



Statistics: Is there an easier way?

Meetup: Data İstanbul

April 5, 2017

H. Sait Ölmez

Deduction & Induction in Statistics

Deduction (probability)



Probabilist: Asks the probability of picking someone who is blond given the proportions in the population?

Induction (statistical inference)



Statistician: Infers the proportion of the blonds by sampling from the population.

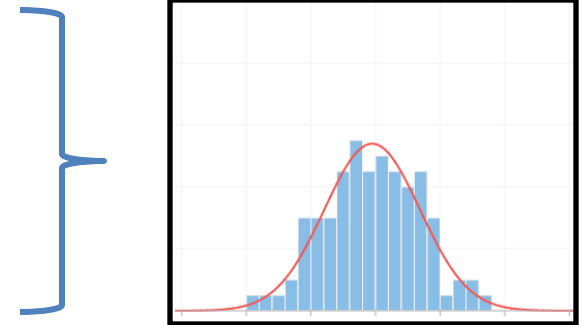
Inferential Statistics

POPULATION



Sample (n) $\rightarrow \bar{x}$
Sample (n) $\rightarrow \bar{x}$
Sample (n) $\rightarrow \bar{x}$
.....

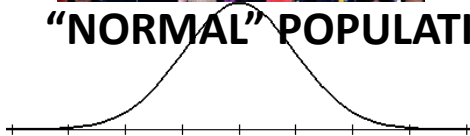
.....



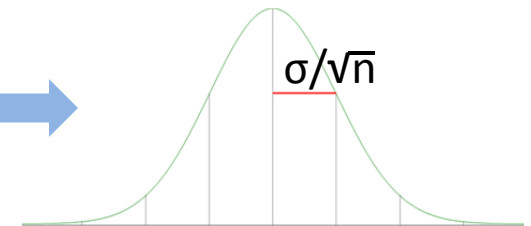
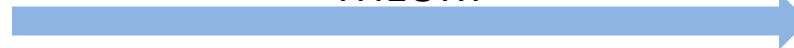
Sampling distribution



"NORMAL" POPULATION



THEORY

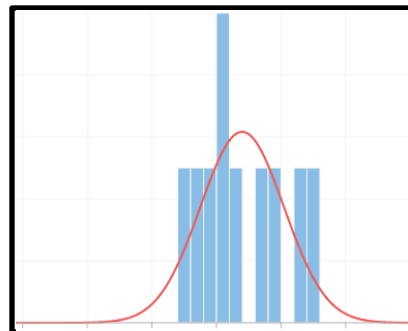
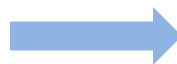


Sampling distribution

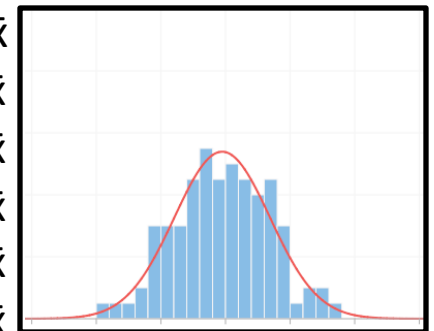
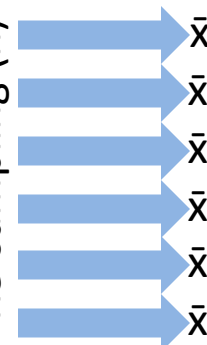
POPULATION



Just one
Sample
(n)



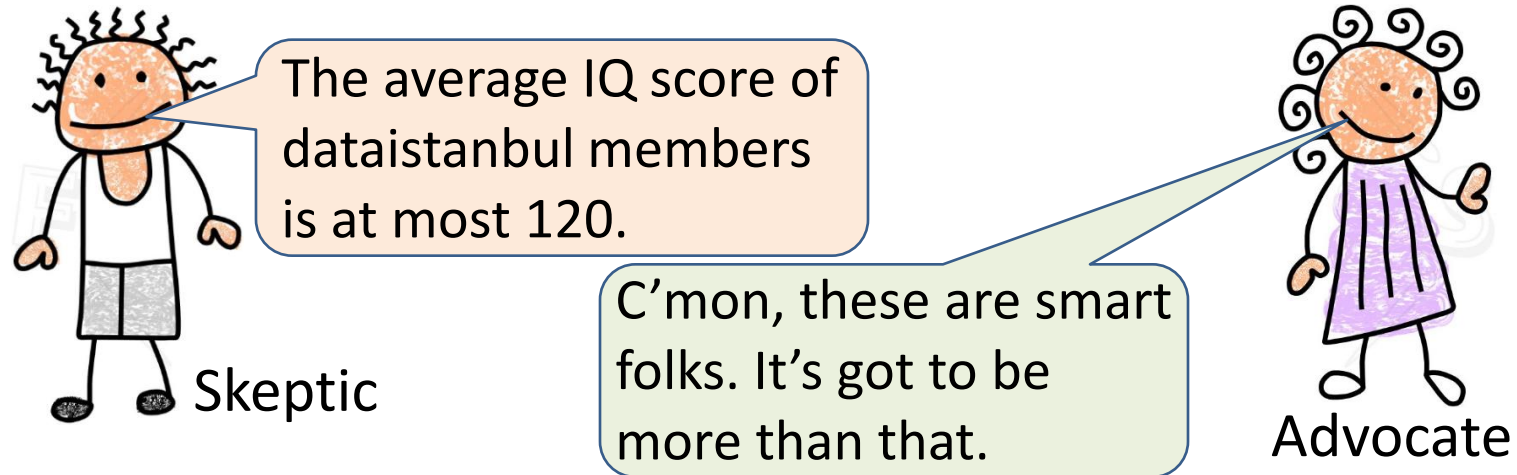
Re-sampling (n)



Bootstrap distribution

- **Testing Hypotheses**
- **Step 1:** Identify the problem and state the hypotheses

Problem: Are **dataistanbulians** smart?



Hypotheses:

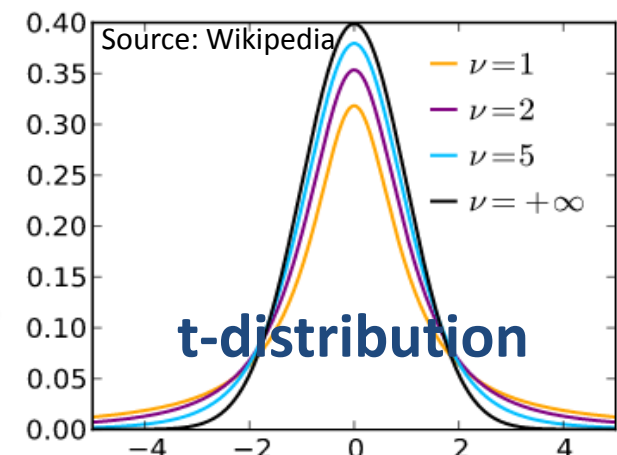
Null Hypothesis $(H_0) : \mu_{IQ} \leq 120$ (assume true)

Alternative Hypothesis $(H_A) : \mu_{IQ} > 120$

- **Step 2:** Set a confidence/significance level: α
- α : Probability of rejecting the Null when it's true. Conventional value is 5% establishing a 95% confidence level. This is the mistake we're willing to make in incorrectly rejecting H_0 when it's true.
- **Step 3:** Collect data (for a sample of 20 let's say)

Örnek = { 129,125,124,120,117,134,122,
123,122,118,123,122,120,124,
119,123,120,121,119,129 } } $\mu_{IQ} = 122.7$

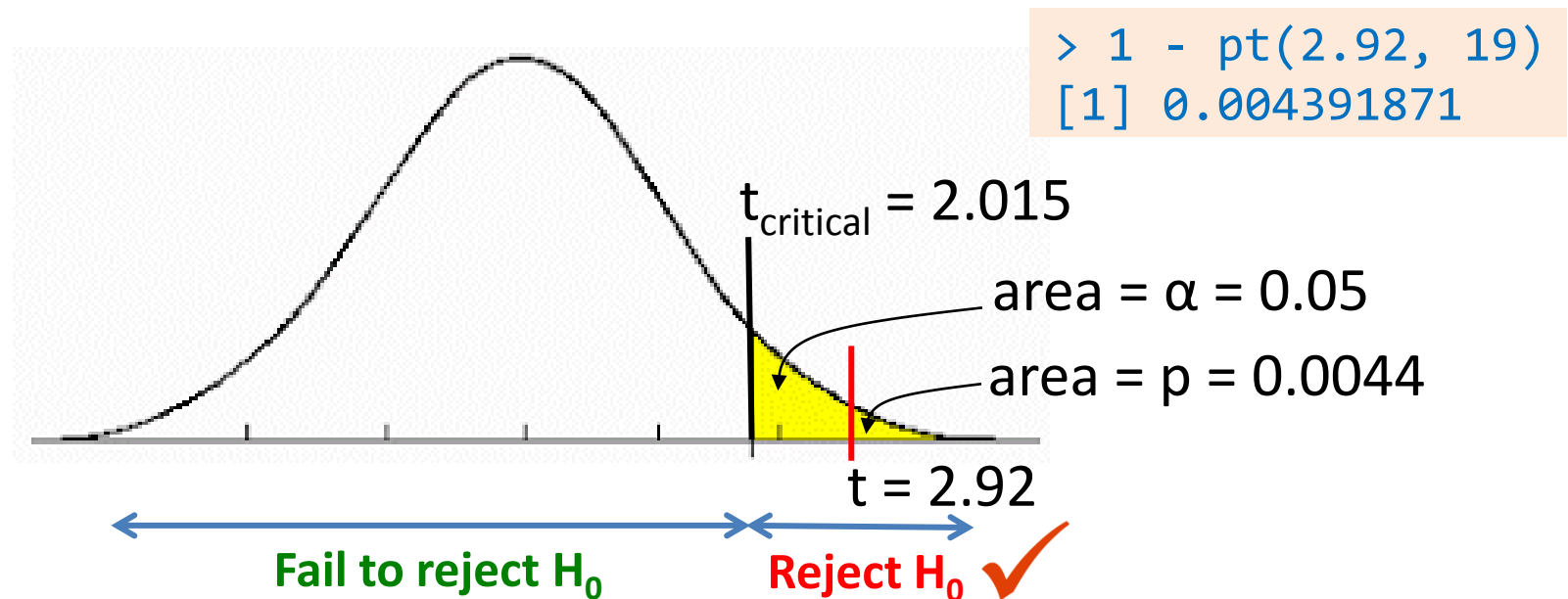
- **Step 4:** Select the sampling dist. and specify the test statistic
 - Population distr. : Unknown
 - Population variance: Unknown
 - Sample size : 20 (small)



- **Step 5:** Calculate the test statistic & critical values

$$\text{t-statistic: } t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{122.7 - 120}{4.131 / \sqrt{20}} = 2.92$$

where \bar{x} is the sample mean, s is the sample standard deviation and n is the sample size.



