UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

GABARITO - PROVA I — MATD44 — 17/06/2025

Professor: Raydonal Ospina
Regras: A prova é individual. Leia com atenção as perguntas. A prova deve ser claramente resolvida.

I ▶ **(Plano amostral geral)** Seja $\mathcal{U} = \{1, 2, 3\}$ uma população finita de tamanho $N = 3$ e $\mathbf{Y} = \{1, 2, 3\}$ o vetor da característica populacional renda bruta (mensal em salários mínimos) familiar. Suponha que o seguinte plano amostral é implementado $p(s_1) = p(\{1, 2\}) = \frac{1}{2}$, $p(s_2) = p(\{1, 3\}) = \frac{1}{4}$ e $p(s_3) = p(\{2, 3\}) = \frac{1}{4}$.

a) Determine as probabilidades de inclusão de primeira e segunda ordem. Determine se o plano amostral induzido pelo esquema de amostragem proposto é mensurável?

Res: 1a. Primeiro, vamos definir as probabilidades de inclusão de primeira ordem. A probabilidade de inclusão de primeira ordem π_i é a probabilidade de um elemento i ser incluído na amostra.

Para $i = 1$:

$$\pi_1 = P(\{1, 2\}) + P(\{1, 3\}) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

Para $i = 2$:

$$\pi_2 = P(\{1, 2\}) + P(\{2, 3\}) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

Para $i = 3$:

$$\pi_3 = P(\{1, 3\}) + P(\{2, 3\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

A probabilidade de inclusão de segunda ordem π_{ij} é a probabilidade de os elementos i e j serem incluídos simultaneamente na amostra.

Para π_{12}:

$$\pi_{12} = P(\{1, 2\}) = \frac{1}{2}$$

Para π_{13}:

$$\pi_{13} = P(\{1, 3\}) = \frac{1}{4}$$

Para π_{23}:

$$\pi_{23} = P(\{2, 3\}) = \frac{1}{4}$$

Um plano amostral é mensurável se todas as probabilidades de inclusão de segunda ordem forem maiores que zero e se $\pi_{ij} \leq \min(\pi_i, \pi_j)$. Neste caso, temos:

$$\pi_{12} = \frac{1}{2} \quad \text{e} \quad \min(\pi_1, \pi_2) = \frac{3}{4}$$
$$\pi_{13} = \frac{1}{4} \quad \text{e} \quad \min(\pi_1, \pi_3) = \frac{1}{2}$$
$$\pi_{23} = \frac{1}{4} \quad \text{e} \quad \min(\pi_2, \pi_3) = \frac{1}{2}$$

Como todas as probabilidades de inclusão de segunda ordem são menores ou iguais ao mínimo das probabilidades de inclusão de primeira ordem correspondentes, podemos concluir que o plano amostral é mensurável.

b) Forneça a distribuição de probabilidades do estimador de Horvitz-Thompson $\hat{t}_\pi$ para o total populacional da renda bruta (mensal em salários mínimos) familiar.

Res: 1b O estimador de Horvitz-Thompson para o total populacional da renda bruta (mensal em salários mínimos) familiar é dado por:

$$\hat{t}_\pi = \sum_{i \in s} \frac{y_i}{\pi_i}$$

Onde: - y_i é o valor da característica para a unidade i, - π_i é a probabilidade de inclusão da unidade i, - s é a amostra.

Dado o plano amostral e as probabilidades de inclusão calculadas anteriormente, vamos calcular $\hat{t}_\pi$ para cada amostra possível:

Para a amostra $s_1 = \{1, 2\}$:

$$\hat{t}_\pi(s_1) = \frac{y_1}{\pi_1} + \frac{y_2}{\pi_2} = \frac{2}{\frac{3}{4}} + \frac{4}{\frac{3}{4}} = \frac{2 \times 4}{3} + \frac{4 \times 4}{3} = \frac{8}{3} + \frac{16}{3} = 8$$

Para a amostra $s_2 = \{1, 3\}$:

$$\hat{t}_\pi(s_2) = \frac{y_1}{\pi_1} + \frac{y_3}{\pi_3} = \frac{2}{\frac{3}{4}} + \frac{6}{\frac{1}{2}} = \frac{8}{3} + 12 = \frac{8}{3} + \frac{36}{3} = \frac{44}{3} \approx 14.67$$

