ME111 - Laboratório de Estatística

Aula 11 - Intervalo de Confiança para p

Profa. Larissa Avila Matos

Intervalo de Confiança para p

- Os intervalos de $100(1-\alpha)\%$ de confiança para p podem então ser de duas formas:

1. Método Conservador

$$IC_1(p, 1-\alpha) = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{1}{4n}}\right].$$

2. Usando $\hat{p}$ para estimar o erro-padrão

$$IC_2(p, 1-\alpha) = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

- Coletamos uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com distribuição de Bernoulli com probabilidade de sucesso igual a p, portanto com média p e a variância $p(1-p)$ e usamos $\bar{X} = \hat{p}$ para estimar p.

- Pelo TCL: $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$. Pela propriedade da Normal:

$$Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \sim N(0,1), \quad P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha.$$

- Então, um intervalo de $100(1-\alpha)\%$ de confiança para p:

$$IC(p, 1-\alpha) = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\,;\, \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right].$$

- Problema: não conhecemos p. Portanto, usamos:

$$IC(p, 1-\alpha) = \left[\hat{p} - z_{\alpha/2}\sqrt{\tfrac{\hat{p}(1-\hat{p})}{n}}\,;\, \hat{p} + z_{\alpha/2}\sqrt{\tfrac{\hat{p}(1-\hat{p})}{n}}\right],$$

ou, pelo método conservador,

$$IC(p, 1-\alpha) = \left[\hat{p} - z_{\alpha/2}\sqrt{\tfrac{1}{4n}}\,;\, \hat{p} + z_{\alpha/2}\sqrt{\tfrac{1}{4n}}\right].$$

Interpretação do Intervalo de Confiança

- Se várias amostras forem retiradas da população e calcularmos um IC de 95% para cada amostra, cerca de 95% desses intervalos irão conter a verdadeira proporção na população, p.

- INCORRETO: Dizer que "a probabilidade de que p esteja dentro do intervalo é 95%".

- Por que incorreto? p é uma constante, não é variável aleatória. Ou p está no intervalo ou não está. O intervalo é que é aleatório.

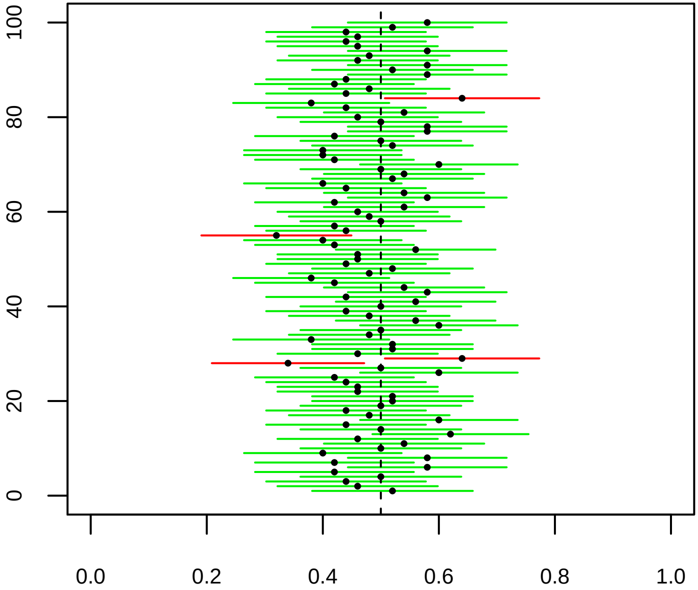
Simulação IC para proporção

```r
p <- 0.5
n <- 50  # tamanho da amostra
B <- 100 # número de simulações
z <- qnorm(.975)

p.hat<-Inflim <- Suplim <- rep(0,B)
for(i in 1:B){
data <- rbinom(n,1,p)
n <- length(data)
p.hat[i] <- mean(data)
Inflim[i] <- p.hat[i] - z*sqrt(p.hat[i]*(1-p.hat[i])/n)
Suplim[i] <- p.hat[i] + z*sqrt(p.hat[i]*(1-p.hat[i])/n)
}
```

```r
Int <- Inflim <= p & p <= Suplim
sum(Int==1)/B
```

```
[1] 0.96
```

```r
cores <- Int
cores[cores==F]<-"red";cores[cores==T]<-"green2"

plot(1,1,xlim=c(0,1),ylim=c(0,B),type='n',xlab ="",ylab ="",
     cex.axis=0.7)
for (i in 1:B) {
  lines(c(Inflim[i],Suplim[i]),c(i,i),col=cores[i])
  points(p.hat[i],i,pch=16,cex=.5)
}
abline(v=p,lty=2)
```


0
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0