PLSC 502 – Autumn 2016
Confidence Intervals

October 13, 2016

Confidence Intervals

A range of values for $\hat{\theta}$ (say, $[\hat{\theta}_L, \hat{\theta}_H]$) for which:

- $\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_H)$ is high, and
- $\hat{\theta}_L - \hat{\theta}_H$ is small

Define:

$$\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha,$$

C.I.s: The “Pivotal” Method

“Pivotal method”: $\hat{\theta}$

- is a function *only* of the sample data and the population parameter θ, and
- whose sampling distribution *does not* depend on θ.

Constructing C.I.s

Recall that:

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2_{\bar{X}})$$

Because $\mathsf{E}(\bar{X}) = \mu$, we use $\bar{X}$ as the "center" of our C.I.

Constructing C.I.s (continued)

Suppose $\alpha = 0.05$, so $1 - \alpha = 0.95$. Then

$$\Pr(\bar{X}_L \leq \mu \leq \bar{X}_U) = 0.95$$

Then choose:

$$\Pr(\mu < \bar{X}_L) = \int_{-\infty}^{\bar{X}_L} \phi_{\bar{X}}(u)\, du = 0.025$$

and

$$\Pr(\mu > \bar{X}_H) = \int_{\bar{X}_H}^{\infty} \phi_{\bar{X}}(u)\, du = 0.025.$$

More Generally

For

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \sim \mathcal{N}(0, 1)$$

we have:

$$\begin{aligned}
1 - \alpha &= \Pr\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right) \\
&= \Pr\left(-z_{\alpha/2}\sigma_{\hat{\theta}} \leq \hat{\theta} - \theta \leq z_{\alpha/2}\sigma_{\hat{\theta}}\right) \\
&= \Pr\left(-\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq -\theta \leq -\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) \\
&= \Pr\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right)
\end{aligned}$$

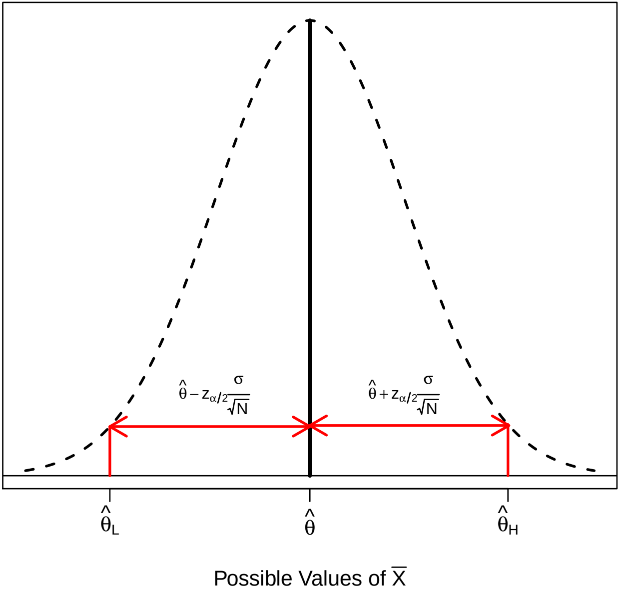
Means...

$$[\hat{\theta}_L, \hat{\theta}_U] = \left[\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}},\ \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right]$$

Steps:

- Select your level of confidence $1 - \alpha$,
- Calculate the sample statistic $\hat{\theta}$,
- Calculate the z-value associated with the $1 - \alpha$ level of confidence,
- Divide that z-value by $\sigma_{\hat{\theta}}$, the standard error of the sampling statistic, and
- Construct the confidence interval according to (1).

