MACS – 11º Ano

5ª Aula

Prof. Álvaro Velosa

Estimativa pontual para a proporção

Já vimos que um conjunto de médias amostrais são distribuídas em torno da média populacional (ou valor médio), veremos que as proporções amostrais ($\hat{p}$) são distribuídas em torno da proporção populacional (p).

Vamos construir a distribuição de amostragem do estimador proporção amostral ($\hat{p}$) tal como fizemos para a média.

Exemplo:

De uma turma de 25 alunos, pretendemos estimar a proporção de alunos do género feminino. Recolheu-se, de forma aleatória, uma amostra de 10 alunos.

Na tabela seguinte temos os dados referentes à amostra.

Nome	Tatiana	Clara	Nuno	Vitória	Bruno	Nélia	André	Ana	Pedro	Maria
Género	Feminino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino

Organizando os dados, obtemos:

Género	Nº de alunos (fi)	Proporção (fri)
Feminino	6	$\frac{6}{10} = 0{,}6 = 60\%$
Masculino	4	$\frac{4}{10} = 0{,}4 = 40\%$
Total	10	1

Proporção amostral Género Feminino: $\hat{p} = \frac{6}{10} = 0{,}6 = 60\%$

Proporção amostral género masculino: $1 - \hat{p} = 1 - \frac{6}{10} = \frac{4}{10} = 0{,}4 = 40\%$

Outro processo para calcular a proporção é recorrer à média.

Se o indivíduo da amostra for do género feminino, anotamos ***1*** e, se for do género masculino anotamos ***0***.

Assim:

Nome	**Tatiana**	**Clara**	**Nuno**	**Vitória**	**Bruno**	**Nélia**	**André**	**Ana**	**Pedro**	**Maria**
Género	***1***	***1***	***0***	***1***	***0***	***1***	***0***	***1***	***0***	***1***

Representamos a população através de uma variável aleatória ***X*** que só assume os valores ***1*** e ***0***, com probabilidades ***p*** e ***1−p***, em que ***p*** é a proporção dos alunos do género feminino.

Podemos afirmar que ***p*** é a frequência relativa com que o ***1*** aparece na população, a média do conjunto de zeros e uns.

A média da amostra é uma <u>estimativa pontual</u> para a proporção populacional.

Assim:

$$\bar{x} = \frac{1+1+0+1+0+1+0+1+0+1}{10} = \frac{6}{10} = 0{,}6 = 60\% \quad \text{ou seja, } \boldsymbol{\bar{x}} = \boldsymbol{\hat{p}}.$$

- Considerando a distribuição de probabilidade da variável aleatória X:

X	0	1
Probabilidade	1−p	p

- Valor médio:

$$\mu = 0 \times (1-p) + 1 \times p = p$$

- Desvio padrão populacional:

$$\sigma = \sqrt{(0-p)^2 \times (1-p) + (1-p)^2 \times p} = \sqrt{p^2 \times (1-p) + (1-2p+p^2) \times p} =$$

$$= \sqrt{p^2 - p^3 + p - 2p^2 + p^3} = \sqrt{p - p^2} = \sqrt{p(1-p)}$$

- Desvio padrão de $\hat{p}$:

$$\sigma_{\hat{p}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

Tal como no caso da estimação do valor médio, também se aplica o TLC na distribuição de amostragem de uma proporção.

Teorema do limite central (TLC)

Seja $\boldsymbol{X}$ uma população, em que cada elemento tem ou não determinada propriedade, da qual se recolhem amostras de dimensão $\boldsymbol{n}$.

Seja $\boldsymbol{p}$ a proporção de elementos da população com essa propriedade e a dimensão da amostra suficientemente grande $(n \geq 30)$. Assim, **a distribuição de amostragem da proporção $\widehat{p}$ pode ser aproximada por uma distribuição normal com valor médio p e desvio padrão $\sqrt{\frac{p(1-p)}{n}}$.**

Exercício:

Uma fábrica produz garrafas de vidro e sabe-se que 8% são defeituosas.

Por dia a fábrica produz 620 garrafas.

