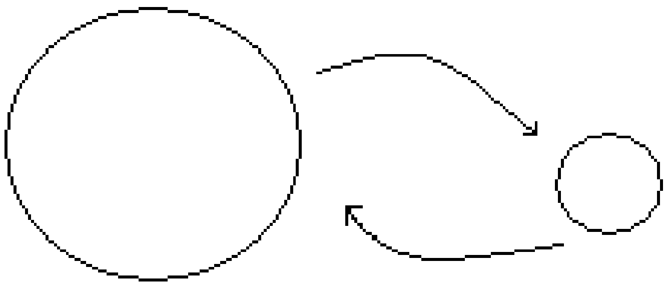
# 4 Amostragem e Distribuições Amostrais

Neste capítulo serão dadas algumas noções e resultados teóricos fundamentais em Estatística. O estudo de características populacionais, a partir de amostras, assenta nestes resultados.

## 4.1 Amostra Aleatória. Estatísticas

**População** $X$ **Amostra Aleatória** $X_1, X_2, ..., X_n$



**Definições**

Uma **amostra aleatória** da v.a. (população) $X$ é um conjunto de $n$ v.a.'s $X_1, X_2, ..., X_n$ independentes e com a mesma distribuição de $X$ (dizem-se independentes e identicamente distribuídas (i.i.d.) com $X$).

**Estatística** é uma função da amostra aleatória, e logo uma variável aleatória, cuja expressão não contém parâmetros desconhecidos.

Dada uma amostra aleatória de $X$, $X_1, X_2, ..., X_n$, podemos definir as seguintes estatísticas:

- A **média da amostra** ou **média amostral**

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- A **variância da amostra** ou **variância amostral** (corrigida)

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

- O **desvio padrão amostral** é dado por $S_n = \sqrt{S_n^2}$.

**Exercício:** Verifique que $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2$.

A média e a variância amostrais são estatísticas (especiais no nosso estudo) e, logo, variáveis aleatórias, com distribuições próprias, com parâmetros de localização e dispersão, que dependem da população em estudo. As suas distribuições são designadas *distribuições amostrais*.

## 4.2 Distribuição da Média Amostral

Considere-se uma amostra aleatória $X_1, X_2, ..., X_n$ da população $X$ com parâmetros $E(X) = \mu$ e $V(X) = \sigma^2$.

Usando propriedades da esperança e da variância, verifica-se que:

$$E(\overline{X}) = \mu \quad \text{e} \quad V(\overline{X}) = \frac{\sigma^2}{n} \Rightarrow \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}} \,.$$

**Teorema do Limite Central (T.L.C.)**

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória da população $X$ com parâmetros $E(X) = \mu$ e $V(X) = \sigma^2$. Para $n$ *suficientemente* grande, tem-se

$$\frac{\overline{X} - E(\overline{X})}{\sigma(\overline{X})} = \sqrt{n}\ \frac{\overline{X} - \mu}{\sigma} \dot\sim \mathcal{N}(0, 1).$$

**Notas**

- A aproximação à Normal centrada e reduzida pode ser efetuada na prática para $n > 30$; no entanto, esta será tanto melhor quanto maior for o valor de $n$.

- Se a população é normal, isto é, $X \sim \mathcal{N}(\mu, \sigma)$ (caso já estudado) então

  $$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \sqrt{n}\ \frac{\overline{X} - \mu}{\sigma} \sim \mathcal{N}(0, 1) \text{ (distribuição exacta!)}$$

  O T.L.C. permite alargar este resultado a qualquer família, desde que se considere uma amostra de tamanho bastante razoável!

**Justificação de alguns resultados apresentados anteriormente**

- Se $X \sim \mathcal{B}(n, p)$ $[E(X) = np$ e $V(X) = np(1 - p)]$, para $n$ *suficientemente grande* e $p \in\ ]0.1, 0.9[$,

  $$X \dot\sim \mathcal{N}\left(np, \sqrt{np(1-p)}\right) \Leftrightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} \dot\sim \mathcal{N}(0, 1)$$

  **Nota**

  Neste caso, $X = \sum_{i=1}^{n} X_i$ com $X_i \sim \mathcal{B}(p)$ para $i = 1, ..., n$.

  **Exercício:** Deduza que

  $$Z_n = \sqrt{n}\frac{\overline{X} - p}{\sqrt{p(1-p)}} \dot\sim \mathcal{N}(0, 1)$$

Anteriormente, vimos que

se $p \leq 0.1$ então $X \dot{\sim}\ \mathcal{P}(np)$;

se $p \geq 0.9$ então $Y = n - X \dot{\sim} \mathcal{P}(n(1-p))$

- Se $X \sim \mathcal{P}(\lambda)$ $[E(X) = V(X) = \lambda]$, para $\lambda$ *suficientemente grande*,

$$X \dot{\sim}\ \mathcal{N}\left(\lambda, \sqrt{\lambda}\right) \ \Rightarrow\ Z = \frac{X - \lambda}{\sqrt{\lambda}} \dot{\sim}\ \mathcal{N}(0,1)$$

**Nota**

Neste caso, $X = \sum_{i=1}^{n} X_i$ com $X_i \sim \mathcal{P}(\frac{\lambda}{n})$ para $i = 1, ..., n$.

**Distribuições da média amostral $\overline{X}$:**

- Para $n$ *suficientemente* grande, $\sqrt{n}\ \frac{\overline{X} - \mu}{\sigma} \dot{\sim}\ \mathcal{N}(0,1)$.
- Se $X \sim \mathcal{N}(\mu, \sigma)$,

$$\sqrt{n}\ \frac{\overline{X} - \mu}{\sigma}\ \sim\ \mathcal{N}(0,1)$$

$$\sqrt{n}\ \frac{\overline{X} - \mu}{S_n}\ \sim\ t_{n-1}$$

## 4.3 Distribuição da Variância Amostral

- Se $X \sim \mathcal{N}(\mu, \sigma)$, $\frac{n-1}{\sigma^2}\ S_n^2 \sim \chi^2_{n-1}$