Capítulo 5 - Estimação Pontual

Estatística

Definição (População)

Uma **população** *consiste no conjunto de elementos sobre o qual incide o estudo estatístico.*

Definição (Característica Estatística ou Atributo)

Característica Estatística ou Atributo *é a característica que se observa nos elementos da população.*

Observação: Numa população podemos ter mais que uma **característica estatística** ou **atributo**.

Definição (Parâmetro populacional)

Um **parâmetro populacional** *é uma medida numérica que descreve uma característica da população.*

Exemplos: a média, a variância e a proporção populacionais (estes são os mais estudados)

Definição (Amostra)

Uma **amostra** *é um subconjunto da* **população**.

Definição (Amostra aleatória)

Vamos admitir que cada valor observado x_i *é a realização da variável aleatória* X_i*, com função de distribuição* F*. O vector* $(X_1, X_2, \ldots, X_n)$ *constitui uma* **amostra aleatória** *se e só se as* n *variáveis aleatórias são independentes e têm todas a mesma distribuição. Os valores que se obtêm por concretização da amostra aleatória são representados por* $(x_1, x_2, \ldots, x_n)$.

Estatística

Definição (Estatística)

Uma **estatística** *é uma qualquer função da amostra aleatória,* $(X_1, X_2, \ldots, X_n)$*, que não depende de qualquer parâmetro desconhecido.*

Definição (Estimador)

Um **estimador** *é uma estatística que estima um determinado parâmetro populacional desconhecido.*

Um estimador diz-se **estimador pontual** *quando estima pontualmente o parâmetro populacional desconhecido.*

Exemplos: a média, a variância, o máximo, o mínimo, a mediana, a moda, ... amostrais

Estimação pontual

Definição (Estimativa)

Seja $X_1, X_2, \ldots, X_n$ uma a.a. de uma população $X \sim F(\theta)$, com F a função de distribuição da população e θ um parâmetro populacional desconhecido (pode ser a média, a variância ou outro qualquer).

Sendo $\hat{\Theta} = h(X_1, \ldots, X_n)$ um estimador pontual de θ, uma vez observadas as concretizações $x_1, \ldots, x_n$ da a.a. acima, o valor particular $\hat{\theta} = h(x_1, x_2, \ldots, x_n)$ diz-se uma **estimativa pontual** *de θ.*

Alguns **parâmetros populacionais** com respectivos **estimadores** e **estimativas**:

Parâmetro Populacional	**Estimador**	**Estimativa**
Média populacional μ	Média amostral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Proporção populacional p	Proporção amostral $\hat{P} = \frac{X}{n}$	$\hat{p} = \frac{x}{n}$
Variância populacional σ^2	Variância amostral $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Desvio padrão populacional σ	Desvio padrão amostral $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Estatística Descritiva

No  iremos agora gerar vários conjuntos de dados usando algumas das funções do `R` já mencionadas em capítulos anteriores, nomeadamente

`rhyper()`, `rbinom()`, `rpois()`, `rexp()` e `rnorm()`.

Iremos depois utilizar o conjunto de instruções abaixo no estudo descritivo desses conjuntos de dados

- `summary()`
  indicadores estatísticos básicos (min, max, mediana, média e 1º e 3º quartis)
- `var()` e `sd()`
  variância e desvio padrão
- `boxplot()`
  diagrama de caixa e bigodes
- `hist()`
  histograma

Estatística Descritiva - Exemplo

Começamos por gerar uma amostra de tamanho 100 da populaçao $X \sim N(0,1)$ que sabemos ser tal que

- $\mu = 0$ (média)
- $\sigma^2 = 1 = \sigma$ (variância e desvio padrão)
- $Q_2 = 0$ (mediana)
- $Q_1 \simeq -0.6745$ e $Q_3 \simeq 0.6745$ (1ª e 3ª quartis)

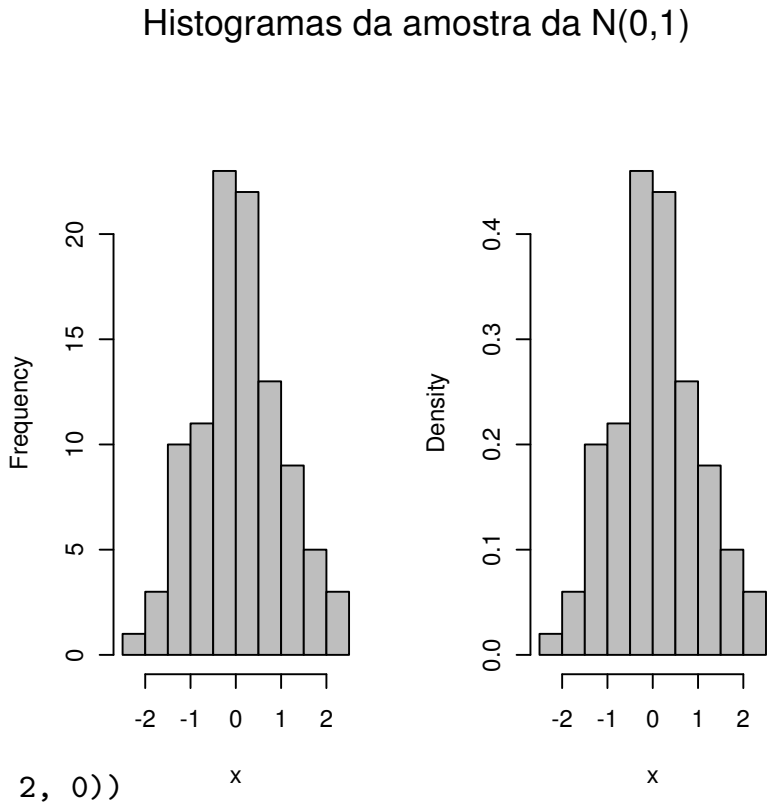
```
# fixamos a semente dos números pseudo-aleatórios
# para podermos reproduzir a amostra que vamos gerar
set.seed(123)
x<-rnorm(100,0,1)
```