Para a amostra $s_3 = \{2, 3\}$:

$$\hat{t}_\pi(s_3) = \frac{y_2}{\pi_2} + \frac{y_3}{\pi_3} = \frac{4}{\frac{3}{4}} + \frac{6}{\frac{1}{2}} = \frac{16}{3} + 12 = \frac{16}{3} + \frac{36}{3} = \frac{52}{3} \approx 17.33$$

Agora, vamos determinar a distribuição de probabilidades do estimador $\hat{t}_\pi$:

$$P\left(\hat{t}_\pi = 8\right) = P(s_1) = \frac{1}{2}$$
$$P\left(\hat{t}_\pi = \frac{44}{3} \approx 14.67\right) = P(s_2) = \frac{1}{4}$$
$$P\left(\hat{t}_\pi = \frac{52}{3} \approx 17.33\right) = P(s_3) = \frac{1}{4}$$

A distribuição de probabilidades do estimador $\hat{t}_\pi$ é dada por:

$$\hat{t}_\pi = \begin{cases} 8, & \text{com probabilidade } \frac{1}{2}, \\ \frac{44}{3} \approx 14.67, & \text{com probabilidade } \frac{1}{4}, \\ \frac{52}{3} \approx 17.33, & \text{com probabilidade } \frac{1}{4}. \end{cases}$$

c) Determine a variância do estimador de Horvitz-Thompson $\hat{t}_\pi$ para o total populacional da renda bruta (mensal em salários mínimos) familiar sob este plano amostral.

Res: 1c Podemos dividir o cálculo da variância em partes menores para facilitar. Cálculo da Variância do Estimador de Horvitz-Thompson $\hat{t}_\pi$ A variância do estimador de Horvitz-Thompson é dada por:

$$\text{Var}(\hat{t}_\pi) = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}\left(\frac{y_i}{\pi_i}\right)\left(\frac{y_j}{\pi_j}\right)(\pi_{ij}-\pi_i\pi_j)$$

Passo 1: Calcular os termos diagonais

Primeiro, calculamos os termos em que $i = j$, ou seja, onde a soma envolve as mesmas unidades:

$$\text{Var}(\hat{t}_\pi)_{\text{diagonal}} = \sum_{i\in\mathcal{U}}\left(\frac{y_i}{\pi_i}\right)^2(1-\pi_i)$$

Os valores que temos são:
- $\pi_1 = \frac{3}{4}, \pi_2 = \frac{3}{4}, \pi_3 = \frac{1}{2}$
- $y_1 = 2, y_2 = 4, y_3 = 6$

Substituindo:

$$\text{Var}(\hat{t}_\pi)_{\text{diagonal}} = \left(\frac{2}{\frac{3}{4}}\right)^2\left(\frac{1}{4}\right)+\left(\frac{4}{\frac{3}{4}}\right)^2\left(\frac{1}{4}\right)+\left(\frac{6}{\frac{1}{2}}\right)^2\left(\frac{1}{2}\right)$$

Agora, vamos calcular esses três termos separadamente:

1. Para $i = 1$:

$$\left(\frac{2}{\frac{3}{4}}\right)^2\times\frac{1}{4} = \left(\frac{8}{3}\right)^2\times\frac{1}{4} = \frac{64}{9}\times\frac{1}{4} = \frac{16}{9}\approx 1.78$$

2. Para $i = 2$:

$$\left(\frac{4}{\frac{3}{4}}\right)^2\times\frac{1}{4} = \left(\frac{16}{3}\right)^2\times\frac{1}{4} = \frac{256}{9}\times\frac{1}{4} = \frac{64}{9}\approx 7.11$$

3. Para $i = 3$:

$$\left(\frac{6}{\frac{1}{2}}\right)^2\times\frac{1}{2} = 12^2\times\frac{1}{2} = 144\times\frac{1}{2} = 72$$

Somando esses termos:

$$\text{Var}(\hat{t}_\pi)_{\text{diagonal}} \approx 1.78+7.11+72 = 80.89$$

Este é o valor parcial para a variância considerando apenas os termos diagonais. Vamos agora calcular os termos cruzados da variância do estimador de Horvitz-Thompson, onde $i \neq j$.

$$\text{Var}(\hat{t}_\pi)_{\text{cruzados}} = \sum_{i\neq j}\left(\frac{y_i}{\pi_i}\right)\left(\frac{y_j}{\pi_j}\right)(\pi_{ij}-\pi_i\pi_j)$$

