Ensino de Estatística na Educação Básica
Inferência: Estimação (2)

Estimação de parâmetros que são funções de dois parâmetros

- Distribuição da diferença entre as médias de duas amostras grandes

- Se quisermos comparar as médias de duas populações, iremos tirar uma amostra de cada, e realizar inferência a partir da diferença entre as médias destas amostras (uma nova estatística, função de duas estatísticas).
- Suponhamos que as duas populações tenham médias $\mu 1$ e $\mu 2$ (desconhecidas), e a diferença entre estas médias seja dada pelo parâmetro Δ:

$$\Delta = \mu 1 - \mu 2$$

Teorema 4: *Distribuição amostral da diferença de médias D*

Se retiramos amostras aleatórias grandes, de tamanhos *n1* e *n2*, de populações de médias $\mu 1$ e $\mu 2$, a diferença *D* entre as médias destas amostras, $D = \bar{X}_1 - \bar{X}_1$, seguirá uma distribuição que tende para a normal, com parâmetros:

$$\mu_D = \Delta \text{ e } \sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

onde σ_1^2 e σ_2^2 são as variâncias das duas populações.

A diferença D poderá então ser padronizada como uma variável *z* de distribuição normal padrão, por meio de:

$$Z = \frac{D - \mu_D}{\sigma_D} = \frac{D - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

A partir da estatística acima chegaremos aos seguintes ICs:

a)amostras independentes de grandes dimensões (n$\geq$30), variáveis nas 2 populações com distribuições quaisquer:

Neste caso, o estimador é $\overline{X_1} - \overline{X_2}$, que tem erro-padrão igual a $\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$. Como as amostras são grandes, podemos usar S_1 e S_2 na estimativa do erro- padrão.

Logo, um IC com (1-α) x 100% de nível de confiança será calculado assim:

$$(\overline{x_1} - \overline{x_2}) - Z_{\alpha/2} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{x_1} - \overline{x_2}) + Z_{\alpha/2} \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Se admitirmos que as **variâncias das duas populações sejam iguais,** a expressão acima pode ser redefinida para:

$$(\overline{x}_1-\overline{x}_2)-Z_{\alpha/2}\times S\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq\mu_1-\mu_2\leq(\overline{x}_1-\overline{x}_2)+Z_{\alpha/2}\times S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

onde

$$S=\sqrt{\frac{(n_1-1)\times S_1^2+(n_2-1)\times S_2^2}{n_1+n_2-2}}$$

Veja que S corresponde à raiz quadrada da média ponderada das variâncias.

E se tivermos duas **amostras independentes pequenas?**

b) Amostras independentes de pequenas dimensões (n≤30), variáveis com distribuições normais nas populações e variâncias iguais.

Neste caso, um IC com (1-α) x 100% de nível de confiança será calculado assim:

$$(\overline{x}_1-\overline{x}_2)-T_{\alpha/2;n1+n2-2GL}\times S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\leq\mu_1-\mu_2\leq(\overline{x}_1-\overline{x}_2)+T_{\alpha/2;n1+n2-2GL}\times S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

- Distribuição da diferença entre proporções de sucessos em 2 amostras grandes

- Se quisermos comparar agora as *proporções de sucessos* de duas populações, iremos tirar uma amostra representativa de cada, e fazer inferência a partir da diferença entre as proporções de sucesso encontradas nestas amostras.

- Suponhamos que as duas populações tenham proporções de sucessos π_1 e π_2, respectivamente, e a diferença entre estas proporções seja denotada por: $\Delta = \pi_1 - \pi_2$

Teorema 5: *Distribuição amostral da diferença de proporções D em amostras grandes*

Se retirarmos amostras aleatórias grandes de duas populações, de tamanhos *n1* e *n2*, a diferença *(D)* entre as proporções de sucessos *p1* e *p2* encontradas nas amostras, isto é, *D = p1 – p2* seguirá uma distribuição que tende para a normal, com parâmetros:

$\mu_D = \Delta$ e $\sigma_D = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}}$

que poderá ser transformada em uma variável z de distribuição normal padrão, por meio de:

$$Z = \frac{D - \Delta}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

A partir da estatística acima chegamos ao IC desejado para um nível de confiança $(1 - \alpha)$:

$$(p_1 - p_2) - Z_{\alpha/2} \times EP_{p1-p2} \le \pi_1 - \pi_2 \le (p_1 - p_2) + Z_{\alpha/2} \times EP_{p1-p2}$$

Façamos agora uma atividade (no. 19)

Exercício:

Supor que você esteja interessado em uma eleição e que as pesquisas mais recentes indicam:

- seu candidato: 36% das intenções de voto
- o outro candidato seria o mais votado: 40% das intenções de voto

Se não há nenhum eleitor indeciso ou que votaria branco / nulo, qual a estimativa mais favorável para o seu candidato:

a) diferença: 4% ± 1% , $1 - \alpha = 0{,}99$
b) diferença: 4% ± 10% , $1 - \alpha = 0{,}99$
c) diferença: 4% ± 10% , $1 - \alpha = 0{,}80$

Justificar a resposta!

The most favorable confidence limit for my candidate would be c., 4% +/- 10%, 1-ALPHA=.80. This limit indicates a larger standard deviation (standard error) than the other two, since t(ALPHA=.20) is smaller than t(ALPHA=.05). Since my candidate is behind, the most favorable situation is that which indicates the greatest uncertainty that the difference is anything more than random sampling variation.

Façamos agora uma outra atividade (no. 20)