Double burden of malnutrition in Peruvian children

Carlos Ballon-Salcedo

2025-06-29

# Load packages

pacman::p\_load(  
 rio,  
 here,  
 tidyverse,  
 gtsummary,  
 finalfit,  
 flextable,  
 bstfun,  
 labelled,  
 janitor,  
 epiDisplay,  
 Hmisc,  
 ggthemes,  
 rfextras,  
 ggpubr,  
 ggdist,  
 gghalves,  
 colorspace,  
 ggsci,  
 showtext,  
 extrafont,  
 patchwork,  
 scales,  
 survey,  
 srvyr,  
 universals)  
  
#source(here::here("scripts", "dictionary.R"))  
rfextras::load\_scripts()

# Import data

data\_ninios <- rio::import(here::here("data", "ninios-2011-2022\_tmp.dta"))

# Process data

## Variable selection

# Inspect dataset structure and variable labels  
labelled::var\_label(data\_ninios) # List of variable labels (if available)  
Hmisc::contents(data\_ninios) # data.frame with variable types and missing values  
dplyr::glimpse(data\_ninios) # Compact summary of variable names and types  
str(data\_ninios) # Detailed structure of the dataset

# Select and clean variables  
data\_dbm <- data\_ninios |>  
 dplyr::select(  
 id1, hc1, hc5, hc8, hc11, hc27, hc57, hc61, hc70, hc71, hc72, hc73,   
 hv025, peso, hv001, hv022, hv024, hc55, hv103, hc13, desnutricion) |>  
 dplyr::filter(hc55 == 0, hv103 == 1, hc13 == 0) |>  
 dplyr::mutate(across(c(hc5, hc8, hc11, hc70, hc71, hc72), ~ replace(.x, .x > 9990, NA)))  
 # dplyr::filter(across(c(hc70, hc71, hc72), ~ .x < 2000))  
 # dplyr::mutate(hc61 = replace(hc61, hc61 == 8, NA\_real\_))

## Dictionary: Recode, relevel, and relabel variables

data\_dbm <- my\_dictionary(data\_dbm) |>  
 dplyr::select(-starts\_with(c("hc", "hv")), -id1) |>  
 dplyr::filter(!is.na(edad\_meses\_minsa\_cat))

## Exploratory Data Analysis

### Categorical data

# Select categorical variables (factors)  
categorical\_vars <- data\_dbm |>  
 dplyr::select(where(\(x) is.character(x) | is.factor(x)))

#### Missing, unique, and frequency analysis without weighting

# Structure  
dplyr::glimpse(categorical\_vars)  
  
# Count of missing values by column (vector)  
colSums(is.na(categorical\_vars))  
  
# List of complete values by column  
lapply(categorical\_vars, function(x) sum(complete.cases(x)))  
  
# Columns without missing values (vector)  
colnames(categorical\_vars)[sapply(categorical\_vars, function(x) all(!is.na(x)))]  
  
# Count of unique values by column  
categorical\_vars |>  
 dplyr::summarise(across(everything(), n\_distinct))  
  
# Detailed list of unique values by column  
lapply(categorical\_vars, function(x) unique(x))  
  
# Count duplicate values by column  
categorical\_vars |>  
 dplyr::summarise(across(  
 where(\(x) is.character(x) | is.factor(x)), ~ sum(duplicated(.))))  
  
# Absolute frequency for categorical variables  
lapply(categorical\_vars, function(x) addmargins(table(x)))  
  
# Relative frequency for categorical variables  
lapply(categorical\_vars, function(x) round(addmargins(prop.table(table(x)) \* 100), 1))  
   
# Absolute frequency for categorical variables by DBM\_anemia\_stunting\_factor  
lapply(categorical\_vars, function(x)  
 addmargins(table(  
 x, categorical\_vars$DBM\_anemia\_stunting\_factor  
 )))

#### Descriptive analysis: Measures of frequency and proportions

# Flextable for sexo\_child\_factor by DBM\_anemia\_stunting\_factor  
categorical\_vars |>  
 janitor::tabyl(area\_residencia\_factor, DBM\_anemia\_stunting\_factor) |>  
 janitor::adorn\_totals(where = "both") |>  
 janitor::adorn\_percentages(denominator = "col") |>  
 janitor::adorn\_pct\_formatting() |>  
 janitor::adorn\_ns(position = "front") |>  
 flextable::flextable() |>  
 flextable::autofit()

| area\_residencia\_factor | No | Si | NA\_ | Total |
| --- | --- | --- | --- | --- |
| Urbano | 64,578 (69.8%) | 3,766 (42.0%) | 143 (81.7%) | 68,487 (67.4%) |
| Rural | 27,922 (30.2%) | 5,201 (58.0%) | 32 (18.3%) | 33,155 (32.6%) |
| Total | 92,500 (100.0%) | 8,967 (100.0%) | 175 (100.0%) | 101,642 (100.0%) |

# Frecuency table for categorical variables by DBM\_anemia\_stunting\_factor  
purrr::map(c(  
 "area\_residencia\_factor", "sexo\_child\_factor", "education\_factor",  
 "edad\_meses\_minsa\_cat\_factor", "year", "stunting", "wasting", "thinness",   
 "exceso\_peso", "anemia\_child\_factor", "anemia\_child\_levels\_factor"), ~ categorical\_vars |>  
 tabyl(.data[[.x]], DBM\_anemia\_stunting\_factor) |>  
 adorn\_totals(where = "both") |>  
 adorn\_percentages(denominator = "col") |>  
 adorn\_pct\_formatting() |>  
 adorn\_ns(position = "front"))  
  
