Aula de Exercícios

Prof. Caio Azevedo

Exercício 3.8

- Uma amostra AASs de tamanho $n = 4 = n_1 + n_2$ é retirada de uma população $\mathcal{U}$ com $N = 6$ elementos, onde $D = (8, 2, 2, 11, 4, 7)$. Uma amostra aleatória simples sem reposição de tamanho $n1 = 2$ é retirada da primeira amostra, apresentando média $\overline{y}_1$. Seja $\overline{y}_2$ a média das n_2 unidades remanescentes na amostra original. Encontre $Var(\overline{Y}_1 - \overline{Y}_2)$ e $Var[\overline{Y}_1 - \overline{Y}]$, onde $\overline{y}$ é a média da amostra original.

Exercício 3.8

- A ideia para a construção das distribuições de $\overline{Y}_1$, $\overline{Y}_2$ e $\overline{Y}$ é:
  1. Retira-se uma amostra sem reposição de tamanho 4, denominada $\mathcal{S}$.
  2. Retira-se uma amostra sem reposição de tamanho 2 de $\mathcal{S}$, denominada $\mathcal{S}_1$, e com o restante das observações forma-se a amostra restante $\mathcal{S}_2$.
  3. Calcula-se $\overline{y}$ a partir de $\mathcal{S}$, $\overline{y}_1$ a partir de $\mathcal{S}_1$ e $\overline{y}_2$ a partir de $\mathcal{S}_2$.
- Para o passo 1, temos $6!/2 = 360$ combinações possíveis para $\mathcal{S}$ e para o passo 2 temos $4!/2 = 12$ combinações possíveis para $\mathcal{S}_1$, totalizando $360 \times 12 = 4320$ combinações de $\boldsymbol{s}$, cada uma com probabilidade $1/4320$.

Exercício 3.8

```
D = c(8,2,2,11,4,7)
aux = expand.grid(1:6, 1:6, 1:6, 1:6) # Indices de S
conti = 0
for(i in 1:nrow(aux)){
  for(j in 1:4){
    if(sum(aux[i,j]==aux[i,-j])>0){
      conti = c(conti,i)
      break
    }
  }
}
conti = conti[-1]
aux = aux[-conti,]
```

Exercício 3.8

```
aux2 = expand.grid(1:4, 1:4) # Indices de S_1
conti = 0
for(i in 1:nrow(aux2)){
  for(j in 1:2){
    if(sum(aux2[i,j]==aux2[i,-j])>0){
      conti = c(conti,i)
      break
    }
  }
}
conti = conti[-1]
aux2 = aux2[-conti,]
```

Exercício 3.8

```
yb1 = 0 # \overline{y}_1
yb2 = 0 # \overline{y}_2
yb = 0 # \overline{y}
cont = 0
for(i in 1:nrow(aux)){
  amo1 = D[as.numeric(aux[i,])]
  for(j in 1:nrow(aux2)){
    cont = cont + 1
    yb[cont] = mean(amo1)
    amo2 = amo1[as.numeric(aux2[j,])]
    amo3 = amo1[as.numeric(-aux2[j,])]
    yb1[cont] = mean(amo2)
    yb2[cont] = mean(amo3)
  }
}
```

Exercício 3.8

- A partir disso, temos $\overline{y}_i$, $\overline{y}_{1i}$ e $\overline{y}_{2i}$ os valores calculados para o i-ésimo elemento de $\boldsymbol{s}$ de $\overline{Y}$, $\overline{Y}_1$ e $\overline{Y}_2$, respectivamente.
- Dito isso, calcula-se duas novas variáveis $x_i = \overline{y}_{1i} - \overline{y}_{2i}$ e $z_i = \overline{y}_{1i} - \overline{y}_i$, que são relacionadas às VAs de interesse do enunciado: $X = \overline{Y}_1 - \overline{Y}_2$ e $Y = \overline{Y}_1 - \overline{Y}$.
- Logo, temos que $E(X) = E(\overline{Y}_1 - \overline{Y}_2) = \frac{1}{4320}\sum_{i=1}^{4320} x_i = 0$ e $E(Y) = E(\overline{Y}_1 - \overline{Y}) = \frac{1}{4320}\sum_{i=1}^{4320} y_i = 0$.