Em seguida calculamos as estimativas dos valores populacionais acima descritos usando a amostra gerada:

```
# usando o summary(), var() e sd()
summary(x)
#     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
# -2.30900 -0.49390  0.06176  0.09041  0.69180  2.18700
var(x)
# 0.8332328
sd(x)
# 0.9128159
```

Estatística Descritiva - Exemplo

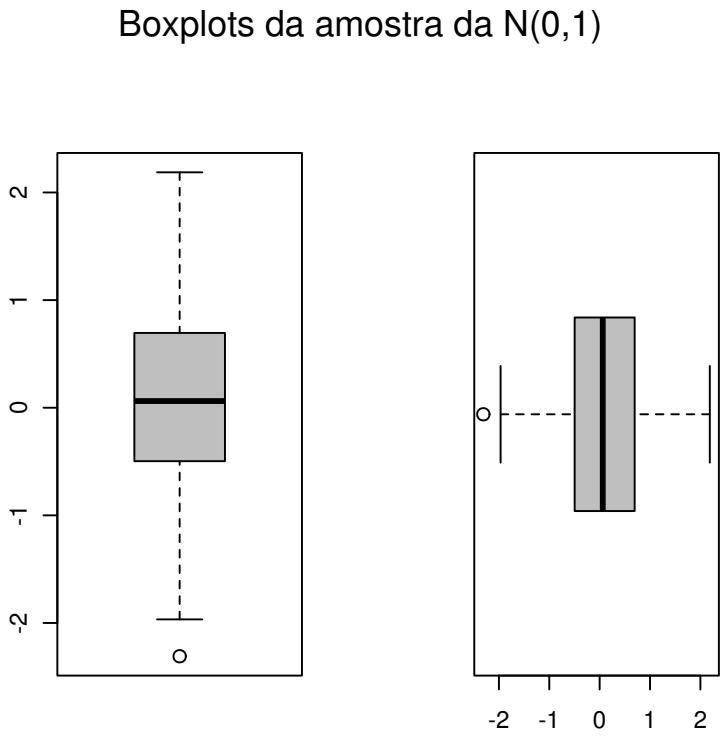
Voltamos agora à nossa amostra inicial de tamanho 100 e usamos a instrução `hist()` para desenhar os histrogramas de frequência e de densidade dos dados gerados




```
par(mfrow=c(1,2),oma = c(0, 0, 2, 0))
set.seed(123)
x<-rnorm(100,0,1)
hist(x,col="grey",main = "")
hist(x,freq = F,col="grey",main = "")
mtext("Histograms of N(0,1) sample", outer = TRUE, cex = 1.5)
```

Estatística Descritiva - Exemplo

Ainda usando a amostra inicial de tamanho 100 usamos a instrução `boxplot()` para desenhar o histrograma de caixa dos dados gerados




```
par(mfrow=c(1,2),oma = c(0, 0, 2, 0))
set.seed(123)
x<-rnorm(100,0,1)
boxplot(x,col="grey",main = "")
boxplot(x,horizontal = T,col="grey",main = "")
mtext("Boxplots da amostra da N(0,1)",
outer = TRUE, cex = 1.5)
```

As linhas da caixa do diagrama representam, respectivamente, de baixo para cima (diag. esquerda) e da esquerda para a direita (diag. direita) o **1º, 2º e 3º quartis**.

Os bigodes do diagrama dizem respeito às **barreiras inferior e superior**. Sendo $IQR = Q_3 - Q_1$ o intervalo interquartil, as barreiras inferior e superior são habitualmente dadas por $B_{inf} = Q1 - 1.5IQR$ e $B_{sup} = Q3 + 1.5IQR$, respectivamente. Os pontos amostrais que caírem fora destas barreiras são considerados suspeitos (**outliers**)!

Estatística Descritiva - Exercícios

Reproduza o exemplo anterior para amostras de tamanho $n = 50$ e $n = 150$ das populações $X \sim P(15)$ e $Y \sim Exp(1/4)$, respectivamente.

Ao nível gráfico, investigue os argumentos das funções `hist()` e `boxplot()` e verifique como pode alterar alguns aspectos gráficos como por exemplo, o tamanho de letra, côr, renomeação dos eixos, etc.