Aqui, precisamos calcular cada combinação de i e j onde $i \neq j$. Lembrando que as probabilidades de segunda ordem π_{ij} foram calculadas anteriormente: $\pi_{12} = \frac{1}{2}, \pi_{13} = \frac{1}{4}$ e $\pi_{23} = \frac{1}{4}$. O Termo para $i = 1$ e $j = 2$:

$$\left(\frac{y_1}{\pi_1}\right)\left(\frac{y_2}{\pi_2}\right)(\pi_{12}-\pi_1\pi_2) = \left(\frac{2}{\frac{3}{4}}\right)\left(\frac{4}{\frac{3}{4}}\right)\left(\frac{1}{2}-\frac{3}{4}\times\frac{3}{4}\right)$$

$$= \frac{8}{3}\times\frac{16}{3}\times\left(\frac{1}{2}-\frac{9}{16}\right) = \frac{128}{9}\times\left(\frac{8}{16}\right) = \frac{128}{9}\times\frac{1}{2} = \frac{64}{9}\approx 7.11$$

O termo para $i = 1$ e $j = 3$:

$$\left(\frac{y_1}{\pi_1}\right)\left(\frac{y_3}{\pi_3}\right)(\pi_{13}-\pi_1\pi_3) = \left(\frac{2}{\frac{3}{4}}\right)\left(\frac{6}{\frac{1}{2}}\right)\left(\frac{1}{4}-\frac{3}{4}\times\frac{1}{2}\right)$$

$$= \frac{8}{3}\times 12\times\left(\frac{1}{4}-\frac{3}{8}\right) = \frac{96}{1}\times\left(\frac{2}{8}-\frac{3}{8}\right) = \frac{96}{1}\times\left(-\frac{1}{8}\right) = -12$$

O termo para $i = 2$ e $j = 3$:

$$\left(\frac{y_2}{\pi_2}\right)\left(\frac{y_3}{\pi_3}\right)(\pi_{23}-\pi_2\pi_3) = \left(\frac{4}{\frac{3}{4}}\right)\left(\frac{6}{\frac{1}{2}}\right)\left(\frac{1}{4}-\frac{3}{4}\times\frac{1}{2}\right)$$

$$= \frac{16}{3}\times 12\times\left(\frac{1}{4}-\frac{3}{8}\right) = \frac{192}{1}\times\left(\frac{2}{8}-\frac{3}{8}\right) = \frac{192}{1}\times\left(-\frac{1}{8}\right) = -24$$

Agora, somamos os termos cruzados:

$$\text{Var}(\hat{t}_\pi)_{\text{cruzados}} = 7.11-12-24 = -28.89$$

Finalmente, somamos os termos diagonais e cruzados para obter a variância total do estimador de Horvitz-Thompson:

$$\text{Var}(\hat{t}_\pi) \approx 80.89-28.89 = 52$$

Este é o valor da variância do estimador de Horvitz-Thompson $\hat{t}_\pi$ para o total populacional da renda bruta sob o plano amostral fornecido.

2 ▶ (**Plano amostral AASs**) Uma amostra aleatória simples e sem substituição de 5 pessoas foi selecionada de uma população de 100 trabalhadores da empresa LInCaTech. Foram coletadas a informações sobre a Renda mensal em miles de reais (Renda) e o sexo do trabalhador. Com as informações da tabela 1 estime:

a) A renda média dos trabalhadores. Estabeleça um intervalo de 95% para a renda média.

b) A renda total dos trabalhadores. Estabeleça um intervalo de 95% para a renda total.

ID	Sexo	Renda
1	Fem	1
2	Mas	2
3	Fem	3
4	Fem	4
5	Mas	5

Tabela 1: Tabela de Informações dos empregados na amostra

Os dados iniciais são o tamanho da população, $N = 100$, o tamanho da amostra, $n = 5$, e os valores de renda (em milhares de reais) coletados na amostra: $y_s = \{1, 2, 3, 4, 5\}$. Calculamos as estatísticas amostrais. A média amostral da renda é: $\bar{y} = \frac{1}{n}\sum_{i\in s} y_i = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$ A variância amostral, que mede a dispersão dos dados na amostra, é calculada como: $s^2 = \frac{1}{n-1}\sum_{i\in s}(y_i-\bar{y})^2 =$

$\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{4} = \frac{4+1+0+1+4}{4} = \frac{10}{4} = 2.5$

Res: 2a Renda Média dos Trabalhadores

A estimativa pontual para a renda média populacional, μ, usando o estimador de Horvitz-Thompson sob um plano AASs, é simplesmente a média amostral. Portanto, a renda média estimada é de R$ 3.000,00.