Exercício 3.8

- Portanto, temos que
  $V(X) = V(\overline{Y}_1 - \overline{Y}_2) = E[(\overline{Y}_1 - \overline{Y}_2)^2] = \frac{1}{4320}\sum_{i=1}^{4320} x_i^2 = 13,0667$
  e $V(Y) = V(\overline{Y}_1 - \overline{Y}) = E[(\overline{Y}_1 - \overline{Y})^2] = \frac{1}{4320}\sum_{i=1}^{4320} y_i^2 = 3,2667.$
- Logo, temos que o desvio de $\overline{Y}_1$ com relação a $\overline{Y}_2$ é maior do que o com relação a $\overline{Y}$, i.e., valores de $\overline{y}_1$ podem ser muito distintos de valores de $\overline{y}_2$ mas não tão distintos com relação a calcular a média amostral da amostra original $(\overline{y})$.
- Por curiosidade, usando a mesma lógica apresentada, defina $W = \overline{Y}_2 - \overline{Y}$, temos que $V(W) = 3,2667$. Portanto, nesse caso não há ganho, em variabilidade da média amostral, em se optar por $\mathcal{S}_1$ ou $\mathcal{S}_2$ quando se comparado um dos dois à amostra original $\mathcal{S}$.

Exercício 3.21

- Um pesquisador deseja estimar a porcentagem de pessoas com sangue do tipo O, entre os 3.200 moradores de uma certa ilha. Ele quer garantir que o coeficiente de variação da estimativa não seja superior a 10%, com 95% de confiança. Ele também sabe que a proporção deve ser um número entre 20% e 30%. Que tamanho da amostra deve ser usado para um plano amostral aleatório simples
    a. com reposição?
    b. sem reposição?

Exercício 3.21

- Primeiramente, vamos reunir algumas informações importantes:
  1. O pesquisador deseja estimar P a porcentagem de pessoas com sangue do tipo O dos $N = 3200$ moradores. Para isso, utiliza-se um estimador $\widehat{P}$.
  2. $\widehat{CV}(\widehat{P}) = \widehat{EP}(\widehat{P})/\widehat{P}$.
  3. Confiança está associada com probabilidade, então com probabilidade de 0.95 queremos que $\widehat{CV}(\widehat{P}) \leq 0,10$.
  4. O pesquisador também afirma que $0,20 \leq p \leq 0,30$.

Exercício 3.21

- Seja $Y = 1$ se o morador tem sangue do tipo O e $Y = 0$ caso contrário.
- Temos sob AASc que $\widehat{P} = \overline{Y}$ é um estimador não viciado para P, e que $\widehat{Var}(\widehat{P}) = \widehat{P}\widehat{Q}/(n-1)$, além disso temos que $\sqrt{n}(\widehat{P} - P)/\sqrt{PQ} \sim \mathcal{N}(0,1)$, portanto:

$$
\begin{aligned}
\widehat{CV}(\widehat{P}) &= \frac{\sqrt{\frac{\widehat{P}\widehat{Q}}{(n-1)}}}{\widehat{P}} = \sqrt{\frac{\widehat{Q}}{(n-1)\widehat{P}}}. \\
P(\widehat{CV}(\widehat{P}) \leq 0,10) &= 0,95 \\
P\left(\sqrt{\frac{\widehat{Q}}{(n-1)\widehat{P}}} \leq 0,10\right) &= 0,95
\end{aligned}
$$

Exercício 3.21

$$
\begin{aligned}
P\left(\frac{\widehat{Q}}{(n-1)\widehat{P}} \leq 0,01\right) &= 0,95 \\
P\left(\frac{\widehat{Q}}{\widehat{P}} \leq 0,01(n-1)\right) &= 0,95 \\
P\left(\frac{1}{\widehat{P}} - 1 \leq 0,01(n-1)\right) &= 0,95 \\
P\left(\frac{1}{\widehat{P}} \leq 0,01n + 0.99\right) &= 0,95 \\
P\left(\widehat{P} > n^*\right) &= 0,95, n^* = \frac{1}{0,01n + 0.99}
\end{aligned}
$$

Exercício 3.21

$$
\begin{aligned}
P\left(\widehat{P} > n^*\right) &= 0,95, \\
P\left(\widehat{P} \leq n^*\right) &= 0,05, \\
P\left(\frac{\widehat{P} - P}{\sqrt{\frac{PQ}{n}}} \leq \frac{n^* - P}{\sqrt{\frac{PQ}{n}}}\right) &= 0,05, \\
\frac{n^* - P}{\sqrt{\frac{PQ}{n}}} &= z_{0,05} \\
n^* &= \sqrt{\frac{PQ}{n}} z_{0,05} + P
\end{aligned}
$$

Exercício 3.21

$$1 = (0,01n + 0.99)\left(\sqrt{\frac{PQ}{n}} z_{0,05} + P\right)$$

$$(0,01n + 0.99)\left(\sqrt{\frac{PQ}{n}} z_{0,05} + P\right) - 1 = 0$$

- Porém, note que o pesquisador tem a informação de que $0,2 \leq p \leq 0,3$, portanto solucionando a equação para $p = 0,2$ e $p = 0,3$ temos que $292 \leq n \leq 489$.