Statistics & Simulation

Calculation of π

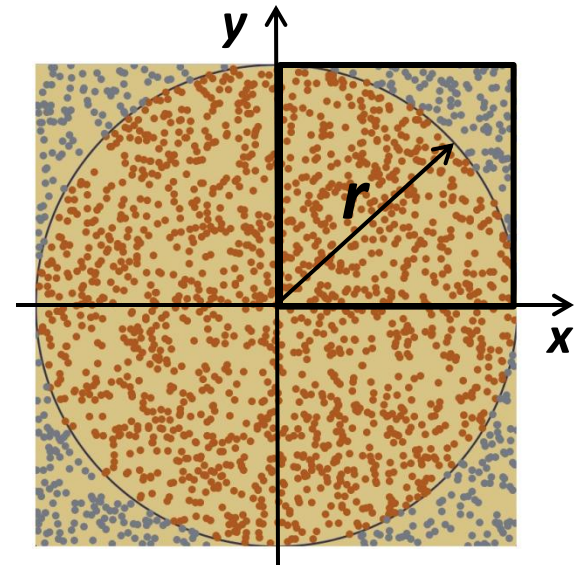
- Randomly sample N points between 0 and 1
- Count the number of points that fall inside the $1/4^{\text{th}}$ circle
- Ratio of this number to N will yield $1/4^{\text{th}}$ of π

```
piR <- function(N) {  
  x <- runif(N,0,1)  
  y <- runif(N,0,1)  
  d <- sqrt(x^2 + y^2)  
  return(4*sum(d < 1.0)/N)  
}
```

```
set.seed(7)  
cat(piR(1000), piR(10000),  
piR(100000), piR(1000000))
```

3.192 3.1312 3.14284 3.141764

- Tried different values of N to see how closely π is approximated.

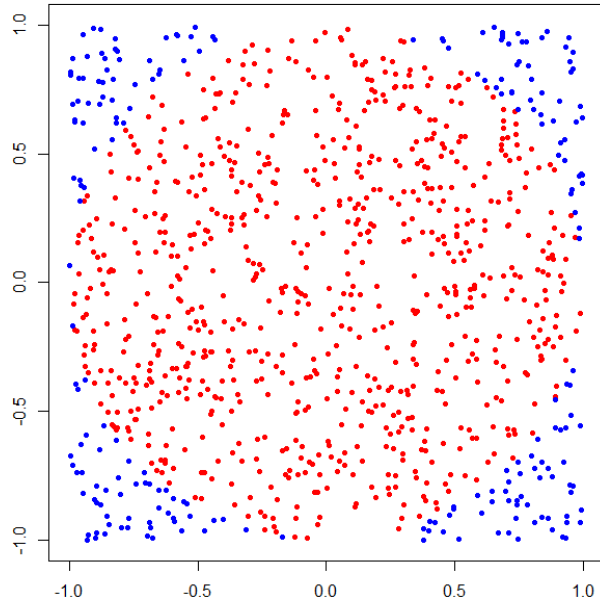


$$\text{Ratio} = \frac{\text{Area}_{1/4\text{circle}}}{\text{Area}_{1/4\text{square}}}$$

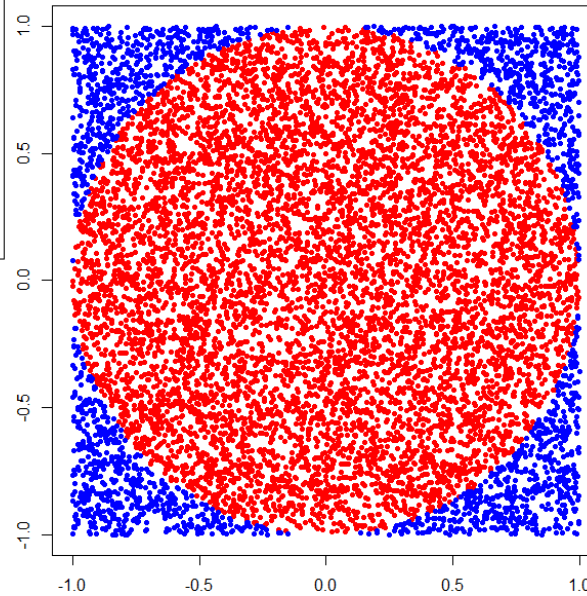
$$\text{Ratio} = \frac{\pi r^2 / 4}{r^2} = \frac{\pi}{4}$$

Calculation of π

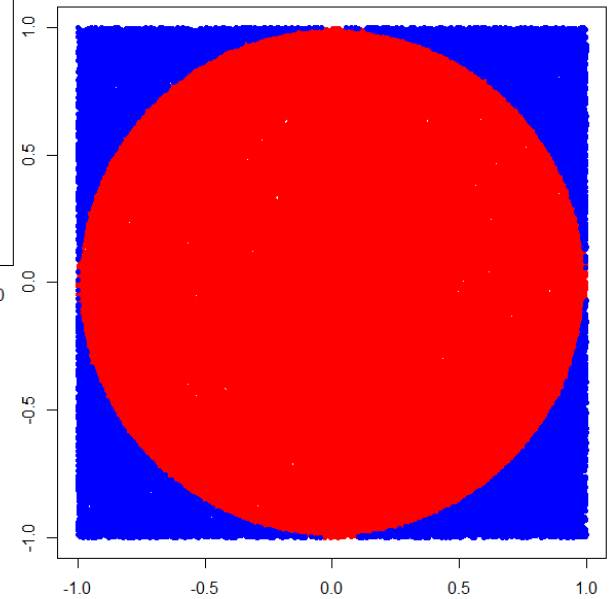
N=1000



N=10000



N=100000



Computational Methods in Statistics

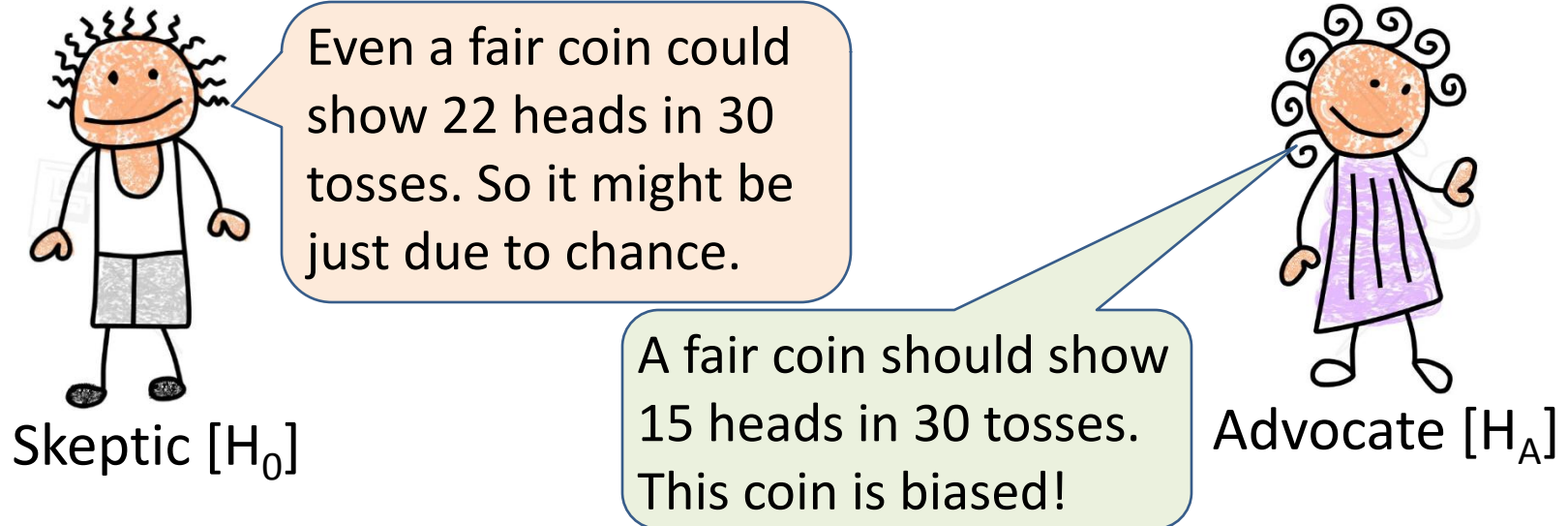
- With the advance of computational methods, many problems of statistics can be solved by using computer simulations.
- Computing the sampling distribution is hard, but simulating the sampling distribution is easy.
- A summary of advantages:
 - Fewer assumptions: These methods don't require that distributions be Normal or that sample size is large
 - Better accuracy: In practice, more accurate than classical methods
 - Generality and easy to understand: Generalizable for a wide range of statistics and help build intuition by solid analogies to theoretical concepts.

Computational Methods in Statistics

- Recipes for easy statistics:
 - Direct simulation
 - Shuffling (random permutation)
 - Simple Random Sampling
 - Bootstrapping
 - Cross Validation

DIRECT SIMULATION

- Problem: You toss a coin 30 times and you see 22 heads. Is this a fair coin?



- Classic method: Assume that the skeptic is right and test the Null Hypothesis [H_0].
- What is the probability of a fair coin showing 22 heads or more in 30 tosses?

Ref: "Statistics for Hackers", Jake Vanderplas, PyCon 2016

Direct simulation

- This problem has a theoretical model for solution (binomial distribution) and the probability of getting 22 heads or more out of 30 trials is:

$$p(k, n) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where} \quad \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{is the number of arrangements (binomial coeff.)}$$

prob. of k heads prob. of (n-k) tails

$$p(k \geq n_H, n) = \sum_{k=n_H}^n \binom{n}{k} p^k (1-p)^{n-k} \quad \text{probability of getting } n_H \text{ heads or more in } n \text{ tosses}$$

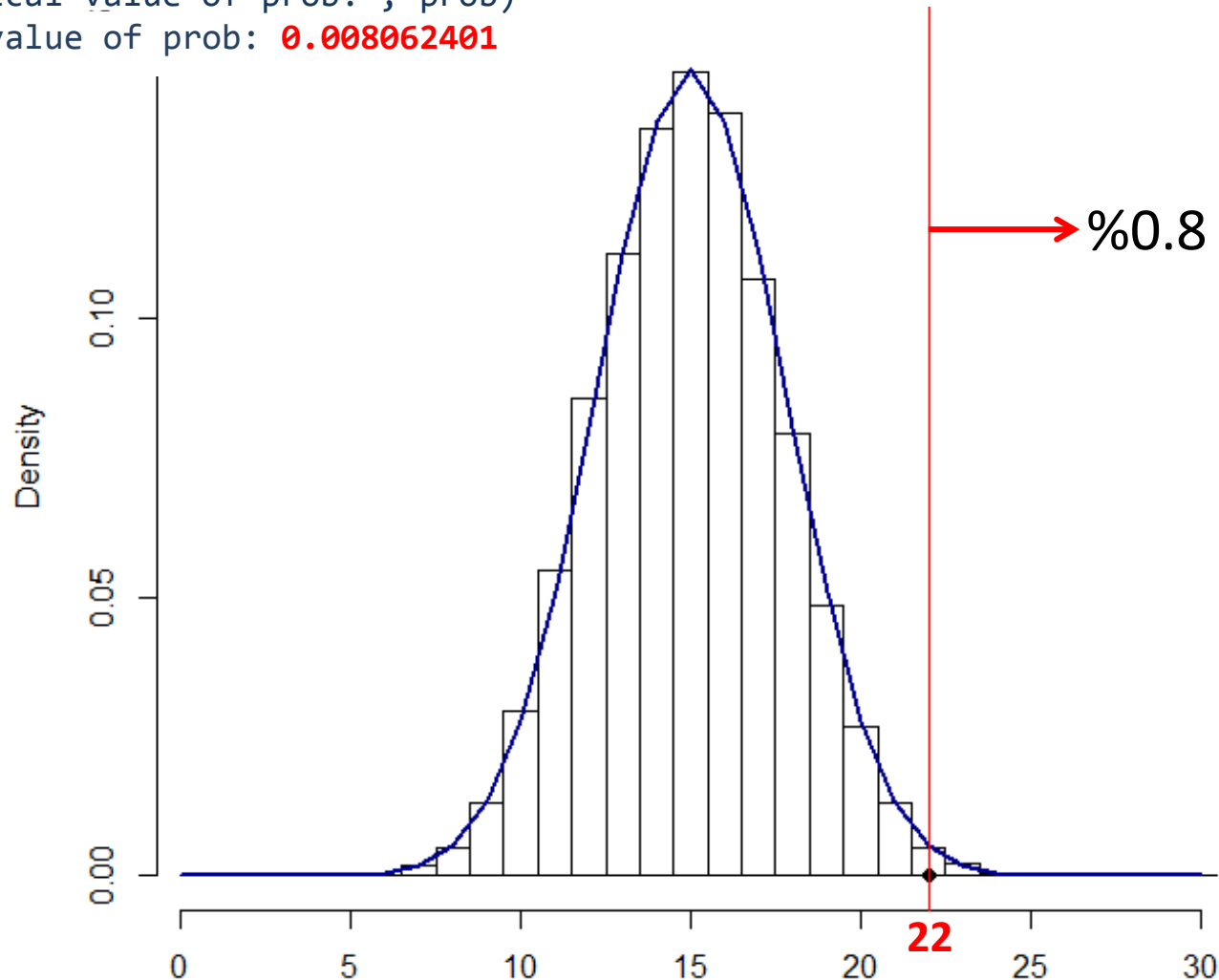
$$p(k \geq 22, 30) = \sum_{k=22}^{30} \binom{30}{k} (0.5)^k (0.5)^{30-k} \approx 0.008 = 0.8\%$$

- Assuming that the Null is true (fair coin), the probability of getting our data simply by chance is 0.8%. This is less than the p-value of 0.05. Therefore, we reject the Null Hypothesis. So, our coin is not a fair coin.

