Amostra Aleatória Simples

Airlane Pereira Alencar

12 de abril de 2023

Notações

População ou Universo $U = \{1, 2, \ldots, N\}$
Elemento Populacional $i \in U$
Característica de interesse $Y_i, i \in U$
Parâmetro populacional $D = (Y_1, \ldots, Y_N)$
Função paramétrica populacional $\theta(D)$

Exemplos de θ

$$\mu = \frac{Y_1 + \ldots + Y_N}{N} \quad (1)$$

$$\sigma^2 = \frac{1}{N} \sum (Y_i - \mu)^2; \quad (2)$$

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y) = E(XY) - \mu_X \mu_Y \quad (3)$$

Notações

População ou Universo $U = \{1, 2, \ldots, N\}$
Elemento Populacional $i \in U$
Característica de interesse $Y_i, i \in U$
Parâmetro populacional $D = (Y_1, \ldots, Y_N)$
Função paramétrica populacional $\theta(D)$

Exemplos de θ

$$\mu = \frac{Y_1 + \ldots + Y_N}{N} \tag{1}$$

$$\sigma^2 = \frac{1}{N}\sum(Y_i - \mu)^2; \tag{2}$$

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y) = E(XY) - \mu_X\mu_Y \tag{3}$$

Exemplos de θ

$$
\begin{aligned}
\mu &= \frac{Y_1 + \ldots + Y_N}{N} \\
\sigma^2 &= \frac{1}{N}\sum (Y_i - \mu)^2; \\
Cov(X,Y) &= E(X-\mu_X)(Y-\mu_Y) = E(XY) - \mu_X\mu_Y \\
Corr(X,Y) &= \frac{Cov(X,Y)}{DP(X)DP(Y)} \\
\tau_X &= X_1 + \ldots + X_N \text{ Total populacional} \\
\theta &= \frac{\mu_Y}{\mu_X} = \frac{\tau_Y}{\tau_X} \text{ Razão populacional}
\end{aligned}
$$

Uma sequência qualquer de n unidades de U é denominada **amostra ordenada** de U, isto é, $s = (k_1, \ldots, k_n)$ tal que $k_i \in U$.
k_i = i-ésimo componente de s.
ex: $U = \{1, 2, 3\}$ $s_1 = (1, 2), s_2 = (2, 2, 1, 3)$

Vamos tratar só de **planos amostrais probabilísticos**, ou seja, aqueles que permitem associar a cada amostra uma probabilidade conhecida desta ser sorteada.

Planejamento amostral ordenado

Def:

Uma função $P(s)$ definida em S(U) (conjunto de todas as amostras), satisfazendo $P(s) \geq 0, \forall s \in S(U)$, tal que $\sum_{s \in S(U)} P(s) = 1$ é um **Planejamento amostral ordenado.**

Exemplos $U = \{1, 2, 3\}$

$S(U) = \{(1), (2), (3), (1, 1), (1, 2), \ldots, (2, 2, 1, 3, 2)\}$
$S(U)$ = todas as N^n amostras de tamanho n com reposição
$S^*(U)$ = todas as $N(N-1)\ldots(N-n+1)$ amostras de tamanho n sem reposição
Para $n = 2$: $S(U) = \{(11), (12), (13), (21), (22), (23), (31), (32), (33)\}$ e $S^*(U) = \{(12), (13), (21), (23), (31), (32)\}$;
$\binom{N}{n}$ conta quantas amostras mas divide por $n!$ para tirar as de ordem diferente.

