Capítulo 1 - Distribuições amostrais

Seção 1 - Introdução à amostragem

Seção 2 - Formas de amostragem e tamanho da amostra

Seção 3 - Distribuição amostral

Capítulo 2 - Estimação de parâmetros

Seção 1 - Estimador e estimativa

Seção 2 - Estimação por ponto

Seção 3 - Estimação por intervalos

Capítulo 3 - Teste de hipóteses

Seção 1 - Introdução aos testes de hipóteses

Seção 2 - Testes para a média populacional

Seção 3 - Testes para a variância e proporção populacional

Seção 4 - Testes para comparação de parâmetros

Nomenclatura

N = tamanho da população.

n = tamanho da amostra.

x_i = valores.

f_i = frequência simples.

μ = média da população.

$\overline{x}$ = média da amostra.

$\sigma^2(x)$ = variância dos valores da população.

$S^2(x)$ = variância dos valores da amostra.

σ = desvio padrão dos valores da população.

$s(x)$ = desvio padrão dos valores da amostra.

p = proporção de elementos da população.

$\hat{p}$ = proporção de elementos da população.

z = grau de confiança (também chamado de valor crítico).

e_o = “erro de origem” = o valor bruto de intervalo multiplicado pelo grau de confiança.

1.1 Introdução à amostragem

Estatística descritiva ou dedutiva;

Estatística inferencial ou indutiva; previsões e conclusões sobre uma população a partir dos resultados descritivos obtidos de uma amostra. Se baseia no cálculo de probabilidades para sustentar essas conjecturas ou inferências, as quais serão possíveis de testar e comprovar sob determinadas circunstâncias.

- Quais relações se podem achar entre as **variáveis** ou entre as **classificações** se comparadas?
- Qual a confiança que se pode ter nessas relações? **Ou seja, qual o grau de probabilidade para que elas não sejam fruto do acaso**?
- Qual a segurança com a qual se pode projetar a amostra sobre a população toda ou sobre outras populações semelhantes?

1.2 Formas de amostragem e tamanho da amostra

Diferentes formas de amostragem

- Aleatória simples
- Aleatória sistemática
- Aleatória estratificada
- Aleatória por conglomerados
- Não aleatória a esmo ou "aproximadamente aleatória"
- Não aleatória intencionada
- Não aleatória voluntária

Tamanho da amostra

Inicialmente apenas para amostras aleatórias simples.

Erro amostral: diferença entre o valor que a estatística acusa ou poderia acusar e o verdadeiro valor do parâmetro estimado.

Erro amostral tolerável: o quanto de erro é admissível por problemas previstos *na própria amostragem*.

Fórmulas de cálculo

- Tamanho da população desconhecido ou infinitamente grande:

$n_0 = \frac{1}{E^2} = E^{-2}$.

n_0: primeira aproximação do tamanho da amostra.

E: erro amostral tolerável da pesquisa (fórmula no livro mostra como E_0).

2.2 Estimação por ponto

Estimador da média populacional: média amostral ($\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$). $\lim_{n\to\infty} \sigma^2(\overline{x}) = 0$ (o limite da variância ser zero conforme tamanho da amostra tende ao infinito) expressa a consistência do estimador.

Estimador da variância populacional: "estatística S": $\frac{\sum_1^n (x_i-\mu)^2}{n}$, caso tenhamos a média populacional; $\frac{\sum_1^n (x_i-\overline{x})^2}{n}$, caso não tenhamos a média populacional; diferenças para as médias multiplicadas pelas frequências dos elementos, e somatórias divididas por $n-1$, caso a "frequência esteja envolvida": $\frac{\sum_1^n (x_i-\mu)^2 \cdot f_i}{n-1}$ e $\frac{\sum_1^n (x_i-\overline{x})^2}{n-1}$.

Estimador do desvio padrão populacional: $\sqrt{S^2}$ (estimador da variância populacional), com amostra grande (> 100); fator de correção para amostras menores que isto: $2 = 2.51;\ 3 = 1.69;\ 4 = 1.45;\ 6 = 1.26;\ 8 = 1.18;\ 10 = 1.14;$.

$12 = 1.11;\ 15 = 1.09;\ 20 = 1.07;\ 25 = 1.05;\ 50 = 1.03$

Estimador de proporção populacional: $p' = \frac{p(1-p)}{n}$. Novamente, $\lim_{n\to\infty} \sigma^2(p') = 0$ é expressão da consistência.

2.3 Estimação por intervalos

Confiança e precisão em sentidos opostos. O aumento da confiança (Z) aumenta o intervalo, o que diminui a precisão. (A amostra menor tem mais confiança e menos precisão.)

95% de confiança: $\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$. Nível de confiança $1-\alpha$.

$\alpha =$ monocaudal, $\frac{\alpha}{2} =$ bicaudal.

Sob a distribuição normal:

$0.99\ (\alpha = 0.01) \Rightarrow z_{\frac{\alpha}{2}} = \pm 2.58$ ($z_{a} = \pm 5.16$).

$0.95\ (\alpha = 0.05) \Rightarrow z_{\frac{\alpha}{2}} = \pm 1.96$ ($z_{a} = \pm 3.92$).

$0.90\ (\alpha = 0.1) \Rightarrow z_{\frac{\alpha}{2}} = \pm 1.65$ ($z_{a} = \pm 3.3$).

Média populacional para σ (desvio padrão populacional) conhecido: $\overline{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

Média populacional para σ desconhecido (independente do tamanho da amostra, pois com $n >= 30$ t-Student se aproxima da normal): $\overline{x} \pm t_{\alpha,n-1} \frac{s(x)}{\sqrt{n}}$. t é encontrado na tabela de t-Student para o α e $n-1$ (graus de liberdade). $s(x)$ é o desvio padrão amostral.

Proporção populacional: $\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

3.2 Testes para a média populacional

Bicaudal: o erro será distribuído em ambas as caudas e por isso trabalharemos com metade dele em cada uma.