1. Como se distribui a proporção, $\hat{p}$, de garrafas em cada dia?

2. Determine a probabilidade da proporção de garrafas defeituosas num dia, escolhido ao acaso, ser superior a 10%.

Proposta de resolução

1.

Dados:

- População: p $= 8\% = 0{,}08$
- Amostra: $n = 620$

> Uma fábrica produz garrafas de vidro e sabe-se que 8% são defeituosas.
> Por dia a fábrica produz 620 garrafas.
>
> 1. Como se distribui a proporção $\hat{p}$ de garrafas em cada dia?

Como $n \geq 30$, pelo TLC, a distribuição de amostragem da proporção pode ser aproximada por um modelo normal de valor médio p $= 8\% = 0{,}08$

e desvio padrão $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}08(1-0{,}08)}{620}} \approx 0{,}011$

Assim, a proporção $\hat{\boldsymbol{p}}$ segue aproximadamente um modelo normal, isto é, N(0,08; 0,011).

2.

> 2. Determine a probabilidade da proporção de garrafas defeituosas num dia, escolhido ao acaso, ser superior a 10%.

Por 1, temos $\hat{p} \sim N(0{,}08; 0{,}011)$

Recorrendo à calculadora gráfica:

- <u>Casio</u>:

$P(\hat{p} > 10\%) = P(\hat{p} > 0{,}1) = 1 - P(\hat{p} \leq 0{,}1) = 1 - P(0 \leq \hat{p} \leq 0{,}1) =$

$= 1-$ normal ncd $\begin{pmatrix} Lower{:}0 \\ Upper{:}0.1 \\ \sigma{:}0.011 \\ \mu{:}0.08 \end{pmatrix} =$ 1- 0,96548182≈ 0,0345≈3,45%

- <u>Texas</u>:

$P(\hat{p} > 10\%) = P(\hat{p} > 0{,}1) = 1 - P(\hat{p} \leq 0{,}1) = 1 - P(0 \leq \hat{p} \leq 0{,}1) =$

$= 1-$ normal cdf $(0, 0.1, 0.08, 0.011) = 1-$ 0,96548182≈ 0,0345≈3,45%

Intervalos de confiança para a proporção

Vimos que a proporção, $\boldsymbol{p}$, também é um valor médio e a proporção amostral, $\widehat{\boldsymbol{p}}$, é uma média amostral.

Para obter os intervalos de confiança para a proporção

no intervalo de confiança para o valor médio $\left]\bar{x} - \mathrm{z} \times \frac{s}{\sqrt{n}}\,; \bar{x} + \mathrm{z} \times \frac{s}{\sqrt{n}}\right[$

substituímos $\overline{\boldsymbol{x}}$ por $\widehat{\boldsymbol{p}}$ e $\boldsymbol{s}$ por $\sqrt{\widehat{\boldsymbol{p}}(\mathbf{1} - \widehat{\boldsymbol{p}})}$.

Se o desvio padrão populacional for conhecido subtituímos $\boldsymbol{\sigma}$ por $\sqrt{\boldsymbol{p}(\mathbf{1} - \boldsymbol{p})}$.

Caso geral:

Intervalo de confiança para uma proporção *p*, admitindo que a amostra tem dimensão superior ou igual a 30

$$\left]\hat{p} - Z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ;\hat{p} + Z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right[$$

n- dimensão da amostra

$\hat{p}$- proporção amostral

z – valor relacionado com o nível confiança

Valores de **z** para os níveis de confiança mais usuais:

Nível de confiança	90%	95%	99%
z	1,645	1,960	2,576

Exercício - Exame 2019-2ª Fase (8)

8. Num encontro de colecionadores de jogos, verificou-se que, numa amostra de 200 colecionadores, 45 colecionavam jogos de tabuleiro.

Determine um intervalo de confiança a 90% para a proporção de colecionadores de jogos de tabuleiro presentes no encontro.

Apresente os extremos do intervalo de confiança, em percentagem, com arredondamento às décimas.

Caso proceda a arredondamentos nos cálculos intermédios, conserve quatro casas decimais.