Conover pag 85 a 88

Intervalos de Confiança aproximados

$X_1, X_2, \ldots X_n$ v.a. independentes com média $\mu$ e variância $\sigma^2$.

$$\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1).$$

A função distribuição de $\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}}$ se aproxima da função distribuição da $N(0,1)$, para $n \to \infty$.

$$\therefore\ P\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \simeq 1-\alpha$$

$$P\left(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}\right) = 1-\alpha = \gamma \qquad Z \sim N(0,1)$$

IC aproximado para $\mu$ com coeficiente de confiança $1-\alpha$, para $n$ grande.

Pag 86

O procedimento Bootstrap

Verifica-se que $E(S^2) = \sigma^2$. Usar o resultado

$$\left[E(X'AX) = tr(A\Sigma) + \mu' A \mu\right]$$

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

Não é simples obter $Var(S^2)$.

Uma possibilidade é estimar $Var(S^2)$ utilizando o método bootstrap.

Sejam $X_1, X_2 \dots X_n$ amostra aleatória da v. a. $X$ com $E(X) = \mu$ e $Var(X) = \sigma^2$. Seja $\theta$ o parâmetro de interesse. O procedimento Bootstrap consiste em

São tomadas várias amostras de tamanho $n$ com reposição da amostra observada.

Calcula-se o estimador $\hat{\theta}$ em cada amostra.

$\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_{1000} \rightarrow$ computador

$$\hat{\theta}_b = \frac{\hat{\theta}_1 + \hat{\theta}_2 + \dots + \hat{\theta}_{1000}}{1000} \rightarrow \text{estimador de } E(\hat{\theta})$$

$$\text{e } \frac{1}{999} \sum_{i=1}^{1000} \left(\hat{\theta}_i - \hat{\theta}_b\right)^2 \rightarrow \text{estimador de } Var(\hat{\theta}).$$

Com base nesse processo é possível construir IC e realizar TH.

Pág 115

Métodos Robustos

São métodos que mesmo construídos para dados com certa distribuição (por ex. normal) continuam eficientes quando a distribuição dos dados não é normal.

O teste $t$ para uma única média ou para igualdade de médias tem essa propriedade. (A distribuição da estatística de teste sob a hipótese nula sem a normalidade dos dados é aproximadamente a mesma que no caso em que existe normalidade).

$T \sim B(n, p)$ - Inferência para $p$

IC otimista $\left[\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$

IC conservador $\left[\hat{p} - z\sqrt{\frac{1}{4n}}, \hat{p} + z\sqrt{\frac{1}{4n}}\right]$

Dimensionamento da amostra

$n = ?$ $\quad P(|\hat{p} - p| \leq \varepsilon) = \gamma \quad 0 < \gamma < 1$

$\hat{p} = \frac{T}{n} \quad E(\hat{p}) = p \quad Var(\hat{p}) = \frac{p(1-p)}{n}$

TLC $\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$

$$\gamma = P(-\varepsilon \leq \hat{p} - p \leq \varepsilon) = P\left(\frac{-\varepsilon}{\sqrt{\frac{p(1-p)}{n}}} \leq \overbrace{\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}}^{Z \sim N(0,1)} \leq \underbrace{\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}}_{z}\right)$$

$$= P(\hat{p} - \varepsilon \leq p \leq \hat{p} + \varepsilon)$$

$$z = \frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}} \iff n = \left(\frac{z}{\varepsilon}\right)^2 p(1-p)$$