Exercício 3.21

- Já sob AASs, $\widehat{P} = \overline{Y}$ é um estimador não viciado para P, e $\widehat{Var}(\widehat{P}) = (1-f)\widehat{P}\widehat{Q}/(n-1)$, $(\widehat{P} - P)/\sqrt{\frac{N-n}{N-1}\frac{PQ}{n}} \sim \mathcal{N}(0,1)$, portanto:

$$
\begin{aligned}
\widehat{CV}(\widehat{P}) &= \frac{\sqrt{(1-f)\frac{\widehat{P}\widehat{Q}}{(n-1)}}}{\widehat{P}} = \sqrt{(1-f)\frac{\widehat{Q}}{(n-1)\widehat{P}}}. \\
P(\widehat{CV}(\widehat{P}) \leq 0,10) &= 0,95 \\
P\left(\sqrt{\frac{(1-f)\widehat{Q}}{(n-1)\widehat{P}}} \leq 0,10\right) &= 0,95
\end{aligned}
$$

Exercício 3.21

$$
\begin{aligned}
P\left(\frac{(1-f)\widehat{Q}}{(n-1)\widehat{P}} \leq 0,01\right) &= 0,95 \\
P\left(\frac{\widehat{Q}}{\widehat{P}} \leq 0,01\frac{(n-1)}{(1-f)}\right) &= 0,95 \\
P\left(\frac{1}{\widehat{P}} - 1 \leq 0,01\frac{(n-1)}{(1-f)}\right) &= 0,95 \\
P\left(\frac{1}{\widehat{P}} \leq 0,01\frac{(n-1)}{(1-f)} + 1\right) &= 0,95 \\
P\left(\widehat{P} > n^{\dagger}\right) &= 0,95, n^{\dagger} = \frac{1}{0,01\frac{(n-1)}{(1-f)} + 1}
\end{aligned}
$$

Exercício 3.21

$$
\begin{aligned}
P\left(\widehat{P} > n^{\dagger}\right) &= 0,95, \\
P\left(\widehat{P} \leq n^{\dagger}\right) &= 0,05, \\
P\left(\frac{\widehat{P} - P}{\sqrt{\frac{N-n}{N-1}\frac{PQ}{n}}} \leq \frac{n^{\dagger} - P}{\sqrt{\frac{N-n}{N-1}\frac{PQ}{n}}}\right) &= 0,05, \\
\frac{n^{\dagger} - P}{\sqrt{\frac{N-n}{N-1}\frac{PQ}{n}}} &= z_{0,05} \\
n^{\dagger} &= \sqrt{\frac{N-n}{N-1}\frac{PQ}{n}}\, z_{0,05} + P
\end{aligned}
$$

Exercício 3.21

$$
\begin{aligned}
1 &= \left(0,01\frac{(n-1)}{(1-f)}+1\right)\left(\sqrt{\frac{N-n}{N-1}\frac{PQ}{n}}z_{0,05}+P\right)\\
&\left(0,01\frac{(n-1)}{(1-f)}+1\right)\left(\sqrt{\frac{N-n}{N-1}\frac{PQ}{n}}z_{0,05}+P\right)-1=0
\end{aligned}
$$

- Lembrando que o pesquisador sabe que $0,2 \leq p \leq 0,3$, portanto solucionando a equação acima para $p = 0,2$ e $p = 0,3$ temos que $268 \leq n \leq 424$.
- Note que aqui precisamos de um n menor do que o achado para AASc, o que é esperado.

Exercício 3.21

- Possíveis soluções alternativas para testar:
- Outra maneira de resolvermos essa questão seria notar que $\widehat{Q} = 1 - \widehat{P} \sim \mathcal{N}(1 - P, Var(\widehat{P}))$ e trabalhar com a distribuição de $\widehat{Q}/\widehat{P}$ que é a razão de duas normais não padrão.
- Para mais informações sobre essa distribuição pode-se o paper *On the Ratio of Two Correlated Normal Random Variables* de D. V. Hinkley de 1969 (https://www.jstor.org/stable/2334671), seu quantil pode ser obtido via métodos numéricos.
- Outra maneira seria obter a distribuição de $\widehat{Q}/\widehat{P} = (1 - \widehat{P})/\widehat{P}$ via transformação de VAs.

Exercício 4.6

- Considere a população do Exemplo 4.1 com a estratificação $\mathcal{U}_1 = \{1, 2, 4, 7\}$ e $\mathcal{U}_2 = \{3, 5, 6, 8\}$. Considere amostras AASc de tamanho 2 de cada um dos estratos. Encontre a variância do estimador $\overline{y}_{es}$ e o erro quadrático médio do estimador $\overline{y}_m$, definido em (4.35). Qual é o melhor estimador?