Direct simulation

```
print(hist(rbinom(10000,30,.5),freq=FALSE, breaks=seq(0.5,30.5,1), ylim=c(0,0.15)))  
print(lines(seq(0,30,1),dbinom(seq(0,30,1),30,0.5), col="darkblue", lwd="2"))  
points(x=22, y=0, pch=16) ; abline(v=22, col="red")  
prob <- print(1-pbinom(21, 30, .5))  
cat("Theoretical value of prob:", prob)  
Theoretical value of prob: 0.008062401
```

Binomial distribution
(analytical)



- Is there an easier way?
- How about a simulation?

```
N = 10000 ; M = 0
set.seed(10)
for (i in 1:N) {
  x1 <- sample(0:1, 30, replace=T)
  if (sum(x1) >= 22) {
    M = M + 1
  }
}
cat("Trials with more than 22 heads : ", M, "\n")
cat("Ratio of M to N                : ", M/N, "\n")
```

```
Trials with more than 22 heads : 81
Ratio of M to N                : 0.0081
```

Less than p-value=0.05,
so reject H_0 (not a fair coin)!

SHUFFLING

(Random Permutations)

Beer Consumption Human Attractiveness to Malaria Mosquitoes



Article

Metrics

Related Content

Comments: 0

Thierry Lefèvre^{1*}, Louis-Clément Gouagna^{2,3}, Kounbo Roch Dabiré^{3,4}, Eric Elguero¹, Didier Fontenille², François Renaud¹, Carlo Costantini^{2,5}, Frédéric Thomas^{1,6}

 To add a note, highlight some text. [Hide notes](#)
 [Make a general comment](#)

- Does drinking beer make you more attractive to mosquitoes?
- We have 25 volunteers who drink beer and 18 volunteers who drink water.
- We then record how many mosquitoes chose to bite the human subjects for each group . Here are the results:

Ref: "Statistics without the agonizing pain", John Rauser, Strata Conf. 2014

- Those who drank beer attracted 4.4 more mosquitoes than the water drinkers. Is this statistically significant? Or could it have happened just by chance?



BEER

27 19 20 20 23
17 21 24 31 26
28 20 27 19 25
31 24 28 24 29
21 21 18 27 20

Mean_B : $\mu_B = 23.6$

WATER

21 19 13
22 15 22
15 22 20
12 24 24
21 19 18
16 23 20

Mean_W : $\mu_W = 19.2$



- Difference in means: $\delta = \mu_B - \mu_W = 4.4$
- Is this statistically significant?

- **Analytical Solution**
- Skeptic => Null Hypothesis $H_0: \mu_B = \mu_W$
- Advocate => Alternative Hyp. $H_A: \mu_B \neq \mu_W$
- This is a problem of hypotheses involving 2 populations when the variances are unknown and unequal. So the t-test is an appropriate test.
- We need to calculate the t-score first:
 - Mean values: $\mu_W = 23.6$ gr and $\mu_B = 19.2$ gr
 - Variances: $S_B^2 = 17.08$, $S_W^2 = 13.48$ where $N_B = 25$ & $N_W = 18$

$$S_B^2 = \frac{\sum_{i=1}^{N_B} (X_i - \mu_B)^2}{N_B - 1} = 17.08$$

$$S_W^2 = \frac{\sum_{i=1}^{N_W} (X_i - \mu_W)^2}{N_W - 1} = 13.48$$

- t-score:

$$t = \frac{(\mu_B - \mu_W)}{\sqrt{(S_B^2 / N_B) + (S_W^2 / N_W)}} \approx 3.68$$

- If the skeptic is right (if Null Hyp is correct), then t is distributed according to the following formula (probability density function for t):

$$p(t, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where Γ is the Gamma function and ν is the degrees of freedom (df) computed according to the following formula:

$$df \approx \frac{[(S_B^2 / N_B) + (S_W^2 / N_W)]^2}{\frac{(S_B^2 / N_B)^2}{N_B - 1} + \frac{(S_W^2 / N_W)^2}{N_W - 1}} = \frac{(0.683 + 0.749)^2}{\frac{0.683^2}{25 - 1} + \frac{0.749^2}{18 - 1}} = \frac{2.051}{0.0194 + 0.033} = 39.14$$

- We need to find the probability that t-score is larger than $t=3.68$ from the given distribution.
- Alternatively, we can look up for the critical t-score from the table of student's t-distribution for $d.f.=39.1$ and $\alpha=0.025$ (significance level) $\Rightarrow t_{\text{critical}}=2.02$

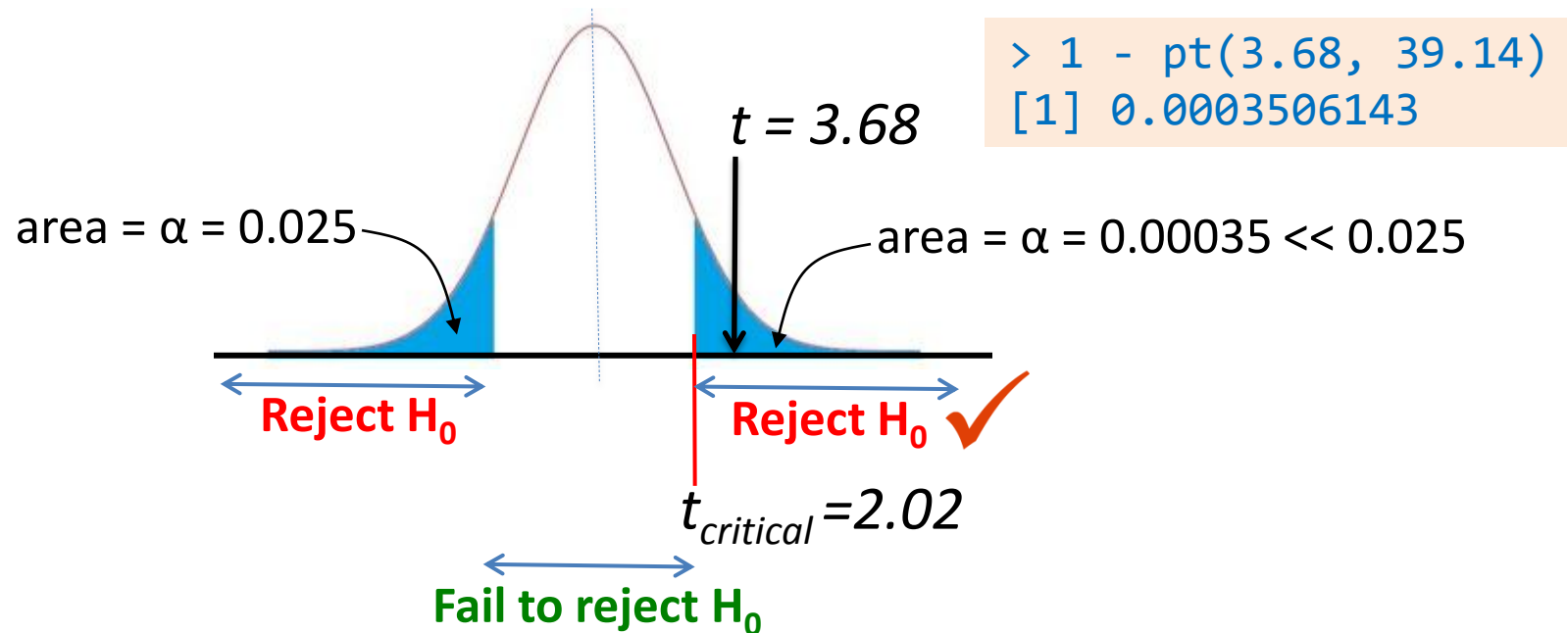
t distribution critical values							Upper-tail probability p
df	.25	.20	.15	.10	.05	.025	.02
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147
39.1	0.681	0.851	1.050	1.303	1.684	2.021	2.123

- If you cannot look this up on the table, you can use the following R command:

```
> qt(p=0.975, df=39.14)
[1] 2.02246
```

which gives you the critical t value for a level of $\alpha=0.025$.

- As the t-score = 3.68 is larger than 2.02, we reject the Null Hypothesis.
- The difference of 4.4 is statistically significant (p-value=0.00035 < 0.025). And yes, beer consumption increases human attractiveness to mosquitoes.



- But you might ask if there is an easier way?
- Let's use a sampling method instead. We don't have a theoretical model (binomial). All we have is a list of human subjects with how many mosquitoes they attracted.

BEER					WATER		
27	19	20	20	23	21	19	13
17	21	24	31	26	22	15	22
28	20	27	19	25	15	22	20
31	24	28	24	29	12	24	24
21	21	18	27	20	21	19	18
					16	23	20

- Null Hypothesis: If the number of mosquitoes attracted are not all that different between the beer and water drinkers, it should not matter what the labels these measurements belong to.
- So shuffling these around shouldn't matter and we should be able to label them any way we want.

- **Idea:** Simulate the distribution by shuffling the labels repeatedly and compute the desired statistic. If the labels really don't matter, then switching them shouldn't change the outcome.
- **Computational Method**
 - We'll keep selecting random records from both batches (beer and water), compute the mean for both beer and water drinkers and then plot the histogram for the difference in means. And we'll do this several times.