$p(1-p)$ desconhecido, utiliza-se $p(1-p) = \frac{1}{4}$

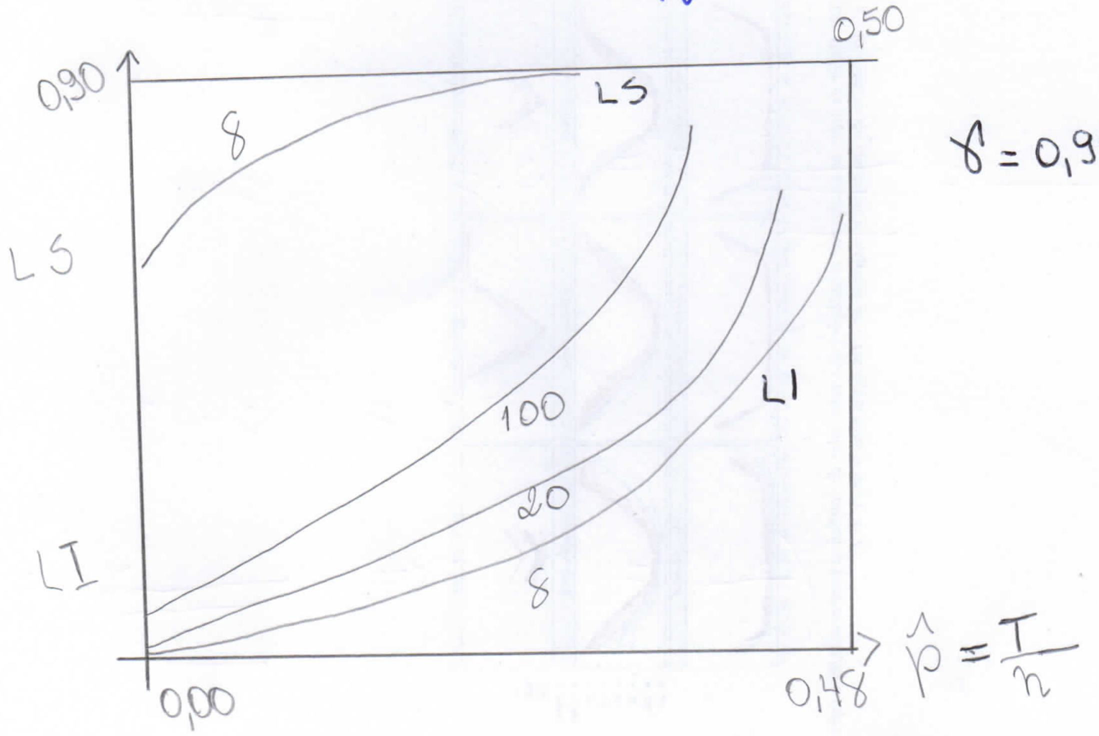
$$n = \left(\frac{z}{\varepsilon}\right)^2 \cdot \frac{1}{4}.$$

$\varepsilon = 0{,}02 \quad \gamma = 0{,}95 \quad z = 1{,}96$

$$n = \left(\frac{1{,}96}{0{,}02}\right)^2 \cdot \frac{1}{4} = 2401$$

## Intervalo de Confiança para p com base em pequenas amostras

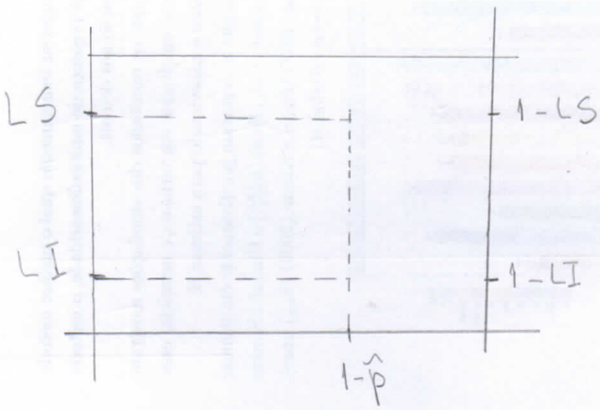
Para $\gamma = 0{,}95$ ou $\gamma = 0{,}99$, a Tabela A.4, Conover apresenta os limites inferior e superior para diferentes valores de $n$ e $\hat{p} = \frac{T}{n}$




$\gamma = 0{,}95$

Se $\hat{p} > 0{,}49$, consultar o eixo horizontal superior e a ordenada da direita. Neste caso, constrói se I C para $1-p$ ao invés de $p$

$$P(LI < 1-p < LS) = \gamma = P(1-LS < p < 1-LI)$$




Construção da tabela:

O intervalo de confiança consiste de todos os valores $p^*$ para os quais, para aquele valor de $T = n\hat{p}$, aceitaríamos a hipótese $H_0: p = p^*$ contra $H_a: p \neq p^*$, no teste exato.

$$RC: \{0, 1, 2 \ldots K\} \cup \{n-K \ldots n-1, n\}$$

Estatística de teste: $T$ sob $H_0 \sim B(n, p^*)$

$$P(T \in \{0, 1, 2 \ldots K\}) = \frac{\alpha}{2} \quad T \sim B(n, p^*).$$

Exemplo

Vinte observações independentes de uma variável aleatória X com função distribuição desconhecida F(x) resultaram em

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 142 | 134 | 98 | 119 | 131 | 103 | 154 | 122 | 93 | 137 |
| 86 | 119 | 161 | 144 | 158 | 165 | 81 | 117 | 128 | 103 |

Determinar um IC com coeficiente 0,95 para F(100).

