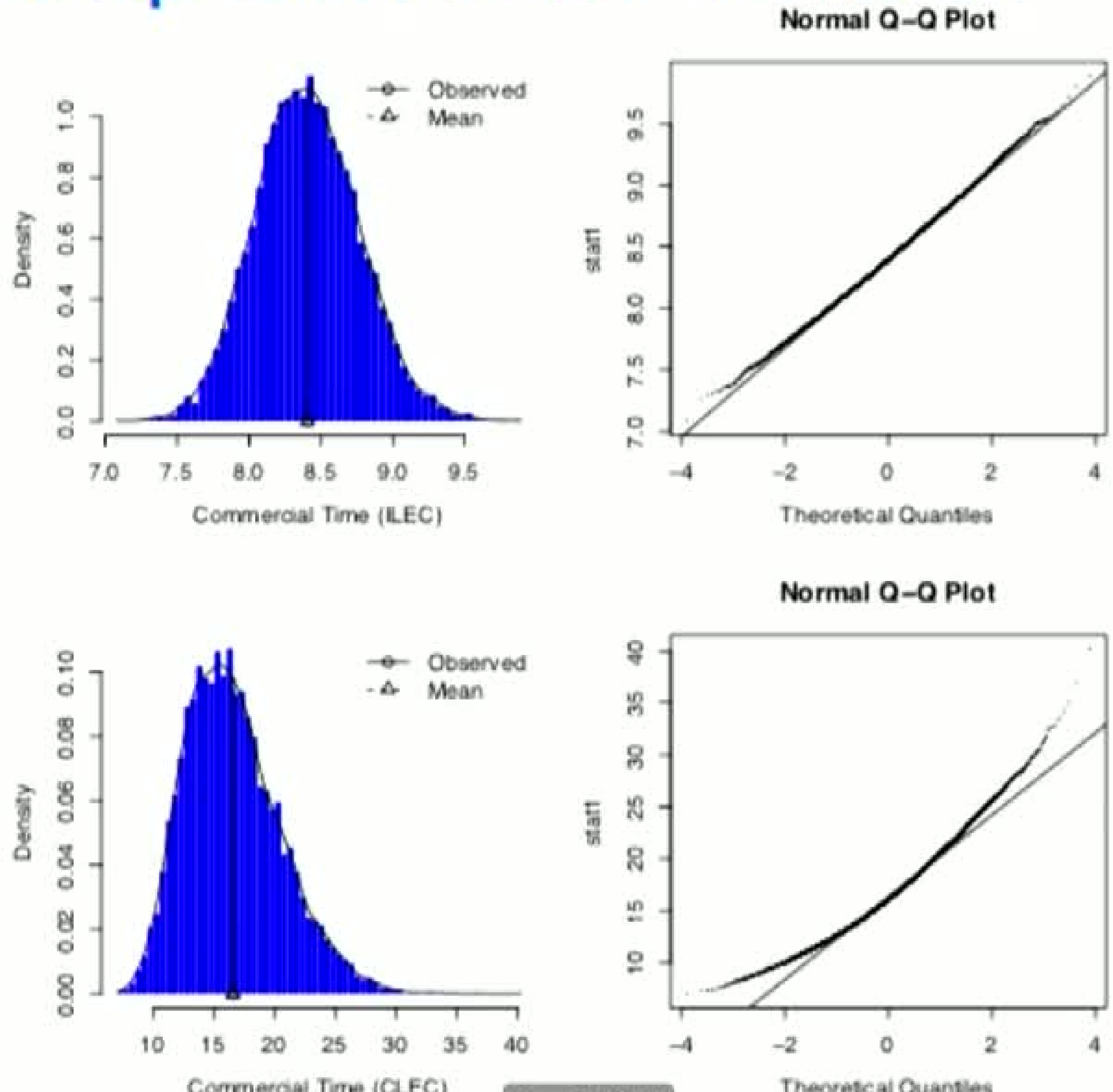
Look at bootstrap distributions for non-normality.

Is the amount of non-normality/asymmetry a cause for concern?

Note – we're looking at a sampling distribution, not data. This is *after* the CLT effect!

Measure – not just eyeball.

Bootstrap Distns for Verizon



Summary So Far

Google

- Reasons to resample
  - Easy
  - Communicate results
  - Flexible – e.g. robust statistics
  - More accurate
    - Bootstrap *t*
  - Diagnostics

Outline

Google

- Case Study, Basics
- Accuracy
- Bootstrap Regression
- Bootstrap Sampling Methods
- Permutation Tests

Meta goals:

Visual Resampling

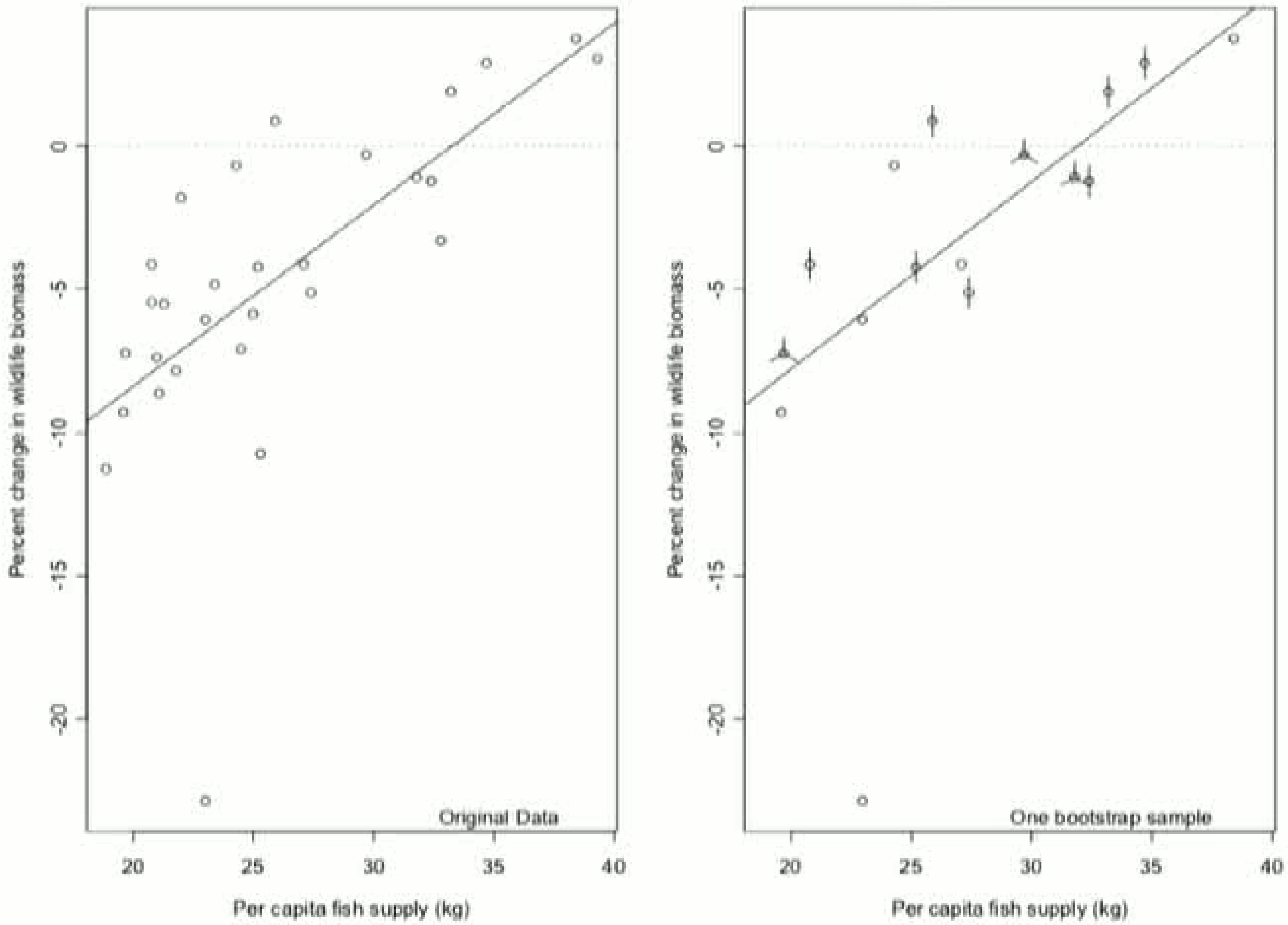
Variety of approaches

Summary So Far

Google

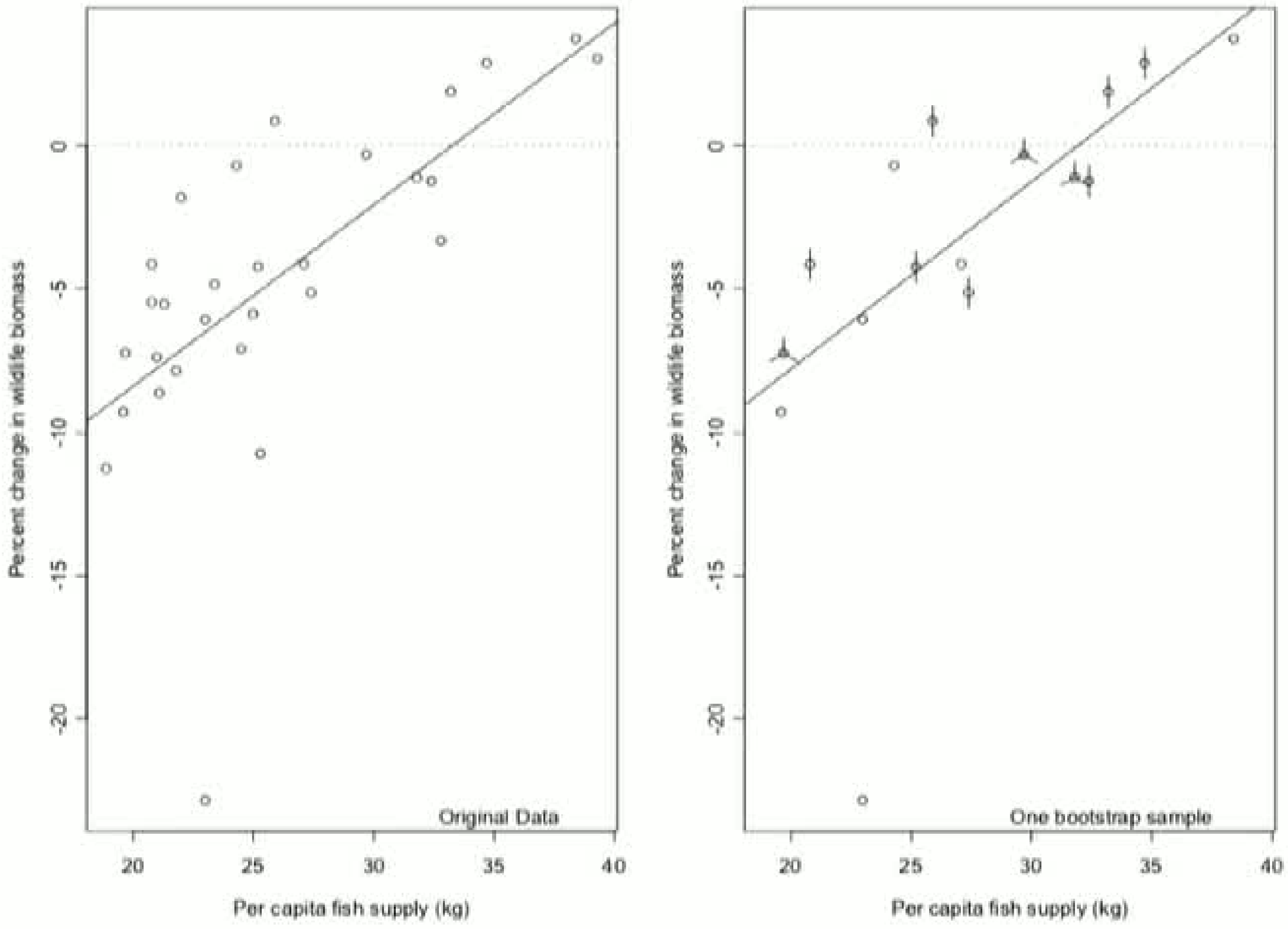
- Reasons to resample
  - Easy
  - Communicate results
  - Flexible – e.g. robust statistics
  - More accurate
    - Bootstrap *t*
  - Diagnostics

