IDA

INSTITUTE FOR DEFENSE ANALYSES

A First Step into the Bootstrap World

Matthew Avery

May 2016
Approved for public release.
IDA Document NS D-5816
Log: H 16-000624

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

This briefing motivates bootstrapping through an intuitive explanation of basic statistical inference, discusses the right way to resample for bootstrapping, and uses examples from operational testing where the bootstrap approach can be applied. Bootstrapping is a powerful and flexible tool when applied appropriately. By treating the observed sample as if it were the population, the sampling distribution of statistics of interest can be generated with precision equal to Monte Carlo error. This approach is nonparametric and requires only acceptance of the observed sample as an estimate for the population. Careful resampling is used to generate the bootstrap distribution. Applications include quantifying uncertainty for availability data, quantifying uncertainty for complex distributions, and using the bootstrap for inference, such as the two-sample t-test.

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-5816

A First Step into the Bootstrap World

Matthew Avery

Executive Summary

Bootstrapping is a powerful nonparametric tool for conducting statistical inference with many applications to data from operational testing. Bootstrapping is most useful when the population sampled from is unknown or complex or the sampling distribution of the desired statistic is difficult to derive. Careful use of bootstrapping can help address many challenges in analyzing operational test data.

Bootstrapping is predicated on the use of the sample data as a plug in estimator for the population. Inference is then conducted in this “bootstrap world” wherein the population of interest is identical to the observed sample. With the population known, repeated sampling can be used to characterize the desired sampling distribution up to Monte Carlo error. This can now be used to calculate exact confidence intervals or perform relevant hypothesis tests within the bootstrap world. These bootstrap world intervals and p-values can be treated as estimates in the real world.

This briefing provide an outline for this approach and include examples applying these principles to synthetic data sets generated to mimic operational test data. The role of the sampling distribution in statistical inference is described, and bootstrapping is motivated intuitively using the metaphor of the bootstrap world introduced above. Examples include confidence intervals for sample means and medians, how to apply the bootstrap to complex statistics involving random variables from multiple distributions (such as Availability calculations), and hypothesis testing via the bootstrap.

The Bootstrap World

Matthew Avery, Institute for Defense Analyses

Bottom Line Up Front

- **Bootstrapping**
  - Powerful tool applicable in a variety of situations
    - Quantify Variance
    - Hypothesis Testing
  - Use for inference not estimation
  - Resample using the same approach that was used to generate your sample
    - For hypothesis testing, resample under the null hypothesis
  - Bootstrap results can only ever be as good as the sample upon which they're based

- **Most useful when:**
  - Distributions unknown or complex
  - Deriving sampling distribution intractable/impractical

Assessing Performance of a Small UAV

- **MQ-8C Fire Scout**
  - Navy Intelligence/Surveillance/ Reconnaissance system
  - Vertical take-off unmanned air vehicle (UAV)
  - Electro-optical/Infrared sensor
  - Mission includes detection of maritime vessels & ability to use sensors to lock on and auto-track targets
- **Questions of interest**
  - What is average detection range?
  - What is the median target lock percentage?
  - What is the system's availability?

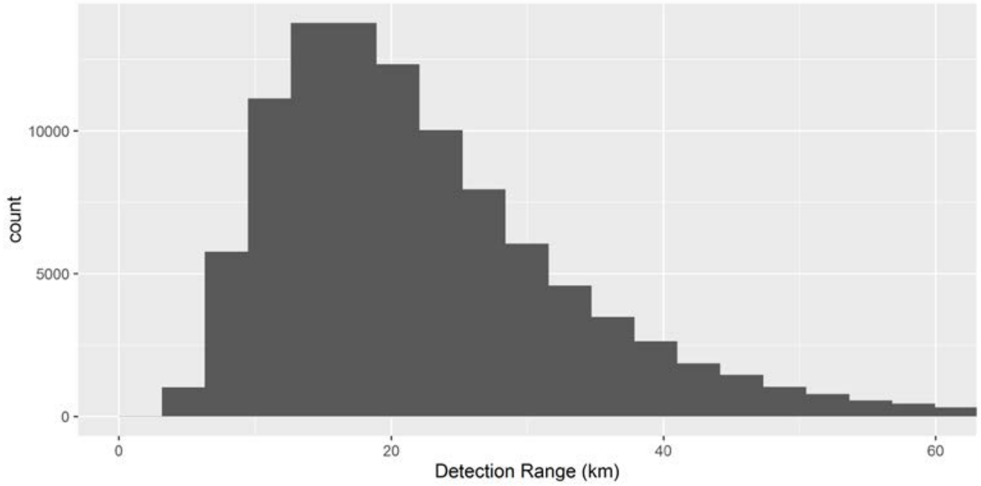
Note: All data and conclusions presented here are strictly notional and are used for illustration purposes only

Outline

- **Background**
  - – Populations & Sampling
  - – Sampling Distributions
  - – Statistical Inference
- **Bootstrap Basics**
  - – Resampling
  - – The Bootstrap World
- **Examples**
  - – Confidence Intervals
    - » Autotrack performance (median)
    - » Availability
  - – Hypothesis testing
    - » Two-sample testing
- **Extensions & Conclusions**

Populations

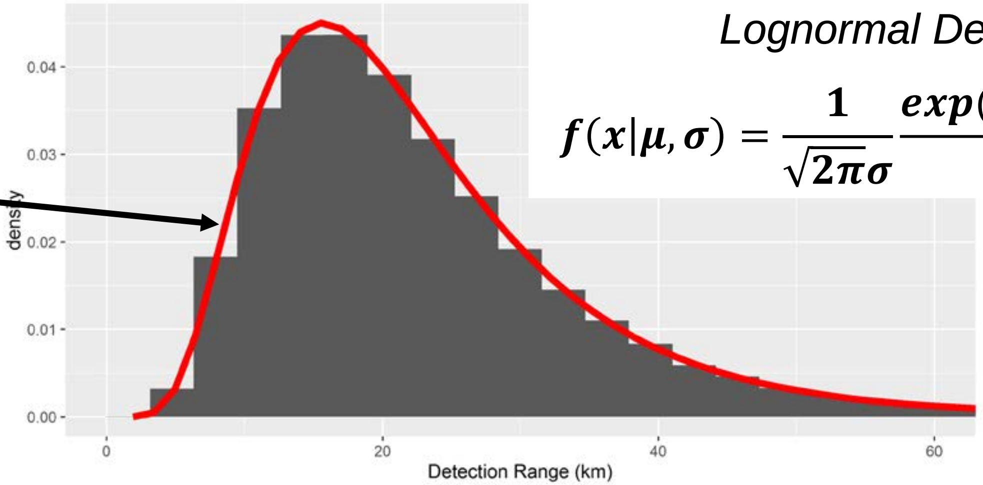
Population of detection ranges



Population: The entire pool of items or events of interest for some question or experiment

- **Population**
  - Can be a group of actually existing objects or a hypothetical group of potential objects/events
- **Population of Detection Ranges for MQ-8C**
  - Hypothetical & infinite
  - Any mission, target vessel, payload operator, etc.

Probability Distribution



Density curve

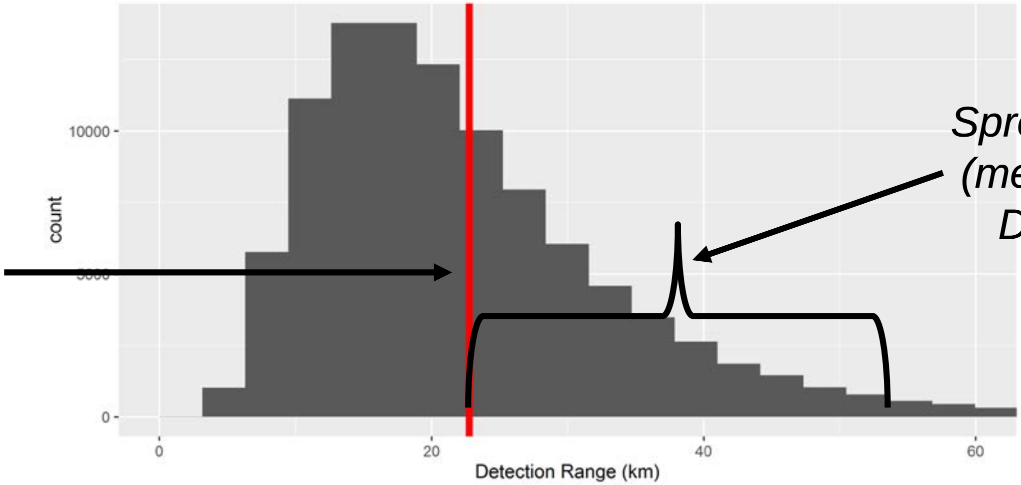
Lognormal Density Function

$$\boldsymbol{f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \frac{exp(-(log x\ - \mu)^2\ /\ (2\sigma^2)}{x}}$$

Probability Distribution: The entire pool of items or events of interest for some question or experiment

- **Probability density function describes how individual objects/events are distributed within a population**
  - Allows calculation of important values
    - » E.g., Probability of detection beyond 10 km
  - Characterized by **<u>parameters</u>**
    - » Mean
    - » Standard deviation

Statistical Parameters



Population Mean

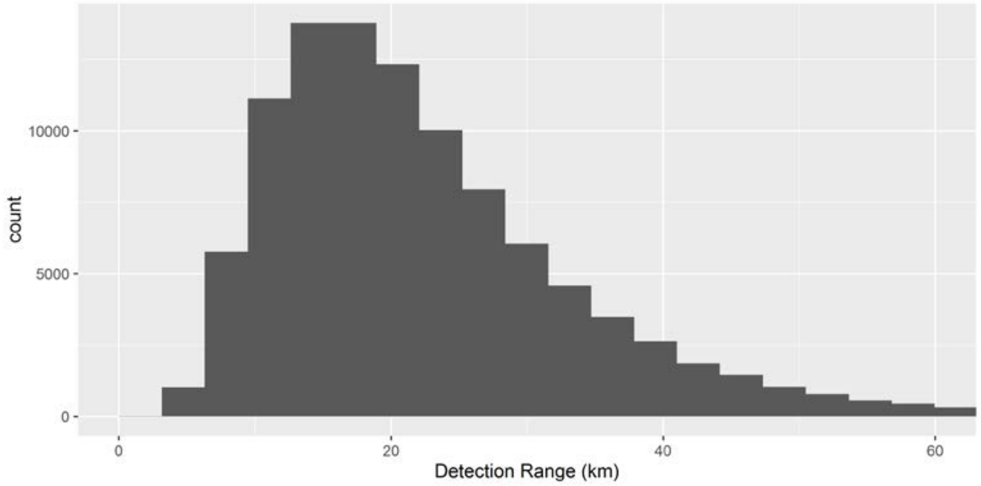
Spread of the population (measured by Standard Deviation/Variance)

