

Método Científico e Técnicas de Pesquisa - Códigos

Gabriella de Oliveira Argenton 255677 e Gabriela Namie Hidaka 204274

```
#library(readr)
#library(dplyr)
#library(ggplot2)
#library(purrr)
#library(tidyr)
#library(stringr)
#library(readxl)
#library(dplyr)
#library(readxl)

dados <- read_excel("Estacionamento.xlsx")
dados <- dados %>%
  mutate(
    Dia = as.factor(Dia),
    `Horário` = factor(`Horário`, levels=c("manhã", "tarde", "noite")),
    Instituto = as.factor(Instituto),
    Quantidade = as.numeric(Quantidade)
  )
N <- nrow(dados)

media_pop <- mean(dados$Quantidade)
var_pop <- var(dados$Quantidade)    ##  $S^2$  com denominador (N-1)
sd_pop <- sd(dados$Quantidade)

tab_h <- dados %>%
  group_by(Horário) %>%
  summarise(N_h = n(),
            ybar_h = mean(Quantidade),
            S2_h = var(Quantidade),
            .groups = "drop") %>%
```

```
mutate(W_h = N_h / N)
```

```
tab_h
```

```
# A tibble: 3 x 5
```

	Horário	N_h	ybar_h	S2_h	W_h
	<fct>	<int>	<dbl>	<dbl>	<dbl>
1	manhã	15	0.756	0.0354	0.333
2	tarde	15	0.814	0.0139	0.333
3	noite	15	0.331	0.0138	0.333

```
set.seed(123)          # reprodutibilidade
n_total <- 6            # TAMANHO DA AMOSTRA (ajuste se quiser testar outros n) - é o tamanho
## Estratificada: alocação igual (mesmo n_h por período)
H <- n_distinct(dados$Horário)
n_h <- rep(n_total / H, H) # alocação igual

## AAS sem reposição: ybar = média simples de 'Quantidade'
## (Se quiser testar com reposição, mude replace=TRUE)
estimador_aas <- function(df, n, replace = FALSE) {
  amostra_idx <- sample(seq_len(nrow(df)), size = n, replace = replace)
  mean(df$Quantidade[amostra_idx])
}

## Estratificada: Horário como estrato.
## Estimador: ybar_st = sum_h W_h * ybar_h_amostral
estimador_ae <- function(df, n_h_vec) {
  ## n_h_vec: vetor com n_h por estrato na ordem dos níveis de Horário
  ## Pesos pelos W_h pop (tamanhos reais do banco)
  ord_niveis <- levels(df$Horário)
  out <- purrr::map2_dfr(ord_niveis, n_h_vec, ~{
    h <- .x; nh <- .y
    sub <- dplyr::filter(df, Horário == h)
    idx <- sample(seq_len(nrow(sub)), size = nh, replace = FALSE)
    tibble(Horário = h, mean_h = mean(sub$Quantidade[idx]))
  })
  ## pesos populacionais verdadeiros
  pesos <- df %>% count(Horário, name = "N_h") %>% mutate(W_h = N_h / nrow(df))
  out <- out %>% left_join(pesos, by = "Horário")
  sum(out$W_h * out$mean_h)
}
```

```

B <- 5000
res <- tibble(
  aas = numeric(B),
  ae = numeric(B)
)

for (b in seq_len(B)) {
  res$aas[b] <- estimador_aas(dados, n = n_total, replace = FALSE) # AAS sem reposição
  res$ae[b] <- estimador_ae(dados, n_h_vec = n_h) # Estratificada
}

emp <- tibble(
  plano = c("AAS (s/ reposição)", "Estratificada"),
  media = c(mean(res$aas), mean(res$ae)),
  desvio_padrao = c(sd(res$aas), sd(res$ae)),
  EQM = c(mean((res$aas - media_pop)^2),
           mean((res$ae - media_pop)^2)),
  viés = c(mean(res$aas) - media_pop,
            mean(res$ae) - media_pop)
) %>%
  mutate(RE = var(res$aas) / var(res$ae)) ## eficiência relativa (ainda baseada na variância)

emp

```

```

# A tibble: 2 x 6
  plano          media desvio_padrao    EQM    viés    RE
  <chr>          <dbl>         <dbl> <dbl>   <dbl> <dbl>
1 AAS (s/ reposição) 0.636         0.0984 0.00969 0.00251 3.14
2 Estratificada    0.635         0.0555 0.00309 0.00118 3.14