Para construir o intervalo de confiança de 95%, precisamos da variância estimada do estimador da média, $\hat{V}(\hat{\mu})$. Esta incorpora o fator de correção para populações finitas, dado que a amostragem é sem reposição: $\hat{V}(\hat{\mu}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} = \left(1 - \frac{5}{100}\right)\frac{2.5}{5} = (0.95)(0.5) = 0.475$

O erro padrão (EP) é a raiz quadrada desta variância: $EP(\hat{\mu}) = \sqrt{0.475} \approx 0.6892$. Com um nível de confiança de 95%, o valor crítico da distribuição normal padrão é $z \approx 1.96$. O intervalo de confiança é, então: $IC(\mu) = \hat{\mu} \pm z \cdot EP(\hat{\mu}) = 3 \pm 1.96 \cdot 0.6892 \approx 3 \pm 1.3508$ Isso nos leva ao intervalo [1.6492, 4.3508].

Res: 2b Renda Total dos Trabalhadores

De forma análoga, a estimativa pontual para a renda total na população, τ, é obtida expandindo a média amostral para o tamanho da população: $\hat{\tau} = N \cdot \bar{y} = 100 \cdot 3 = 300$ A estimativa da renda total é de R$ 300.000,00.

A variância estimada do estimador do total, $\hat{V}(\hat{\tau})$, está diretamente relacionada à variância do estimador da média: $\hat{V}(\hat{\tau}) = N^2 \cdot \hat{V}(\hat{\mu}) = 100^2 \cdot 0.475 = 10000 \cdot 0.475 = 4750$

O erro padrão correspondente é $EP(\hat{\tau}) = \sqrt{4750} \approx 68.92$. O intervalo de 95% de confiança para o total é: $IC(\tau) = \hat{\tau} \pm z \cdot EP(\hat{\tau}) = 300 \pm 1.96 \cdot 68.92 \approx 300 \pm 135.083$ O intervalo resultante é [164.917, 435.083].

3

▶ **(Amostragem Bernoulli)** Seja s uma amostra obtida de um plano amostral de tipo Bernoulli com probabilidades de inclusão $\pi_k = \pi$ para todo $k \in U$ (população). Seja n_s o tamanho de amostra da amostra s. Mostre que a probabilidade condicional de se obter s dado n_s é a mesma que a probabilidade obtida por uma amostragem aleatória simples sem substituição de tamanho fixado n_s de N (Tamanho da população).

Res: 3 Considere um plano amostral Bernoulli com probabilidades de inclusão $\pi_k = \pi$ para todo $k \in U$. Se s é uma amostra de tamanho n_s, a probabilidade de observar essa amostra específica s é dada por:

$$P(s \mid n_s) = \pi^{n_s}(1-\pi)^{N-n_s}.$$

Onde π^{n_s} é a probabilidade de incluir exatamente essas n_s unidades na amostra, e $(1-\pi)^{N-n_s}$ é a probabilidade de não incluir as restantes $N - n_s$ unidades.

Por outro lado, na amostragem aleatória simples sem substituição (AASs), a probabilidade de obter uma amostra específica s de tamanho n_s é dada por:

$$P(s) = \frac{1}{\binom{N}{n_s}},$$

onde $\binom{N}{n_s}$ é o número total de combinações possíveis de tamanho n_s a partir da população de tamanho N. Para amostras de tamanho fixo n_s, a probabilidade de se obter uma amostra s é a mesma tanto para um plano amostral Bernoulli quanto para a amostragem aleatória simples sem substituição:

$$P(s \mid n_s) = \frac{1}{\binom{N}{n_s}} = P(s).$$

Isso ocorre porque, para amostras de tamanho fixo n_s, todas as amostras possíveis têm a mesma probabilidade de serem selecionadas, seja em um plano amostral Bernoulli ou em uma amostragem aleatória simples sem substituição.

4

▶ **(Amostragem Sistemática)** Suponha uma população de 7 elementos cujos valores para a característica de interesse sejam dados por $\mathbf{Y} = \{1, 3, 5, 7, 6, 4, 2\}$. Calcular a variância do estimador de Horvitz-Thompson para o total populacional em um plano amostral sistemático com $a = 2$ grupos. Para esse caso específico, o plano amostral sistemático é mais eficiente do que o plano amostral aleatório simples sem reposição? Explique.