Exercício 4.6

- A partir do enunciado temos que $\boldsymbol{D}_1 = \{13, 17, 5, 19\}$ e $\boldsymbol{D}_2 = \{6, 10, 12, 6\}$, além disso, $N_1 = N_2 = 4$, $N = 8$ e portanto $W_1 = W_2 = 1/2$.

Tabela: Distribuição amostral de $\overline{Y}_1$ e $\overline{Y}_2$.

$\boldsymbol{s}$	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
$P(\boldsymbol{s})$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
$\overline{y}_{1i}$	13	15	9	16	15	17	11	18	9	11	5	12	16	18	12	19
$\overline{y}_{2i}$	6	8	9	6	8	10	11	8	9	11	12	9	6	8	9	6

Exercício 4.6

- Dito isso, temos que $E(\overline{Y}_1) = \frac{1}{16}\sum_{i=1}^{16} \overline{y}_{1i} = 13,5$,
  $E(\overline{Y}_1^2) = \frac{1}{16}\sum_{i=1}^{16} \overline{y}_{1i}^2 = 196,625$,
  $Var(\overline{Y}_1) = 196,625 - (13,5)^2 = 14,375$.
- Além disso, $E(\overline{Y}_2) = \frac{1}{16}\sum_{i=1}^{16} \overline{y}_{2i} = 8,5$,
  $E(\overline{Y}_2^2) = \frac{1}{16}\sum_{i=1}^{16} \overline{y}_{2i}^2 = 75,625$,
  $Var(\overline{Y}_2) = 75,625 - (8,5)^2 = 3,375$.
- Em posse dessas informações, podemos calcular
  $E(\overline{y}_{es}) = \frac{1}{2}E(\overline{Y}_1) + \frac{1}{2}E(\overline{Y}_2) = \frac{1}{2}13,5 + \frac{1}{2}8,5 = 11$ e
  $Var(\overline{y}_{es}) = \frac{1}{4}Var(\overline{Y}_1) + \frac{1}{4}Var(\overline{Y}_2) = \frac{1}{4}14,375 + \frac{1}{4}3,375 = 4,4375$.
- Sabendo que $\mu = (13 + 17 + 5 + 19 + 6 + 10 + 12 + 6)/8 = 11$,
  temos que o viés B de $\overline{y}_{es}$ é dado por $B(\overline{y}_{es}) = E(\overline{y}_{es}) - \mu = 0$.

Exercício 4.6

- Por fim, temos que
  $EQM(\overline{y}_{es}) = Var(\overline{y}_{es}) + B(\overline{y}_{es})^2 = Var(\overline{y}_{es}) = 4,4375.$
- Note que $\overline{y}_m = \sum_{h=1}^{H} \frac{n_h}{n} \overline{y}_h$, como $n_1 = n_2 = 2$ e $n = 4$, portanto $\frac{n_h}{n} = \frac{1}{2} = \frac{N_h}{N} = W_h$, para $h = 1, 2$. Logo, temos que $\overline{y}_m = \overline{y}_{es}$, e temos que esses estimadores são equivalentes e
  $EQM(\overline{y}_{es}) = EQM(\overline{y}_m)$.

Exercício 5.1 - Elementos de amostragem

- Considere a população $\mathcal{U}$ do Exemplo 5.2. Queremos estimar $R = \mu_X/\mu_Y$. Considere os estimadores

$$\widehat{R}_1 = \frac{\overline{y}}{\overline{x}}, \quad \widehat{R}_2 = \frac{\overline{y}}{\mu_X} \quad \text{e} \quad \widehat{R}_3 = \frac{1}{n}\sum_{i \in \boldsymbol{s}} \frac{Y_i}{X_i}$$

Encontre as distribuições de $\widehat{R}_i$, $i = 1, 2, 3$, seus vícios e EQM, para AASc e AASs. Qual dos estimadores você prefere?

Exercício 5.1 - Elementos de amostragem

- Observando o Exemplo 5.2 citado. Temos que Y é a renda bruta familiar e X é o número de trabalhadores. Portanto $\mathcal{U} = (1, 2, 3)$, $\boldsymbol{Y} = (12, 30, 18)$ e $\boldsymbol{X} = (1, 3, 2)$. A partir disso, temos que $\mu_X = 2$, e para AASc com $n = 2$:

Exercício 5.1 - Elementos de amostragem

Tabela: Distribuição amostral de $\widehat{R}_i$ usando o plano AASs.