- Procedure (1)**

Shuffle the labels:

$N_B = 25$

BEER				
27	19	20	20	23
17	21	24	31	26
28	20	27	19	25
31	24	28	24	29
21	21	18	27	20

$N_W = 18$

WATER		
21	19	13
22	15	22
15	22	20
12	24	24
21	19	18
16	23	20

- **Procedure (2)**

Re-arrange the labels and re-create the beer and water groups:

BEER				
27	20	23	21	24
26	28	19	25	24
28	21	21	18	20
21	19	22	22	12
24	21	19	16	20

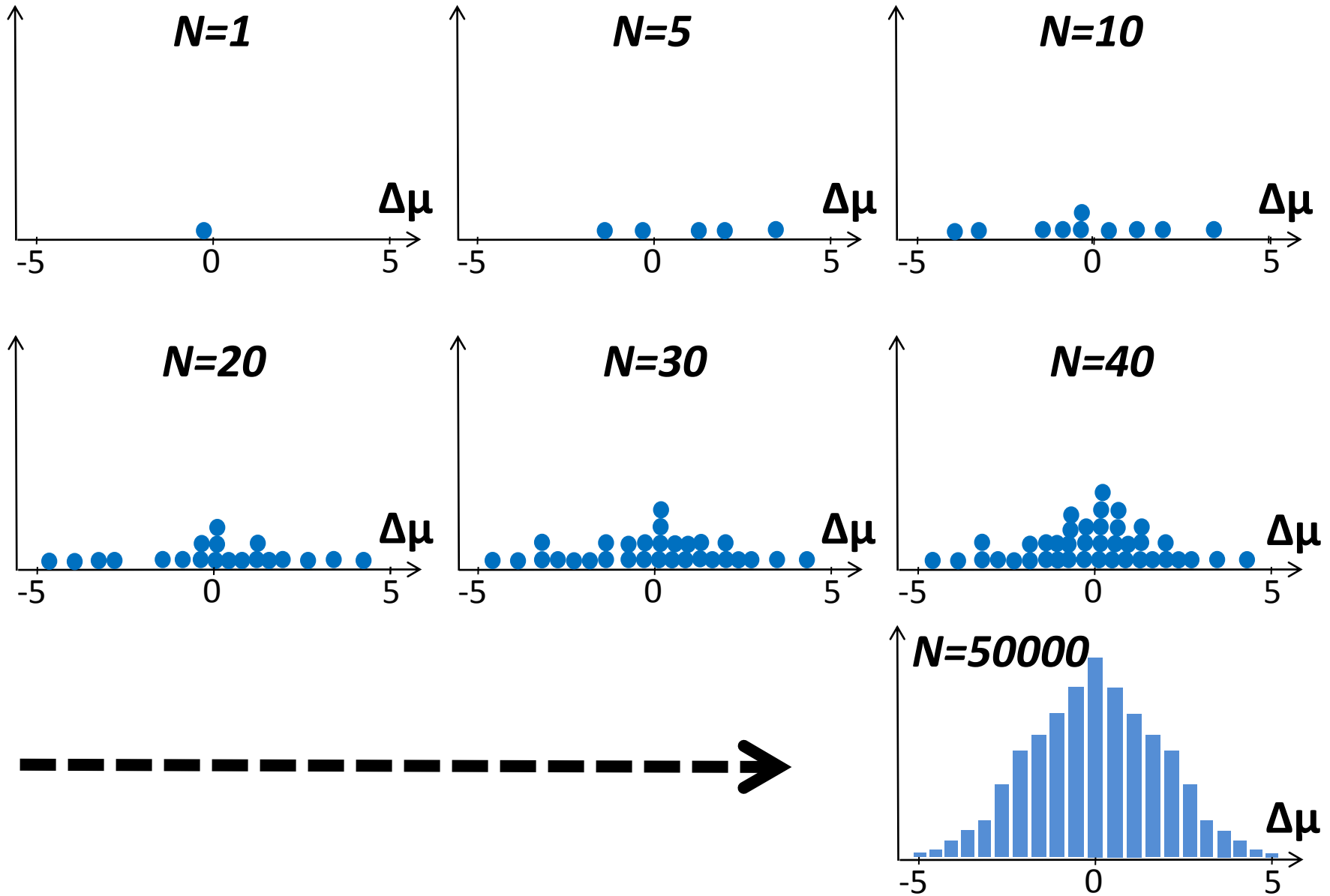
WATER		
19	20	17
31	20	27
31	24	29
27	13	22
15	15	20
24	18	23

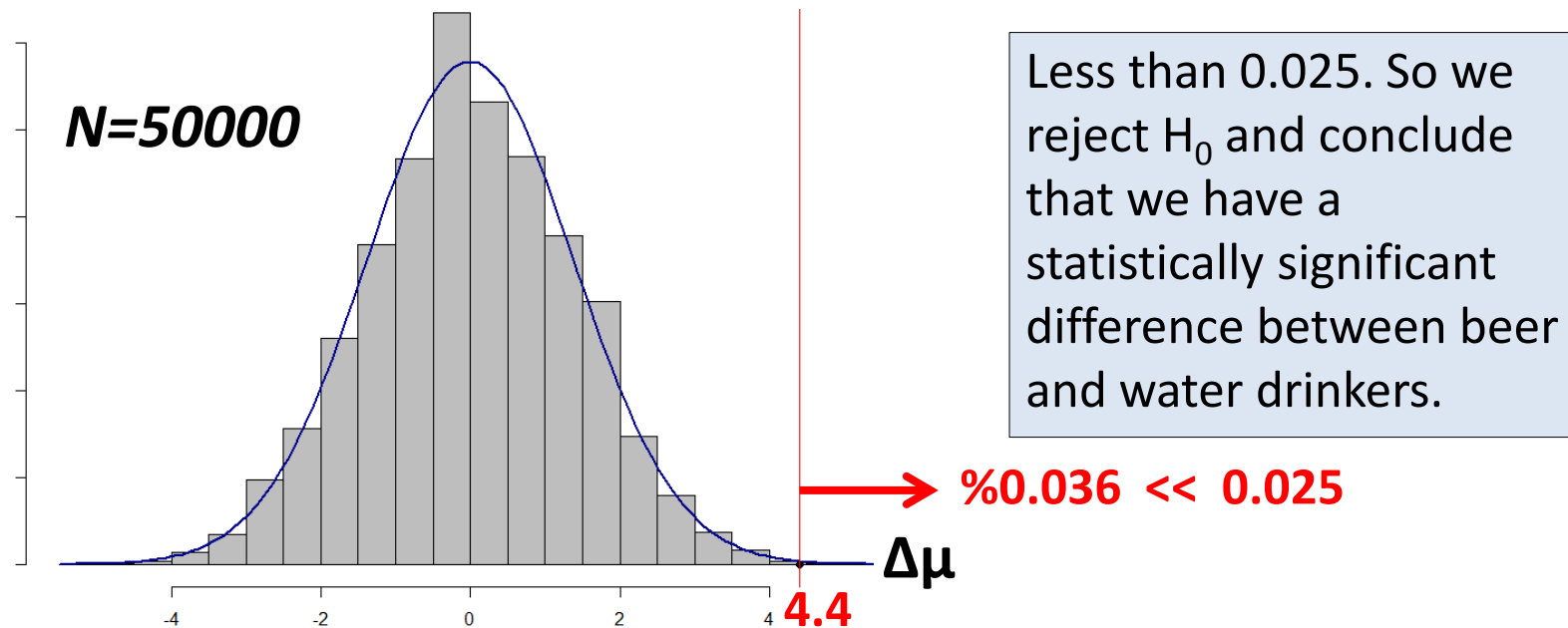
Difference in means: $\Delta\mu = 21.64 - 21.94 = -0.3$

- **Procedure (3)**

- Keep repeating this (by looping over 1-2) for N times
- Find $\Delta\mu_i$ for $i=1,\dots,N$ and plot the **frequency distribution**

Shuffling



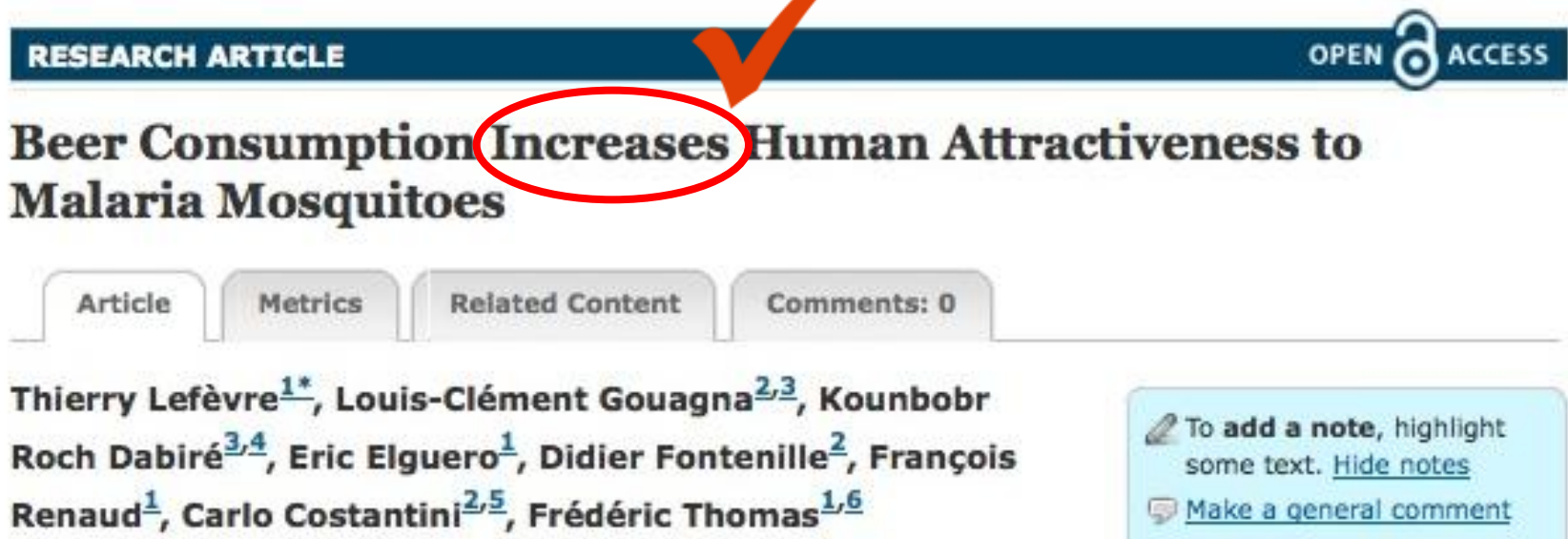


- This is the sample proxy for the real distribution. The ratio of trials where $\Delta\mu > 4.4$ to the total number of iterations:

$$\frac{N_{>4.4}}{N_{tot}} = \frac{18}{50000} = 0.00036$$

- If the skeptic is right, if no significant difference exists, we'd obtain the observed difference or more only in 18 times in 50000 trials due to random error in sampling which is extremely rare. So it's unlikely that this might've happened by chance.

- Verdict:



The screenshot shows the header of a research article. At the top, a dark blue bar contains the text "RESEARCH ARTICLE" on the left and "OPEN ACCESS" with a padlock icon on the right. Below this bar, the title "Beer Consumption Increases Human Attractiveness to Malaria Mosquitoes" is displayed. The word "Increases" is circled in red, and a large red checkmark is placed over the top right of the title. Below the title, there are four tabs: "Article", "Metrics", "Related Content", and "Comments: 0". The authors' names are listed below the tabs: "Thierry Lefèvre^{1*}, Louis-Clément Gouagna^{2,3}, Kounbobr Roch Dabiré^{3,4}, Eric Elguero¹, Didier Fontenille², François Renaud¹, Carlo Costantini^{2,5}, Frédéric Thomas^{1,6}". On the right side of the article header, there is a light blue box with two options: "To add a note, highlight some text. [Hide notes](#)" and "Make a general comment".