```

```

## AAS s/ reposição:
## Var_teor(ybar_AAS) = (1 - n/N) * S^2 / n
var_teor_aas <- (1 - n_total/N) * var_pop / n_total #formula da var aas sem rep

## Estratificada:
## Var_teor(ybar_st) = sum_h W_h^2 * (1 - f_h) * S_h^2 / n_h
## onde f_h = n_h/N_h
tab_h_calc <- dados %>%
  group_by(Horário) %>%
  summarise(N_h = n(),
            S2_h = var(Quantidade),

```

```

    .groups = "drop") %>%
mutate(W_h = N_h / N) %>%
arrange(Horário)

## alinhar com ordem dos níveis/ n_h
tab_h_calc$n_h <- n_h
tab_h_calc$f_h <- tab_h_calc$n_h / tab_h_calc$N_h

var_teor_ae <- sum(tab_h_calc$W_h^2 * (1 - tab_h_calc$f_h) * tab_h_calc$S2_h / tab_h_calc$n_h)

## Tabela teórica x empírica
comp_var <- tibble(
  plano = c("AAS (s/ reposição)", "Estratificada"),
  var_empirica = c(var(res$aas), var(res$ae)),
  var_teorica = c(var_teor_aas, var_teor_ae),
  RE_empirica = c(1, var(res$aas)/var(res$ae)),
  RE_teorica = c(1, var_teor_aas/var_teor_ae)
)
comp_var

```

```

# A tibble: 2 x 5
  plano          var_empirica var_teorica RE_empirica RE_teorica
  <chr>          <dbl>         <dbl>         <dbl>         <dbl>
1 AAS (s/ reposição)  0.00969    0.00975         1             1
2 Estratificada      0.00308    0.00304        3.14          3.21

```

```

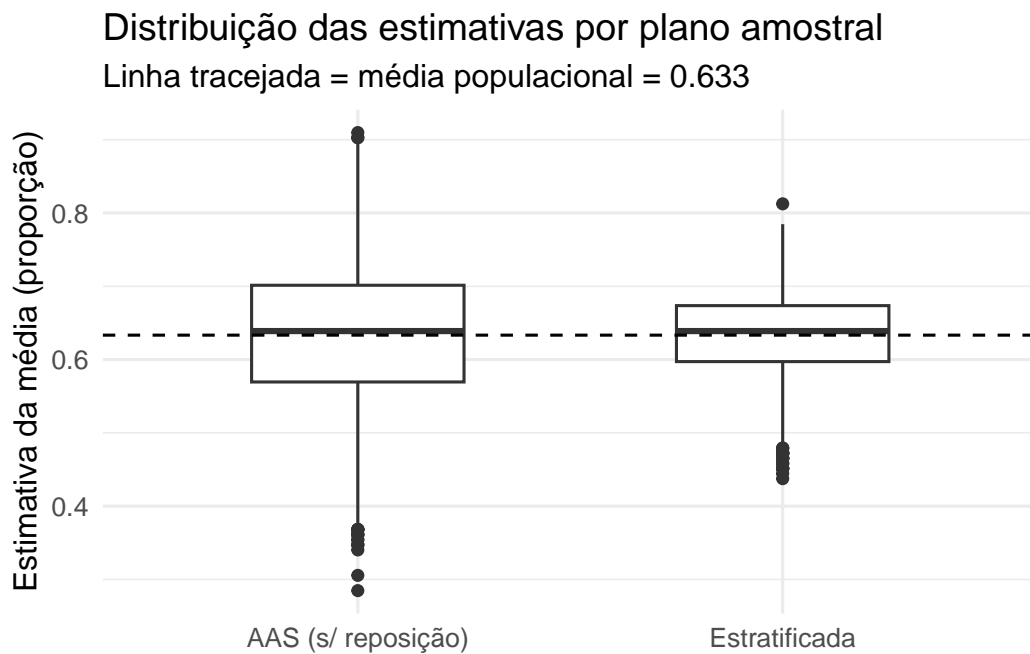
long_res <- res %>%
  pivot_longer(everything(), names_to = "plano", values_to = "estimativa") %>%
  mutate(plano = recode(plano,
    "aas" = "AAS (s/ reposição)",
    "ae" = "Estratificada"))

## Gráficos (Densidade e Boxplot)
g1 <- ggplot(long_res, aes(x = plano, y = estimativa)) +
  geom_boxplot(width = 0.5) +
  geom_hline(yintercept = media_pop, linetype = 2) +
  labs(title = "Distribuição das estimativas por plano amostral",
    subtitle = paste0("Linha tracejada = média populacional = ",
      round(media_pop, 3)),
    x = NULL, y = "Estimativa da média (proporção)") +
  theme_minimal(base_size = 12)