Parameter: Numerical quantity that characterizes a statistical distribution, such as a population

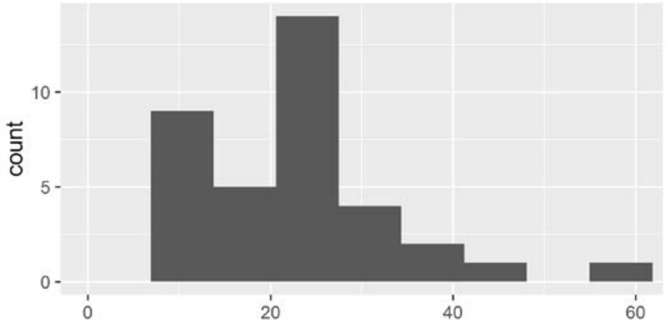
- **Knowing the parameters of the distribution is equivalent to knowing the distribution**
- **Normal (Two parameters)**
  - – Mean (μ)
  - – Variance (σ^2)
- **Exponential (One parameter)**
  - – Mean (λ)

Sampling from Populations

Population



Sample (n = 36)

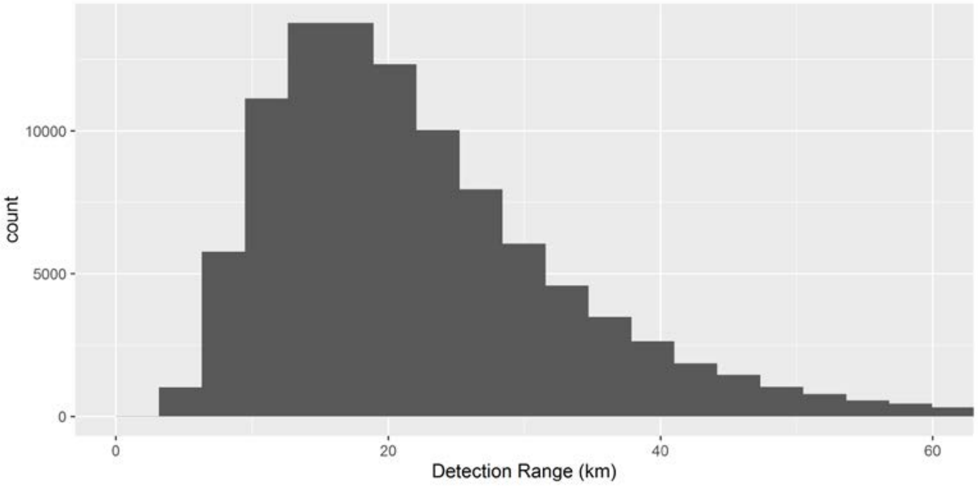


To understand the sample, it is necessary to understand the population and the procedure by which the sample is selected

Sample: A subset of a population selected by a defined procedure ("Simple random sample", etc.)

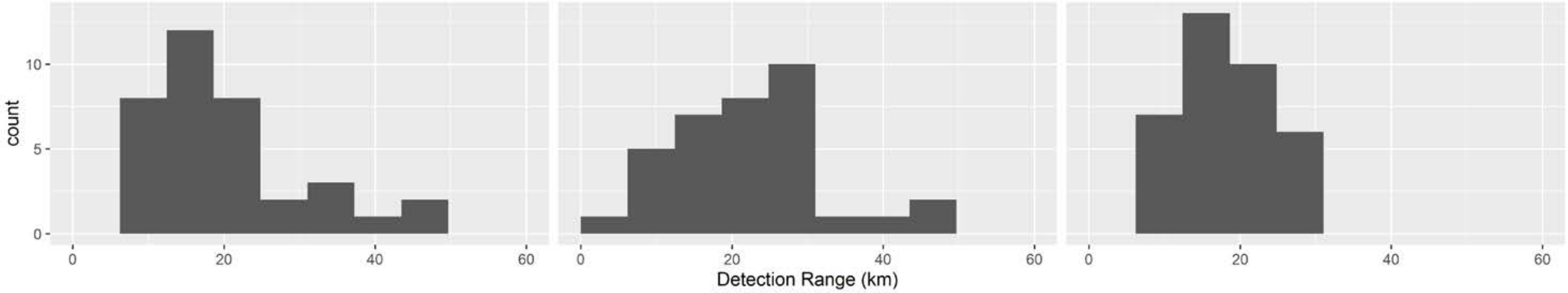
Sampling from Populations

Population



There are many samples that can be generated from a single population. Which one is generated is dependent on the <u>sampling procedure</u> and is typical random.

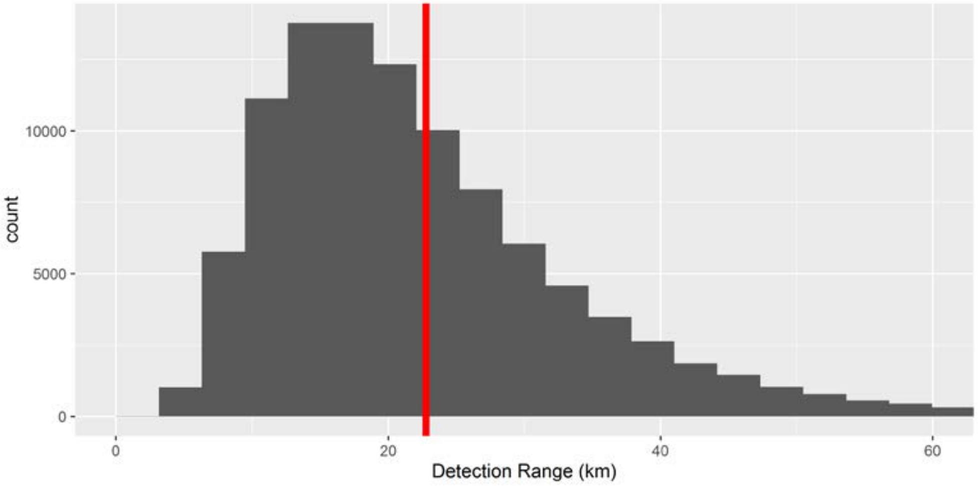
Potential Samples (n = 36)



Individual samples

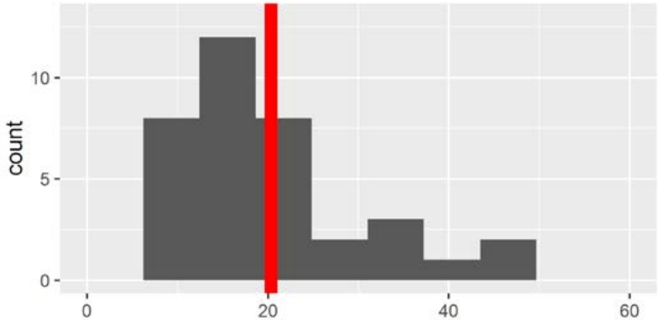
The Sample Mean is an Example of an Estimator

Population



Sample mean: Mean of the observed sample

Sample



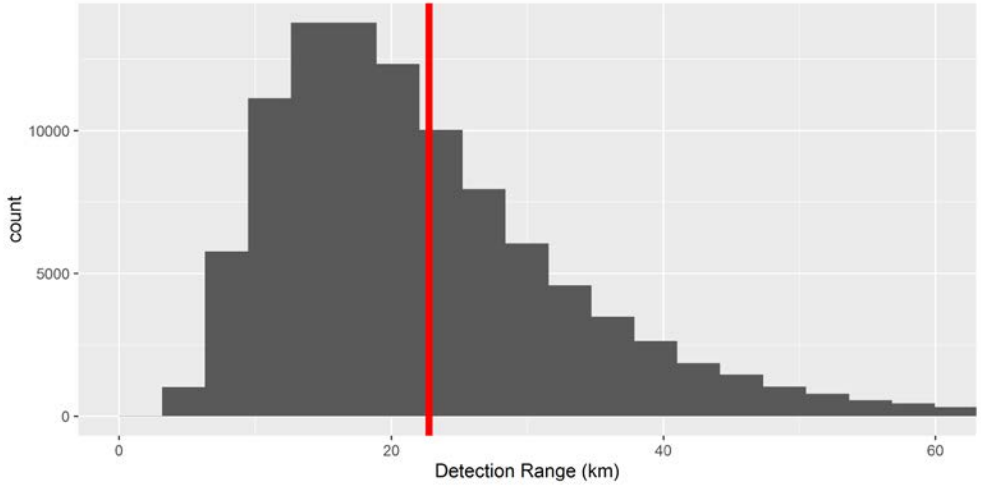
Statistic: A function based on a sample

Estimator: A statistic used to estimate a quantity of interest, such as a parameter.

- The <u>sample mean</u> is <u>statistic</u> often used to estimate the mean of a population.
- Since it is based a specific sample, a sample mean won't be equal to the <u>population mean</u>.
- Estimators can be derived for different values of interest (median, variance, quantiles, etc.).
- The quality of an estimator depends on the distribution of the population.