$\boldsymbol{s}$	11	12	13	21	22	23	31	32	33
$P(\boldsymbol{s})$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\overline{y}$	12	21	15	21	30	24	15	24	18
$\overline{x}$	1	2	1,5	2	3	2,5	1,5	2,5	2
$\widehat{R}_{1i}$	12	10,5	10	10,5	10	9,6	10	9,6	9
$\widehat{R}_{2i}$	6	10,5	7,5	10,5	15	12	7,5	12	9
$\widehat{R}_{3i}$	12	11	10,5	11	10	9,5	10,5	9,5	9

Exercício 5.1 - Elementos de amostragem

- Sabendo que $R = \mu_Y / \mu_X = 20/2 = 10$.
- Temos que $E(\widehat{R}_1) = (1/9)\sum_{i=1}^{9} \widehat{R}_{1i} = 10,1333$, $E(\widehat{R}_2) = (1/9)\sum_{i=1}^{9} \widehat{R}_{2i} = 10$ e $E(\widehat{R}_3) = (1/9)\sum_{i=1}^{9} \widehat{R}_{3i} = 10,3333$.
- O vicio B é então dado por $B(\widehat{R}_1) = 0,1333$, $B(\widehat{R}_2) = 0$ e $B(\widehat{R}_3) = 0,3333$.
- $E(\widehat{R}_1^2) = (1/9)\sum_{i=1}^{9} \widehat{R}_{1i}^2 = 103,3133$,
  $E(\widehat{R}_2^2) = (1/9)\sum_{i=1}^{9} \widehat{R}_{2i}^2 = 107$ e
  $E(\widehat{R}_3^2) = (1/9)\sum_{i=1}^{9} \widehat{R}_{3i}^2 = 107,5556$.

Exercício 5.1 - Elementos de amostragem

- A variância é dada por $Var(\widehat{R_1}) = 0,6295$, $Var(\widehat{R_2}) = 7$ e $Var(\widehat{R_3}) = 0,7785$.
- O EQM é dado por $EQM(\widehat{R_1}) = 0.6473$, $EQM(\widehat{R_2}) = 7$ e $EQM(\widehat{R_3}) = 0,8896$.

Exercício 5.1 - Elementos de amostragem

- Note que o menor EQM pertence a $\widehat{R}_1$, logo, este é preferível.
- Apesar de $\widehat{R}_2$ ser não viesado, este possui alta variância. Possível motivo: note que se apenas o indivíduo 1 ou 2 forem selecionados, por consequência $\widehat{R}_{2i}$ é muito distante de R.
- Além disso, temos que nesse caso, utilizar a razão das médias amostrais é melhor (em EQM) do que utilizar a média da razão amostral.

Exercício 5.1 - Elementos de amostragem

- Fazendo agora para AASs com $n = 2$, temos que:

Tabela: Distribuição amostral de $\widehat{R}_i$ usando o plano AASs.

$\boldsymbol{s}$	12	13	21	23	31	32
$\mathrm{P}(\boldsymbol{s})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$\overline{y}$	21	15	21	24	15	24
$\overline{x}$	2	1,5	2	2,5	1,5	2,5
$\widehat{R}_{1i}$	10,5	10	10,5	9,6	10	9,6
$\widehat{R}_{2i}$	10,5	7,5	10,5	12	7,5	12
$\widehat{R}_{3i}$	11	10,5	11	9,5	10,5	9,5

Exercício 5.1 - Elementos de amostragem

- Logo, temos que $E(\widehat{R}_1) = (1/6)\sum_{i=1}^{6}\widehat{R}_{1i} = 10,0333$, $E(\widehat{R}_2) = (1/6)\sum_{i=1}^{6}\widehat{R}_{2i} = 10$ e $E(\widehat{R}_3) = (1/6)\sum_{i=1}^{6}\widehat{R}_{3i} = 10,3333$.
- O vicio B é então dado por $B(\widehat{R}_1) = 0,0333$, $B(\widehat{R}_2) = 0$ e $B(\widehat{R}_3) = 0,3333$.
- $E(\widehat{R}_1^2) = (1/6)\sum_{i=1}^{6}\widehat{R}_{1i}^2 = 100,8033$,
  $E(\widehat{R}_2^2) = (1/6)\sum_{i=1}^{6}\widehat{R}_{2i}^2 = 103,5$ e
  $E(\widehat{R}_3^2) = (1/6)\sum_{i=1}^{6}\widehat{R}_{3i}^2 = 107,1667$.