C.I.s, Illustrated



Example: Proportions

We have

$$\hat{\theta} = \hat{\pi} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

and

$$\sigma_{\hat{\pi}}^2 = \frac{\pi(1-\pi)}{N}$$

so that

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{N}}.$$

Proportions (continued)

We know that:

$$\hat{\pi} \sim \mathcal{N}(\pi, \sigma^2_{\hat{\pi}})$$

Implies:

$$\hat{\pi}_L = \hat{\pi} - z_{\alpha/2}\left[\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}}\right]$$

and

$$\hat{\pi}_U = \hat{\pi} + z_{\alpha/2}\left[\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}}\right].$$

Proportions: Example

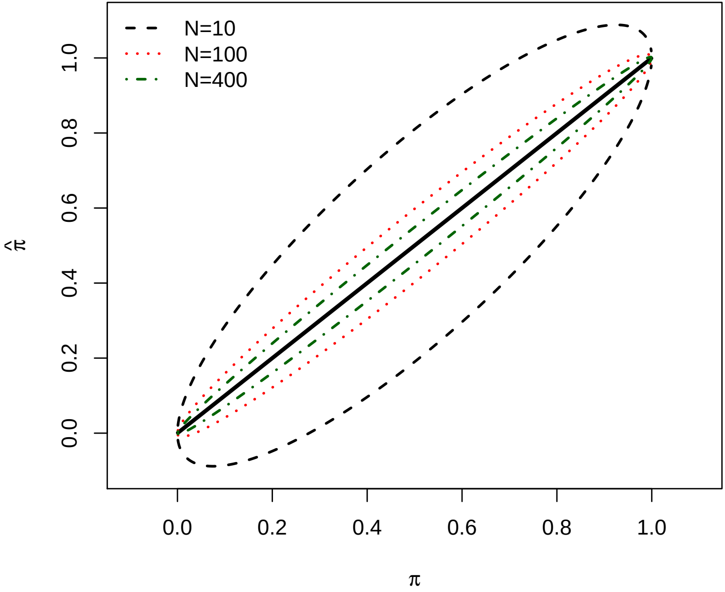
For $N = 20$ and $\hat{\pi} = 0.390$, we have:

$$
\begin{aligned}
\hat{\pi}_L &= 0.390 - 1.96\left[\sqrt{\frac{0.39(0.61)}{20}}\right] \\
&= 0.390 - 0.214 \\
&= \mathbf{0.176}
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\pi}_U &= 0.390 + 1.96\left[\sqrt{\frac{0.39(0.61)}{20}}\right] \\
&= 0.390 + 0.214 \\
&= \mathbf{0.604}.
\end{aligned}
$$

C.I.s for Proportions



Consider:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{N}}$$

As $N \to \infty$, $s \to \sigma$.

In small samples,

$$[\bar{X}_L, \bar{X}_U] = \bar{X} \pm t_{\alpha/2} \left(\frac{s}{\sqrt{N}} \right)$$

Talking About C.I.s

"$[(1-\alpha)\times 100]\%$ of all confidence intervals constructed from independent simple random samples will contain the population parameter θ, and $(\alpha \times 100)\%$ of them will not."

Never "There is a 95% chance that our confidence interval contains the true population value θ."