Sample Statistics From Multiple Samples

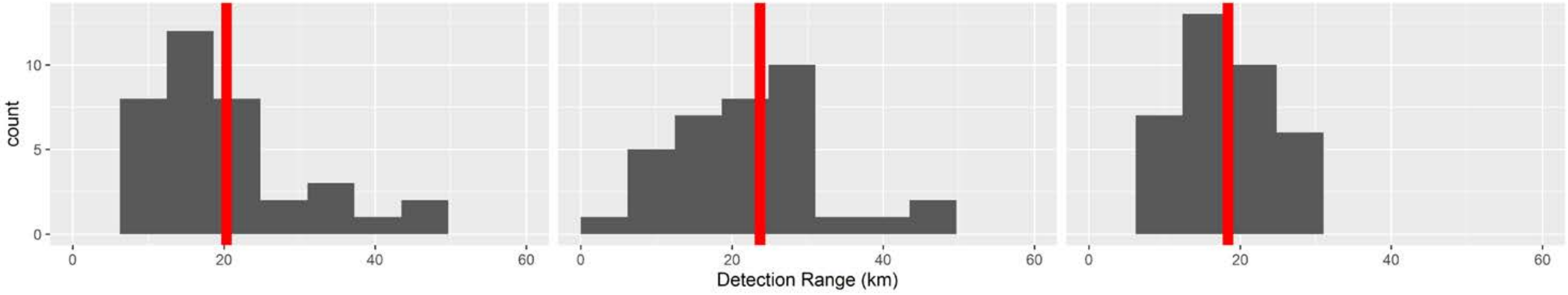
Population



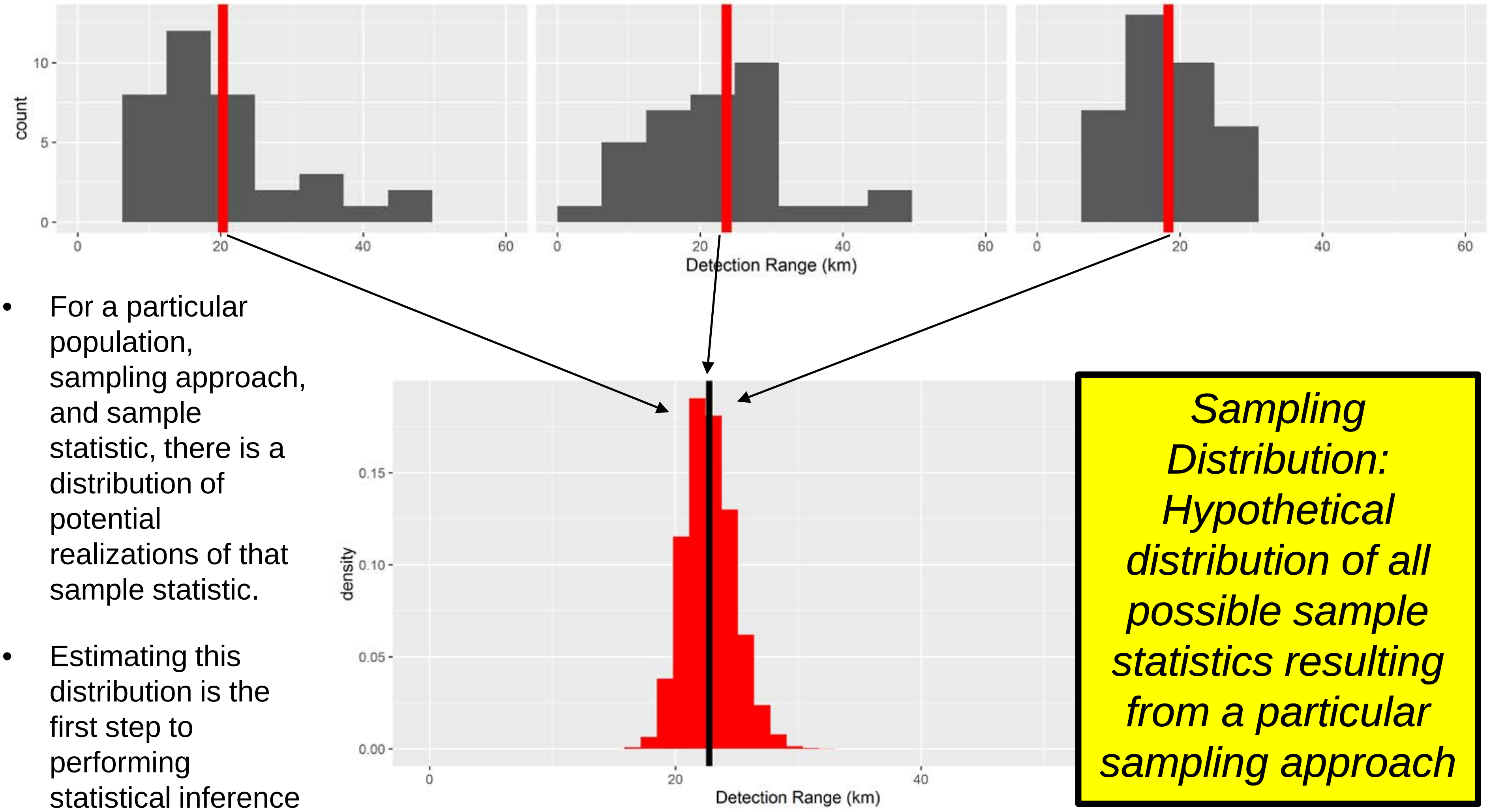
Each potential sample will be different. Statistics associated with those samples will also vary.

Just because your sample statistic doesn't equal your population parameter doesn't mean that your sample or statistic is invalid

Potential Samples
(n = 36)



Sampling Distributions

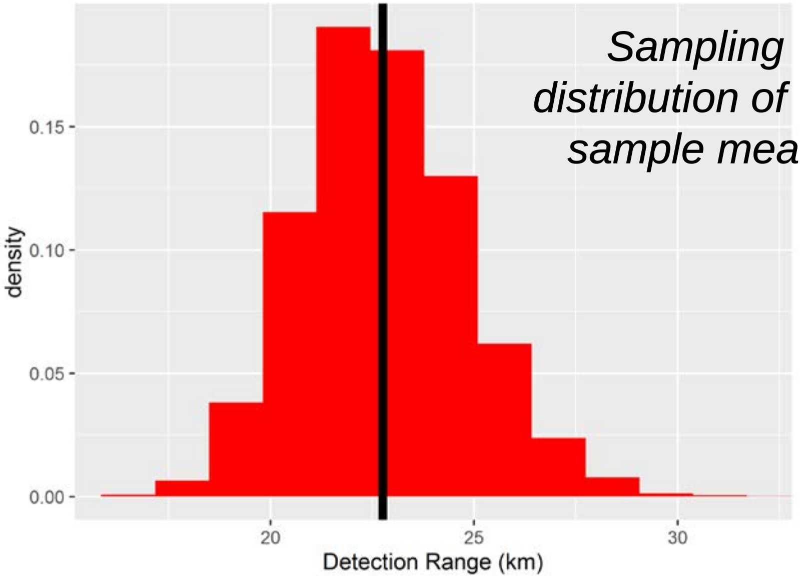


- For a particular population, sampling approach, and sample statistic, there is a distribution of potential realizations of that sample statistic.
- Estimating this distribution is the first step to performing statistical inference

Sampling Distribution: Hypothetical distribution of all possible sample statistics resulting from a particular sampling approach

P-values and Sampling Distributions

P-value: The probability of observing a sample as extreme or more extreme than the observed sample under a particular null hypothesis.



Sampling distribution of the sample mean

- **XTREME!**
  - “More extreme” meaning “Less likely under the null hypothesis”
  - Need to estimate sampling distribution of sample statistic under the null

- **Further information on p-values: see recent ASA statement**

Estimating the Sampling Distribution

- **Case 1: We know the population distribution perfectly**

  → Estimate the sampling distribution via Monte Carlo

  – Almost never see this

- **Case 2: We are willing to make some assumptions about the nature of the population distribution**

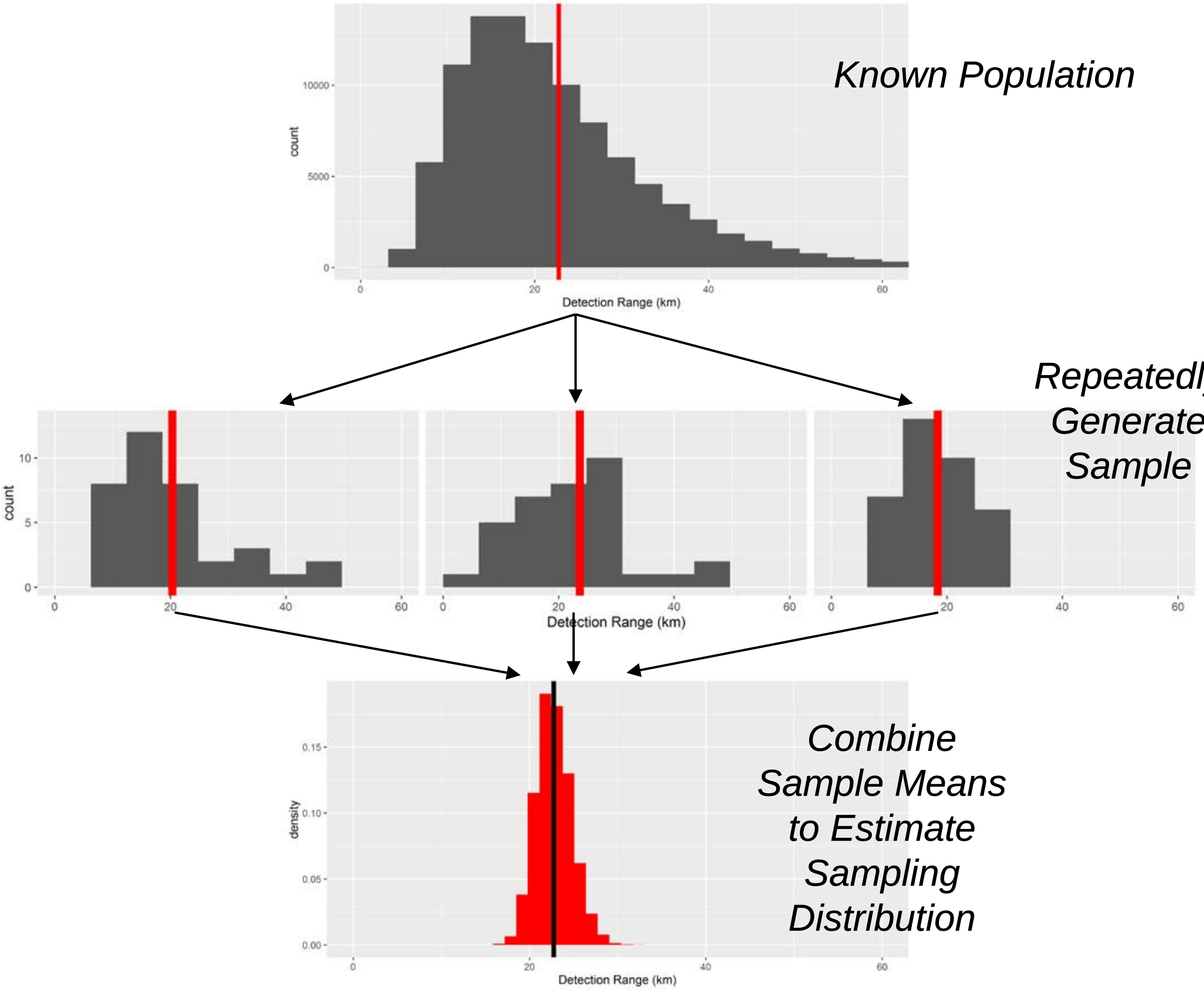
  → Estimate population parameters and derive sampling distribution mathematically using estimated population parameters

  – Most common case when statistical inference is applied

- **Case 3: We have little information about the population and no basis for making credible assumptions**

  → Estimate the sampling distribution via <u>bootstrapping</u>

Case 1: We know the sampling distribution perfectly → Estimate via Monte Carlo simulation



Known Population

Repeatedly Generate Sample

Combine Sample Means to Estimate Sampling Distribution

Case 2: We derive or approximate the sampling distribution based on assumptions → Estimate population parameters as necessary

Properties of Normal Random Variables:

1) *The sum of independent Normal RVs is Normal.*
2) *Multiplying a Normal RV by a constant will result in a Normal RV with scaled mean and variance*

Sometimes, we can approximate the distribution of a statistic, such the mean of a sample drawn from a non-normal distribution, which will follow a normal distribution according to the Central Limit Theorem.