Bushmeat, ΔBiomass vs Fish

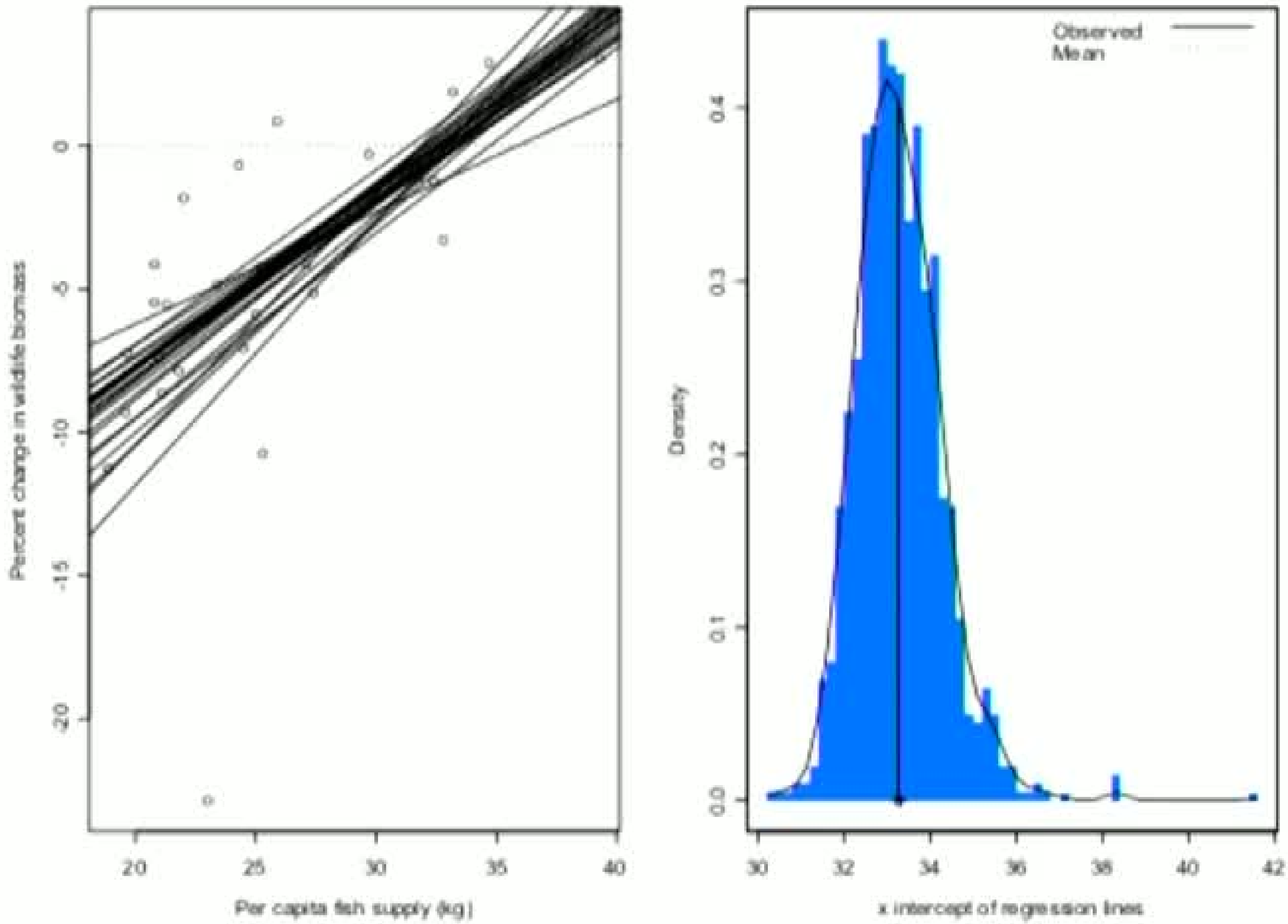


Bushmeat, ΔBiomass vs Fish

Google

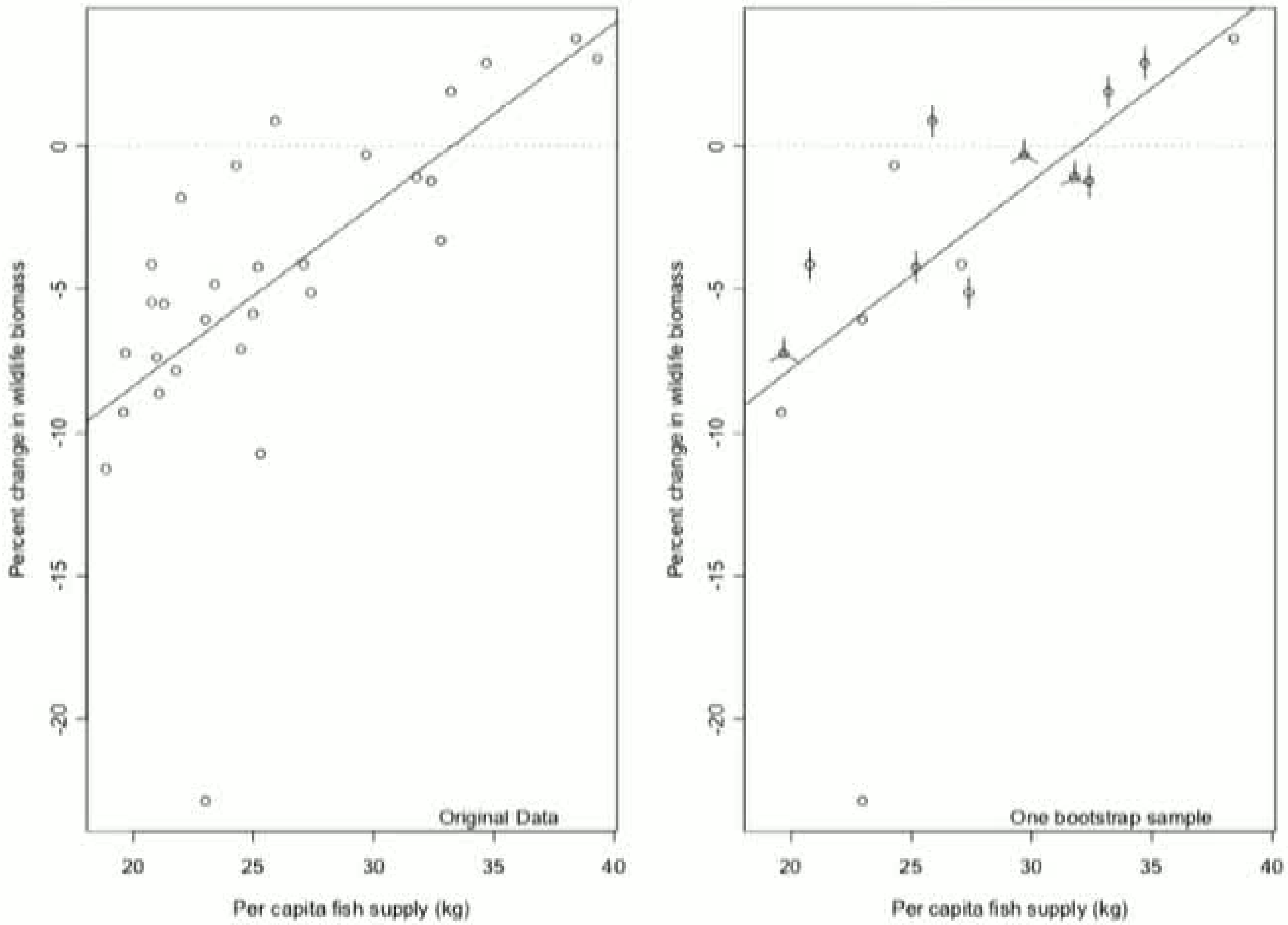


Bootstrap Bushmeat

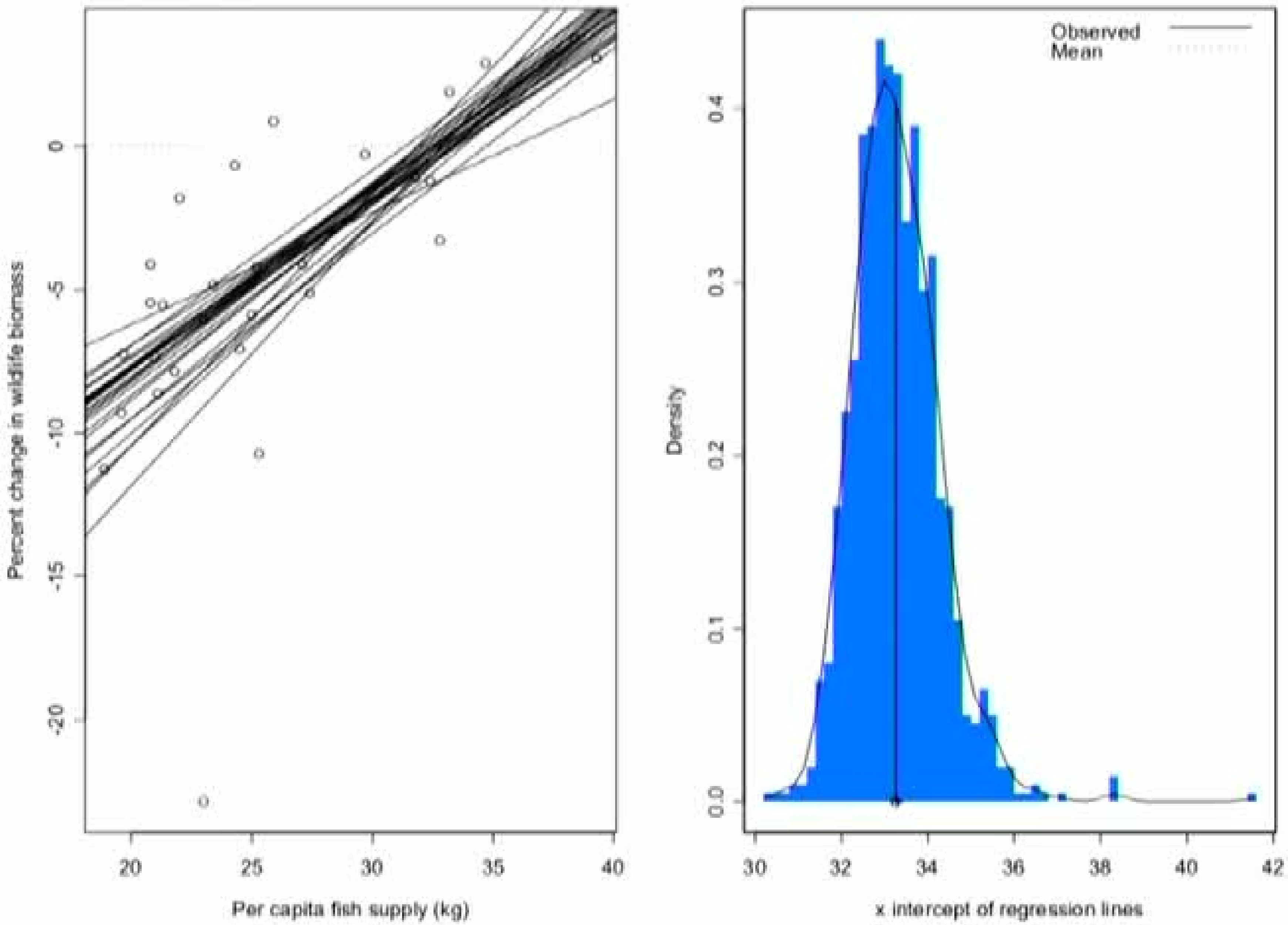


Compare: $\hat{y} \pm ts\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum (x_i - \overline{x})^2}}$

Bushmeat, ΔBiomass vs Fish



Bootstrap Bushmeat



Compare: $$\hat{y} \pm ts\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum(x_i-\overline{x})^2}}$$

