

Confidence Intervals

Randy Johnson

Confidence Intervals

Two results will help us understand the derivation of confidence intervals:

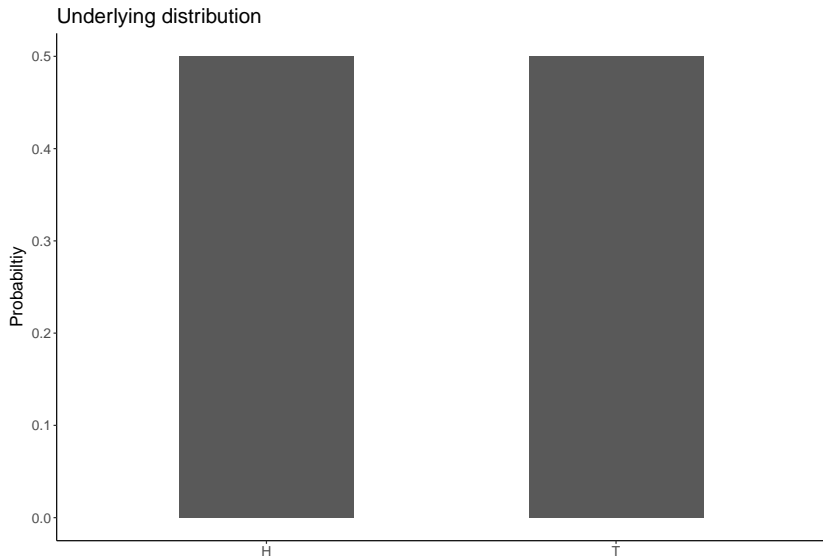
- ▶ Central Limit Theorem
- ▶ Law of Large Numbers

Central Limit Theorem

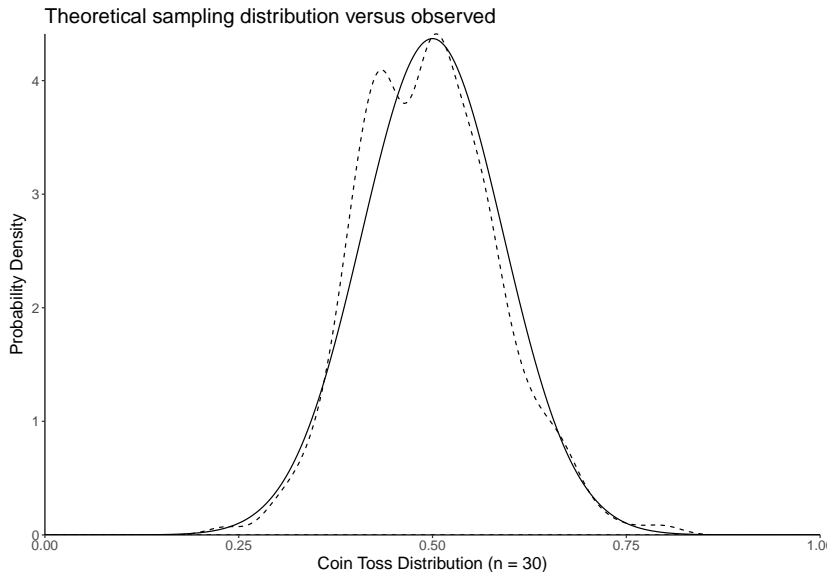
The mean, \bar{x} , of n independent, identically distributed random variables, X , with well defined expected value, $E(X) = \mu$, and variance, $\text{Var}(X) = \sigma^2$, will be approximately normally distributed when n is sufficiently large:

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

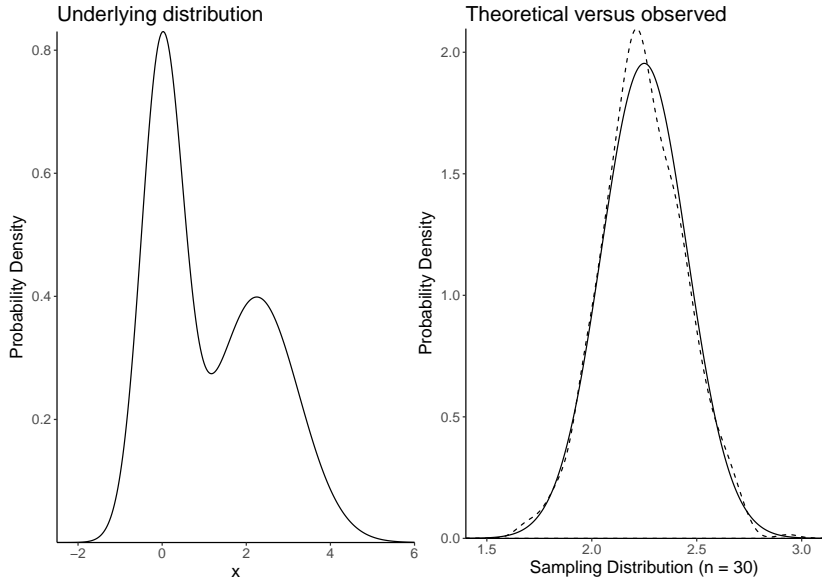
Central Limit Theorem: Simulation - Tossing Coins



Central Limit Theorem: Simulation - Tossing Coins



Central Limit Theorem: Simulation - Bimodal DISTR

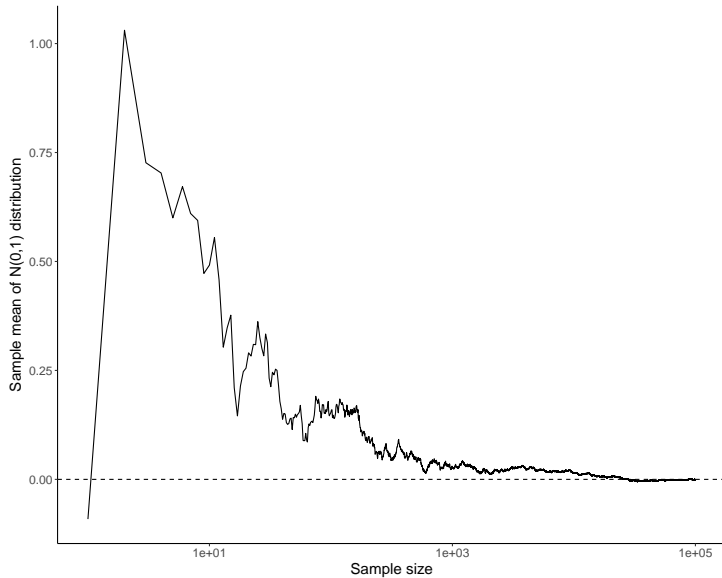


Law of Large Numbers

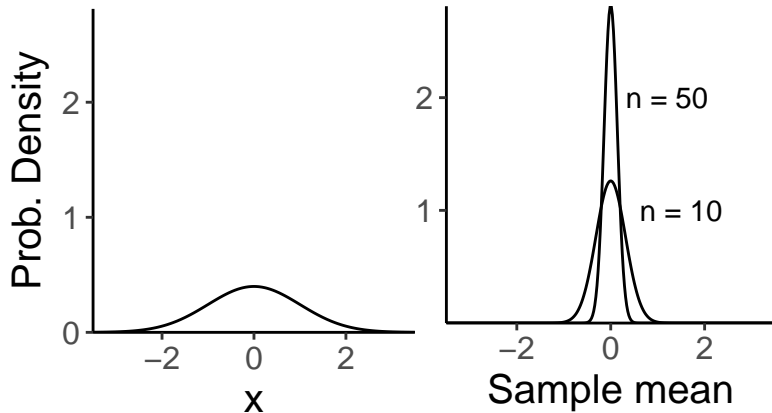
Given our sample mean, \bar{x} will converge to the true population mean as the sample size increases, assuming the sample, X , are independent, identically distributed random variables.