- **Assume: Population, $\boldsymbol{x_1, x_2}$, … is Normally distributed with mean μ and variance σ²**

Sampling distribution for $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\rightarrow \sum_{i=1}^{n} x_i \sim N(n\mu, n\sigma^2)$$

$$\rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\frac{1}{n} * n\mu, \left(\frac{1}{n}\right)^2 * n\sigma^2\right)$$

$$\rightarrow \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- **Using the observed sample, we can estimate μ to generate confidence intervals and perform hypothesis tests***

**Note: This assumes a known value for σ². If the variance is unknown, it can be shown that the sampling distribution of $\bar{x}$ is a t distribution*

Case 3: We have little information about the population and no basis for making credible assumptions
→ Take a Magical Journey into the Bootstrap World!

- **Bootstrap World**
  - The observed sample from the real world is the population in the Bootstrap world.
  - Analyst has perfect knowledge of the bootstrap world
  - In the Bootstrap World, we’re in Case 1 instead of Case 3!

Real World	**Bootstrap World**
• Underlying distributions unknown	• Distributions can be fully characterized
• Finite samples	• Take as many samples as you like
• Interval estimates must be derived through complex math	• Interval estimates fall out from sampling distribution
• Reality	• Estimate of reality

If your Estimator is Based on an Unknown Parameter, "Plug In" an Estimator for the Unknown Parameter

Plug-In Estimator Example

Want to estimate the variance of some distribution

$$\boxed{\hat{\sigma}}^2 = \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \boxed{\bar{x}})^2}_{\textit{Common variance estimator}}$$

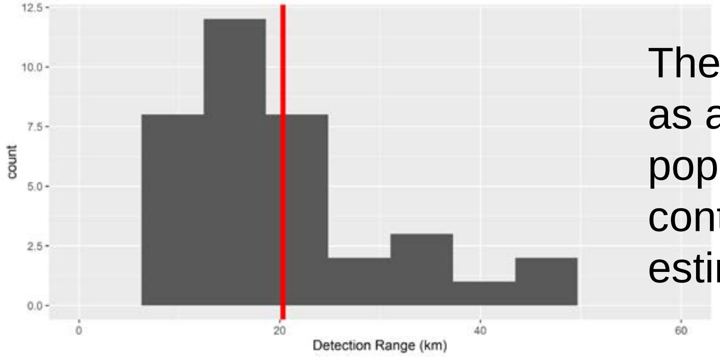
Plug-In estimator of population mean

- **Plug it in, plug it in!**
  - Widely-used approach
  - Resulting estimates depend on the quality of the plug-in estimator

Plug-in Principle: When a value of interest depends on something unknown (a parameter, distribution, etc.), plug in an estimator for it.

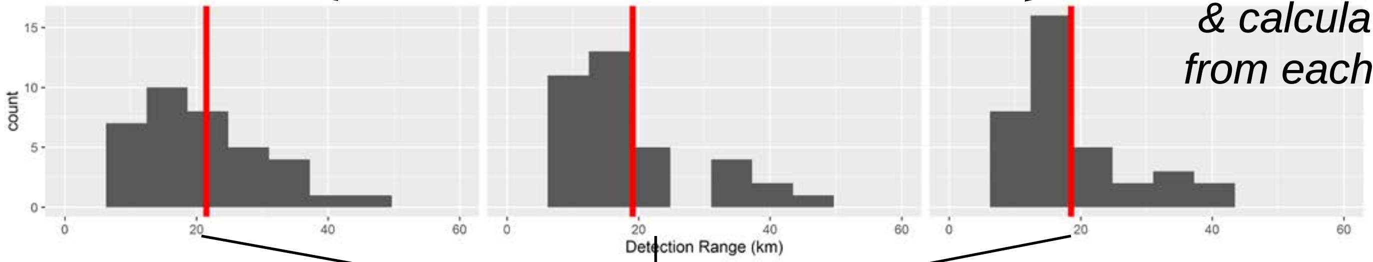
Estimating the Sampling Distribution via Bootstrap

Resampling: Drawing with replacement from observed sample

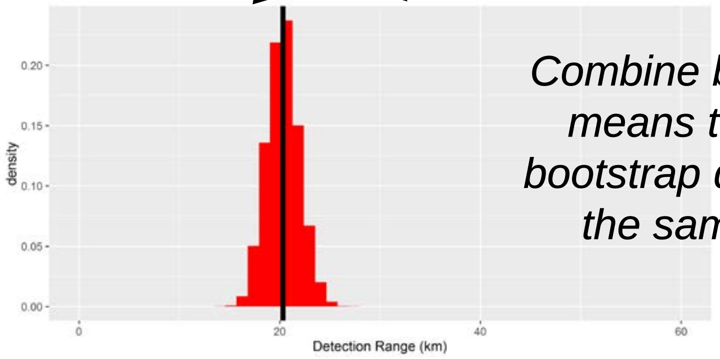


The Observed Sample is treated as a plug-in estimator for the true population. The process then continues like the Monte Carlo estimation described in Case 1.

Repeatedly resample & calculate means from each resample



"With Replacement" means that sampled values can repeat. Each observation in the original sample has an equal probability selected for each draw



Combine bootstrapped means to generate bootstrap distribution of the sample mean

Estimating the Sampling Distribution in the Bootstrap World By Resampling

- **Resampling approach**
  - Repeatedly re-sample observed data
    - Draw resamples from observed data <u>with replacement</u>
  - Calculate statistic of interest on each resample
  - Combine these resampled statistics to generate <u>bootstrap distribution</u>
- **Resampling Appropriately**
  - Complex statistics require a more careful approach
    - System availability
  - Ensure that the resampling is done using the same sampling approach that was used to generate the original sample
    - Simple Random Sample
    - Sampling from multiple populations
    - Relevant factors?
    - Complex statistics

Bootstrapping: Statistical inference accomplished by estimation of a particular sampling distribution through resampling an observed data set.

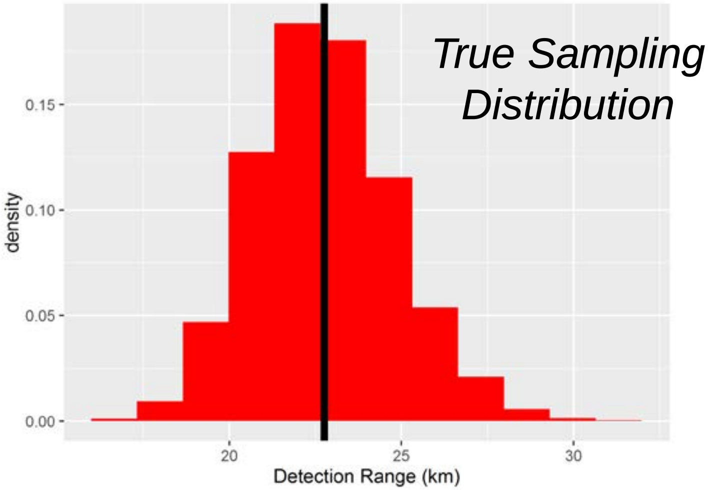
<u>System Availability</u>

$$A_O = \frac{\sum Up\ Times_i}{\sum (Up\ Time_i + Down\ Time_i)}$$

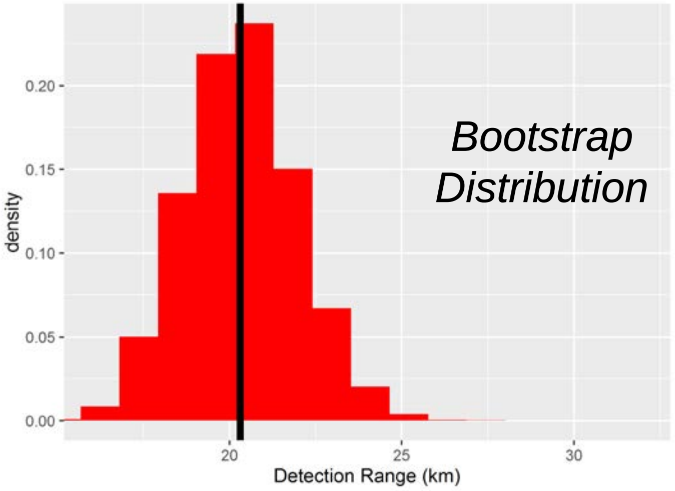
Bootstrap Distribution as an Estimator for the Sampling Distribution

- **Bootstrap for <u>inference</u> not for better <u>estimates</u>**
  - Mean of bootstrap distribution is still your <u>sample estimate</u>, not the mean of your true sampling distribution
  - Tells you how accurate your estimates are (confidence intervals)

- **Bootstrap distribution can be fully known**
  - n^n possible bootstrap resamples
  - Typically use a smaller number for estimating bootstrap distribution (10,000 for example)
  - Draw as many resamples as you need based on how precise an estimate you require