Example: SCOTUS Cases

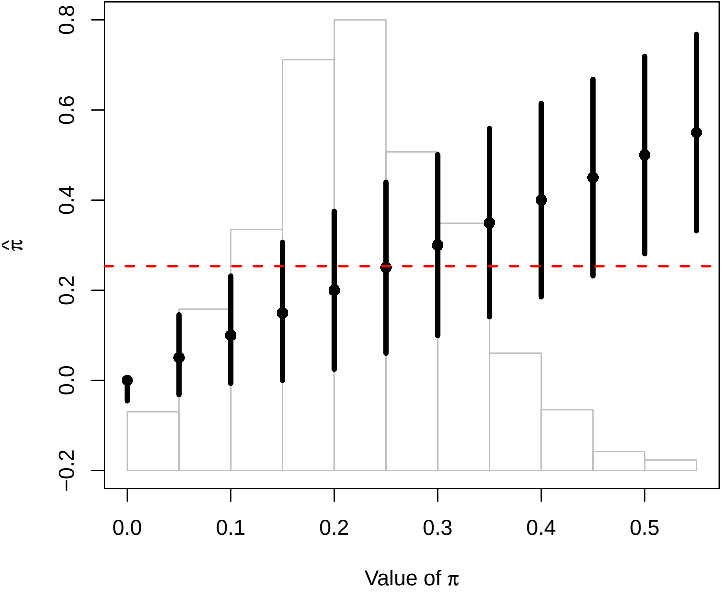
```
> summary(WB)
       us                 id            amrev           amaff
 394/0310:  15   Min.   :   1   Min.   : 0   Min.   : 0
 390/0747:  14   1st Qu.:1791   1st Qu.: 0   1st Qu.: 0
 389/0486:  12   Median :3581   Median : 0   Median : 0
 375/0002:  10   Mean   :3581   Mean   : 0   Mean   : 0
 375/0032:   9   3rd Qu.:5371   3rd Qu.: 0   3rd Qu.: 0
 391/0009:   9   Max.   :7161   Max.   :33   Max.   :37
 (Other) :7092
     sumam         fedpet          constit           sgam
 Min.   : 0   Min.   :0.00   Min.   :0.00   Min.   :0.00
 1st Qu.: 0   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
 Median : 0   Median :0.00   Median :0.00   Median :0.00
 Mean   : 1   Mean   :0.17   Mean   :0.25   Mean   :0.08
 3rd Qu.: 1   3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:0.00
 Max.   :39   Max.   :1.00   Max.   :1.00   Max.   :1.00
```

Sample, $N = 20$

```
> set.seed(7222009)
> WBsample <- with(WB, sample(constit,20,replace=F))
> summary(WBsample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0     0.0     0.0     0.1     0.0     1.0
> LB <- mean(WBsample) - 1.96*sqrt((mean(WBsample)*(1-mean(WBsample)))/(20))
> UB <- mean(WBsample) + 1.96*sqrt((mean(WBsample)*(1-mean(WBsample)))/(20))
> print(c(LB,UB))
[1] -0.031  0.231
```

“Coverage”

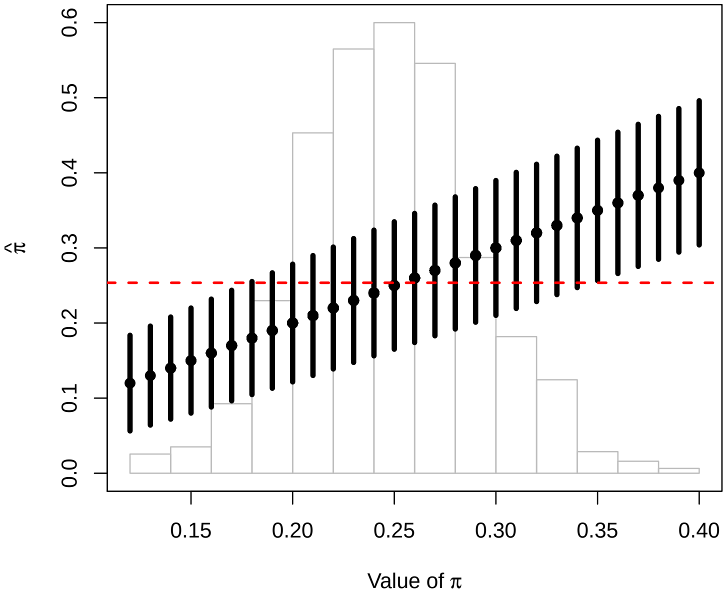
```
program define CI20, rclass
set.seed(7222009)
N <- 20
reps <- 1000
PI20 <- numeric(reps)
UB20<-numeric(reps)
LB20<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(WB, sample(constit,N,replace=F))
  bar <- prop.test(sum(foo),length(foo),correct=FALSE)
  PI20[i] <- bar$estimate
  LB20[i] <- PI20[i] - 1.96 * sqrt((PI20[i] * (1-PI20[i]))/(N))
  UB20[i] <- PI20[i] + 1.96 * sqrt((PI20[i] * (1-PI20[i]))/(N))
  }
```