Res: 4

Para a população finita de $N = 7$ num plano amostral sistemático com um intervalo de amostragem $k = 2$. Este procedimento gera $k = 2$ amostras potenciais. A primeira, partindo do elemento de índice 1, é $S_1 = \{y_1, y_3, y_5, y_7\} = \{1, 5, 6, 2\}$. A segunda, partindo do elemento de índice 2, é $S_2 = \{y_2, y_4, y_6\} = \{3, 7, 4\}$. A probabilidade de seleção de qualquer uma dessas amostras é $P(S_j) = 1/k = 1/2$. Como cada elemento da população pertence a exatamente uma dessas amostras, a probabilidade de inclusão de primeira ordem para qualquer elemento i é constante e igual a $\pi_i = 1/2$.

O estimador de Horvitz-Thompson para o total populacional, τ, é dado por $\hat{\tau}_{HT} = \sum_{i \in s} y_i/\pi_i$. O verdadeiro total populacional é $\tau = 1 + 3 + 5 + 7 + 6 + 4 + 2 = 28$. Calculando o valor do estimador para cada amostra possível, obtemos $\hat{\tau}_1 = (1 + 5 + 6 + 2)/(1/2) = 14 \cdot 2 = 28$ e $\hat{\tau}_2 = (3 + 7 + 4)/(1/2) = 14 \cdot 2 = 28$.

A variância do estimador é calculada como a variabilidade das estimativas em torno do verdadeiro parâmetro, ponderada pela probabilidade de cada amostra: $V(\hat{\tau}_{SYS}) = \sum_{j=1}^{k} P(S_j)(\hat{\tau}_j - \tau)^2 = \frac{1}{2}(28-28)^2 + \frac{1}{2}(28-28)^2 = 0$ A variância do estimador para este plano amostral é zero, indicando que, independentemente da amostra selecionada, a estimativa será perfeitamente acurada.

Para avaliar se este plano é mais eficiente que um plano amostral aleatório simples sem reposição (AASs), comparamos sua variância com a de um estimador AAS (sem reposição) para um tamanho de amostra equivalente. O tamanho esperado da amostra sistemática é $E[n] = (4 \cdot 1/2) + (3 \cdot 1/2) = 3.5$. A variância do estimador do total sob AASs é $V(\hat{\tau}_{AAS}) = N^2(1 - n/N)S^2/n$. A variância populacional é $S^2 = \frac{1}{N-1}\sum(y_i - \mu)^2 = \frac{28}{6} = 14/3$. Assim, $V(\hat{\tau}_{AAS}) = 7^2\left(1 - \frac{3.5}{7}\right)\frac{14/3}{3.5} = 49 \cdot (0.5) \cdot \frac{14}{10.5} \approx 32.67$ Como $V(\hat{\tau}_{SYS}) = 0 < V(\hat{\tau}_{AAS}) \approx 32.67$, concluímos que o plano sistemático é mais eficiente.

A análise de variância (ANOVA) nos permite decompor a variabilidade total da população em duas componentes: a variância entre as amostras sistemáticas e a variância dentro delas. Neste caso, como as estimativas de total para cada amostra foram idênticas ($\hat{\tau}_1 = \hat{\tau}_2$), a variância entre as amostras é nula, resultando em $V(\hat{\tau}_{SYS}) = 0$. Toda a variabilidade da população está, na verdade, contida dentro de cada amostra.

Assim, o grau de homogeneidade das amostras dada pela ordenação da população ($\{1, 3, 5, 7, 6, 4, 2\}$) exibe uma tendência periódica que o intervalo de amostragem $k = 2$ captura perfeitamente. Isso resulta em amostras que são internamente muito heterogêneas, pois cada uma contém valores de toda a faixa da população. Consequentemente, as amostras são extremamente homogêneas entre si. Isto indicaque o coeficiente de correlação intraclasse, ρ, seria fortemente negativo, indicando que elementos dentro da mesma amostra são mais diferentes entre si do que o esperado ao acaso, o que é o cenário ideal para a eficiência deste tipo de amostragem.

O efeito de desenho (deff), definido como a razão entre a variância do plano atual e a de um plano AASs de mesmo tamanho: deff $= \frac{V(\hat{\tau}_{SYS})}{V(\hat{\tau}_{AAS})} =$

$\frac{0}{32.67} = 0$ Um deff < 1 indica maior eficiência que a AASs. Um deff de zero representa a máxima eficiência teórica, significando que o desenho amostral eliminou completamente o erro de amostragem ao se alinhar perfeitamente com a estrutura latente na população.