Exercício 5.1 - Elementos de amostragem

- A variância é dada por $Var(\widehat{R}_1) = 0,1362$, $Var(\widehat{R}_2) = 3,5$ e $Var(\widehat{R}_3) = 0,3896$.
- O EQM é dado por $EQM(\widehat{R}_1) = 0.1373$, $EQM(\widehat{R}_2) = 3,5$ e $EQM(\widehat{R}_3) = 0,5007$.

Exercício 5.1 - Elementos de amostragem

- Note que aqui também o menor EQM pertence a $\widehat{R}_1$, logo, este é preferível.
- Também $\widehat{R}_2$ é não viesado e possui alta variância quando comparado aos outros.
- Além disso, temos que nesse caso também, utilizar a razão das médias amostrais é melhor (em EQM) do que utilizar a média da razão amostral.

Exercício 5.1 - Elementos de amostragem

- Percebe-se que em comparação ao plano AASc, aqui temos um viés menor para $\widehat{R}_1$ e variância muito mais baixas para todos, por exemplo, para $\widehat{R}_2$ a redução é pela metade. Mostrando assim, a vantagem de se utilizar AASs nesse caso.

Exercício 6.8 - Elementos de amostragem

- Para verificar a influência de uma nova marca de ração, um criador de galinhas pesou 10 de seus frangos ao comprá-los (X_i) e depois de 30 dias (Y_i). Os resultados estão na tabela abaixo.

Frango:	1	2	3	4	5	6	7	8	9	10
X_i	1,50	1,60	1,4	5 1,40	1,40	1,55	1,60	1,45	1,55	1,50
Y_i	2,14	2,16	2,10	1,95	2,05	2,10	2,26	2,00	2,20	2,04

O peso médio de todos os frangos na hora da compra era de 1,54.

a) Qual o estimador regressão mais indicado? Justifique.

b) E qual seria a estimativa?

c) E o erro padrão?

Exercício 6.8 - letra a) - Elementos de amostragem

- Nesse caso, temos que Y e X são duas variáveis correlacionadas, pois tratam-se do peso dos frangos coletados em momentos diferentes.
- Além disso, assume-se uma relação linear que não passa pela origem, pois não temos peso 0, i.e., $Y_i = \alpha + \beta X_i + e_i$.
- Logo, de forma a verificar a influência da nova marca de ração podemos utilizar um estimador do tipo regressão para a média de Y tal que:

$$\overline{y}_{reg} = \overline{y} + b(\mu_X - \overline{x}),$$

Exercício 6.8 - letra a) - Elementos de amostragem

- Em que $\overline{x}$ e $\overline{y}$ são as médias amostrais de X e Y, respectivamente, μ_X é a verdadeira média de X e b é um estimativa para β que representa o impacto no peso médio do frango depois de 30 dias da compra (Y) pela variação de uma unidade no peso do frango no dia da compra (X).
- Do enunciado temos que $\overline{x} = 1,5$, $\overline{y} = 2,1$ e $\mu_X = 1,54$. Portanto, o estimador regressão fica dado por:

$$\overline{y}_{reg} = 2,1 + 0,04b. \tag{1}$$

Exercício 6.8 - letra b) - Elementos de amostragem

- A partir da equação (6.5) do livro, temos que:

$$\widehat{b} = \frac{\sum_{i \in \boldsymbol{s}} (Y_i - \overline{y})(X_i - \overline{x})}{\sum_{i \in \boldsymbol{s}} (X_i - \overline{x})^2} = \frac{s_{XY}}{s_X^2} = \frac{0,052}{0,05} = 1,04.$$

- Portanto, temos que: $\overline{y}_{reg} = 2,1 + 0,04 * 1,04 = 2,1416$. Como $\widehat{b} > 0$, temos indícios de que há um aumento na média do peso dos frangos depois de 30 dias da compra.

Exercício 6.8 - letra c) - Elementos de amostragem

- A partir da equação (6.6) do livro, temos que:

$$\widehat{V}_{reg} = \frac{s_Y^2}{n}(1 - \widehat{\rho}^2[X, Y]) = \frac{0.0794}{10}(1 - 0,8253^2) = 0,0025.$$

- Portanto, temos que o erro padrão é dado por $\sqrt{\widehat{V}_{reg}} = 0,0503$.
- Além disso, é interessante notar que $IC_{95\%}(\mu_Y) = \overline{y}_{reg} \pm 1,9599 \times 0,0530 = (2,0430; 2,2402)$, cujos limites superior e inferior são maiores que μ_X, ou seja, temos mais um indício do aumento no peso médio do frango.