Planejamento amostral ordenado

Def:

Uma função $P(s)$ definida em S(U) (conjunto de todas as amostras), satisfazendo $P(s) \geq 0, \forall s \in S(U)$, tal que $\sum_{s \in S(U)} P(s) = 1$ é um **Planejamento amostral ordenado.**

Exemplos $U = \{1, 2, 3\}$

$S(U) = \{(1), (2), (3), (1, 1), (1, 2), \ldots, (2, 2, 1, 3, 2)\}$
$S(U)$ = todas as N^n amostras de tamanho n com reposição
$S^*(U)$ = todas as $N(N-1)\ldots(N-n+1)$ amostras de tamanho n sem reposição
Para $n = 2$: $S(U) = \{(11), (12), (13), (21), (22), (23), (31), (32), (33)\}$ e $S^*(U) = \{(12), (13), (21), (23), (31), (32)\}$;
$\binom{N}{n}$ conta quantas amostras mas divide por $n!$ para tirar as de ordem diferente.

- Plano A: Amostra Aleatória Simples com reposição (AASc) - n=2

$$
\begin{aligned}
P(11) &= P(12) = P(13) = P(21) = P(22) = P(23) = \\
&= P(31) = P(32) = P(33) = \frac{1}{9}
\end{aligned}
$$

- Plano B: Amostra Aleatória Simples sem reposição (AASs) - n=2

$$
P(12) = P(13) = P(21) = P(23) = P(31) = P(32) = \frac{1}{6}
$$

- Plano C: Amostra Aleatória até observar o número 2 até n=3.

$$
\begin{aligned}
P(2) &= 1/3 \\
P(12) &= P(32) = 1/9 \\
P(112) &= P(132) = P(312) = P(332) = 1/27 \\
P(111) &= P(113) = P(131) = P(311) = 1/27 \\
P(133) &= P(313) = P(331) = P(333) = 1/27
\end{aligned}
$$

- Plano D: Outro planejamento

$$\begin{array}{lclllcl} P(12) & = & 1/10 & & P(21) & = & 1/6 \\ P(13) & = & 1/15 & & P(31) & = & 1/12 \\ P(23) & = & 1/3 & & P(32) & = & 1/4 \end{array}$$

- Exemplo 2.10 (BB): $U = \{1, 2, 3\}$

Variável	Valores			Notação
Unidade	1	2	3	i
Renda	12	30	18	F_i
Número de pessoas trabalhando	1	3	2	T_i

Sorteia-se 2 elementos sem reposição com probabilidade proporcional ao número de trabalhadores.

A distribuição amostral de uma estatística $h(d_s)$, segundo um plano amostral A, é a dist. de prob. $H(d_s)$ definida sobre S_A com função densidade de probabilidade

$$p_h = P_A(s \in S_A | H(d_s) = h) = P(h).$$

O valor esperado de H será

$$E_A(H) = \sum h p_h = \sum_{s \in S_A} h(d_s) P_A(s \in S_A | H(d_s) = h).$$

A variância de H é

$$Var_A(H) = \sum_{s \in S_A} [H(d_s) - E_A(H)]^2 P_A(s).$$

A covariância entre H e G é
$Cov_A(H, G) = \sum_s [h(d_s) - E_a(H)][g(d_s) - E_a(G)] P_A(s).$

Variável de interesse: Renda (F)
$U = \{1, 2, 3\}$ com rendas $F = (12, 30, 18)$
$\mu = E(F) = \frac{12+30+18}{3} = 20$ é a renda média populacional
$\sigma^2 = Var(F) = \frac{(12-20)^2+(30-20)^2+(18-20)^2}{3} = 56$ variância pop.

Plano A: Todas as amostras com n=2 com reposição (importa a ordem, temos $N * N = 9$ amostras)

s	11	12	13	21	22	23	31	32	33
prob	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9
$\bar{f}$	12	21	15	21	30	24	15	24	18

$$E(\bar{F}) = \sum_s \bar{f} P(s) = 12\frac{1}{9} + \ldots = 20$$

$$Var(\bar{F}) = E(\bar{F} - 20)^2 = (12 - 20)^2\frac{1}{9} + \ldots = \frac{252}{9} = 28 = \left(\frac{\sigma^2}{n}\right)$$