- A few notes on Shuffling:
 - This method works only when the two groups are assumed to be equivalent (as in the Null Hyp)
 - We also assume that samples are representative of their respective populations

RANDOM SAMPLING

Point estimators in Statistics

- A point estimate consists of a single value or point. Suppose we have an unknown population parameter such as population mean:
 - Mean height of women in Turkey
- To estimate these parameters, we take a random sample of size n from the population.
- An estimate is the value obtained when the observations X_i have been substituted into a formula (estimator) such as computing the mean.
- Examples:
 - Use the sample mean to estimate the population mean
 - Use the **sample variance** to estimate the **population variance**

Point estimators in Statistics – cont'd

- What is a good estimator?
- A crucial question: Does the estimator (sample variance) differ from the population variance in a systematic manner? Any bias?
- When the population size **N** is large, we take a sample of **n** observations and compute the variance:

	Population (parameter)	Sample (statistic)
<i>Variance</i>	$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$

biased

unbiased

$$s^2_{unbiased} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Point estimators in Statistics – cont'd

- Sample variance has a tendency to under-estimate the population variance. So the sample variance with the $(n-1)$ adjustment is an unbiased estimator. Proof:

$$\begin{aligned} E[\sigma^2 - S_{biased}^2] &= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right] \\ &= \frac{1}{n} E\left[\sum_{i=1}^n \left((x_i^2 - 2x_i\mu + \mu^2) - (x_i^2 - 2x_i\bar{x} + \bar{x}^2)\right)\right] \\ &= E\left[\mu^2 - \bar{x}^2 + \frac{1}{n} \sum_{i=1}^n (2x_i(\bar{x} - \mu))\right] = E[\mu^2 - 2\bar{x}\mu + \bar{x}^2] \\ &= E[(\bar{x} - \mu)^2] = Var(\bar{x}) = \frac{\sigma^2}{n} > 0 \end{aligned}$$

So $S_{biased}^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n$
underestimates the variance.

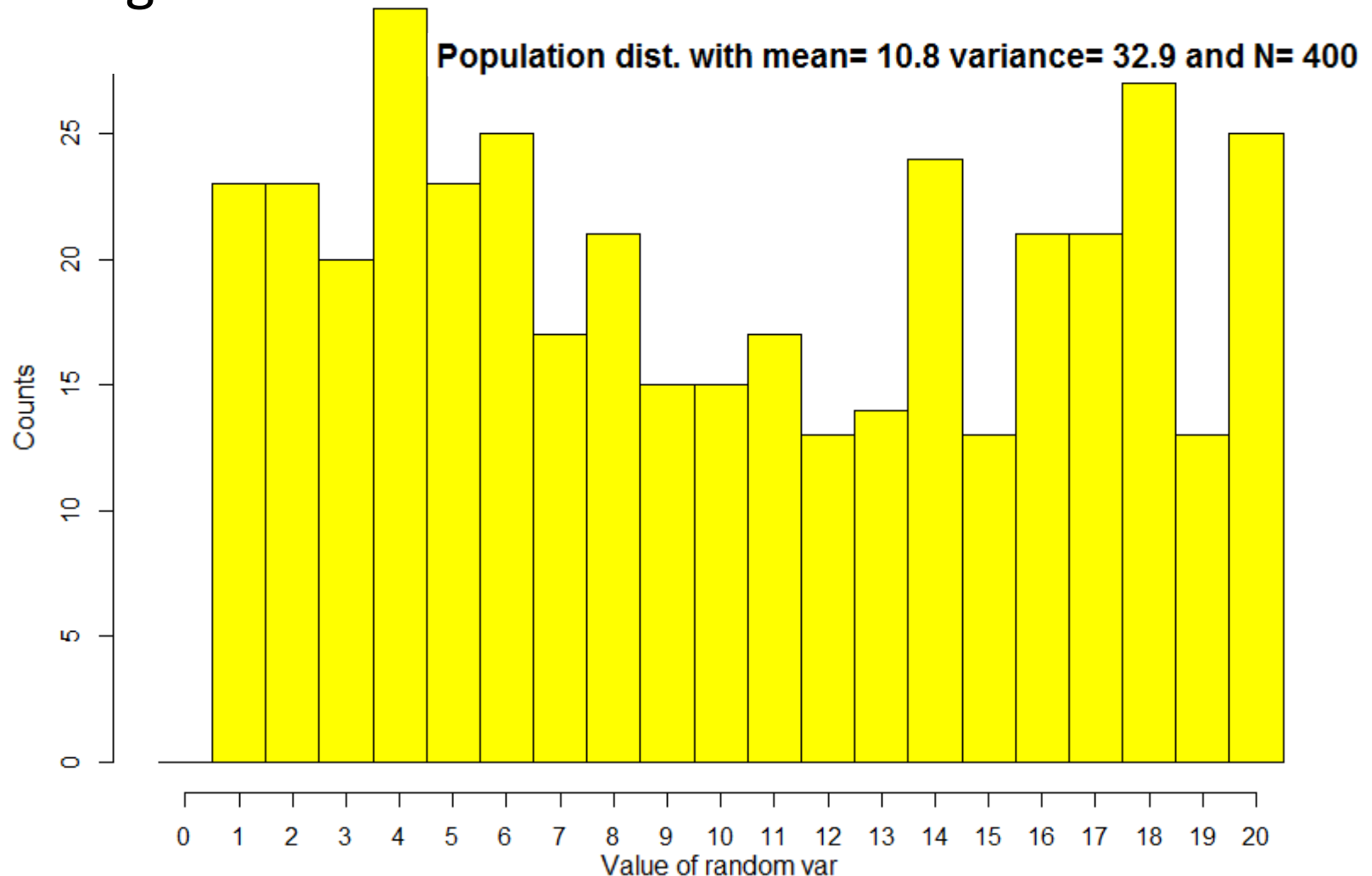
- So, the expected value of the estimators will be:

$$E[S_{biased}^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \quad S_{unbiased}^2 = \frac{n}{n-1} S_{biased}^2$$

This is why S^2 with $n-1$ is an unbiased estimator.

Point estimators in Statistics – cont'd

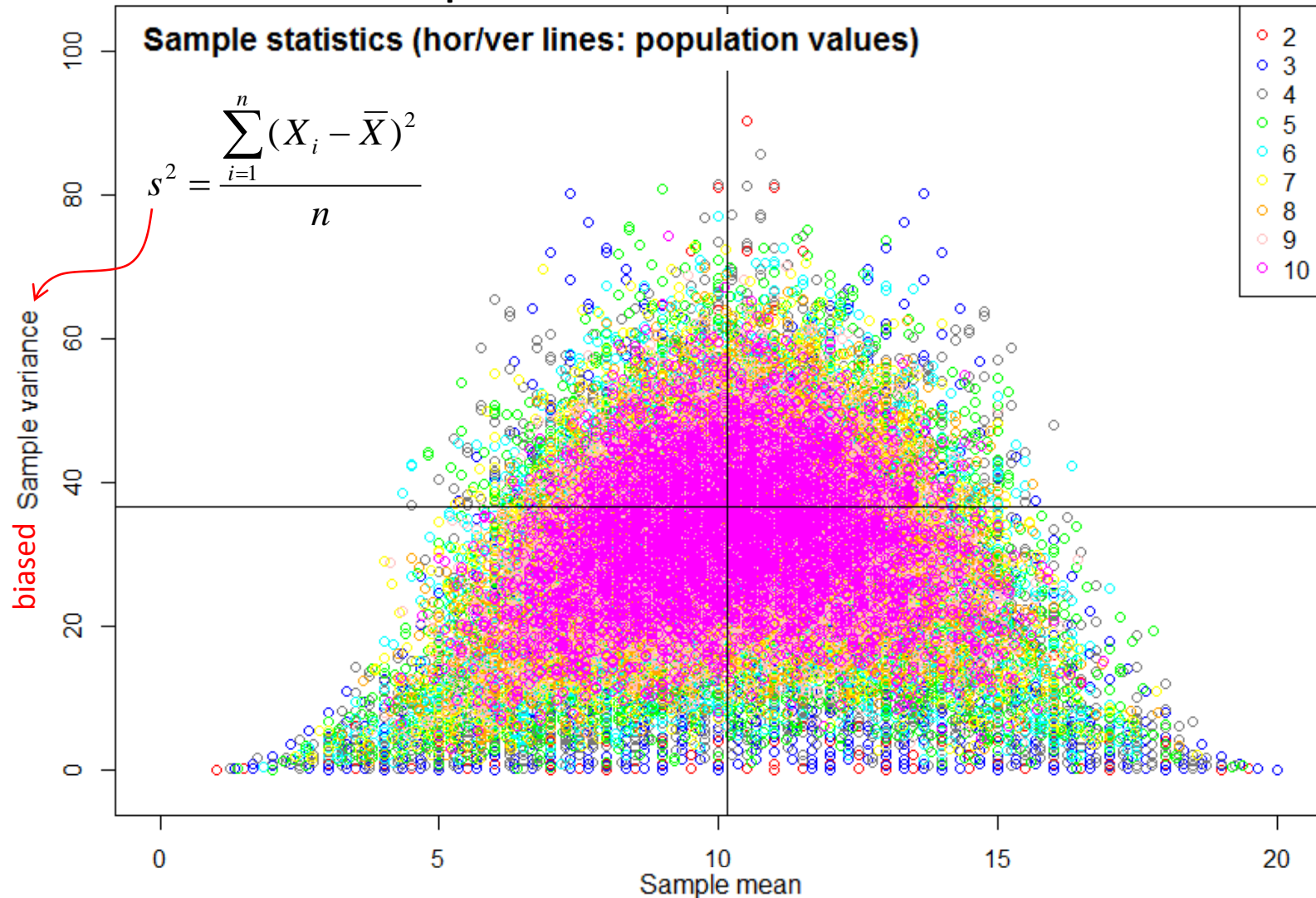
- Let's run a simulation using a population of size 400 in a range between 1 and 20



Ref: Simulation showing bias in sample variance | Probability and Statistics | Khan Academy