Resampling Linear Regression

Google

Resample observations

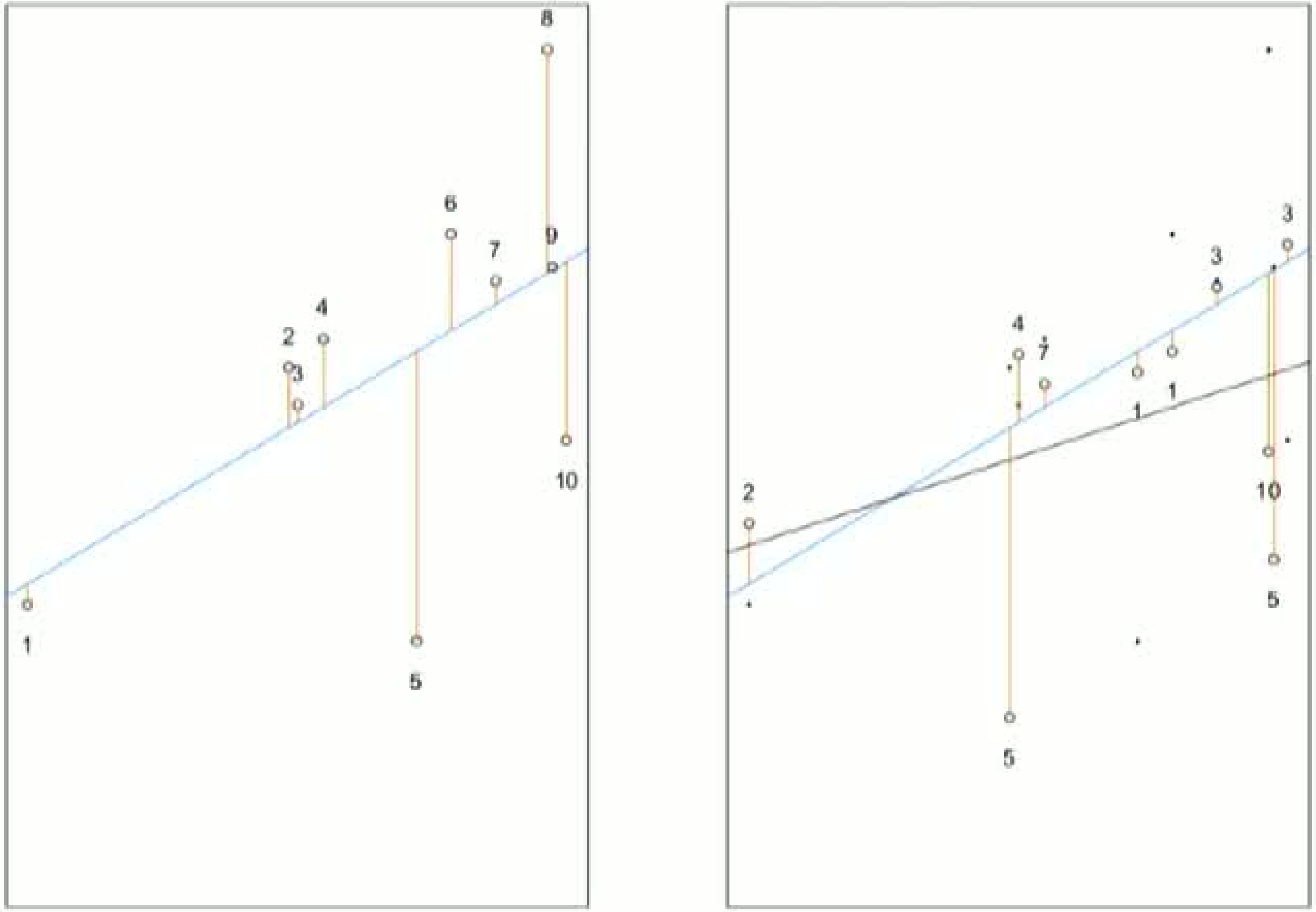
- Problem with factors

Resample residuals

- Fit model
- Resample residuals, with replacement
- Add to fitted values
- Problems with heteroskedasticity, lack of fit

Resample Residuals

Google



Resampling Linear Regression

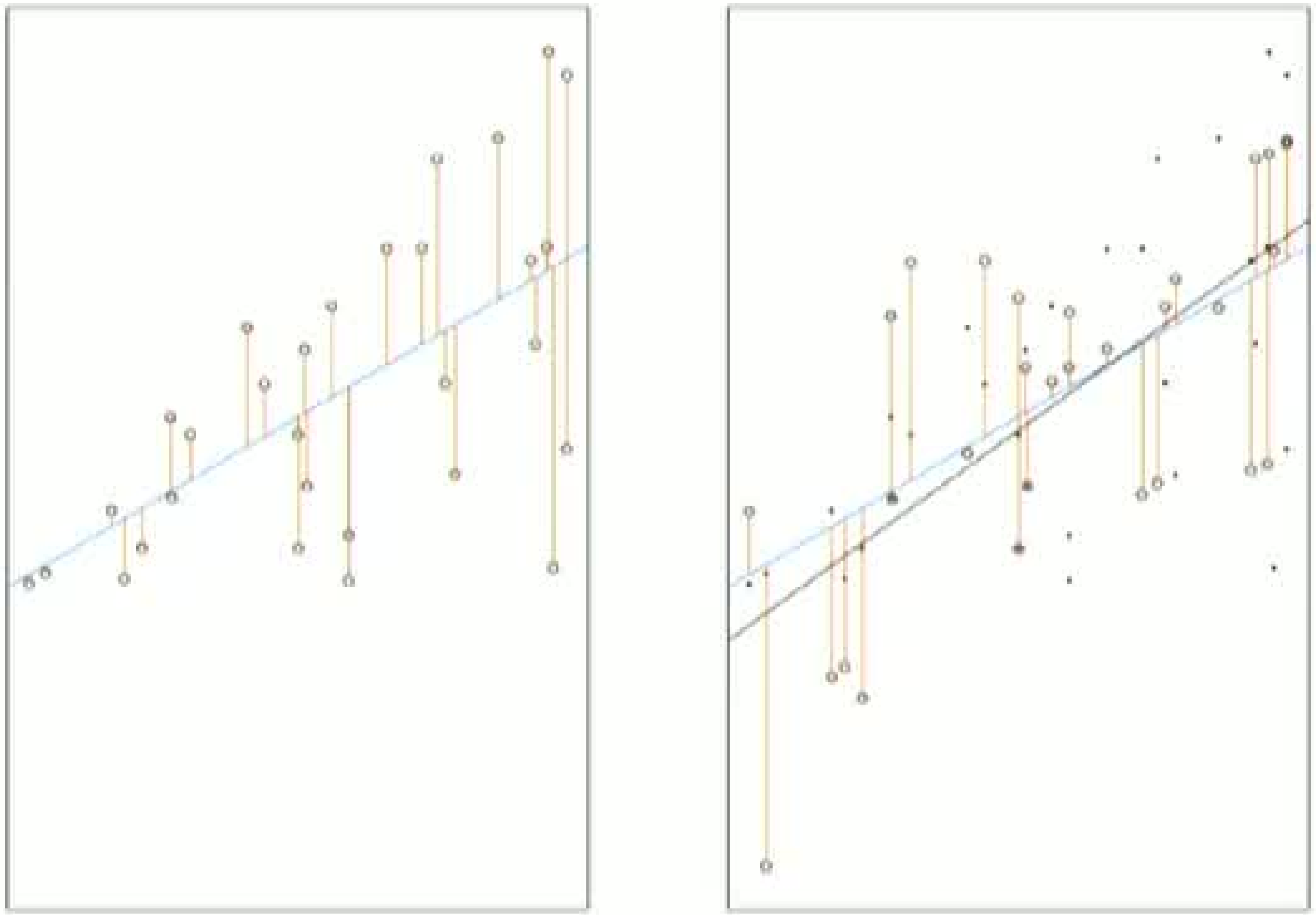
Resample observations

- Problem with factors

Resample residuals

- Fit model
- Resample residuals, with replacement
- Add to fitted values
- Problems with heteroskedasticity, lack of fit