$U = \{1, 2, 3\}$ com rendas $F = (12, 30, 18)$
Plano B: Todas as amostras com n=2 **sem** reposição (importa a ordem, temos $N(N-1) = 6$ amostras)

s	12	13	21	23	31	32
prob	1/6	1/6	1/6	1/6	1/6	1/6
$\bar{f}$	21	15	21	24	15	24

$$
\begin{aligned}
E(\overline{F}) &= \sum_s \bar{f} P(s) = \frac{21+15+24}{3} = 20 \\
Var(\overline{F}) &= E(\overline{F} - 20)^2 \\
&= \frac{(21-20)^2 + (15-20)^2 + (24-20)^2}{3} = \frac{42}{3} = 14 \\
Var_{AAS_s}(\overline{F}) &= (1-f)\frac{S^2}{n} \\
&= \left(1 - \frac{n}{N}\right)\frac{1}{n}\frac{\sum(F_i - \mu)^2}{N-1} = \left(1 - \frac{2}{3}\right)\frac{1}{2}\frac{168}{2} = 14
\end{aligned}
$$

$U = \{1, 2, 3\}$ com $D = \begin{pmatrix} F_i \\ T_i \end{pmatrix} = \begin{pmatrix} 12 & 30 & 18 \\ 1 & 3 & 2 \end{pmatrix}$
sendo $r = h(d_S)$ a razão entre o total de renda e o número de trabalhadores $E(R) = \frac{12+30+18}{1+3+2} = \frac{60}{6} = 10$
Plano A: AAS_c

s	11	12	13	21	22	23	31	32	33
prob	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9
r	12	10.5	10	10.5	10	9.6	10	9.6	9

ex: $r(12) = \frac{12+30}{1+3} = \frac{42}{4} = 10.5$
A distribuição amostral de R é

r	9	9.6	10	10.5	12
prob	1/9	2/9	3/9	2/9	1/9

Obtenha $E(R)$ e $Var(R)$.

Plano B: AAS_s

s	12	13	21	23	31	32
prob	1/6	1/6	1/6	1/6	1/6	1/6
$\overline{F}$	10.5	10	10.5	9.6	10	9.6

$\overline{F}$	9.6	10	10.5
prob	1/3	1/3	1/3

$$
\begin{aligned}
E(\overline{F}) &= \sum_s \bar{f} P(s) = \frac{9.6+10+10.5}{3} = \frac{30.1}{3} = 10.0\bar{3} \\
Var(\overline{F}) &= E(\overline{F} - 10.0\bar{3})^2 \\
&= \frac{(9.6-10.0\bar{3})^2 + (10-10.0\bar{3})^2 + (10.5-10.0\bar{3})^2}{3} = \\
&= \frac{0.40\bar{6}}{3} = 0.13\bar{5}
\end{aligned}
$$

Exemplo de totais em Levy e Lemeshow p. 25

Parâmetros θ

$$\begin{aligned}
\mu &= \frac{Y_1 + \ldots + Y_N}{N} \\
\tau &= Y_1 + \ldots + Y_N \\
\sigma^2 &= \frac{1}{N}\sum (Y_i - \mu)^2; \\
S^2 &= \frac{1}{N-1}\sum (Y_i - \mu)^2; \\
Cov(X,Y) &= E(X-\mu_X)(Y-\mu_Y) = E(XY) - \mu_X\mu_Y \\
\theta &= \frac{\mu_Y}{\mu_X} = \frac{\tau_Y}{\tau_X} \text{ Razão populacional}
\end{aligned}$$

Notem que há duas variâncias populacionais σ^2 e S^2.