$F(x) = P(X \leq x)$ desconhecida para todo $x$.

Estimar $F(100) = P(X \leq 100)$.

Amostra de $n = 20$ observações independentes.

Categoria 1 - a observação é menor ou igual a 100.

$p = P(\text{categoria } 1) = P(X \leq 100) = F(100)$.

Problema - estimar p

T = nº de elementos na amostra $\leq 100$

$T = 4 \quad \hat{p} = \frac{T}{n} = \frac{4}{20} = 0,2$

$\gamma = 0,95$ Tabela A.4 $LI = 0,057$ $LS = 0,44$

IC : $[0,057; 0,440]$

Usando a aproximação normal

$$\left[0{,}2 - 1{,}96\sqrt{\frac{1}{80}} \;;\; 0{,}2 + 1{,}96\sqrt{\frac{1}{80}}\right]$$

$$[0{,}000 \;;\; 0{,}419]$$

Aplicações do teste binomial

1 - Teste do Quantil para v.a. contínua

$X$ v.a. contínua

$x$ é o quantil de ordem $p^*$ da distribuição de $X$ se

$P(X \leq x) = P(X < x) = p^*$.

$x_{p^*} = x$

$H_0$: O quantil de ordem $p^*$ de $X$ é $x$ $\Leftrightarrow$ $H_0: P(X \leq x) = p^*$

Seja $p = P(X \leq x)$ então $H_0: p = p^*$, hipótese nula do teste binomial.

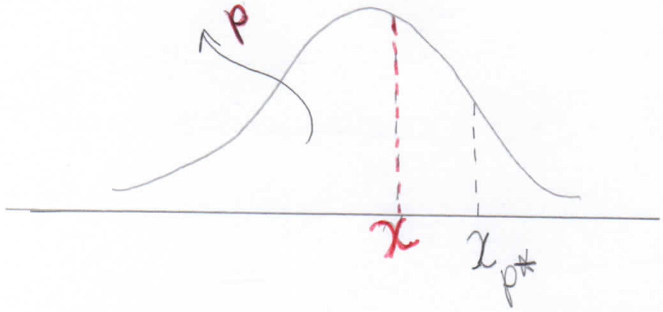
Amostra aleatória: $X_1 \; X_2 \ldots X_n$

$T$ = nº de observações na amostra $\leq x$.

$T \sim B(n, p)$ e sob $H_0$, $T \sim B(n, p^*)$

A região crítica depende da hipótese alternativa

$H_a$: O quantil de ordem $p^*$ é maior que $x$.




$p = P(X \leq x) < p^*$

$H_a \iff p < p^*$

Ex: O intervalo de tempo entre as erupções de um geyser (nascente termal que entra em erupção periodicamente) é marcado 112 vezes pois deseja-se verificar se o intervalo mediano é maior que 60 minutos.

X - intervalo de tempo  Mediana$(X) = x_{0,50}$  $p^* = 0,50$

$x = 60$

$H_0$: $x_{0,50} = 60 \iff P(X \leq 60) = 0,5 \iff p = 0,5$

$H_a$: $x_{0,5} > 60 \iff P(X \leq 60) < 0,5 \iff p < 0,5$

T - nº de observações $\leq 60$

RC: $T \leq t$  $P(Y \leq t) = \alpha$  $Y \sim B(112, 0,5)$

$$\simeq P\left(Z \leq \overset{1,96}{\frac{t - 112.0,5}{\sqrt{112.0,5.0,5}}}\right) = 0,05 \quad Z \sim N(0,1)$$

Se $T = 30$

$$\text{Pvalor} = P(T \leq 30 \mid p = 0{,}5) = P\left(Z \leq \frac{30 - 56}{\sqrt{28}}\right) =$$

$$= P(Z \leq -4{,}9) \simeq 0.$$ Rejeita-se $H_0$.

56 56

Mediana = 60

(30 obs ≤ 60) ⟹ mediana > 60

Ex: Os valores a seguir correspondem a uma amostra aleatória dos graus obtidos por 15 estudantes em um determinado exame.

189 233 195 160 212 176 231 185 199

213 202 193 174 166 248

Teste, ao nível de significância de 0,05, a hipótese de que os graus vêm de uma população cujo 3º quartil é 193.