$$\bar{X} \xrightarrow{n \rightarrow \infty} \mu$$

Law of Large Numbers: Simulation

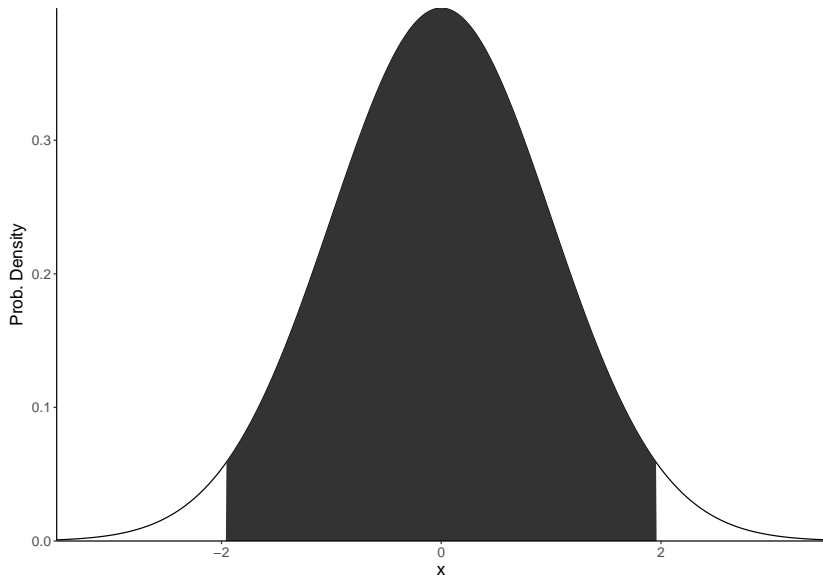


The sampling distribution



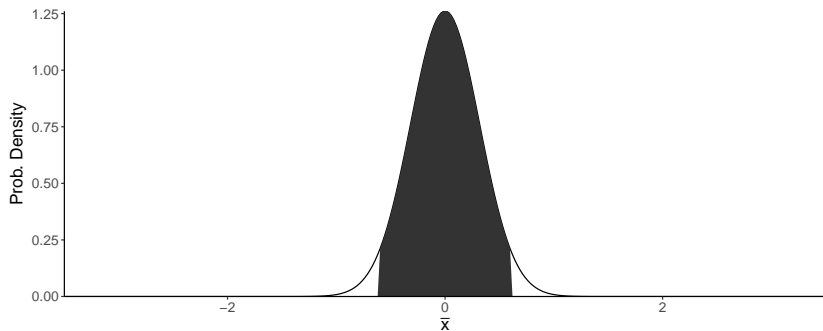
95% Confidence Region

95% of the samples of x we collect will fall in $\mu \pm 1.96\sigma$.



95% Confidence Interval construction

$$\begin{aligned} P\left(\bar{x} \in \left[\mu \pm \frac{1.96 * \sigma}{\sqrt{n}}\right]\right) &= 0.95 \\ \rightarrow P\left(\mu \in \left[\bar{x} \pm \frac{1.96 * sd}{\sqrt{n}}\right]\right) &= 0.95 \end{aligned}$$



95% CI: Example

- ▶ Sample size: $n = 100$
- ▶ Mean: $\bar{x} = 123$
- ▶ SD: $sd = 12$

$$\begin{aligned} 95\% \text{ CI}(\bar{x}) &= 123 \pm \frac{1.96 * 12}{\sqrt{100}} \\ &= (120.66, 125.34) \end{aligned}$$