Verifique ainda no código fornecido acima para o display gráfico, como pode juntar mais do que dois gráficos numa única janela.

Algumas Propriedades dos Estimadores

Definição (Estimador centrado)

Um estimador pontual $\hat{\Theta}$ diz-se centrado para o parâmetro θ se $E(\hat{\Theta}) = \theta$.

Observação: Caso $E(\hat{\Theta}) \neq \theta$, o estimador diz-se enviesado. A diferença

$$b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$$

corresponde ao valor do enviesamento ou viés de $\hat{\Theta}$.

Se $E(\hat{\Theta}) \neq \theta$, e $\lim_{n\to\infty} E(\hat{\Theta}) = \theta$, diz-se que o estimador é assintoticamente centrado.

Exemplo

Seja, $X_1, \ldots, X_n$ uma a.a. de uma população X de média $E(X) = \mu$ e variância $V(X) = \sigma^2$. A média $\bar{X}$ é um estimador centrado de μ.

Relembre que

$$E(\bar{X}) = E\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \underset{X_i ident. dist.}{=} \frac{1}{n}\sum_{i=1}^{n} E(X) = \frac{1}{n} \times n\mu = \mu$$

e logo temos que $\bar{X}$ é um estimador centrado para μ.

Algumas Propriedades dos Estimadores

Definição (Erro Padrão de um estimador)

Dado um estimador pontual $\hat{\Theta}$, *centrado, define-se o seu erro padrão, que se designa* $SE_{\hat{\Theta}}$, *como a raiz quadrada da sua variância, caso exista:*

$$SE_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$$

Caso o erro padrão envolva parâmetros desconhecidos, que possam ser estimados dos dados, a substituição destes valores estimados no erro padrão produz o chamado **erro padrão estimado**, *denotado por* $\widehat{SE}_{\hat{\Theta}}$.

Exemplo

Seja, $X_1, \ldots, X_n$ *uma a.a. de uma população* X *de média* $E(X) = \mu$ *e variância* $V(X) = \sigma^2$. *Calcule o erro padrão do estimador* $\bar{X}$ *de* μ. *Considerando uma concretização* $4\ 7\ 5\ 7\ 8\ 2\ 5\ 8\ 5\ 5$ *da a.a, estime o erro padrão do estimador.*

Temos então o erro padrão dado por

$$SE(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{V(X)}{n}} = \sigma/\sqrt{n}$$

pelo que uma estimativa do erro padrão será dada por $\widehat{SE(\bar{X})} = \hat{\sigma}/\sqrt{10} = s/\sqrt{10} \simeq 0.6$.

Algumas Propriedades dos Estimadores

Definição (Erro Quadrático Médio)

O erro quadrático médio de um estimador pontual $\hat{\Theta}$ de um parâmetro θ é definido por

$$EQM(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2] \underset{TPC}{=} V(\hat{\Theta}) + b^2(\hat{\Theta}).$$

Observação: Se o estimador for centrado, então $EQM(\hat{\Theta}) = V(\hat{\Theta})$. Como verificámos anteriormente que $\bar{X}$ é um estimador centrado para μ então $EQM(\bar{X}) = V(\bar{X}) = \sigma^2/n$.

Definição (Estimador consistente em média quadrática)

Um estimador pontual $\hat{\Theta}$ de um parâmetro θ, diz-se consistente em média quadrática se

$$\lim_{n\to\infty} EQM(\hat{\Theta}) = 0.$$

Exemplo

Seja, $X_1, \ldots, X_n$ uma a.a. de uma população X de média $E(X) = \mu$ e variância $V(X) = \sigma^2$. Tem-se que $\bar{X}$ é um estimador consistente em média quadrática de μ.

Facilmente se observa que $\lim_{n\to\infty} EQM(\bar{X}) \underset{obs.}{=} \lim_{n\to\infty} \sigma^2/n \underset{\sigma>0}{=} 0.$

Algumas Propriedades dos Estimadores

Definição (Eficiência)

Sejam $\hat{\Theta}_1$ e $\hat{\Theta}_2$ dois estimador pontuais de um parâmetro θ. Diz-se que $\hat{\Theta}_1$ é mais eficiente que $\hat{\Theta}_2$, se e só se,

$$EQM(\hat{\Theta}_1) \leq EQM(\hat{\Theta}_2).$$

No caso de ambos os estimadores serem centrados para θ esta condição simplifica para

$$V(\hat{\Theta}_1) \leq V(\hat{\Theta}_2).$$

Consistência: Exemplo

Apresentamos abaixo os gráficos animados das estimativas $\bar{x}_n$ obtidas de amostras de tamanhos $n = 1, ..., 200$ das populações $N(\mu = 0, \sigma = 1)$ (esquerda) e $P(\lambda = 5 = \mu = \sigma^2)$ (direita), respectivamente. (para a correcta visualização do mesmo é necessário abrir este pdf com recurso ao software Adobe Reader).

Para qual das populações a média populacional converge mais depressa para o verdadeiro parâmetro?