Exemplo 2.2 Lohrs

$U = \{1, 2, 3, 4, 5, 6, 7, 8\}$
$y = (1, 2, 4, 4, 7, 7, 7, 8)$
Considere todas as amostras de tamanho 4 sem reposição. Lohrs considera amostra (1,2) igual a amostra (2,1).
AASs em R

AASc: Teorema 3.1 BB

A variável número de vezes que o elemento i aparece na amostra $f_i \sim Binomial(n, 1/N)$

$$
\begin{aligned}
E(f_i) &= \frac{n}{N} \\
Var(f_i) &= \frac{n}{N}\left(1 - \frac{1}{N}\right) \\
\pi_i &= P(f_i \neq 0) = 1 - \left(1 - \frac{1}{N}\right)^n \\
\pi_{ij} &= P(f_i \neq 0 \cap f_j \neq 0) = 1 - 2\left(1 - \frac{1}{N}\right)^n + \left(1 - \frac{2}{N}\right)^n \\
Cov(f_i, f_j) &= -\frac{n}{N^2}
\end{aligned}
$$

$(f_1, \ldots, f_N)$ Multinomial$(n; 1/N, \ldots, 1/N)$.
Estudar a distribuição multinomial.

AASc: Propriedades

A estatística total amostral $t = \sum_{i \in s} Y_i$ tem

$$\begin{aligned} E(t) &= n\mu \\ Var(t) &= n\sigma^2 \end{aligned}$$

$(f_1, \ldots, f_N)$ Multinomial$(n; 1/N, \ldots, 1/N)$.

$$\begin{aligned} t &= \sum_{i \in s} Y_i = \sum_{i=1}^{N} f_i Y_i \\ E(t) &= E(f_i) \sum_{i=1}^{N} Y_i = \frac{n}{N} N\mu = n\mu \\ Var(t) &= Var\left(\sum_{i=1}^{N} Y_i f_i\right) = \sum_{i=1}^{N} Y_i^2 Var(f_i) + \sum_{i \neq j} Y_i Y_j Cov(f_i, f_j) \end{aligned}$$

Para planos simétricos com n fixo (AASc é caso particular):

$$
\begin{aligned}
Var(n) &= Var\left(\sum_{i=1}^{N} f_i\right) = \sum_{i=1}^{N} Var(f_i) + \sum_{i\neq j} Cov(f_i, f_j) \\
0 &= NVar(f) + N(N-1)Cov(f, f') \Longrightarrow \\
Cov(f, f') &= -\frac{Var(f)}{N-1}
\end{aligned}
$$

$$
\begin{aligned}
Var(t) &= Var(f)\sum_{i=1}^{N} Y_i^2 + \sum_{i\neq j} Y_i Y_j Cov(f_i, f_j) \\
&= (\sum_{i=1}^{N} Y_i^2) Var(f_i) + (\sum_{i\neq j} Y_i Y_j)\frac{-Var(f)}{N-1} \\
&= Var(f)\left[(\sum_{i=1}^{N} Y_i^2) - \frac{1}{N-1}(\sum_{i\neq j} Y_i Y_j)\right]
\end{aligned}
$$

$$\begin{aligned}
\sum_{i \neq j} Y_i Y_j &= -\sum Y_i^2 + N^2 \mu^2 \\
\sum_{i \neq j} Y_i Y_j + \sum Y_i^2 &= +N^2 \mu^2 \\
\left(\sum_i Y_i \right)^2 &= +N^2 \mu^2
\end{aligned}$$

Voltando

$$\begin{aligned}
Var(t) &= Var(f) \left[\left(\sum_{i=1}^{N} Y_i^2 \right) - \frac{1}{N-1} \left(\sum_{i \neq j} Y_i Y_j \right) \right] \\
&= Var(f) \left[\left(\sum_{i=1}^{N} Y_i^2 \right) - \frac{1}{N-1} \left(-\sum Y_i^2 + N^2 \mu^2 \right) \right]
\end{aligned}$$

