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

Professor: Raydonal Ospina

TD44 — GABARITO

I ▶ (**Plano amostral geral**) Seja $\mathcal{U} = \{1, 2, 3\}$ uma população finita de tamanho $N = 3$ e $\mathbf{Y} = \{2, 4, 6\}$ 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?

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.

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: 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.

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:

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

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 56 pessoas foi selecionada de uma população de 1000 trabalhadores da empresa LInCaTech. Foram coletadas as informações sobre a Renda mensal em miles de reais (Renda) e o sexo do trabalhador. Com as informações da tabela **??** 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.

Res: 2a Cálculo da renda média e intervalo de confiança de 95% para a renda média. O Cálculo da renda média amostral ($\bar{X}$) é

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

onde n é o número de observações e X_i é a renda de cada trabalhador. O desvio padrão amostral é calculado como:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

O intervalo de confiança de 95% para a média amostral é dado por:

$$\bar{X} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$$

onde $t_{\alpha/2,n-1}$ é o valor crítico da distribuição t de Student com $n-1$ graus de liberdade para um intervalo de confiança de 95

Vamos calcular esses valores passo a passo:
- Número de observações (n) = 56
- Rendas (X_i) fornecidas na tabela

ID	Sexo	Renda	ID	Sexo	Renda
1	Fem	2094.90	15	Mas	2939.21
2	Mas	2386.14	16	Mas	1722.62
3	Mas	1562.82	17	Fem	2739.79
4	Mas	1781.41	18	Mas	1821.61
5	Mas	1603.14	19	Mas	1742.40
6	Mas	479.73	20	Mas	1845.22
7	Mas	2196.85	21	Mas	1916.60
8	Mas	2365.21	22	Mas	1329.28
9	Mas	2016.32	23	Mas	2143.04
10	Mas	1322.23	24	Mas	2618.97
11	Fem	2589.08	25	Fem	1399.85
12	Fem	2896.28	26	Mas	1610.41
13	Mas	1370.55	27	Mas	2300.84
14	Mas	975.94	28	Mas	1192.90
29	Mas	2715.62	43	Mas	1802.16
30	Mas	2042.58	44	Mas	2444.68
31	Mas	2235.73	45	Mas	2644.75
32	Mas	2223.33	46	Fem	1431.53
33	Mas	2618.16	47	Mas	1094.02
34	Mas	2206.57	48	Mas	1548.96
35	Mas	2432.08	49	Mas	2410.51
36	Mas	1340.94	50	Mas	2286.37
37	Mas	2321.37	51	Mas	1589.07
38	Mas	1922.91	52	Fem	1646.21
39	Mas	2520.83	53	Mas	3358.63
40	Mas	2063.78	54	Mas	1369.42
41	Mas	2335.66	55	Mas	2047.63
42	Mas	2357.94	56	Mas	1719.34

Tabela 1: Tabela de Informações dos empregados na amostra

1. Renda média amostral ($\bar{X}$):

$$\bar{X} = \frac{1}{56}\sum_{i=1}^{56} X_i$$

Então,

$$\bar{X} = \frac{110,538.54}{56} = 1,970.67$$

2. Desvio padrão amostral (s): Calculamos a variância amostral e depois tiramos a raiz quadrada:

$$s^2 = \frac{1}{55}\left[(2094.90 - 1970.67)^2 + \cdots + (1719.34 - 1970.67)^2\right]$$

Portanto,

$$s = \sqrt{1,618,856.16} = 1,272.35$$

3. Intervalo de confiança de 95%: Utilizando o valor crítico da distribuição t com 55 graus de liberdade ($t_{0.025,55} \approx 2.004$):

$$\text{Margem de erro} = 2.004 \cdot \frac{1272.35}{\sqrt{56}} = 2.004 \cdot 170.84 = 341.44$$

O intervalo de confiança é:

$$\bar{X} \pm \text{Margem de erro} = 1970.67 \pm 341.44$$

$$\text{Intervalo de confiança de 95\%} = [1,629.23, 2,312.11]$$

Res: 2b Cálculo da renda total e intervalo de confiança de 95% para a renda total

Para calcular a renda total e seu intervalo de confiança, usamos: Renda total amostral ($\hat{T} = n \cdot \bar{X}$):

$$\hat{T} = 56 \cdot 1,970.67 = 110,538.54$$

Utilizamos a fórmula para o intervalo de confiança da renda total:

$$\hat{T} \pm t_{\alpha/2,n-1} \cdot \frac{s \cdot \sqrt{n}}{\sqrt{N-n}}$$

O erro padrão da renda total é calculado como:

$$\text{Erro padrão} = t_{\alpha/2,n-1} \cdot \frac{s \cdot \sqrt{56}}{\sqrt{1000-56}} = 564.46$$

O intervalo de confiança para a renda total é:

$$\hat{T} \pm \text{Erro padrão} = 110,538.54 \pm 564.46$$

$$\text{Intervalo de confiança de 95\% para a renda total} = [109,974.08, 111,102.99]$$

3

▶ (**Tamanho de amostra**) A variância da média amostral num plano amostral aleatório simples sem substituição (AASs) está dada por

$$\text{Var}_{\text{AASs}}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{S^2_{Y_U}}{n},$$

em que $S^2_{Y_U}$ é a variância populacional da variável de interesse, n o tamanho de amostra e N o tamanho populacional. Demonstre que o tamanho de amostra n que garante uma variância máxima, estabelecida de antemão, $\text{Var}_{\text{AASs}}(\bar{y})$ é

$$n = \frac{\frac{S^2_{Y_U}}{\text{Var}_{\text{AASs}}(\bar{y})}}{1 + \frac{S^2_{Y_U}}{\text{Var}_{\text{AASs}}}}.$$