0.8
0.6
0.4
0.2
0.0
−0.2
$\hat{\pi}$
0.0
0.1
0.2
0.3
0.4
0.5
Value of π

Coverage, $N = 100$

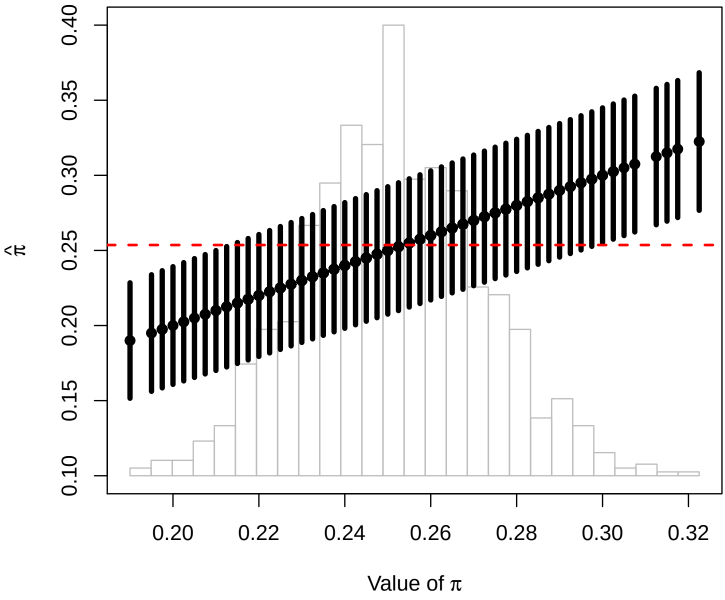
```
set.seed(7222009)
N <- 100
reps <- 1000
PI100 <- numeric(reps)
UB100<-numeric(reps)
LB100<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(WB, sample(constit,N,replace=F))
  bar <- prop.test(sum(foo),length(foo),correct=FALSE)
  PI100[i] <- bar$estimate
  LB100[i] <- PI100[i] - 1.96 * sqrt((PI100[i] * (1-PI100[i]))/(N))
  UB100[i] <- PI100[i] + 1.96 * sqrt((PI100[i] * (1-PI100[i]))/(N))
}
```



0.0
0.1
0.2
0.3
0.4
0.5
0.6
$\hat{\pi}$
0.15
0.20
0.25
0.30
0.35
0.40
Value of π

Coverage, $N = 400$

```
set.seed(7222009)
N <- 400
reps <- 1000
PI400 <- numeric(reps)
UB400<-numeric(reps)
LB400<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(WB, sample(constit,N,replace=F))
  bar <- prop.test(sum(foo),length(foo),correct=FALSE)
  PI400[i] <- bar$estimate
  LB400[i] <- PI400[i] - 1.96 * sqrt((PI400[i] * (1-PI400[i]))/(N))
  UB400[i] <- PI400[i] + 1.96 * sqrt((PI400[i] * (1-PI400[i]))/(N))
}
```



0.10
0.15
0.20
0.25
0.30
0.35
0.40
$\hat{\pi}$
0.20
0.22
0.24
0.26
0.28
0.30
0.32
Value of π

Coverage...

```
> popmean<-mean(WB$constit)

> prop.table(table(ifelse(UB20>popmean & LB20<popmean,1,0)))

   0    1
0.12 0.88

> prop.table(table(ifelse(UB100>popmean & LB100<popmean,1,0)))

    0     1
0.048 0.952

> prop.table(table(ifelse(UB400>popmean & LB400<popmean,1,0)))

    0     1
0.045 0.955
```