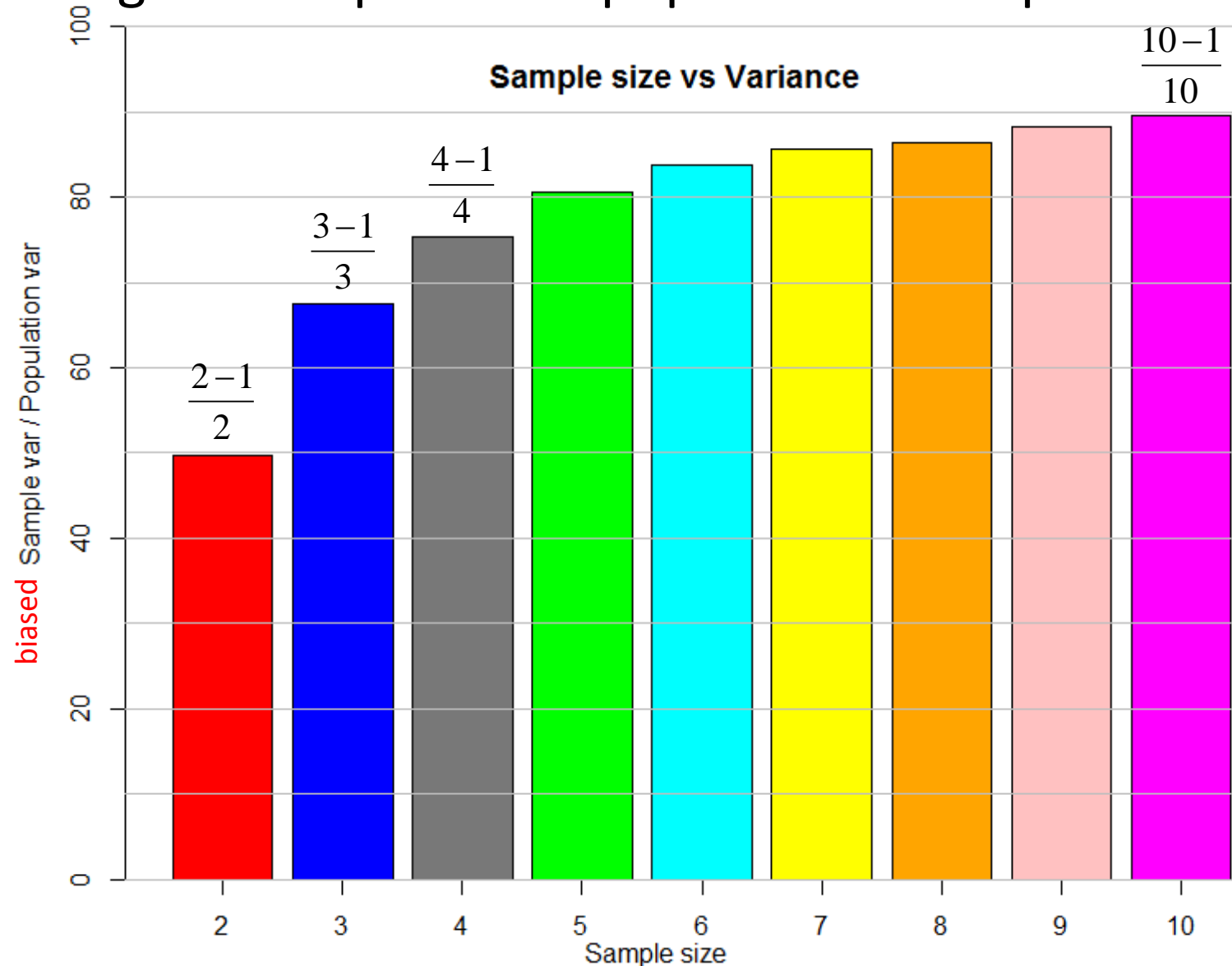
Point estimators in Statistics – cont'd

- We continuously sample the population: 5000 times each with a sample size between 2 and 10



Point estimators in Statistics – cont'd

- Change in sample var vs pop. var wrt sample size



Point estimators in Statistics – cont'd

- When we use the following equation to compute the sample variance

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

- We're not approaching the population variance but approaching (n-1)/n times the population variance, which is the biased estimate:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \Rightarrow \frac{n-1}{n} \sigma^2$$

- How do we unbiased this? To get the best estimate for the true population variance, multiply both sides by n/(n-1) to get the unbiased estimate:

$$\frac{\cancel{n}}{n-1} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\cancel{n}} \Rightarrow \frac{\cancel{n}}{n-1} \frac{n-1}{\cancel{n}} \sigma^2$$

$$S_{unbiased}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

BOOTSTRAPPING

- **What is Bootstrapping?**
 - Resampling process on the original sample by replacement
 - Population – sample \leftrightarrow Sample – bootstrap samples
- **Where do we use it?**
 - In finding sampling distributions (useful when the distribution of a statistic is complicated or unknown)
 - In computation of valid means, standard deviations and confidence intervals
 - For assesment of the uncertainty in an estimator or a learning method
 - Teorik birikimin yetersiz olduğu durumlarda bir yöntemin performans değerlendirmesinde

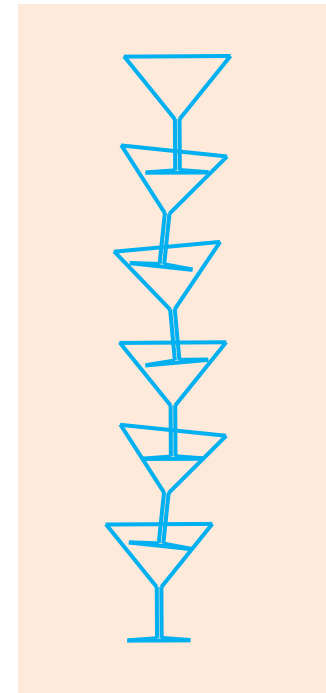
Boostrapping mimics the sampling process from the population without any assumptions about the population distribution (normality, symmetry, outliers etc.)
- Well studied and has solid theoretical grounds...

- **How many times and what sample size?**
 - Sampling is repeated (typically thousands) by drawing the same number of observations (with replacement) from the original sample each time
- **When does it fail?**
- If the sample is not representative of your underlying distribution (selection bias, measurement errors etc.)
- In the presence of dependent data
- When used for rank-based estimates (such as finding the maximum)
 - Always underestimates the maximum in the population
- When the sample size is too small
 - $N > 20$ is a good rule of thumb

Bootstrapping

- How high a stack can you get placing glasses on top of one another?
- We tried it and here is the number of glasses in a stack for 20 observations:

48	24	32	61	51	12	32	18	19	24
21	41	29	21	25	23	42	18	23	13
- What is the mean of the number of glasses in the stack?
- What is the uncertainty on this estimate?
- You want to know, in the long run, if you observe the height of the stacks, what would be the spread of that and how would you characterize the distribution of the glass stack heights?



Example taken from: "Statistics for Hackers", Jake Vanderplas, PyCon 2016

- **Classical method:**

- Compute the sample mean and standard deviation of the mean: $\bar{X} = 28.85$

$$\sigma_{\bar{X}} = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2} = 2.97 \quad \text{(standard error for the mean)}$$

Belirsizlik: $X = \bar{X} \mp t_{0.025, df=19} \sigma_{\bar{X}}$

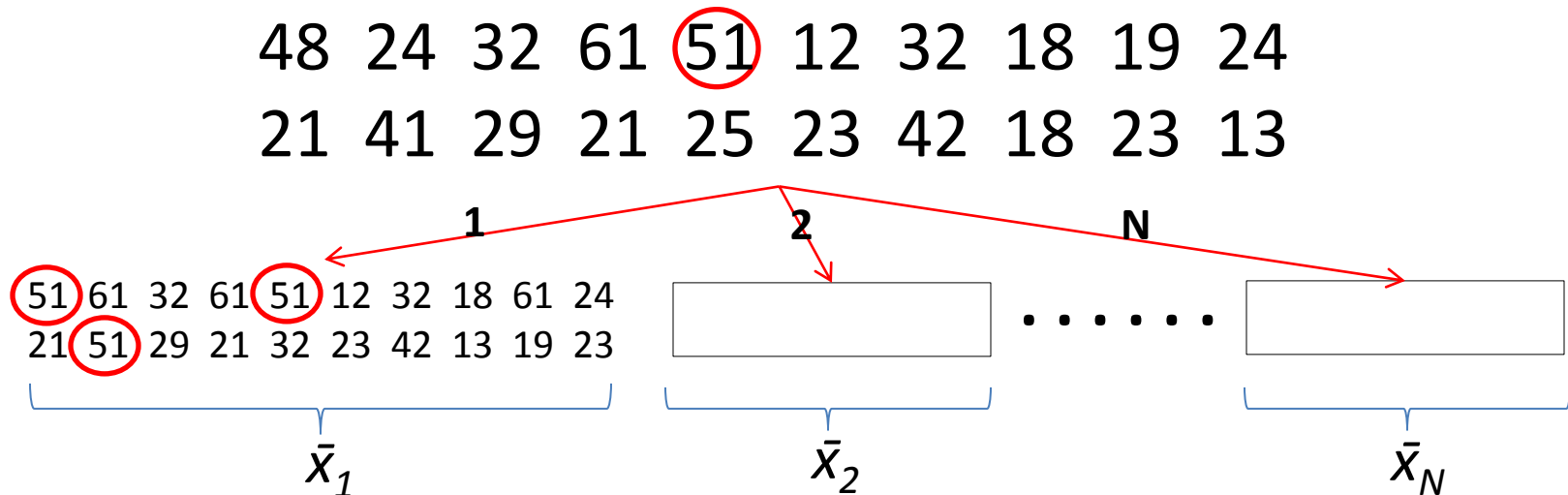
$$X = 28.85 \mp 6.216$$

```
> qt(0.975, 19)
[1] 2.093024
```

$CI[22.633, 35.067]$ (95% confidence interval)

- We don't know what assumptions go into these formulae. We don't have a generating model as before. And unlike before, we don't have two groups to compare. So we cannot use the shuffling approach.

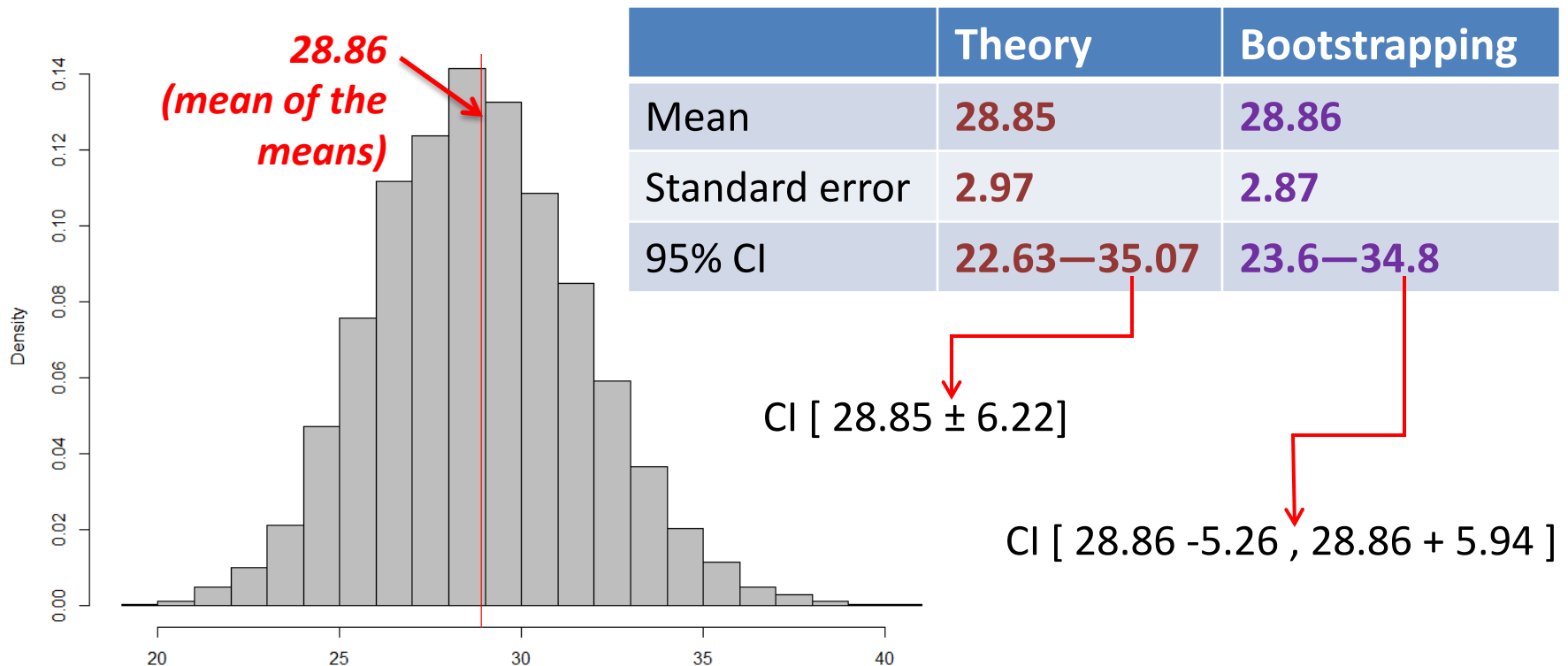
- **Solution:** Bootstrap resampling
- **Method:** Simulate the distribution by drawing samples with replacement
 - Keep resampling from the original data set (some values may repeat as it's done with replacement) and keep computing the mean for the new sample at each iteration
 - Repeat this several thousand times and you end up with the distribution of the means



