

Técnicas de Inferência Estatística baseadas na Distribuição Binomial

Hélio Radke Bittencourt, Simone Echeveste

Laboratório de Estatística, ULBRA

Av. Miguel Tostes, 101

92420-280, Canoas, RS

E-mail: heliorb@brturbo.com, echeveste@redemeta.com.br

Arno Bayer, Josy Rocha[†]

Programa de Pós Graduação em Ensino de Ciências e

Matemática, ULBRA

Av. Miguel Tostes, 101

92420-280, Canoas, RS

E-mail: arnob@ulbra.br

As técnicas de inferência lecionadas na maior parte das disciplinas de Estatística no Ensino Superior baseiam-se nas distribuições Normal e *t* de Student. Estes dois modelos probabilísticos são próprios para descrever o comportamento de variáveis aleatórias contínuas como, por exemplo, a média amostral (\bar{X}) e o estimador da proporção populacional (\hat{p}) para grandes amostras.

Todo modelo probabilístico pode ser encarado como uma descrição simplificada da realidade. A célebre frase de Box - “Todos modelos são errados, mas alguns são úteis.” - apresenta muito bem esta idéia. A prática de utilizar distribuições contínuas para representar situações discretas é amplamente utilizada na Estatística. A título de exemplificação, considere uma população formada por N elementos onde p é a proporção de elementos que apresentam uma determinada característica, representada pela variável binária X, assim definida:

$$X = \begin{cases} 1, & \text{característica presente} \\ 0, & \text{característica ausente} \end{cases}$$

Desta forma, numa amostra de n elementos desta população ($n < N$) podemos definir o valor do estimador \hat{p} :

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n}$$

O número de possíveis resultados para o estimador \hat{p} será igual a $n+1$, dependendo, portanto,

do tamanho amostral. Quando $n \rightarrow \infty$ o comportamento probabilístico de \hat{p} segue uma distribuição aproximadamente Normal, valendo-se da convergência da distribuição Normal para a Binomial, encontrada em [1]:

$$X \underset{n \rightarrow \infty}{\sim} \text{Normal} (\text{Média} = np; \text{Variância} = npq)$$

É possível encontrar, portanto, intervalos de confiança para o parâmetro p e realizar testes de hipóteses valendo-se da relação entre a Binomial e a Normal, entretanto este comportamento, de acordo [2], só é verificado para situações onde $np \geq 5$ e $np(1-p) \geq 5$. Para casos onde o tamanho amostral é reduzido ou, ainda, o evento investigado é muito raro ($p \cong 0$), as condições de convergência não serão atendidas. A proposta deste estudo é, portanto, apresentar de forma didática técnicas de inferência estatística baseadas na distribuição Binomial, salientando as vantagens e desvantagens da utilização de uma distribuição contínua para modelar um fenômeno notoriamente discreto.

Referências

- [1] M. R. Spiegel. “Probabilidade e Estatística”. Makron Books, São Paulo, 1978.
- [2] P. L. O. Costa Neto, “Estatística”, 2a Ed, Edgard Blücher, São Paulo, 2002.

[†] Bolsista de iniciação científica Proict / Ulbra