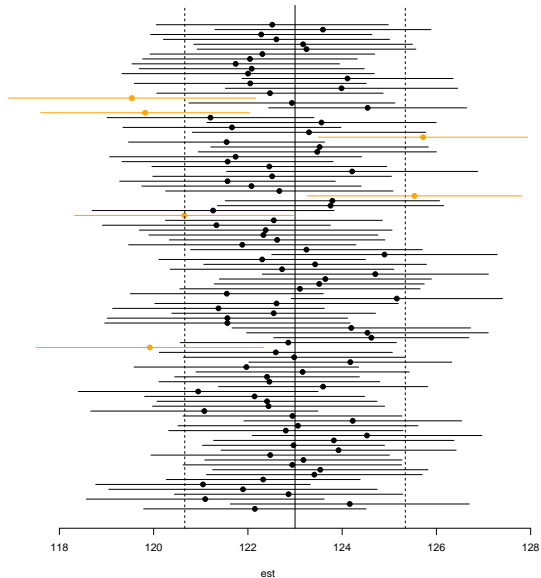
95% CI: Simulation

Let's simulate a similar data set in R and use the gmodels package to calculate the CI.

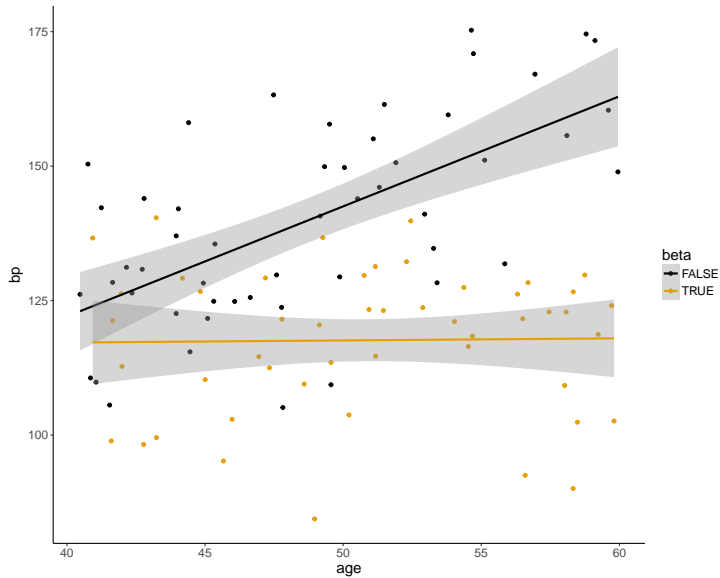
```
set.seed(29874)
rnorm(100, 123, 12) %>%
  ci()
```

```
##      Estimate    CI lower    CI upper Std. Error
## 124.023961 121.599383 126.448540    1.221932
```

95% CI: Simulation



95% Confidence Interval: Practice



95% Confidence Interval: Practice

```
##  
## Call:  
## lm(formula = bp ~ age * beta, data = dat)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -33.147  -9.743   2.708   9.836  27.018   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   40.1783    17.1898   2.337 0.021499 *      
## age           2.0467     0.3508   5.835 7.26e-08 ***    
## betaTRUE      75.4380    24.8314   3.038 0.003067 **     
## age:betaTRUE  -2.0073     0.4957  -4.050 0.000104 ***    
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 14.17 on 96 degrees of freedom  
## Multiple R-squared:  0.4983, Adjusted R-squared:  0.4826   
## F-statistic: 31.78 on 3 and 96 DF,  p-value: 2.346e-14
```


95% Confidence Interval: Practice

```
ci(lm0)[,-4]
```

##	Estimate	CI lower	CI upper	p-value
## (Intercept)	40.178325	6.056900	74.299750	2.149891e-02
## age	2.046744	1.350478	2.743011	7.261338e-08
## betaTRUE	75.437980	26.148032	124.727928	3.067416e-03
## age:betaTRUE	-2.007271	-2.991179	-1.023363	1.040238e-04

```
estimable(lm0, c(0, 1, 0, 1), conf.int = 0.95)[,c(1,6,7,5)]
```

##	Estimate	Lower.CI	Upper.CI	Pr(> t)
## (0 1 0 1)	0.03947368	-0.6557158	0.7346631	0.910496

95% Confidence Interval: Practice

```
estimable(lm0, c(1, 52, 0, 0), conf.int = 0.95)[,c(1,6,7)]
```

```
##              Estimate Lower.CI Upper.CI  
## (1 52 0 0)   146.609 142.0423 151.1757
```