$$
\begin{aligned}
Var(t) &= Var(f)\left[\frac{N-1}{N-1}\sum_{i=1}^{N} Y_i^2 + \frac{\sum Y_i^2}{N-1} - \frac{N^2\mu^2}{N-1}\right] \\
&= Var(f)\frac{N}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\mu^2\right] \\
&= Var(f)\frac{N}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - 2\mu(\sum Y_i) + N\mu^2\right] \\
&= Var(f)\frac{N}{N-1}\left[\sum_{i=1}^{N}(Y_i^2 - 2Y_i\mu + \mu^2)\right] \\
&= Var(f)\frac{N}{N-1}\left[\sum_{i=1}^{N}(Y_i-\mu)^2\right] = n\frac{1}{N}\frac{N-1}{N}\frac{N}{N-1}\sum_{i=1}^{N}(Y_i-\mu)^2 \\
&= n\sigma^2, \quad \sigma^2 = \frac{\sum_{i=1}^{N}(Y_i-\mu)^2}{N}
\end{aligned}
$$

AASc: Propriedades

Total amostral $t(s) = \sum_{i \in s} Y_i$:

$$E(t) = n\mu \quad Var(t) = n\sigma^2$$

Média amostral $\bar{y} = \frac{\sum_{i \in s} y_i}{n} = \frac{t}{n}$:

$$E(\bar{y}) = \mu, \quad Var(\bar{y}) = \frac{\sigma^2}{n}$$

Estimador do Total $T = \hat{\tau} = N\bar{y}$:

$$E(T) = N\mu = \sum_{i=1}^{N} Y_i, \quad Var(T) = N^2 \frac{\sigma^2}{n}$$

Vamos estimar σ^2.

AASc: Estimador da variância

$s^2 = \frac{1}{n-1}\sum_{i\in s}(Y_i - \bar{y})^2$ é estimador não viesado de σ^2.

$$(n-1)s^2 = \sum_{i\in s}(Y_i - \bar{y})^2 = \sum_{i\in s} Y_i^2 - n\bar{y}^2 = s_q^2 - \frac{t^2}{n}$$

Usando diversos resultados do cap. 2 (BB),

$$\begin{aligned} E(s_q^2) &= E(\sum_{i\in s} Y_i^2) = N(\sigma^2+\mu^2)E(f) = N(\sigma^2+\mu^2)\frac{n}{N} = n\sigma^2 + n\mu^2 \\ E(t^2) &= Var(t) + E(t)^2 = n\sigma^2 + n^2\mu^2 \end{aligned}$$

$$E((n-1)s^2) = = n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2$$

Variável de interesse: Renda

$U = \{1, 2, 3\}$ com rendas $Y = (12, 30, 18)$

$\mu_Y = \frac{12+30+18}{3} = 20$ e $\sigma^2_Y = \frac{(12-20)^2+(30-20)^2+(18-20)^2}{3} = 56$

AASc s	11	12	13	21	22	23	31	32	33
Ys	12	12	12	30	30	30	18	18	18
	12	30	18	12	30	18	12	30	18
$\bar{y}$	12	21	15	21	30	24	15	24	18
$(Y - \bar{Y})^2$	0	81	9	81	0	36	9	36	0
	0	81	9	81	0	36	9	36	0
s^2	0	162	18	162	0	72	18	72	0

$$E(\overline{Y}) = \sum_s \overline{Y} P(s) = 12\frac{1}{9} + \ldots = 20$$

$$Var(\overline{Y}) = E(\overline{Y} - 20)^2 = (12 - 20)^2 \frac{1}{9} + \ldots = \frac{252}{9} = 28 = \left(\frac{\sigma^2}{n}\right)$$

$$E(s^2) = \frac{504}{9} = 56 = \sigma^2$$

AASc: IC e tamanho da amostra

Como $s^2 = \frac{1}{n-1}\sum_{i \in s}(Y_i - \bar{y})^2$ é estimador não viesado de σ^2, então $\widehat{Var}(\bar{y}) = \frac{s^2}{n}$ é estimador não viesado de $Var(\bar{y})$;
Se Y *Normal* e estimando σ^2, usaremos

$$\frac{\bar{Y} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t_{n-1} \quad IC = \left[\bar{y} \mp t\sqrt{\frac{s^2}{n}}\right]$$