5 ► **(Amostragem estratificada)** A Tabela 2 contem as informações do gasto mensal em serviços públicos de uma amostra aleatória estratificada de 120 famílias na cidade de Salvador a qual foi geograficamente dividida em três estratos: Norte, Centro e Sul.

	Estratos		
Estatísticas	Norte (1)	Centro (2)	Sul (3)
N_h	4000	6000	10000
W_h	0,3	0,2	0,5
n_h	40	36	44
$\bar{y}_h$	1,2	2,4	0,6
$\bar{Y}_h$	9600	7200	6000
s_h^2	0,36	1,21	0,04
$\text{Var}(\bar{y}_h)$	0,000993	0,004404	0,000226

Tabela 2: Informações do gasto familiar mensal em serviços públicos (em salários mínimos) a partir de uma amostra aleatória simples estratificada na cidade de Salvador.

Estime o gasto total de toda a população e estabeleça um intervalo de confiança de 95% para o total populacional.

O gasto médio da população (em salários mínimos), $\bar{y}$, é dado por:

$$\bar{y} = \sum_{h=1}^{H} W_h \bar{y}_h = 0,3 \times 1,2 + 0,2 \times 2,4 + 0,5 \times 0,6 = 1,14$$

O gasto total estimado da população (em salários mínimos), $\hat{Y}$, é:

$$\hat{Y} = N\bar{y} = (4000+6000+10000) \times 1,14 = 20.000 \times 1,14 = 22.800$$

O erro padrão da média é dado pela fórmula ajustada, incluindo o fator de correção finita:

$$\text{Var}(\bar{y}) = \sum_{h=1}^{H} W_h^2 \frac{s_h^2}{n_h} \left(\frac{N_h - n_h}{N_h} \right).$$

Substituindo os valores dos estratos:

$$\text{Var}(\bar{y}) = (0,3)^2 \times \frac{0,36}{40} \times \frac{4000-40}{4000} + (0,2)^2 \times \frac{1,21}{36} \times \frac{6000-36}{6000}$$

$$+ (0,5)^2 \times \frac{0,04}{44} \times \frac{10000-44}{10000}.$$

$$\text{Var}(\bar{y}) \approx 0,002944.$$

Logo,

$$\text{EP}(\bar{y}) = \sqrt{0,002944} = 0,0543.$$

Para construir o intervalo de confiança usando a distribuição t-Student, devemos calcular os graus de liberdade aproximados. Como o plano é estratificado, podemos usar a fórmula aproximada para graus de liberdade em amostragem estratificada:

$$df \approx \frac{\left(\sum_{h=1}^{H} W_h^2 \frac{s_h^2}{n_h} \right)^2}{\sum_{h=1}^{H} \left(\frac{W_h^4 s_h^4}{n_h^2 (n_h - 1)} \right)}.$$

Calculando este valor para os dados fornecidos, obtemos:

$$df \approx 78.$$

O valor crítico da distribuição t para $df = 78$ e um nível de confiança de 95% é aproximadamente $t_{0,025} \approx 1,99$. O intervalo de confiança de 95% para o gasto médio da população é dado por:

$$\bar{y} \pm t_{0,025} \times \text{EP}(\bar{y}) = 1,14 \pm 1,99 \times 0,0543 = [1,032, 1,248] \text{ salários mínimos.}$$

Para o gasto total da população, o intervalo de confiança de 95% é:

$$\hat{Y} \pm t_{0,025} \times \text{EP}(\hat{Y}) = 22.800 \pm 1,99 \times (20.000 \times 0,0543) = [21.724, 23.876]$$

Se usamos uma aproximação apenas normal, supondo que o n é suficientemente grande, temos que o intervalo de confiança de 95% para o gasto médio é:

$$\bar{y} \pm z_{0,025} \times \text{EP}(\bar{y}) = 1,14 \pm 1,96 \times 0,0543 = [1,033, 1,264]$$

Para o gasto total da população, o intervalo de confiança de 95% é:

$$\hat{Y} \pm z_{0,025} \times \text{EP}(\hat{Y}) = 22.800 \pm 1,96 \times (20.000 \times 0,0543) = [20.671, 24.928]$$

BOA PROVA