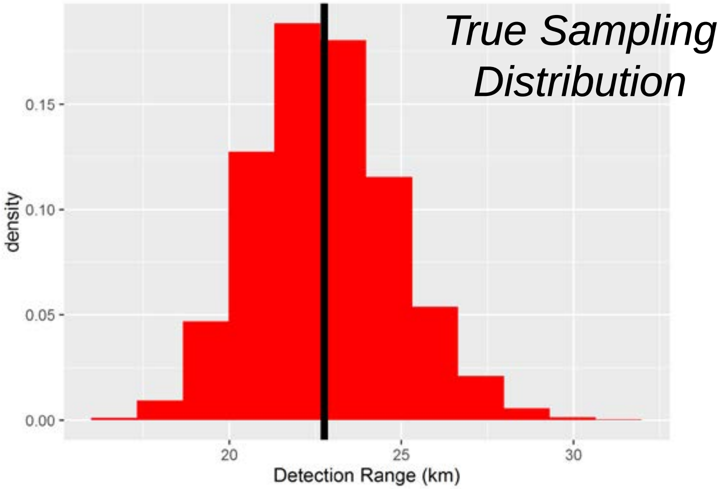
True Sampling Distribution



Bootstrap Distribution

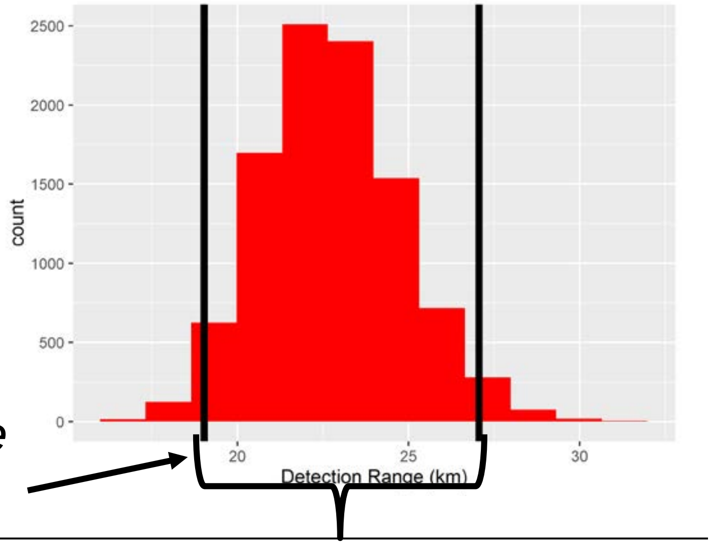
Confidence Intervals

Confidence Interval: A range of values that will contain a particular parameter with a specified probability



True Sampling Distribution

- **Confidence interval for sample mean in the real world**
  - Sampling distribution known
  - Interval around mean that will contain the mean 100*(1-α)% of the time
  - Monte Carlo approach: Generate 10,000 samples from population, drop the smallest 250 and largest 250

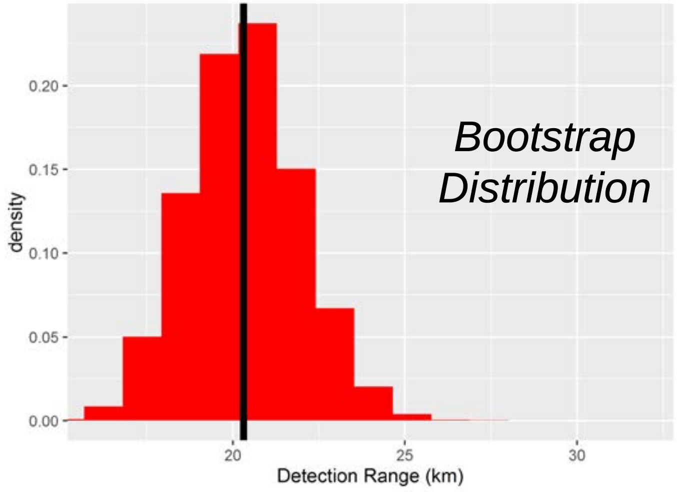


95 percent confidence interval for the mean

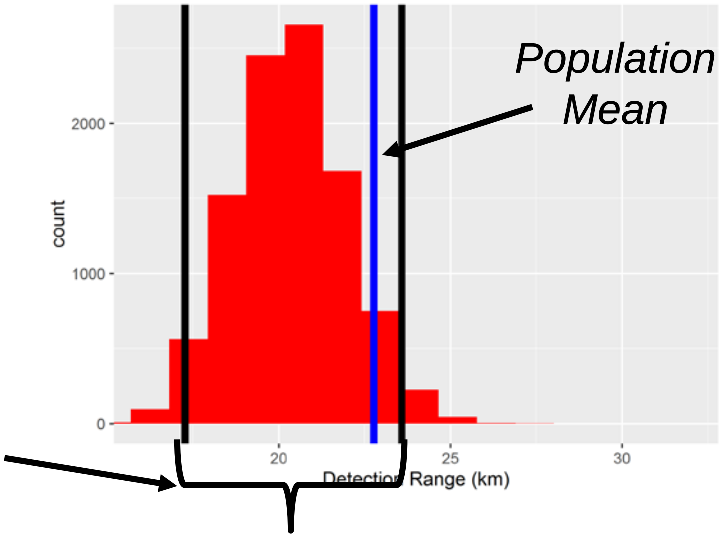
Bootstrap Confidence Intervals

Percentile Interval: Bootstrap confidence interval using percentiles of the bootstrap distribution to define an interval for the parameter of interest

- **Percentile Interval (Bootstrap World)**
  - Use bootstrap distribution for the sample mean as estimator for true sampling distribution
  - Monte Carlo approach:
    - » Generate 10,000 bootstrap resamples
    - » Calculate mean for each
    - » The 250th and 9,751st largest observations are lower and upper confidence bounds for a 95% confidence interval



Bootstrap Distribution



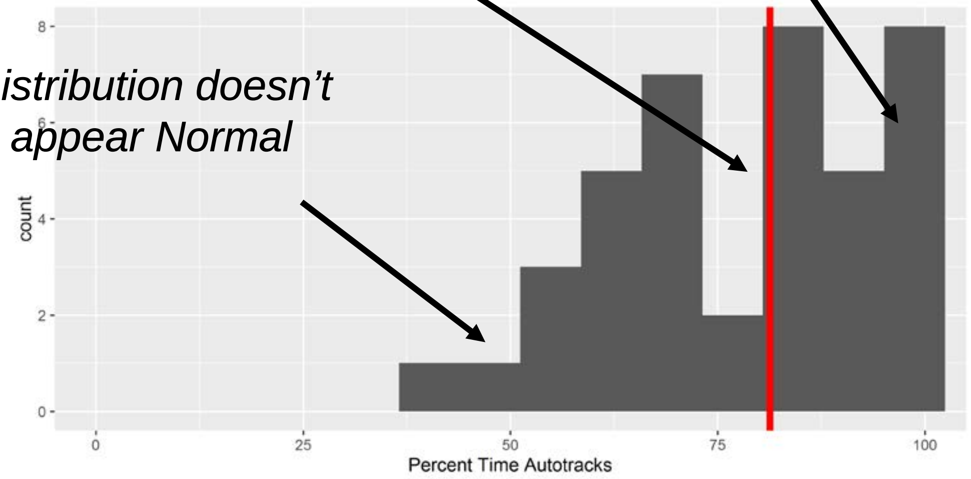
Population Mean

95 percent confidence bootstrap percentile interval for the mean

Bootstrap CI Example: MQ-8C Autotrack performance

- **Evaluate MQ-8C payload’s capability to lock onto particular targets & auto-track them**
  - Percent Time Autotrack:

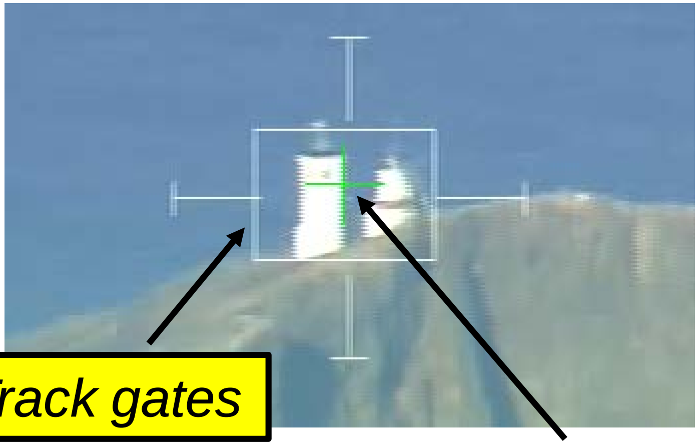
$$100 * \frac{Time\ Locked\ on\ Target}{Total\ Time\ Attempting\ to\ Lock\ on\ Target}$$