Problem: heteroskedasticity



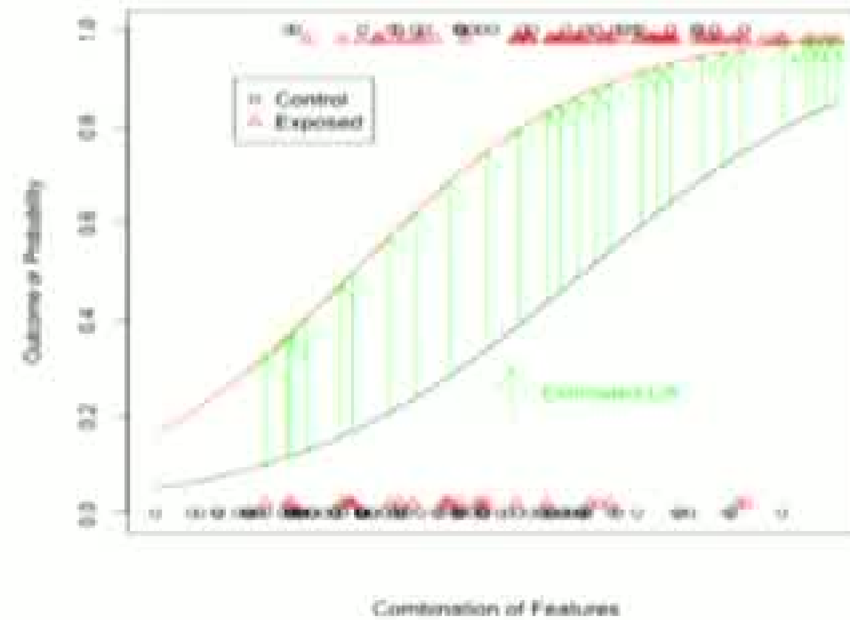
Remedy: "Wild bootstrap" = resample $\pm$ residual

Resampling Regression

Resample observations (random effects)

Resample *Y* from conditional distribution given *X*'s (fixed effects)

- Special case: resampling residuals
- For logistic regression, P(*Y*=1) = prediction
- Condition on observed information



Recommendations for Regression

Google

Coefficients / Predictions at fixed X

- (see next page)

Causal Modeling / Predictions at Random X

- Resample observations

Recommendations for Regression (coefficients)

Google

Quick-and-dirty

- Resample observations

Large samples

- Any method

Medium samples

- Resample *Y* from conditional distribution given *X*'s

Small samples

- Parametric bootstrap

Outline

- Case Study / Basics
- Accuracy
- Bootstrap Regression
- Bootstrap Sampling Methods
- Permutation Tests

Meta goals:

Variety of options

Big Data

Recommendations for Regression (coefficients)

Google

Quick-and-dirty

- Resample observations

Large samples

- Any method

Medium samples

- Resample *Y* from conditional distribution given *X*'s

Small samples

- Parametric bootstrap

Recommendations for Regression (coefficients)

Quick-and-dirty

- Resample observations

Large samples

- Any method

Medium samples

- Resample *Y* from conditional distribution given

Small samples

Recommendations for Regression (coefficients)

Google

Quick-and-dirty

- Resample observations

Large samples

- Any method

Medium samples

- Resample Y from conditional distribution given X's

Small samples

- Parametric bootstrap

Outline

Google

- Case Study / Basics
- Accuracy
- Bootstrap Regression
- Bootstrap Sampling Methods
- Permutation Tests

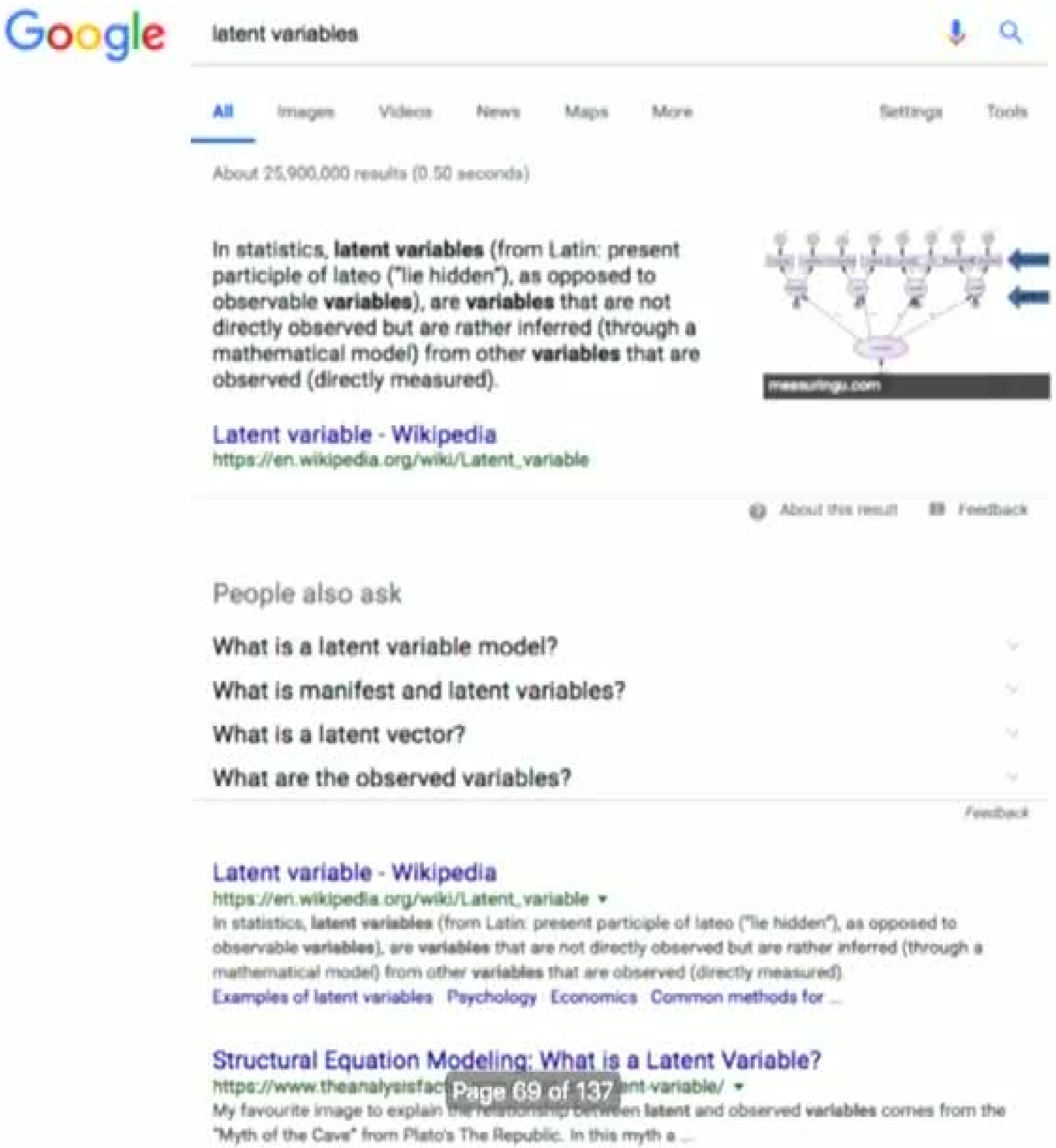
Meta goals:

Variety of options

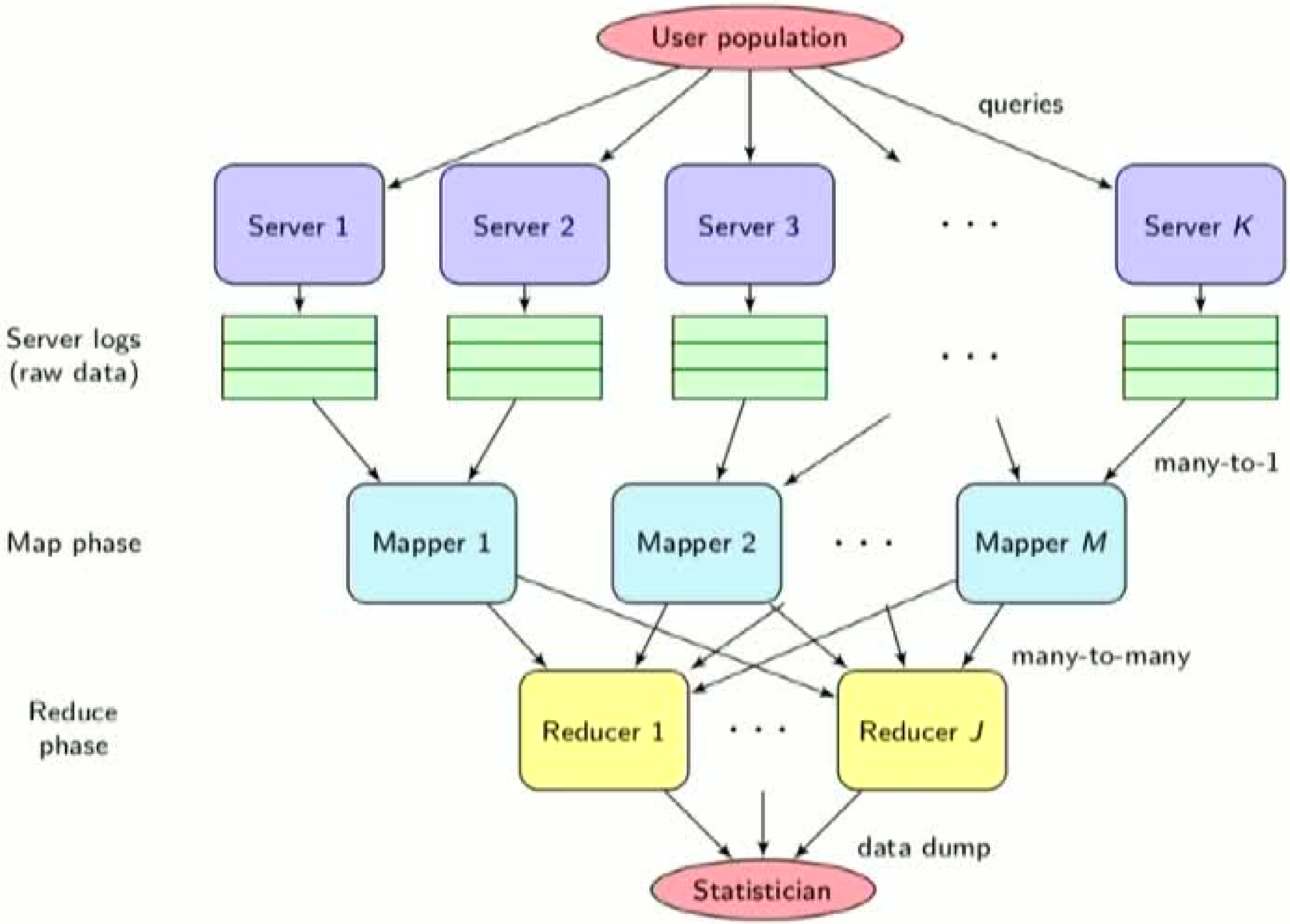
Big Data

In contrast…

Google



A cartoon Google data stream



Cost per Click Standard Error

Cost per click (by country, time period, ...)

= sum(cost) / number(clicks)

= sum(Y_i) / sum(N_i)

where Y_i = cost, N_i = clicks for user *i*.

Can't compute!

Can't calculate formula standard error.

Bootstrap!

Var(sum($f_i Y_i$) / sum($f_i N_i$))

where f ~ Poisson(1)

$$\begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,r} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n,1} & f_{n,1} & \cdots & f_{n,r} \end{bmatrix}$$

Resampling Variations

Google

- Ordinary bootstrap
  - Sample with replacement
  - Frequencies are Bi(n, 1/n)
- Poisson bootstrap
  - Frequencies are Poisson(lambda = 1)
  - Independent
- Half-Sampling
  - Sample without replacement, size n/2
  - Same SE as bootstrap (but not skewness)
- Bag of Little Bootstraps
- Jackknife, group jackknife, cross-validation, …

Cost per Click Standard Error

Google

Cost per click (by country, time period, …)

= sum(cost) / number(clicks)

= sum(Y_i) / sum(N_i)

where Y_i = cost, N_i = clicks for user i.

Can't compute!

Can't calculate formula standard error.

Bootstrap!

Var(sum($f_i Y_i$) / sum($f_i N_i$))

where f ~ Poisson(1)

$$\begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,r} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n,1} & f_{n,1} & \cdots & f_{n,r} \end{bmatrix}$$

Resampling Variations

Google

- Ordinary bootstrap
  - Sample with replacement
  - Frequencies are Bi(n, 1/n)
- Poisson bootstrap
  - Frequencies are Poisson(lambda = 1)
  - Independent
- Half-Sampling
  - Sample without replacement, size n/2
  - Same SE as bootstrap (but not skewness)
- Bag of Little Bootstraps
- Jackknife, group jackknife, cross-validation, …

Different Sampling Procedures

Google

Two-sample

Regression

Stratified sampling

Finite Population

Hierarchical

Time Series (in reserve)

Other Sampling Situations

Google

Stratified Sampling

- Resample within strata

Finite Population

- Create finite population, resample without replacement (repeat each observation N/n times)

General Idea:

- Resampling should mirror real life
- Except condition on observed information
  - Like sample sizes, x in regression

Complex Sampling

Google

Resampling should mimic how the data was produced

There are some twists

- Downward bias in bootstrap SE
- Finite-population sampling
- Regression
- Time Series (omit in this talk)

Bootstrap SE too small

Google

Usual SE for mean is $s / \sqrt{n}$

Theoretical bootstrap corresponds to using divisor of *n* instead of *n-1*.

Bias factor for each sample, each stratum

$$s^2 = (n-1)^{-1} \sum (x_i - \overline{x})^2$$

$$\hat{\sigma}^2 = n^{-1} \sum (x_i - \overline{x})^2$$

Example - TV

Google

Student data, commercial minutes in cable TV, comparing standard and extended channels.

```
Basic    mean = 9.21
7.0 10.0 10.6 10.2 8.6 7.6 8.2 10.4 11.0 8.5
Extended mean = 6.87
3.4  7.8  9.4  4.7 5.4 7.6 5.0  8.0  7.8 9.6
```

Usual standard error for difference: .798

Bootstrap Standard error: .758 (5% smaller)

Remedies for small SE

Multiply SE by $\sqrt{n/(n-1)}$

- Equal strata sizes only

Sample with reduced size, (n-1)

Bootknife sampling

- Omit random observation
- Sample size n from remaining n-1

Smoothed bootstrap

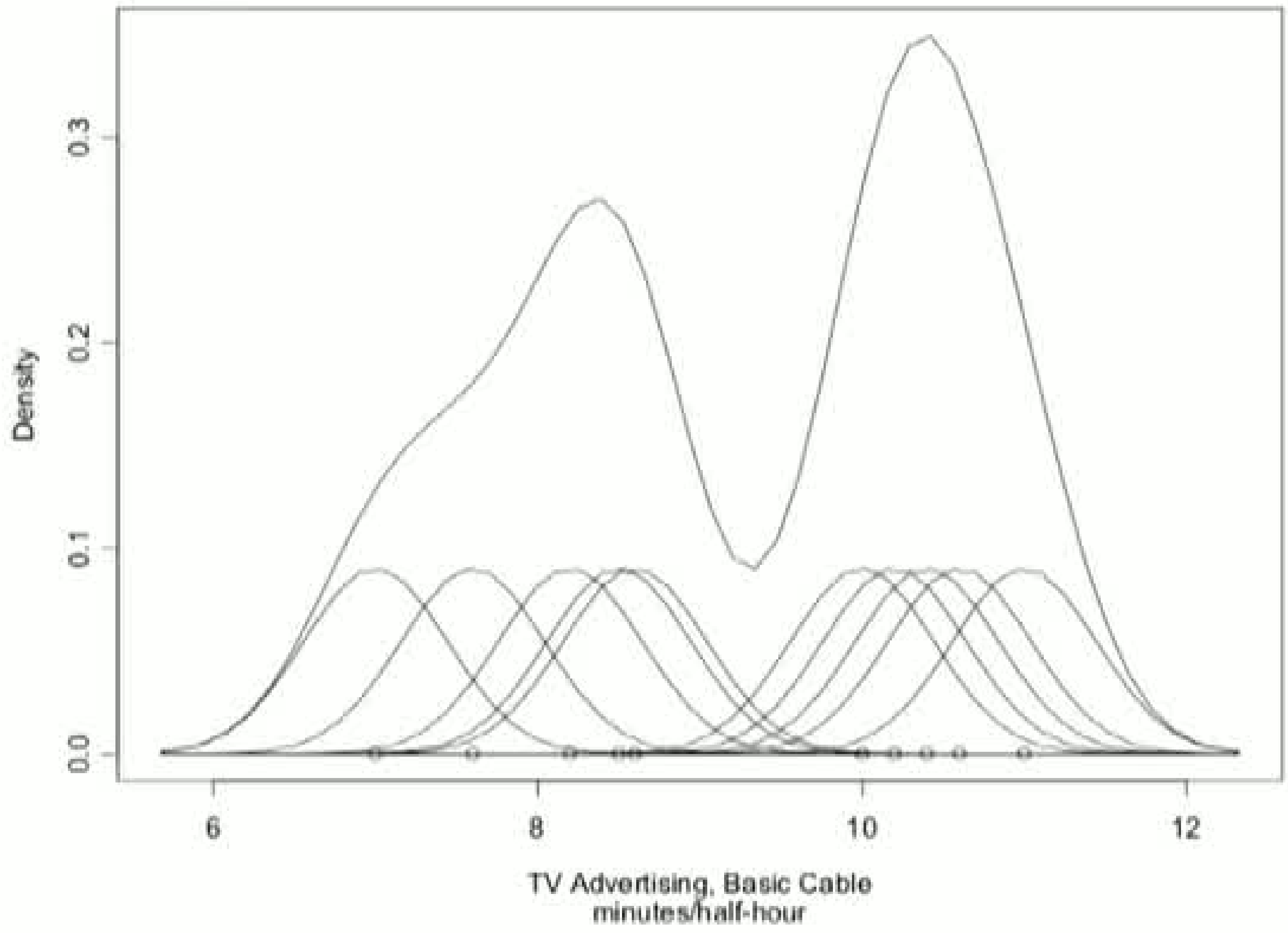
- Choose smoothing parameter to correct variance
- Continuous data only.