Se n for grande, podemos usar a aproximação para a dist. normal.
Quando estimamos σ^2, o erro amostral é estimado por $e = t\sqrt{\frac{s^2}{n}}$, mas para calcular tamanho de amostra, o erro amostral é $e = z\sqrt{\frac{\sigma^2}{n}}$ e

$$n = \frac{z^2\sigma^2}{e^2}. \tag{4}$$

AASc: ex3.2

Suponha que uma AASc com n=10, obteve média amostral igual a 18 e variância amostral igual a 48. Obtenha o intervalo de confiança para a média populacional com coeficiente de confiança igual a 95%. Calcule também o tamanho da amostra para ter erro amostral igual a 1.5.

Como n=10, vamos pegar os quantis da distribuição t_9 e $t = 2,262$ e

$$IC = \left[18 \mp 2.262\sqrt{\frac{48}{10}}\right] = [18 \mp 4,956] = [13,044; 22,956].$$

Se o erro amostral desejado é e=1,5 e assumimos que $\sigma^2 = 48$, temos

$$n = \frac{z^2\sigma^2}{e^2} = \frac{2^2 48}{1,5^2} = 85,3. \tag{5}$$

Para n=86, temos t=1,99 e esse tamanho n é suficiente.

AASc: Proporções

Seja Y_i igual a 1 se ocorre uma característica e 0 caso contrário.
A proporção populacional é

$$P = \frac{1}{N}\sum_{i=1}^{N} Y_i = \mu$$

Usando o teorema que vimos para $\bar{y}$, temos que

$$p = \hat{P} = \bar{y} = \frac{1}{n}\sum_{i\in s} Y_i \qquad (6)$$

$$Var(\hat{P}) = \frac{P(1-P)}{n} = \frac{\sigma^2}{n} \qquad (7)$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - P)^2 = P(1-P)$$

AASc: Proporções

A proporção amostral $\hat{P} = \bar{y} = \frac{1}{N}\sum_{i \in s} Y_i$ é estimador não viesado para P e com

$$Var(\hat{P}) = \frac{P(1-P)}{n} = \frac{\sigma^2}{n}$$

E um estimador não viesado dessa variância é

$$\widehat{Var}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{n-1}$$

A distribuição **assintótica** de $\hat{P}$ é normal (TCL), então para n grande:

$$IC = \left[\hat{P} \mp z\sqrt{\frac{\hat{P}(1-\hat{P})}{n-1}}\right]$$

AASc: Proporções

Usando a normalidade assintótica e para P conhecido, o erro amostral é $e = z\sqrt{\frac{P(1-P)}{n}}$ e

$$n = \frac{z^2 P(1-P)}{e^2}.$$

Se não souber o valor de P, pode usar $P = 1/2$ (por quê?)

Agora os sorteios são sem reposição, então cada elemento i $(1 = 1, \ldots, N)$ só pode aparecer uma vez na amostra de tamanho n.
$f_i = 1$, se o elemento i está na amostra ou 0 caso contrário.

$$P(f_i = 1) \quad = \quad \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{\frac{(N-1)!}{(n-1)!(N-n)!}}{\frac{N.(N-1)!}{n(n-1)!(N-n)!}} = \frac{n}{N}$$