Exercício 7.5 - Elementos de amostragem

- Uma companhia que fornece carros a seus vendedores quer uma estimativa do número médio de quilômetros percorridos pelos seus carros no ano passado. A companhia tem 12 filiais. O número de carros (B_α), a média (μ_α) e a variância S_α^2 do número de quilômetros percorridos (em milhares), para cada filial, são dados por:

Exercício 7.5 - Elementos de amostragem

Filial	B_α	μ_α	S^2_α
1	6	24.32	5.07
2	2	27.06	5.53
3	11	27.60	6.24
4	7	28.01	6.59
5	8	27.56	6.21
6	14	29.07	6.12
7	6	32.03	5.97
8	2	28.41	6.01
9	2	28.91	5.74
10	5	25.55	6.78
11	12	28.58	5.87
12	6	27.27	5.38

Exercício 7.5 - Elementos de amostragem

- Selecione uma AASc de 4 filiais e estime o número médio de quilômetros percorridos por carro utilizando a informação sobre todos os carros nas filiais selecionadas. Encontre a variância de sua estimativa e também uma estimativa para a variância. Compare a variância de sua estimativa com a variância correspondente à utilização de uma AASc de tamanho $n = 27$.

Exercício 7.5 - Elementos de amostragem

- Nesse caso, como temos informações sobre o número total de unidades $N = \sum_{\alpha=1}^{12} B_\alpha = 81$ e o tamanho dos conglomerados são diferentes, podemos utilizar como estimador $\widehat{\mu}_{C_1} = \overline{T}/\overline{B}$, em que $\overline{T} = (1/4)\sum_{\alpha=1}^{4} T_\alpha = (1/4)\sum_{\alpha=1}^{4} B_\alpha \mu_\alpha$.
- utilizando como semente 2018 no R, os conglomerados sorteados são 2, 3, 7 e 12.
- Dito isso, temos que $\overline{B} = 81/12 = 6,75$, $\overline{T} = (1/4)(2 \times 27,06 + 11 \times 27,6 + 6 \times 32,03 + 6 \times 27,27) = 178,38$, e portanto $\widehat{\mu}_{C_1} = 26,4267$.

Exercício 7.5 - Elementos de amostragem

- Do teorema 7.1 do livro temos que $Var(\widehat{\mu}_{C_1}) = \sigma^2_{ect}/4$, em que $\sigma^2_{ect} = (1/12)\sum_{\alpha=1}^{12}\left(\frac{B_\alpha}{6{,}75}\mu_\alpha - \mu\right)^2$, em que $\mu = (1/12)\sum_{\alpha=1}^{12}\frac{b_\alpha}{6{,}75}\mu_\alpha = 28,0039$, logo $\sigma^2_{ect} = 264,3386$ e portanto $Var(\widehat{\mu}_{C_1}) = 66,0846$.
- Do Corolário 7.1 temos que $\widehat{Var}(\widehat{\mu}_{C_1}) = (1/12)\sum_{\alpha=1}^{12}\left(\frac{B_\alpha}{6{,}75}\overline{y}_\alpha - \widehat{\mu}_{C_1}\right)^2 = 57,6660$.

Exercício 7.5 - Elementos de amostragem

- Sabe-se que, sob AASc temos que $Var_{A_1}(\widehat{\mu}) = \frac{\sigma^2}{27}$. Nesse caso, podemos utilizar o fato de que $\sigma^2 = \sigma^2_{dc} + \sigma^2_{ec}$. Portanto, calcula-se:

$$\begin{aligned}
\sigma^2_{dc} &= \frac{1}{12}\sum_{\alpha=1}^{12}\frac{B_\alpha}{6,75}S^2_\alpha = 6,0191 \\
\sigma^2_{ec} &= \frac{1}{12}\sum_{\alpha=1}^{12}\frac{B_\alpha}{6,75}(\mu_\alpha - \mu)^2 = 2,9511 \\
\sigma^2 &= 6,0191 + 2,9511 = 8,9702. \\
Var_{A_1}(\widehat{\mu}) &= 8,9702/27 = 0.3322.
\end{aligned}$$

- Portanto note que $Var(\widehat{\mu}_{C_1})/Var_{A_1}(\widehat{\mu}) = 198.9302$, nesse caso a perda em se utilizar AC é muito grande.