Bootstrapping

- Resampling for N=10000 iterations
- Compute the mean for each iteration

```
x<-c(48,24,32,61,51,12,32,18,19,24,21,41,29,21,25,23,42,18,23,13)
randx <- replicate(10000, mean(sample(x, length(x), replace=T)))
bs_mean <- mean(randx) ; bs_sd <- sd(randx)
cat("Mean_bs: ", bs_mean, " Std.dev_bs: ", bs_sd, "\n")
CI <- quantile(randx, c(0.025,0.975))
cat("CI for bootstrapped samples:", CI)
```



Bootstrapping for Linear Regression

- Bootstrapping can be applied to even more complicated problems...
- Bootstrapping on Linear Regression:
 - What is the relationship between speed of wind and the height of the glass stack tower?
- Data: Height vs Wind speed

```
> summary(windsp)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 8.100   9.050   9.600   9.832  10.550  12.600

> summary(height)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 8.00   12.00   15.00  14.58   17.00   21.00
```

Windspeed	Height
8.1	21
8.4	19
8.7	16
8.8	18
9	15
9.1	17
9.2	17
9.3	17
9.4	19
9.6	14
9.9	14
10	15
10	11
10.5	12
10.6	12
10.6	13
11.2	10
11.9	8
12.6	9

```
...  
fit0 <- lm(height ~ windsp)  
print(summary(fit0))
```

Call:

lm(formula = height ~ windsp)

Residuals:

Min	1Q	Median	3Q	Max
-3.1043	-0.8767	0.3592	0.7684	3.2047

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	42.2879	3.0928	13.673	1.33e-10	***
windsp	-2.8184	0.3125	-9.019	6.87e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

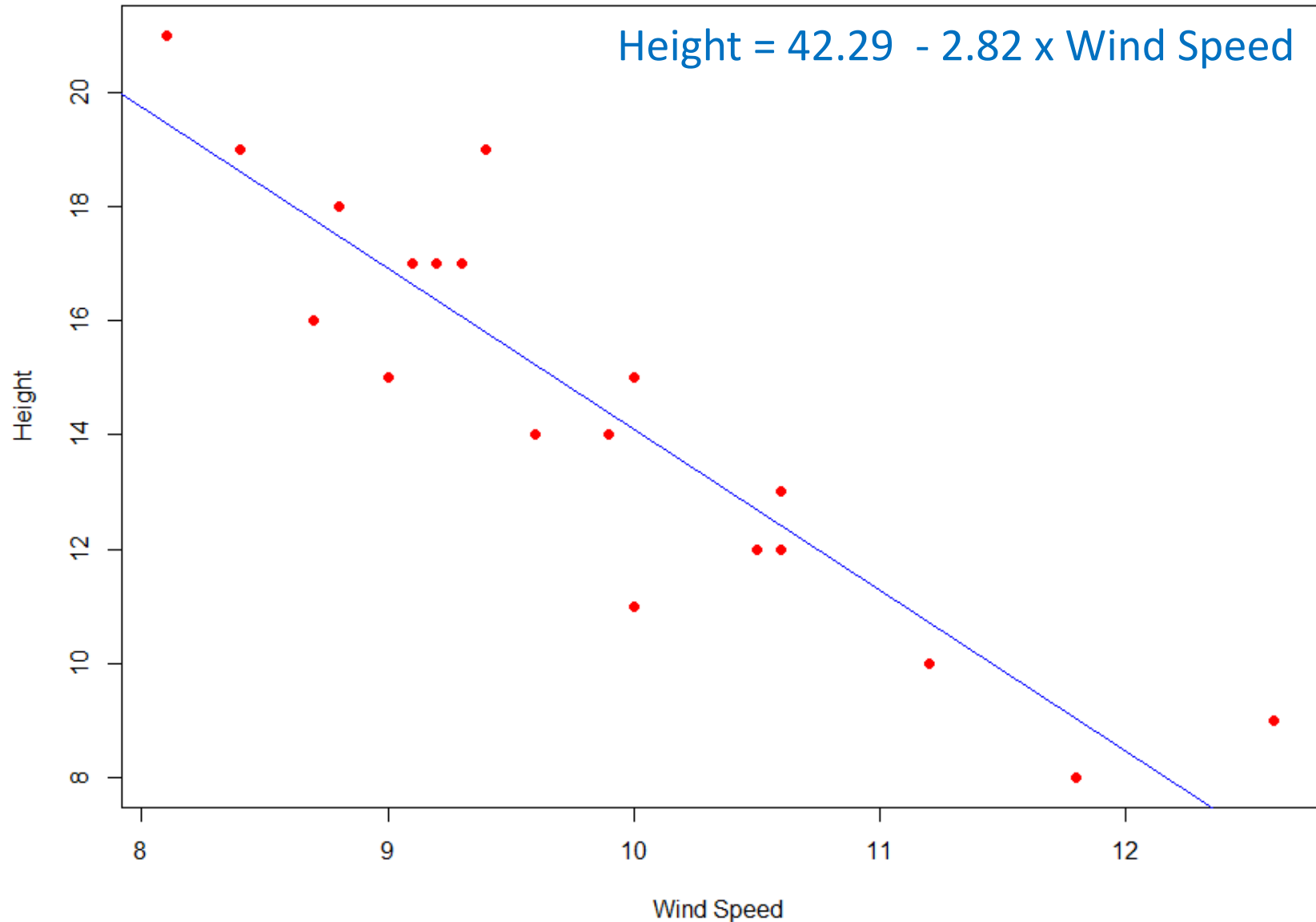
Residual standard error: 1.551 on 17 degrees of freedom

Multiple R-squared: 0.8271, Adjusted R-squared: 0.817

F-statistic: 81.35 on 1 and 17 DF, p-value: 6.875e-08

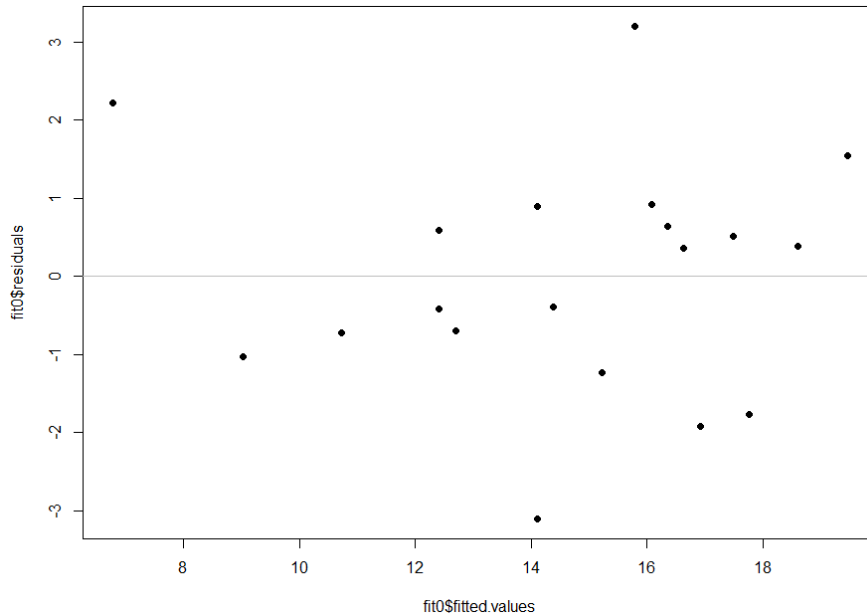
Bootstrapping for Linear Regression

- Linear regression fit:

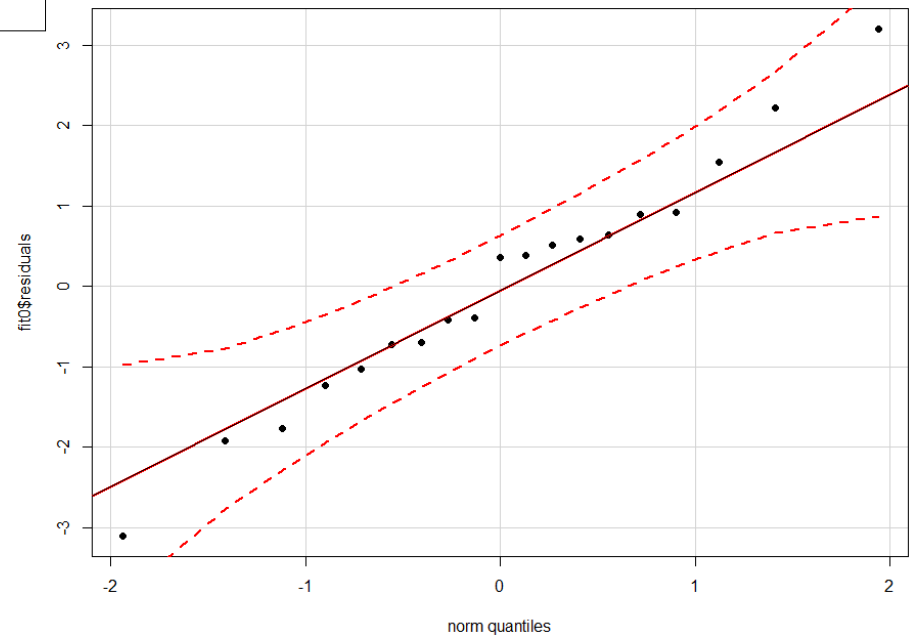


Bootstrapping for Linear Regression

Plot of the residuals

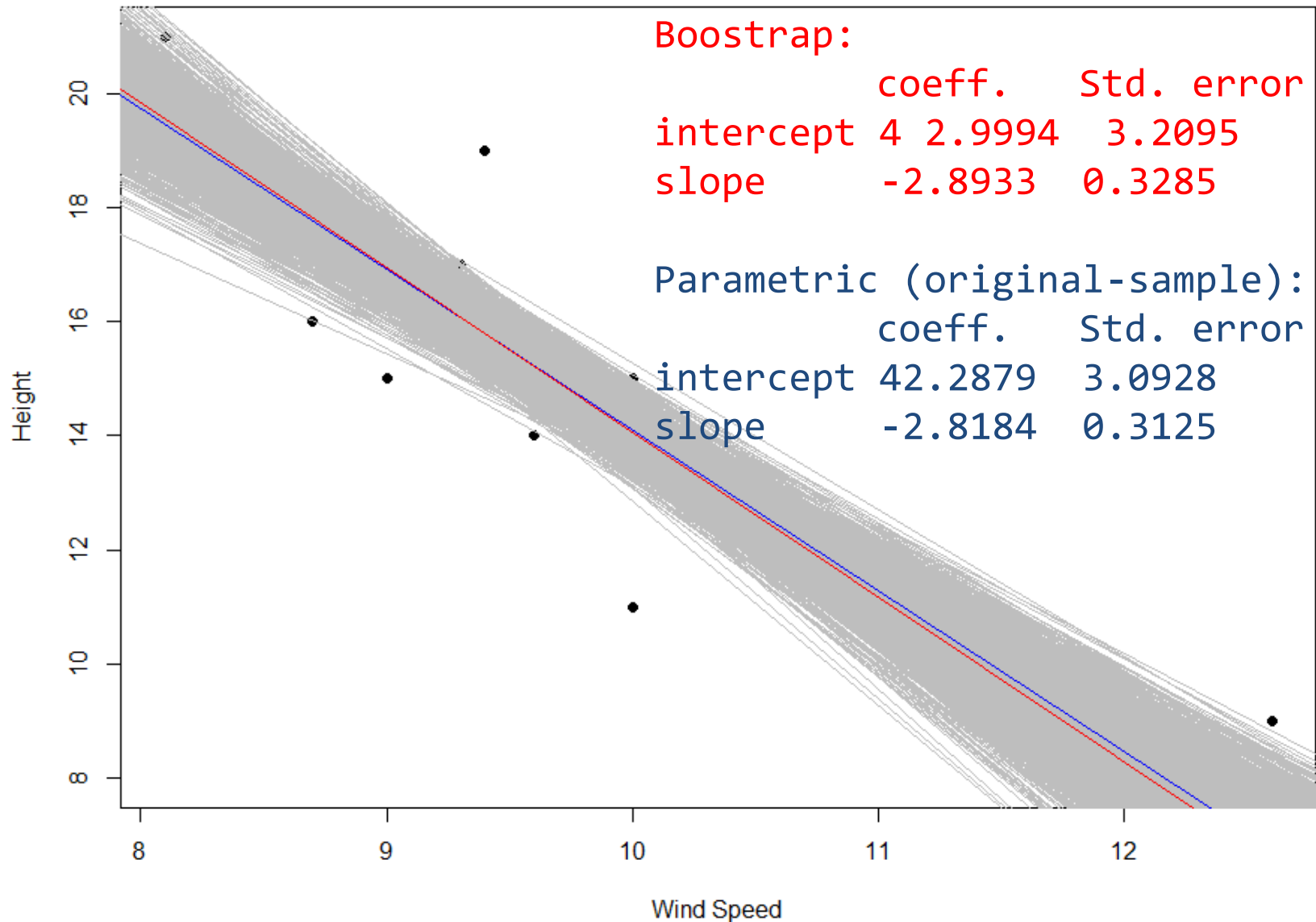


Test for normality



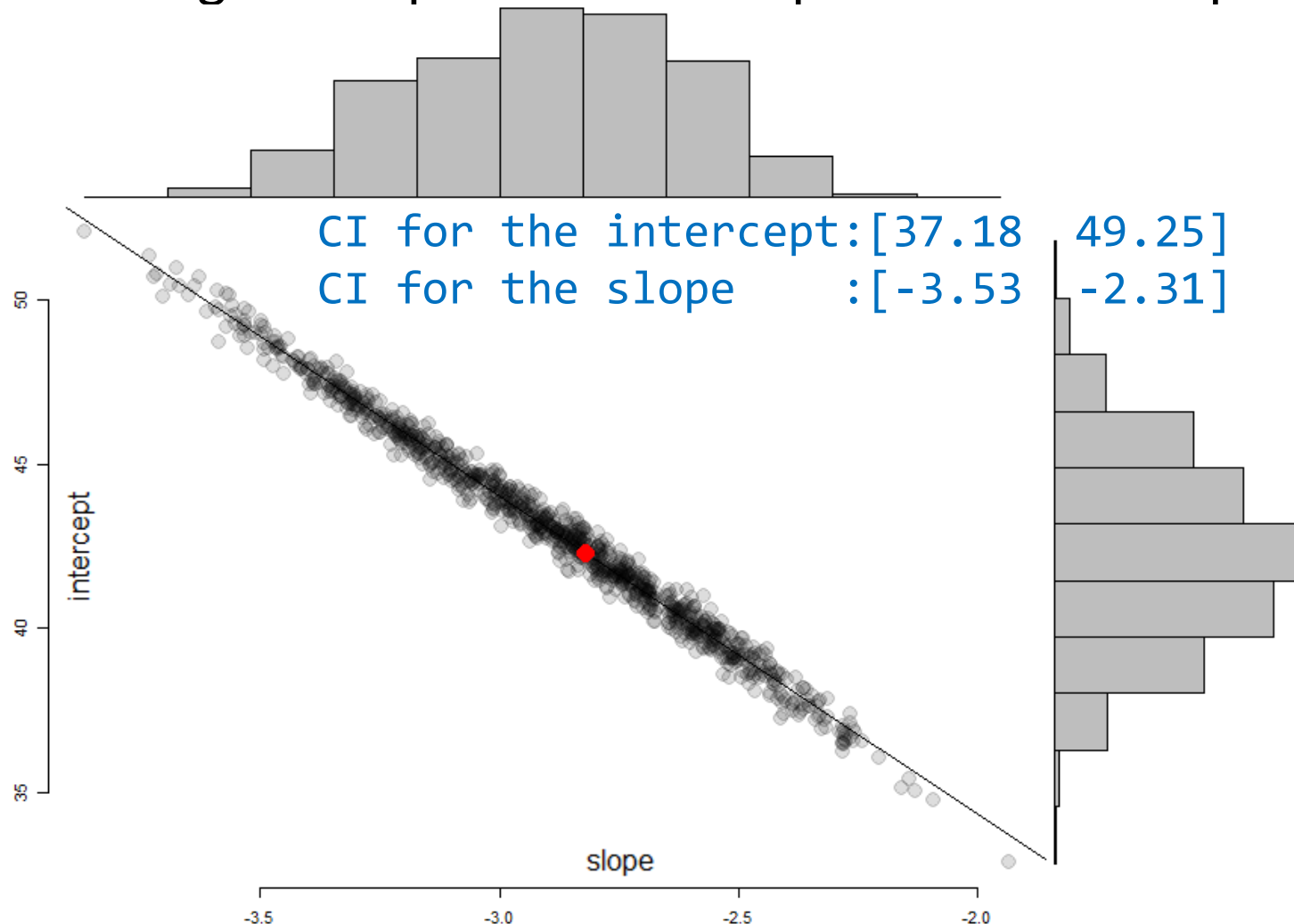
Bootstrapping for Linear Regression

- Plot of the best fits ($i=1,1000$)



Bootstrapping for Linear Regression

- This joint sampling distribution gives us an idea about what range of slopes and intercepts we should expect.



Intercept and slope from the original data: 42.29 and -2.82

CROSS VALIDATION

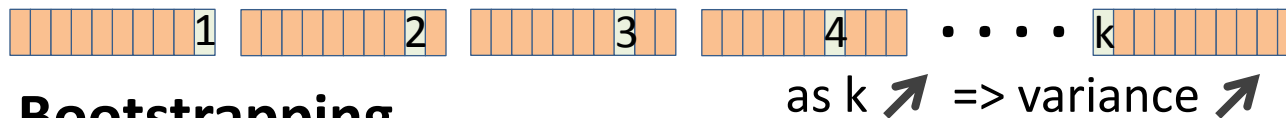
Bootstrapping for CV

- We need to split the data into training and test sets for model validation
- Methods:

- Holdout evaluation

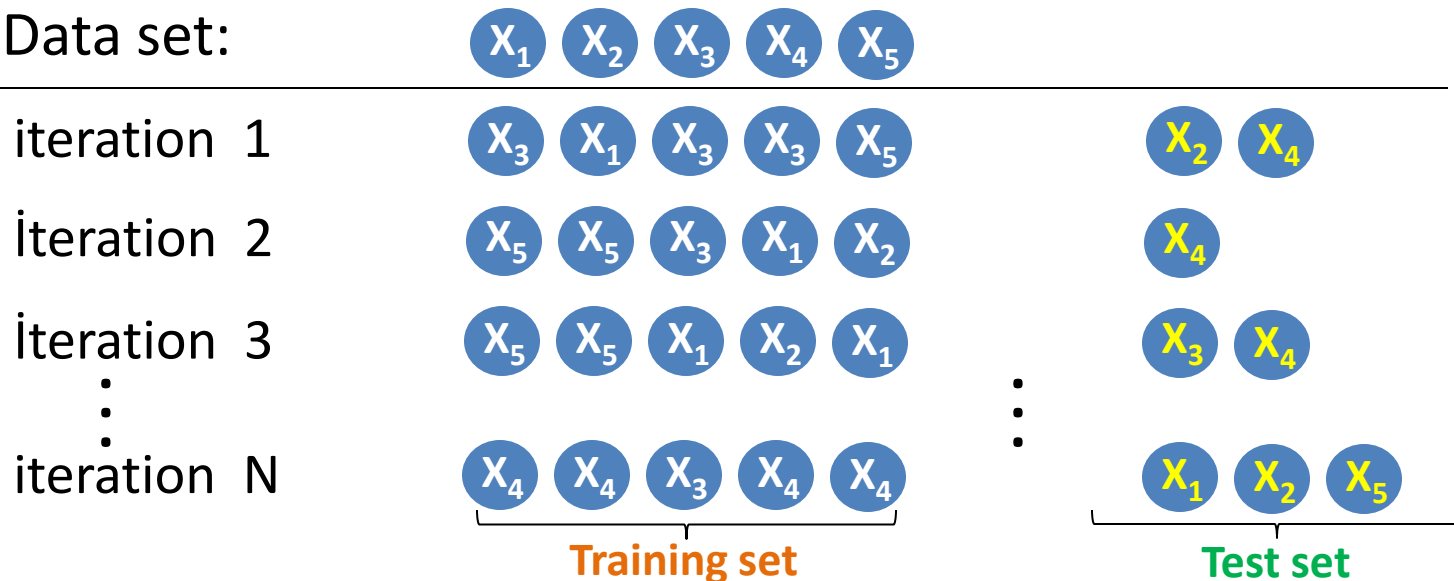


- k-fold cross validation

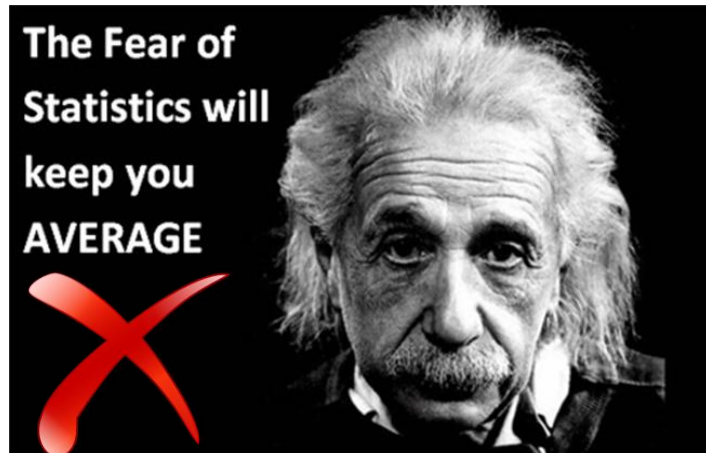


- **Bootstrapping**

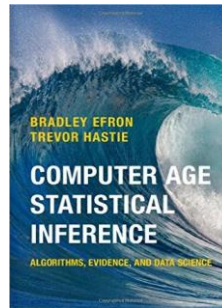
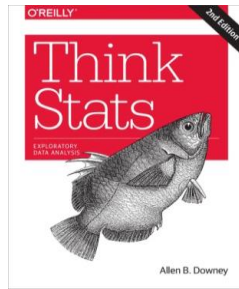
Data set:



- If you followed the methodology presented here...
- If you can produce random numbers...
- If you can put a simple loop into work (with Python, R, whatever)...



- Think Stats: Probability and Statistics for Programmers, Allen Downey
- Computer Age Statistical Inference, Bradley Efron, Trevor Hastie
- Resampling: The new statistics, Julian L. Simon



- Statistics for Hackers, Jake Vanderplas, Pycon 2016
- Statistics without the agonizing pain, John Rauser, Strata+Hadoop World, 2014
- Presentation w/ the R code: github.com/solmez

H. Sait Ölmez, PhD

Sabanci University

Faculty of Eng. and Natural Sci.



olmez@sabanciuniv.edu



@saitolmez



solmez



solmez



Veri Analitiği Araştırma ve Uygulama Merkezi
Center of Excellence in Data Analytics (CEDA)