Want to estimate the median

Many observations at 100%

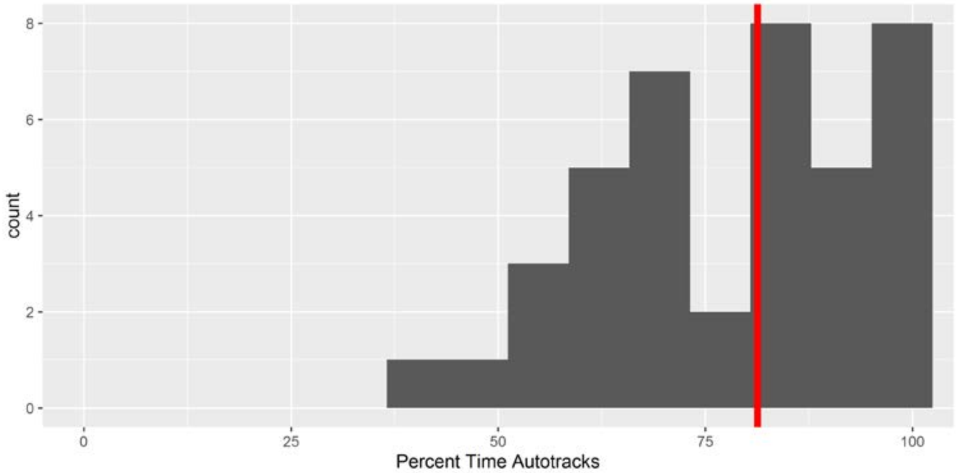
Distribution doesn’t appear Normal



Track gates

Targeting reticle

Sampling Distribution for Median Autotrack Times



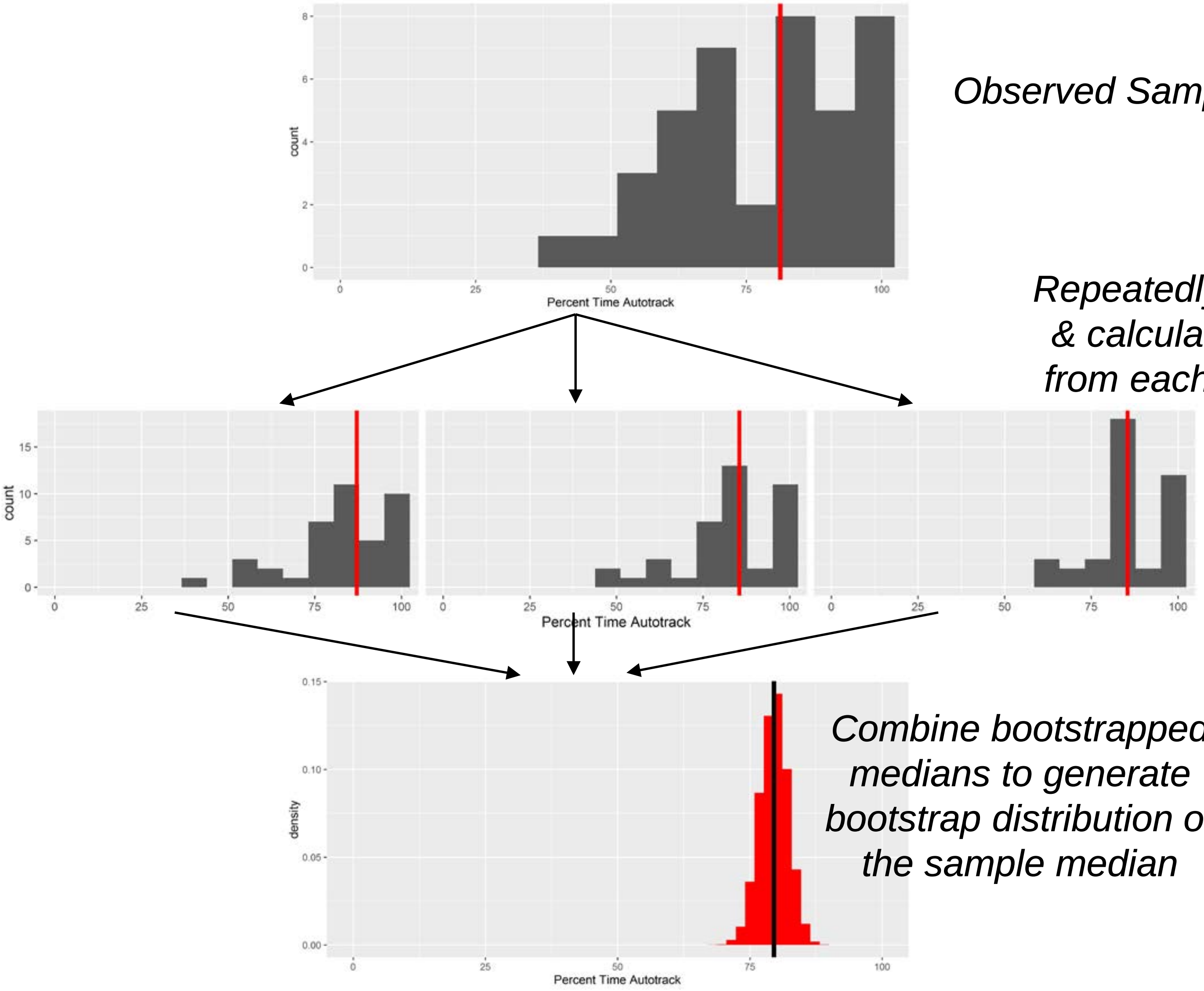
- ~~Case 1: We know the population distribution perfectly~~

- ~~Case 2: We are willing to make some assumptions about the nature of the population distribution~~

- **Case 3: We have little information about the population,✓ and no basis for making credible assumptions✓**

  → Estimate the sampling distribution of the median via <u>bootstrapping</u>

Estimating the Sampling Distribution via Bootstrap



Observed Sample

Repeatedly resample & calculate median from each resample

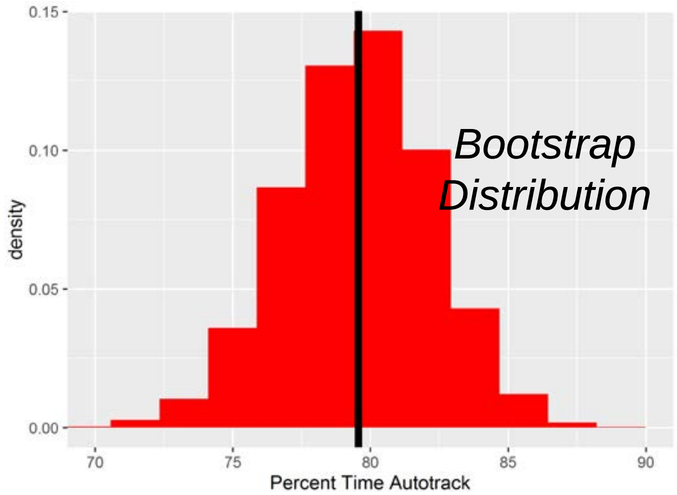
Combine bootstrapped medians to generate bootstrap distribution of the sample median

Bootstrap Confidence Interval for the Sample Median

- **Different Statistic, Same Approach**
  - Methodology for estimating median identical to methodology for mean
  - Generate bootstrap distribution of median & pick off the relevant quantiles

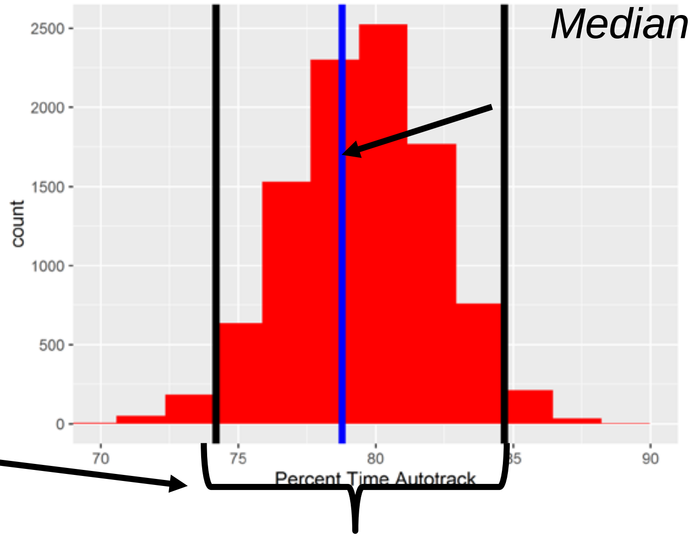
- **Nonparametric estimate**
  - No model specified
  - Able to quantify variance of our estimate of the median

- **Works with other quantiles, too!**
  - Remember: Must have sufficient data to estimate quantile to begin with



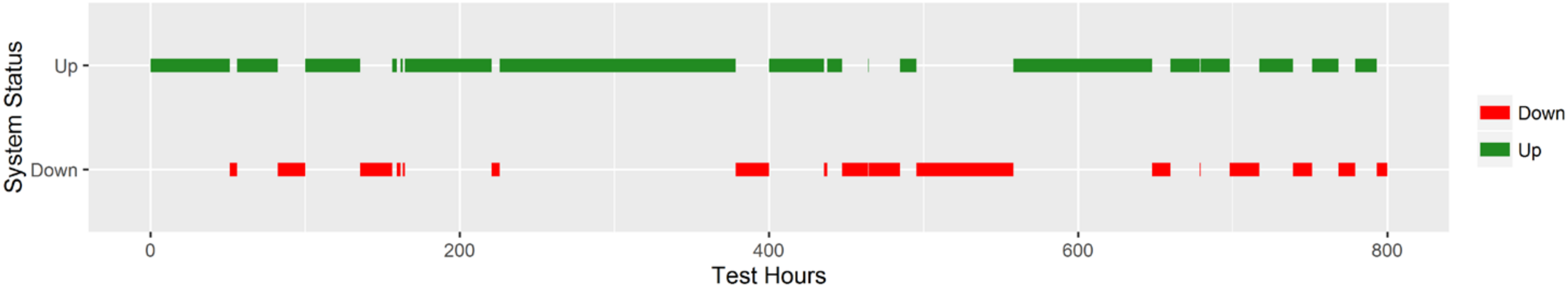
Bootstrap Distribution

Population Median



95 percent confidence bootstrap percentile interval for the median

Bootstrapping Offers an Alternative to Parametric Methods for Availability Confidence Intervals



$$A_O = \frac{\sum Up\ Times_i}{\sum(Up\ Time_i + Down\ Time_i)}$$

- **System Availability**
  - – Function of observations from *two* distributions

Parametric Approach

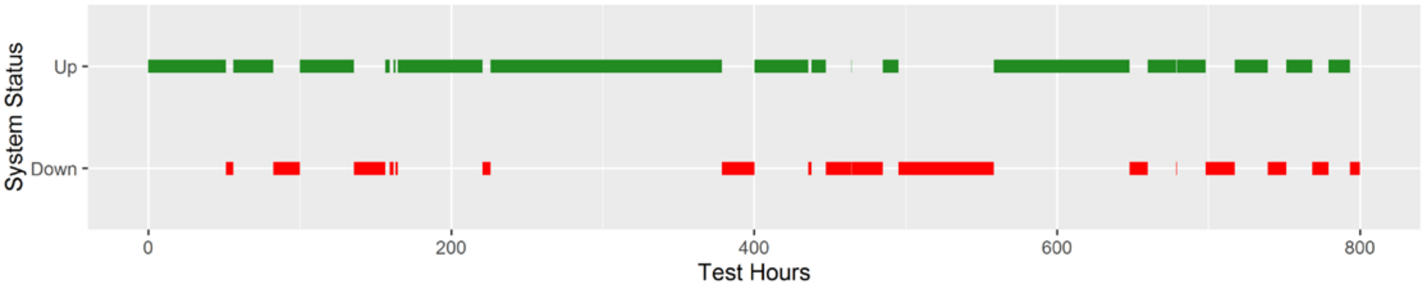
- Specify model for each distribution
- Derive distribution of statistic
- Estimate confidence interval

Bootstrap Approach

- Re-sample *entire test* (up times and downtimes)
- Compute statistic for each iteration
- Generate bootstrap distribution