```

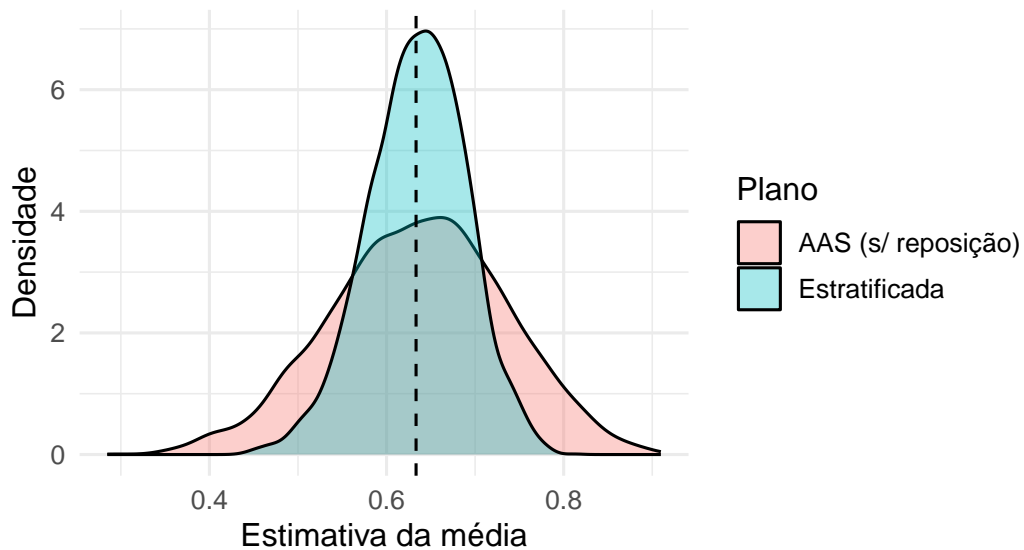
```
g2 <- ggplot(long_res, aes(x = estimativa, fill = plano)) +
  geom_density(alpha = 0.35) +
  geom_vline(xintercept = media_pop, linetype = 2) +
  labs(title = "Densidades das estimativas",
        subtitle = paste0("Linha tracejada = média populacional = ",
                           round(media_pop, 3)),
        x = "Estimativa da média", y = "Densidade", fill = "Plano") +
  theme_minimal(base_size = 12)

print(g1); print(g2)
```



Densidades das estimativas

Linha tracejada = média populacional = 0.633



```
cat("\n--- Verdade populacional ---\n")
```

```
--- Verdade populacional ---
```

```
cat("N          :", N, "\n")
```

```
N          : 45
```

```
cat("Média pop :", round(media_pop, 4), "\n")
```

```
Média pop : 0.6333
```

```
cat("SD pop    :", round(sd_pop, 4), "\n\n")
```

```
SD pop    : 0.2598
```

```
cat("--- Simulação (B =", B, ", n_total =", n_total, ", n_h =", paste(n_h, collapse = ","),
```

```
--- Simulação (B = 5000 , n_total = 6 , n_h = 2,2,2 ) ---
```

```
print(emp)
```

```
# A tibble: 2 x 6
```

	plano <chr>	media <dbl>	desvio_padrao <dbl>	EQM <dbl>	viés <dbl>	RE <dbl>
1	AAS (s/ reposição)	0.636	0.0984	0.00969	0.00251	3.14
2	Estratificada	0.635	0.0555	0.00309	0.00118	3.14

```
cat("\n--- Variâncias: teórica vs empírica ---\n\n")
```

```
--- Variâncias: teórica vs empírica ---
```

```
print(comp_var)
```

```
# A tibble: 2 x 5
```

	plano <chr>	var_empirica <dbl>	var_teorica <dbl>	RE_empirica <dbl>	RE_teorica <dbl>
1	AAS (s/ reposição)	0.00969	0.00975	1	1
2	Estratificada	0.00308	0.00304	3.14	3.21