Res: 4 A variância da média amostral em um plano amostral aleatório simples sem substituição (AASs) é dada por:

$$\mathrm{Var}_{\text{AASs}}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{S^2_{Y_U}}{n},$$

onde $S^2_{Y_U}$ é a variância populacional, n é o tamanho da amostra, e N é o tamanho populacional. Queremos encontrar o tamanho de amostra n que garante uma variância máxima estabelecida de antemão, denotada por $\sigma^2_{\bar{y}}$. Substituindo $\sigma^2_{\bar{y}}$ na fórmula, temos:

$$\sigma^2_{\bar{y}} = \left(1 - \frac{n}{N}\right)\frac{S^2_{Y_U}}{n}.$$

Multiplicando ambos os lados por n:

$$n\sigma^2_{\bar{y}} = \left(1 - \frac{n}{N}\right)S^2_{Y_U}.$$

Distribuindo $S^2_{Y_U}$ no lado direito:

$$n\sigma^2_{\bar{y}} = S^2_{Y_U} - \frac{nS^2_{Y_U}}{N}.$$

O que implica,

$$n\sigma^2_{\bar{y}} + \frac{nS^2_{Y_U}}{N} = S^2_{Y_U}.$$

logo, resolvendo para n:

$$n = \frac{S^2_{Y_U}}{\sigma^2_{\bar{y}} + \frac{S^2_{Y_U}}{N}}.$$

e simplificando:

$$n = \frac{\frac{S^2_{Y_U}}{\sigma^2_{\bar{y}}}}{1 + \frac{S^2_{Y_U}}{\sigma^2_{\bar{y}}N}}.$$

4 ▶ (**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: 4 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.

5 ▶ (**Amostragem Sistemática**) Suponha uma população de 9 elementos cujos valores para a característica de interesse sejam dados por $\mathbf{Y} = \{23, 20, 24, 31, 24, 29, 25, 33, 21\}$. Use a análise de variância (ANOVA) para 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: 5 Temos a população:

$$\mathbf{Y} = \{23, 20, 24, 31, 24, 29, 25, 33, 21\}$$

A média populacional é:

$$\bar{Y} = \frac{1}{9}\sum_{i=1}^{9} Y_i = \frac{23 + \cdots + 21}{9} = \frac{187}{9} = 20.77$$

A Variância da População

$$S^2 = \frac{1}{9-1}\sum_{i=1}^{9}(Y_i - \bar{Y})^2$$

$$S^2 = \frac{4.58 + 5.23 + \cdots + 0.64}{8} = 43.62$$

Plano Amostral Sistemático. Para um plano amostral sistemático com $a = 2$ grupos, a variância do estimador é:

$$\mathrm{Var}(\hat{T}_S) = \frac{N-n}{n}\left(\frac{S^2_B}{a} + \frac{S^2_W}{n}\right)$$

em que $N = 9$, $n = 2$ e $a = 2$. A Variância Entre os Grupos S_B^2 é

$$S_B^2 = \frac{1}{a-1}\left[(\bar{Y}_1 - \bar{Y})^2 + (\bar{Y}_2 - \bar{Y})^2\right]$$

Com médias dos grupos:
- Grupo 1: $\{23, 24, 24, 25, 21\}$ com média $\bar{Y}_1 = 23.4$
- Grupo 2: $\{20, 31, 29, 33\}$ com média $\bar{Y}_2 = 28.25$

$$S_B^2 = \frac{1}{2-1}\left[(23.4 - 20.77)^2 + (28.25 - 20.77)^2\right]$$

$$S_B^2 = \left[7.78^2 + 7.48^2\right] = 60.5$$

Por outro lado, a Variância Dentro dos Grupos S_W^2:

$$S_W^2 = \frac{1}{N-a}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2\right]$$

Para o Grupo 1 e Grupo 2, calculamos:
- Grupo 1: $S_{W1}^2 = 6.2$
- Grupo 2: $S_{W2}^2 = 40.7$

$$S_W^2 = \frac{5 \cdot 6.2 + 4 \cdot 40.7}{9-2} = \frac{31 + 162.8}{7} = 27.8$$

Assim, a Variância do Estimador Sistemático é

$$\text{Var}(\hat{T}_S) = \frac{9-2}{2}\left(\frac{60.5}{2} + \frac{27.8}{2}\right) = \frac{7}{2} \cdot 44.15 = 154.52$$

Para o plano Amostral Aleatório Simples Sem Reposição

$$\text{Var}(\hat{T}_S) = \frac{N-n}{N-1} \cdot \frac{S^2}{n}$$

$$\text{Var}(\hat{T}_S) = \frac{9-2}{9-1} \cdot \frac{43.62}{2} = \frac{7}{8} \cdot 21.81 = 19.21$$

Quando comparamos as variâncias:
- Plano Sistemático: 154.52
- Plano Aleatório Simples: 19.21

Portanto, a variância do estimador de Horvitz-Thompson no plano amostral sistemático é significativamente maior do que no plano amostral aleatório simples sem reposição. Assim, o plano amostral aleatório simples sem reposição é mais eficiente neste caso específico.