Exercício 7.5 - Elementos de amostragem

- Porém, note que se ignorarmos algumas informações, como o tamanho total populacional e considerar os outros estimadores temos que:

$$
\begin{aligned}
\sigma^2_{eq} &= \frac{1}{12}\sum_{\alpha=1}^{12}\left(\frac{B_\alpha}{6,75}\right)^2(\mu_\alpha - 28,0039)^2 = 2,8394, \\
\sigma^2_{em} &= \frac{1}{12}\sum_{\alpha=1}^{12}(\mu_\alpha - 27,8642)^2 = 3,3177, \\
\overline{\mu} &= \frac{1}{12}\sum_{\alpha=1}^{12}\mu_\alpha = 27,8642.
\end{aligned}
$$

Exercício 7.5 - Elementos de amostragem

$$
\begin{aligned}
Var(\widehat{\mu}_{C_2}) &= \frac{\sigma^2_{eq}}{4} = 0,7098, \\
Var(\widehat{\mu}_{C_3}) &= \frac{\sigma^2_{em}}{4} + (27,8642 - 28,0039)^2 = 0,8489.
\end{aligned}
$$

- Logo, utilizar $\widehat{\mu}_{C_2}$ seria muito mais eficiente do que utilizar $\widehat{\mu}_{C_1}$, apesar de teoricamente este último ser não viesado. Dito isso podemos calcular $\overline{b} = (1/4)\sum_{\alpha=1}^{4} B_\alpha = 6,25$, e portanto $\widehat{\mu}_{C_2} = \overline{T}/\overline{b} = 28,5408$.

Exercício 7.5 - Elementos de amostragem

- Além disso, é interessante notar que o viés estimado nesse caso de $\widehat{\mu}_{C_2}$ é muito menor do que o obtido por $\widehat{\mu}_{C_1}$.

Como obter intervalos de confiança?

- Uma coisa importante para relembrarmos é: como obter intervalos de confiança (IC) uma vez que conhecemos a distribuição de origem do estimador.
- Por exemplo, para AAS com e sem reposição usualmente utiliza-se estimadores baseados em $\overline{y}$, e sabemos, uma vez que as condições sejam satisfeitas, que pelo Teorema Central do Limite que $\overline{y}$ se aproxima de uma normal. Logo, é importante sabermos construir IC's normais.

Como obter intervalos de confiança?

- Na maioria dos exemplos e exercícios pode-se utilizar o método da quantitade pivotal.
- Seja $\boldsymbol{X}$ o vetor de VAs de uma população cujo vetor de parâmetros é $\boldsymbol{\theta}$, uma quantidade pivotal é uma função Q de $\boldsymbol{X}$ e $\boldsymbol{\theta}$ cuja distribuição não depende de $\boldsymbol{\theta}$.
- Dito disso, seja $\gamma \in (0,1)$ podemos construir um IC a partir de $P(q_1 \leq Q \leq q_2) = \gamma$, em que q_1 e q_2 são quantis apropriados.

Como obter intervalos de confiança?

- No caso dos estimadores sob algum plano amostral, temos na maioria dos casos resultados como este:

$$W = \frac{\widehat{\mu} - \mu}{\widehat{EP}(\widehat{\mu})} \sim \mathcal{N}(0,1),$$

em que W é uma quantidade pivotal para μ, portanto podemos utilizá-lo para construir um IC.

$$\begin{aligned} P(q_1 \leq W \leq q_2) &= \gamma \\ P\left(q_1 \leq \frac{\widehat{\mu} - \mu}{\widehat{EP}(\widehat{\mu})} \leq q_2\right) &= \gamma \end{aligned}$$

Como obter intervalos de confiança?

$$
\begin{aligned}
P\left(q_1\widehat{EP}(\widehat{\mu}) \leq \widehat{\mu} - \mu \leq q_2\widehat{EP}(\widehat{\mu})\right) &= \gamma \\
P\left(q_1\widehat{EP}(\widehat{\mu}) - \widehat{\mu} \leq -\mu \leq q_2\widehat{EP}(\widehat{\mu}) - \widehat{\mu}\right) &= \gamma \\
P\left(\widehat{\mu} - q_2\widehat{EP}(\widehat{\mu}) \leq \mu \leq \widehat{\mu} - q_1\widehat{EP}(\widehat{\mu})\right) &= \gamma
\end{aligned}
$$

em que q_1 e q_2 são quantis da normal padrão.

- Logo, um IC para μ com $100\gamma\%$ de confiança é dado por $IC_{100\gamma\%}(\mu) = [\widehat{\mu} - q_2\widehat{EP}(\widehat{\mu}); \widehat{\mu} - q_1\widehat{EP}(\widehat{\mu})]$.

Como obter intervalos de confiança?

- Pode-se mostrar que o IC de menor comprimento é dado quando $q_2 = -q_1$, logo, o IC de menor comprimento é dado por $IC_{100\gamma\%}(\mu) = [\widehat{\mu} - q_2\widehat{EP}(\widehat{\mu}); \widehat{\mu} + q_2\widehat{EP}(\widehat{\mu})]$, em que q_2 é o quantil γ de uma normal padrão. Ver aqui
- Raciocínio similar pode ser feito para o estimadores do total, proporção, etc.