AASs: Teorema 3.7 BB

A variável número de vezes que o elemento i aparece na amostra $f_i \sim Bernoulli(n/N) = Binomial(1, n/N)$:

$$
\begin{aligned}
E(f_i) &= \frac{n}{N} \\
Var(f_i) &= \frac{n}{N}\left(1 - \frac{n}{N}\right) \\
\pi_i &= P(f_i = 1) = \frac{n}{N} \\
\pi_{ij} &= P(f_i = 1 \cap f_j = 1) = P(f_i = 1)P(f_j = 1 | f_i = 1) = \frac{n}{N}\frac{n-1}{N-1} \\
E(f_i f_j) &= P(f_1 = 1 \cap f_j = 1) = \pi_{ij} \\
Cov(f_i, f_j) &= E(f_i f_j) - E(f_1)E(f_j) = -\frac{n}{N^2}\frac{N-n}{N-1}
\end{aligned}
$$

AASs: Propriedades

A estatística total $t = \sum_{i \in s} Y_i$ tem

$$\begin{aligned} E(t) &= n\mu \\ Var(t) &= n(1-f)S^2, \end{aligned}$$

com $f = n/N$ sendo a fração amostral e o parâmetro $S^2 = \frac{\sum_{i=1}^{N}(Y_i-\mu)^2}{N-1}$.

No slide 20 obtivemos:

$$Var(t) = Var(f)NS^2 = \frac{n}{N}\left(1-\frac{n}{N}\right)NS^2 = n(1-f)S^2$$

Então $\bar{y}$ é estimador não viesado da média pop. μ e

$$Var(\bar{y}) = (1-f)\frac{S^2}{n}.$$

A variância amostral s^2 é estimador não viesado de S^2.

Exemplo 3.4
$U = \{1, 2, 3\}$, com os dados $Y = (12, 30, 18)$, $\mu = 20$, $S^2 = \frac{168}{2} = 84$ e n=2.

s	12	21	13	31	23	32
prob	1/6	1/6	1/6	1/6	1/6	1/6
$\bar{y}$	21	21	15	15	24	24
s^2	162	162	18	18	72	72

$\bar{y}$	15	21	24
s^2	18	72	162
prob	1/3	1/3	1/3

$E(\bar{y}) = 20$, $Var(\bar{y}) = \left(1 - \frac{2}{3}\right) \frac{84}{2} = 14$
$E(S^2) = \frac{18+72+162}{3} = \frac{252}{3} = 84.$

AASs: Tamanho amostral

$$
\begin{aligned}
IC &= \bar{y} \mp t\sqrt{\frac{(1-f)s^2}{n}} \\
Var(\bar{y}) &= (1-f)\frac{S^2}{n} = \frac{S^2}{n/(1-f)} = \frac{S^2}{n'}
\end{aligned}
$$

De modo semelhante ao obtido para AASc:

$$
n' = \frac{z^2 S^2}{e^2}
$$

Como $n' = \frac{n}{1-f} = \frac{n}{1-n/N} = \frac{nN}{N-n} \Longrightarrow n'(N-n) = nN \Longrightarrow n'N - n'n = nN \Longrightarrow n(N+n') = n'N$, então

$$
n = \frac{n'N}{N+n'} = \frac{Nz^2 S^2}{Ne^2 + z^2 S^2}
$$

Exemplo 3.5. p. 78
Uma pesquisa com mil funcionários quer estimar o número médio de faltas dos funcionários nos últimos 6 meses em uma empresa com 36 mil (N) funcionários.

Faltas	0	1	2	3	4	5	6	7	8
Func.	451	162	187	112	49	21	5	11	2
$(y_i - \bar{y})^2$	1,68	0,09	0,5	2,9	7,31	13,7	22,1	32,5	44,9

A média amostral é de 1,296 falta por funcionário e $s^2 = 2,397$. O valor t é 1,96.

O erro amostral é $1,96\sqrt{(1 - \frac{1000}{36000})s^2/1000} = 0,095$.

O IC para μ é $[1,201; 1,391]$.

Como fica o tamanho da amostra se quero erro amostral 0,05?

Referências

- Bolfarine e Bussab. Elementos de amostragem. Ed. Edgar-Blucher.
- Lohrs. Sampling.