$$\sigma_{\text{kernel}} = s / \sqrt{n}$$

Expanded percentile interval

Smoothed bootstrap

Kernel Density Estimate = Nonparametric bootstrap + random noise



Stratified Sampling

Google

Bootstrap Sampling: independent within each stratum

Caution – bootstrap SE may be very small

- Sample with reduced stratum size
- Bootknife sampling
- Smoothing
- ?Does statistic depend on sample size?

Outline

Google

- Case Study / Basics
- Accuracy
- Bootstrap Regression
- Bootstrap Sampling Methods
- Permutation Tests

Permutation Tests

Google

Hypothesis test: sample consistent with H_0

- 2-sample example

General procedure

- Matched pairs
- Bivariate relationship
- Regression

Meta goals:

Handle some situations

Understand when it doesn't work

Permutation Test for 2-samples

H_0: no real difference between groups; observations could come from one group as well as the other

Resample: randomly choose n_1 observations for group 1, rest for group 2.

Equivalent to permuting all n, first n_1 into group 1.

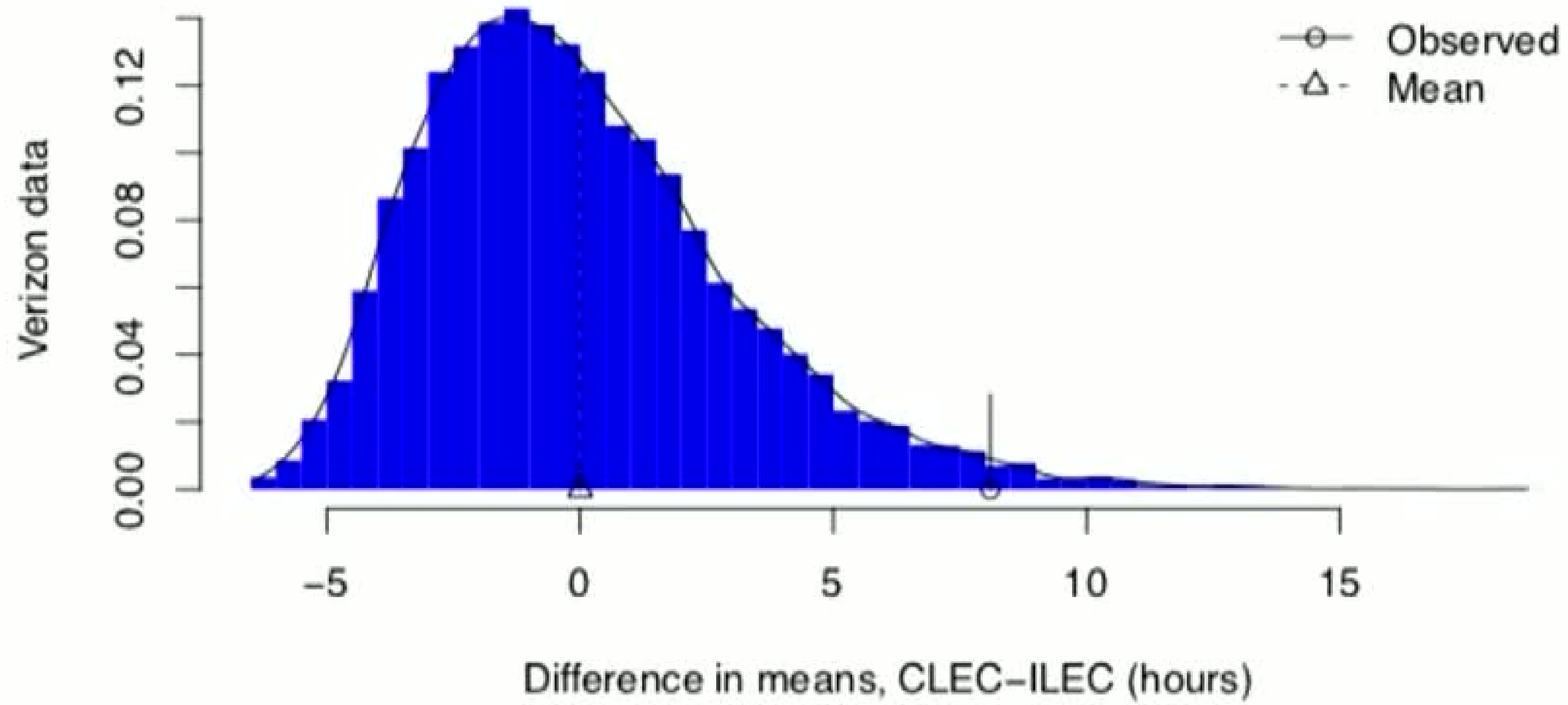
Permutation Test for 2-samples

H_0: no real difference between groups; observations could come from one group as well as the other

Resample: randomly choose n_1 observations for group 1, rest for group 2.

Equivalent to permuting all n, first n_1 into group 1.

Verizon permutation test



General Permutation Tests

Google

Compute Statistic for data

Resample consistent with H_0 and study design

Construct permutation distribution

P-value = percentage of resampled statistics that exceed original statistic

One-sided *P*: $(\# \geq \text{original} + 1) / (R + 1)$

Two-sided *P*: $2 * $ smaller of one-sided *P*

Perm Test for Matched Pairs or Stratified Sampling

Google

Permute within each pair

Permute within each stratum

Permutation Test of Relationship

To test H_0: X and Y are independent

Permute either X or Y

Test statistic may be correlation, slope, chi-square statistic (Fisher's exact test), …