Resample Over the Test Period Rather Than the Number of Up and Down Times

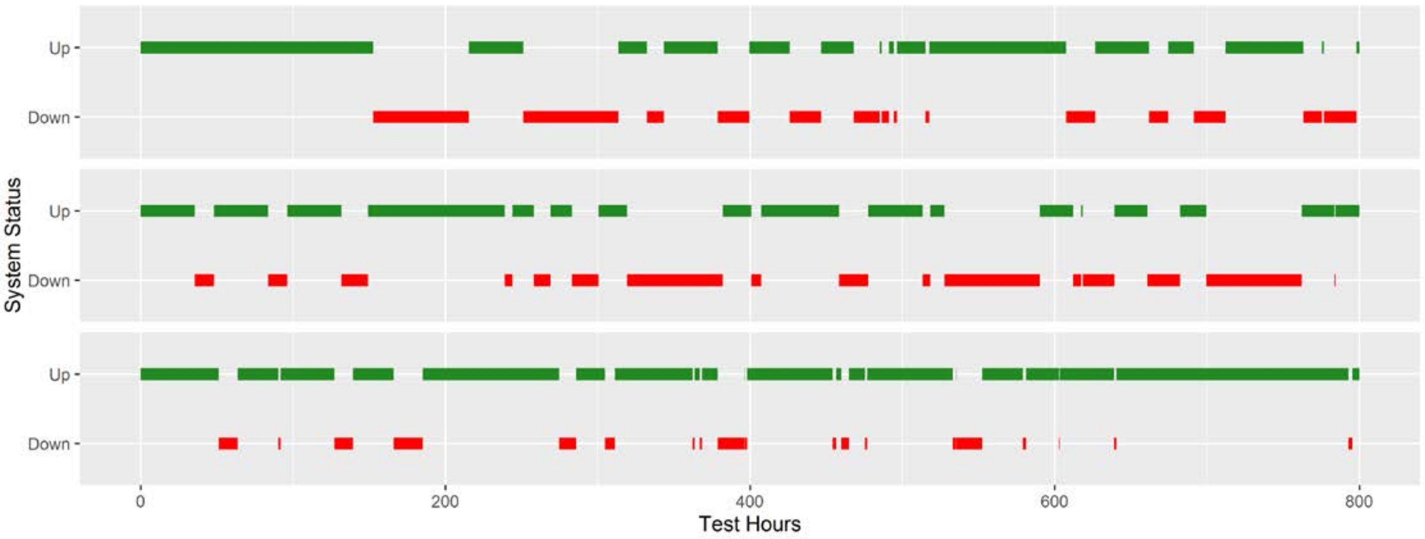
Original Data



$$A_O = \frac{562.2}{800} = 0.703$$

🟥 Down
🟩 Up

Resample 800 hours of testing rather than n uptimes and m downtimes. Draw individual up/down times from data and continue drawing until the resampled test has the same length as the actual test



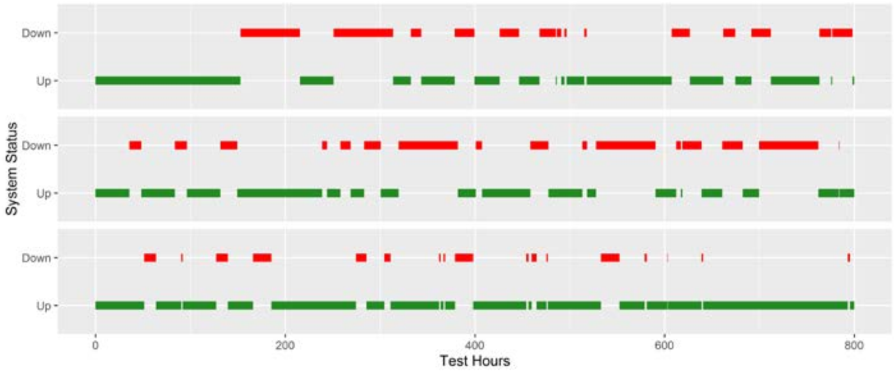
$$A_O^* = \frac{510.5}{800} = 0.638$$

$$A_O^* = \frac{456.8}{800} = 0.571$$

$$A_O^* = \frac{680.2}{800} = 0.850$$

Bootstrap resamples of the test

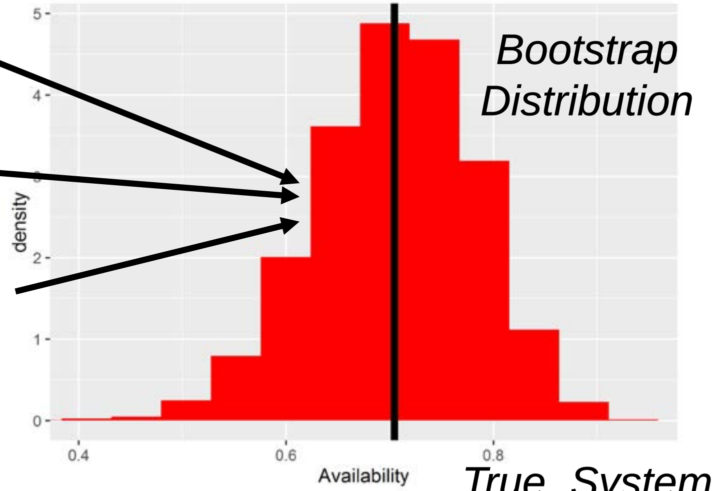
Bootstrap Confidence Interval for System Availability



$A_O^* = 0.638$

$A_O^* = 0.571$

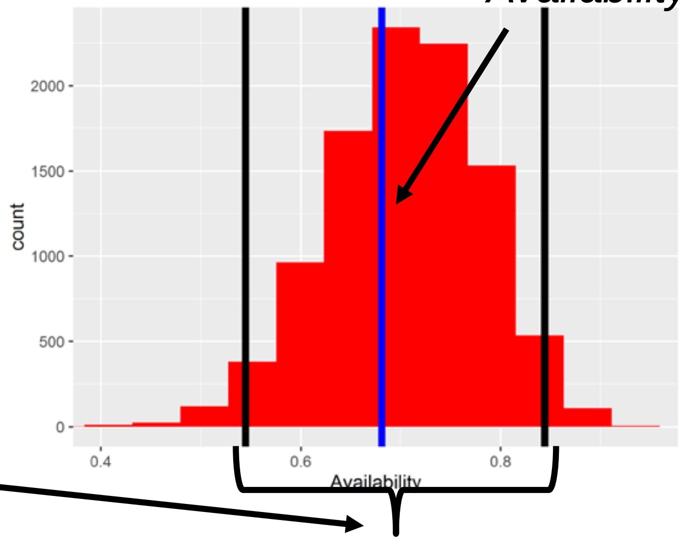
$A_O^* = 0.850$



Bootstrap Distribution

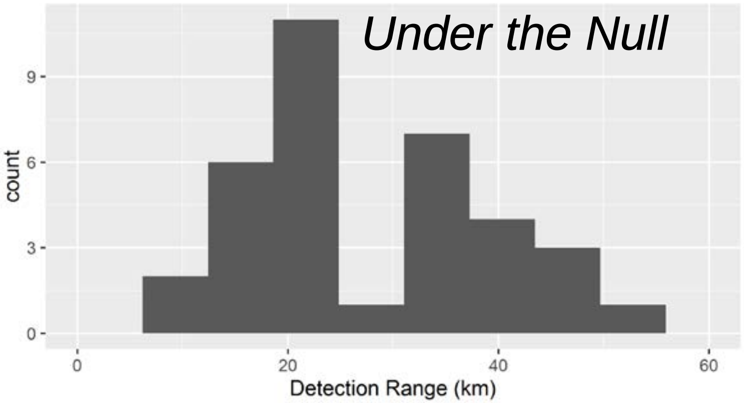
True System Availability

- **Generate bootstrap distribution using the same approach as for the original sample**
  - Draw Up Times and Down Times from *sample* Up & Down Times instead of population
  - Are Up & Down Times independent?
    - Based on correlation and/or understanding of system engineering and maintenance
    - If not, my need to draw as *pairs*

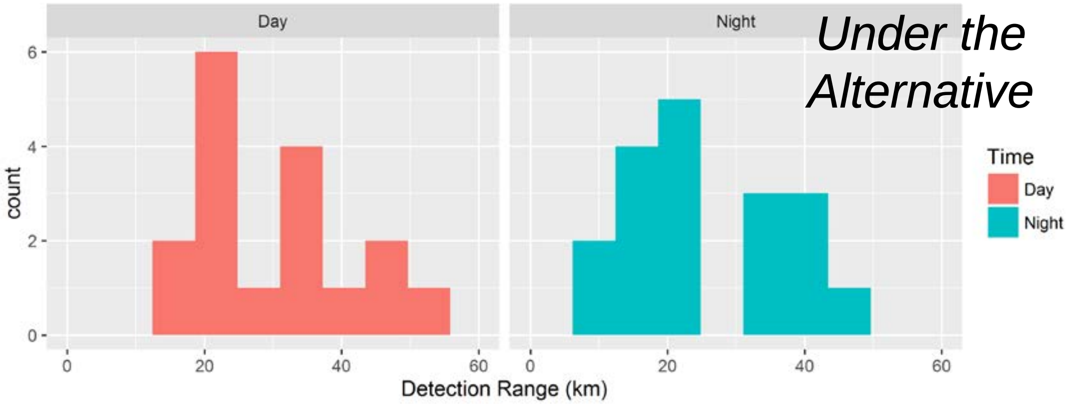


95 percent bootstrap percentile interval for the median

Two Sample Hypothesis Testing



Under the Null



Under the Alternative

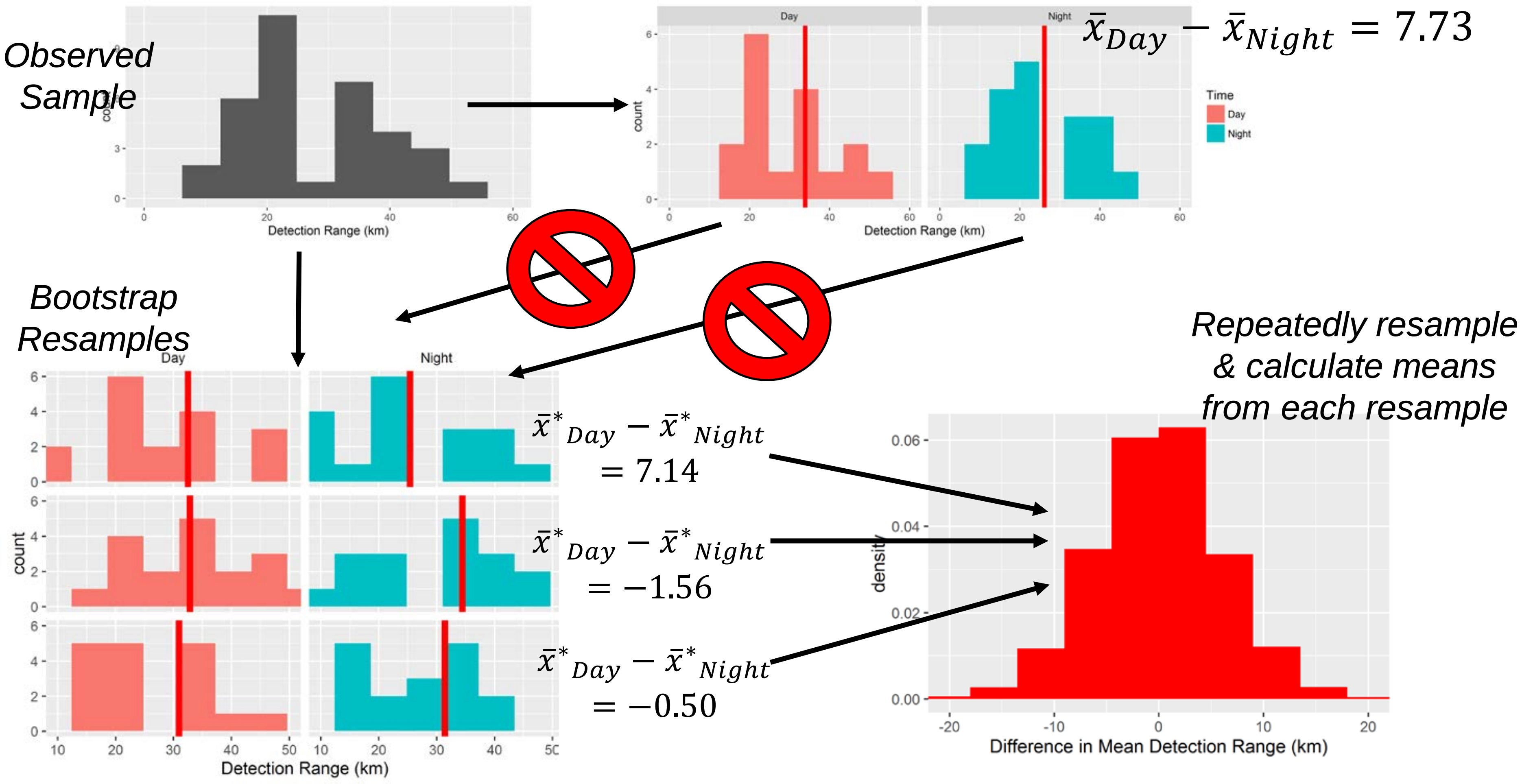
Hypothesis Test:

$$H_0: Mean\ Detection\ Range_{Day} = Mean\ Detection\ Range_{Night}$$
$$H_1: Mean\ Detection\ Range_{Day} \neq Mean\ Detection\ Range_{Night}$$

Phrased differently:

$$H_0: Mean\ Detection\ Range_{Day} - Mean\ Detection\ Range_{Night} = 0$$
$$H_1: Mean\ Detection\ Range_{Day} - Mean\ Detection\ Range_{Night} \neq 0$$

Estimating the Sampling Distribution via Bootstrapping



Observed Sample

$\bar{x}_{Day} - \bar{x}_{Night} = 7.73$

Bootstrap Resamples

$\bar{x}^*{}_{Day} - \bar{x}^*{}_{Night} = 7.14$

$\bar{x}^*{}_{Day} - \bar{x}^*{}_{Night} = -1.56$

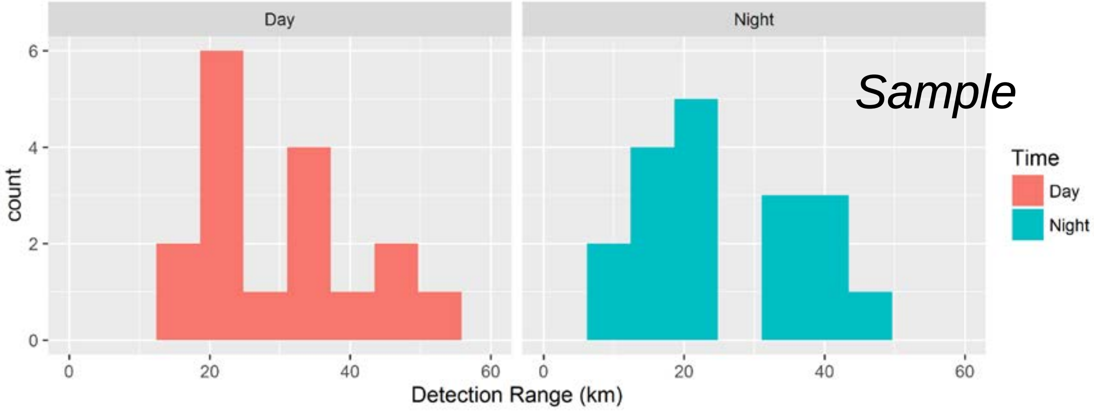
$\bar{x}^*{}_{Day} - \bar{x}^*{}_{Night} = -0.50$

Repeatedly resample & calculate means from each resample

Two Sample Hypothesis Test via Bootstrapping

Hypothesis Test:

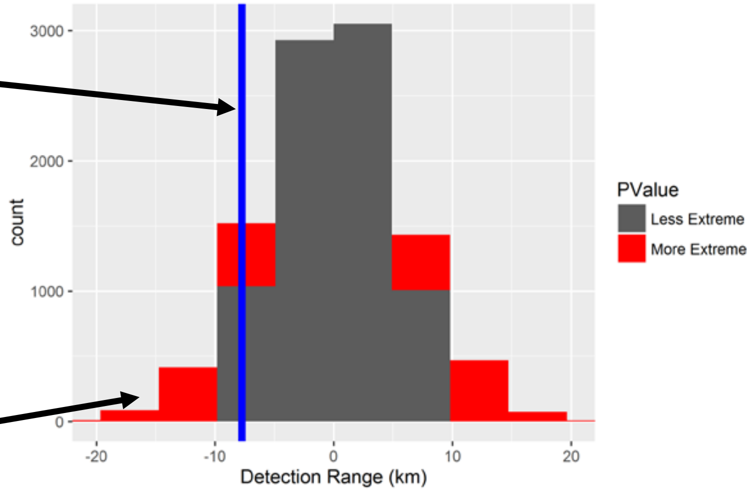
$$H_0: \mu_{Day} - \mu_{Night} = 0$$
$$H_1: \mu_{Day} - \mu_{Night} \neq 0$$



Sample

- **Calculate p-value by determining proportion of sampling distribution more extreme than observed sample mean**
  - P-value = 0.1974

Observed difference in detection range



Portion of sampling distribution "more extreme" (according to alternative) than observed sample mean

More Things to Explore in the Bootstrap World

- **Parametric Bootstrap**
  - Assume population distribution & estimate parameters with sample. Then re-sample from estimated population to characterize sampling distribution of parameter of interest.
- **Other kinds of bootstrap confidence intervals**
  - Bias-corrected
  - Accelerated bootstrap
  - Bootstrap t
  - Etc., etc., etc.
- **Bootstrap confidence intervals in regression**
  - Simple Linear Regression
  - Generalized Linear Models
  - Mixed Models
- **Comparisons with permutation testing**

Summary and Cautions

- **Bootstrapping**
  - Powerful tool applicable in a variety of situations
    - Quantify Variance
    - Hypothesis Testing
- **Most useful when:**
  - Distributions unknown or complex
  - Deriving sampling distribution intractable/impractical
- **Always remember:**
  - Use for inference not estimation
  - Resample using the same approach that was used to generate your sample
    - For hypothesis testing, resample under the null hypothesis
  - Bootstrap results can only ever be as good as the sample upon which they're based, since you're using the sample as a plug-in estimator for your population.

References

- **“Introduction to the Bootstrap World,” Dennis Boos; *Statistical Science*, 2003, Vol. 18, No. 2 168-174**

- ***Essential Statistical Inference: Theory and Methods.* Dennis Boos & Leonard Stefanksi. Springer Texts, 2013.**

- **“Bootstrap Methods: Another look at the Jackknife”, Bradley Efron. *The Annals of Statistics*. 1979 Vol 7, No 1 1-26.**

- **“Some Asymptotic Theory for the Bootstrap,” Peter Bickel and David Freedman. *The Annals of Statistics.* 1981 Vol 9, No 6, 1196-1217.**

- **“What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum”, Tim C. Hesterberg. The American Statistician, 2015, 69:4, 371-386**

Questions?

???

Backup

The Sampling Distribution is the Basis for Statistical Inference

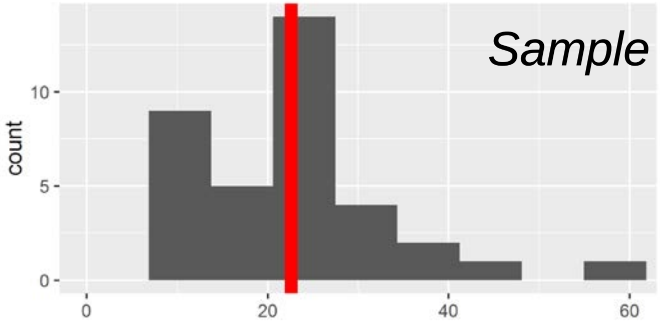
Most common approaches for estimating the sampling distribution of a sample statistic:

- **Known (or assumed) properties of population distributions**
  - If population has a Normal distribution, sample mean will have a normal distribution
    - » $\frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)\ if\ x_1, \ldots, x_n \sim N(\mu, \sigma^2)$
- **Known properties of estimators**
  - Confidence interval for the mean based on the Central Limit Theorem
    - » $\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu\right) \xrightarrow{d} N(0, \sigma^2),\ \text{where}\ \mathrm{Var}(x_i) = \sigma^2 < \infty$
- **In some cases, these approaches break down**
  - Don’t know or can’t easily characterize population distribution
  - Interested in quantities that don’t have nice properties/easily applicable theorems

One Sample Hypothesis Testing
Monte Carlo Approach

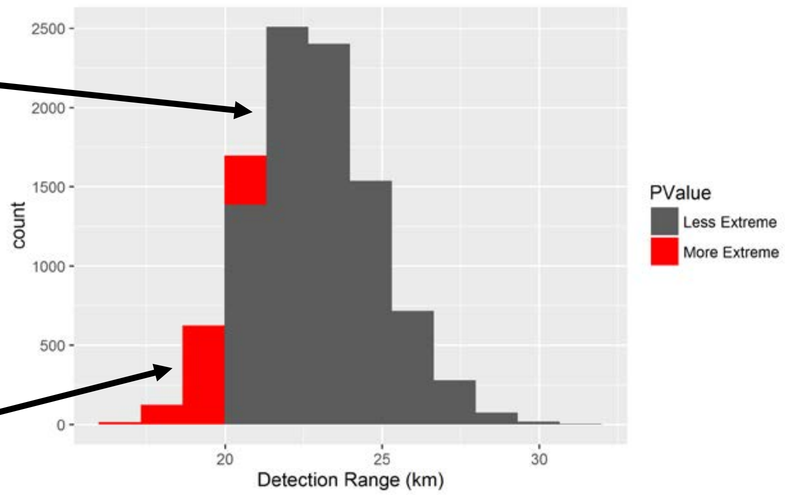
<u>Hypothesis Test:</u>

$$H_0: Mean\ Detection\ Range = 20.3\ km$$
$$H_1: Mean\ Detection\ Range < 20.3\ km$$



- **Calculate p-value by determining proportion of sampling distribution lower than observed sample mean**
  - P-value = 0.1072

Sampling distribution for sample mean under null distribution



Portion of sampling distribution "more extreme" (according to alternative) than observed sample mean