In contrast, to bootstrap (for SE or CI), resample whole rows, X and Y together

Permutation Test of Relationship

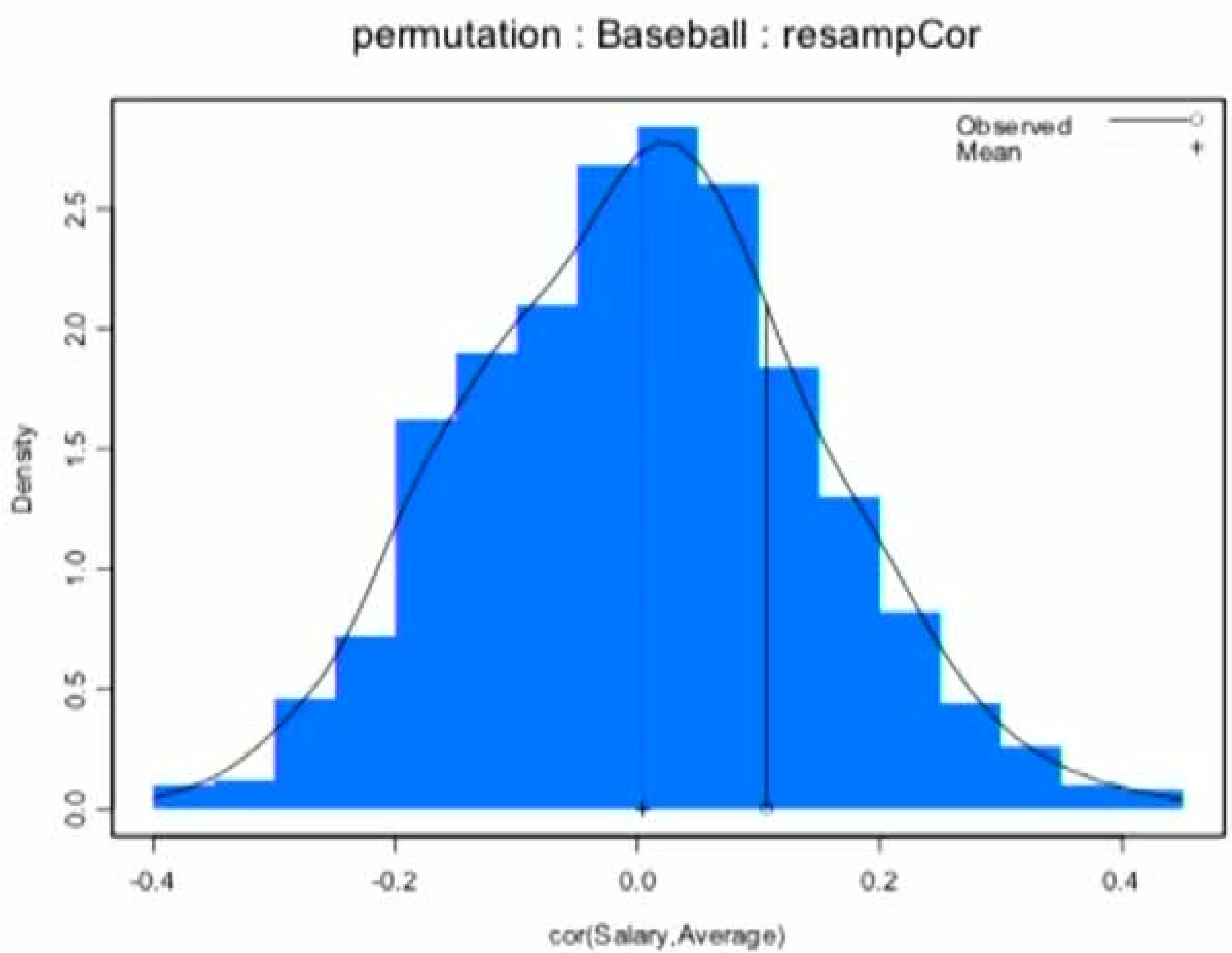
To test H_0: *X* and *Y* are independent

Permute either *X* or *Y*

Test statistic may be correlation, slope, chi-square statistic (Fisher's exact test), …

In contrast, to bootstrap (for SE or CI), resample whole rows, X and Y together

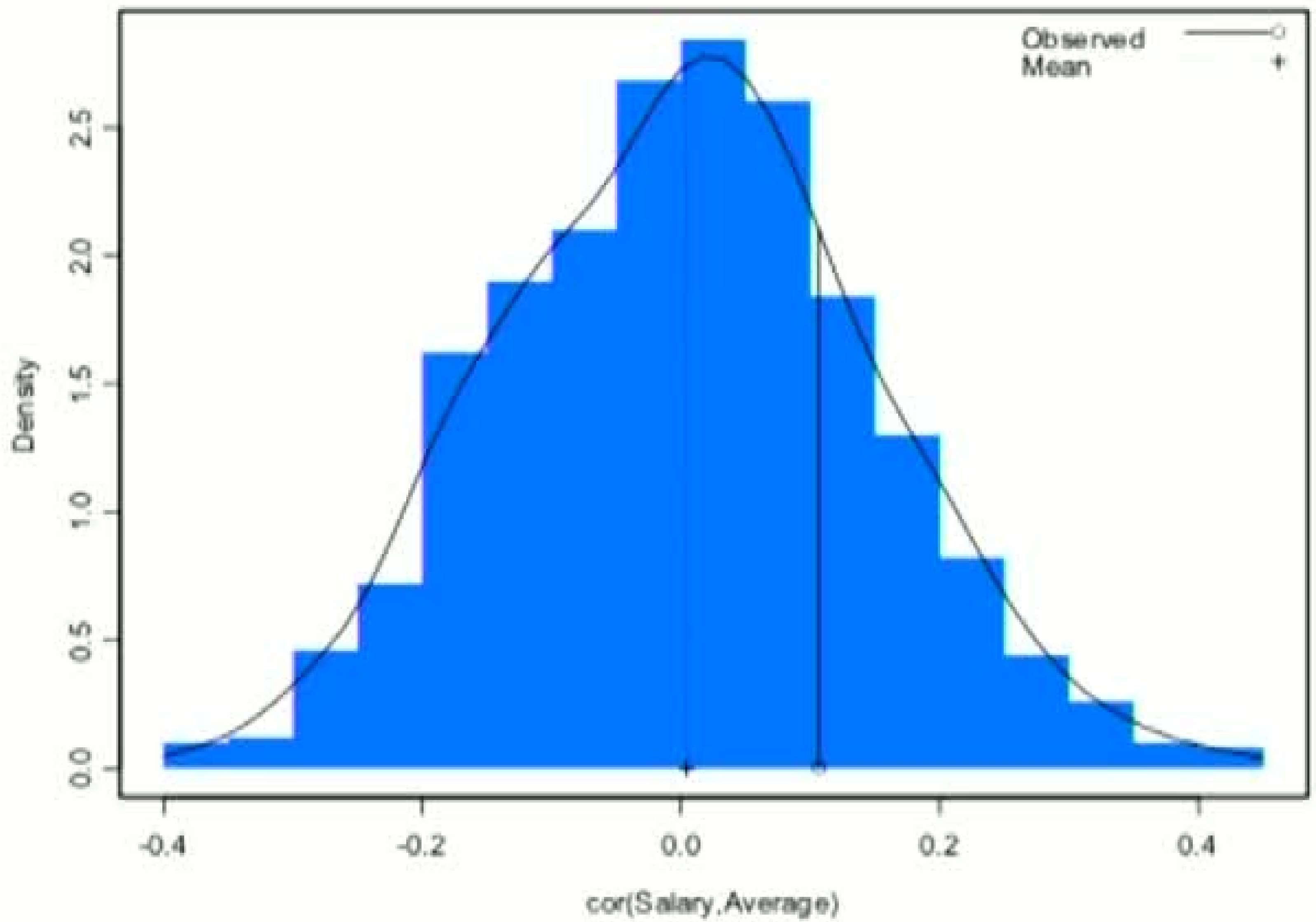
Test: Baseball correlation



Test: Baseball correlation

Google

permutation : Baseball : resampCor



Test terms in Multiple Regression

Google

To test incremental contribution of X_1

- No exact test exists
- Best (usually) approximate test:
  - Permute residuals of reduced model
  - Test statistic ~pivotal: *t* or *F* statistic
  - Depends on model correctness,
    - $Y - (\beta_{0:1} + \beta_{2:1}X_2 + \ldots + \beta_{p:1}X_p)$ are exchangeable
  - Anderson 2001 Canadian J. Fish. Aquat. Sci.

Summary

Google

- Reasons to Resample
  - Easy
  - Communicate results
  - Flexible – e.g. robust statistics
  - More accurate
- Examples
  - Easy to use for variety of statistics
  - Easy for independent data; less easy otherwise
- Sampling Methods
  - Non-iid situations
- Permutation Tests
  - Two samples, relationship

Further Information

Google

http://www.timhesterberg.net/bootstrap

Mathematical Statistics with Resampling and R, Chihara & Hesterberg

Intro stats

Software