# Frecuency table for categorical variables by DBM\_anemia\_wasting\_factor  
purrr::map(c(  
 "area\_residencia\_factor", "sexo\_child\_factor", "education\_factor",  
 "edad\_meses\_minsa\_cat\_factor", "year", "stunting", "wasting", "thinness",   
 "exceso\_peso", "anemia\_child\_factor", "anemia\_child\_levels\_factor"), ~ categorical\_vars |>  
 tabyl(.data[[.x]], DBM\_anemia\_wasting\_factor) |>  
 adorn\_totals(where = "both") |>  
 adorn\_percentages(denominator = "col") |>  
 adorn\_pct\_formatting() |>  
 adorn\_ns(position = "front"))  
  
# Frecuency table for categorical variables by DBM\_anemia\_thinness\_factor  
purrr::map(c(  
 "area\_residencia\_factor", "sexo\_child\_factor", "education\_factor",  
 "edad\_meses\_minsa\_cat\_factor", "year", "stunting", "wasting", "thinness",   
 "exceso\_peso", "anemia\_child\_factor", "anemia\_child\_levels\_factor"), ~ categorical\_vars |>  
 tabyl(.data[[.x]], DBM\_anemia\_thinness\_factor) |>  
 adorn\_totals(where = "both") |>  
 adorn\_percentages(denominator = "col") |>  
 adorn\_pct\_formatting() |>  
 adorn\_ns(position = "front"))  
  
# Frecuency table for categorical variables by DBM\_anemia\_overweight\_factor  
purrr::map(c(  
 "area\_residencia\_factor", "sexo\_child\_factor", "education\_factor",  
 "edad\_meses\_minsa\_cat\_factor", "year", "stunting", "wasting", "thinness",   
 "exceso\_peso", "anemia\_child\_factor", "anemia\_child\_levels\_factor"), ~ categorical\_vars |>  
 tabyl(.data[[.x]], DBM\_anemia\_overweight\_factor) |>  
 adorn\_totals(where = "both") |>  
 adorn\_percentages(denominator = "col") |>  
 adorn\_pct\_formatting() |>  
 adorn\_ns(position = "front"))

### Numerical data

# Selection of variables  
numerical\_vars <- data\_dbm |>  
 dplyr::select(  
 edad\_meses, talla\_edad, talla\_edad\_oms, peso\_edad, peso\_edad\_oms, peso\_talla,  
 peso\_talla\_oms, DBM\_anemia\_stunting\_factor)

#### Descriptive statistics: Summary statistics

numerical\_vars |> epiDisplay::summ(digits = 2, na.rm = FALSE)

No. of observations = 101642  
  
 Var. name obs. mean median s.d. min. max.   
1 edad\_meses 101642 20.72 21 8.64 6 35   
2 talla\_edad 101515 -0.78 -0.78 1.03 -6 5.88   
3 talla\_edad\_oms 101468 -0.94 -0.94 1.08 -6 5.97   
4 peso\_edad 101515 -0.43 -0.51 1.15 -5.72 5.9   
5 peso\_edad\_oms 101468 -0.15 -0.18 1.08 -5.36 4.93   
6 peso\_talla 101536 0.12 0.06 1.02 -3.97 5.94   
7 peso\_talla\_oms 101468 0.46 0.43 1.01 -4.98 4.98   
8 DBM\_anemia\_stunting\_factor 101467 1.088 1 0.284 1 2

#### Descriptive statistics: Grouped summary statistics

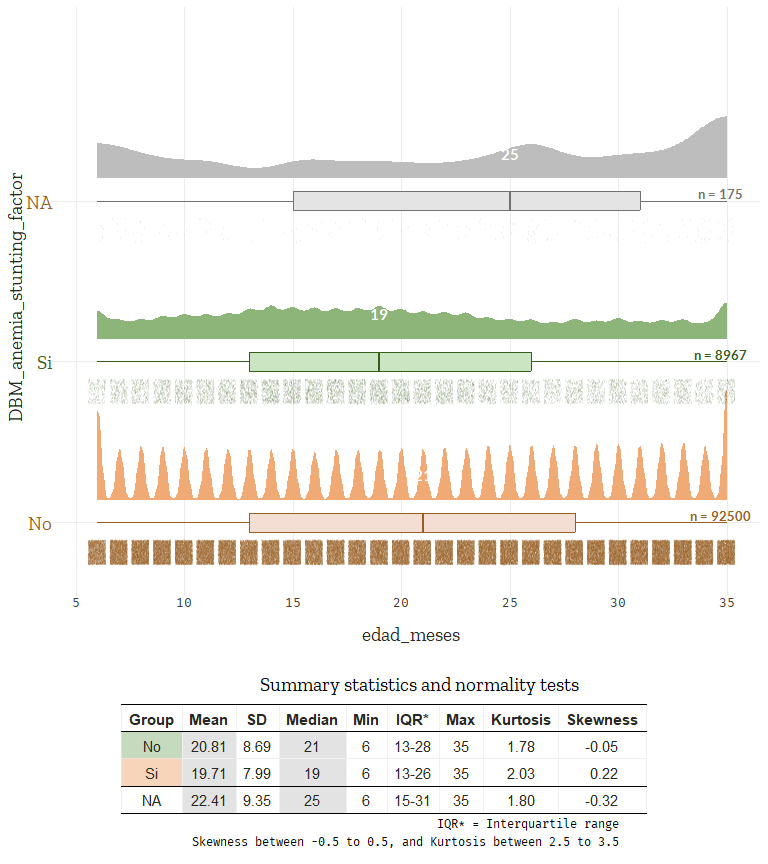
numerical\_vars |>   
 group\_split(DBM\_anemia\_stunting\_factor) |>   
 lapply(function(df) summ(df, digits = 2, na.rm = TRUE))

[[1]]  
  
No. of observations = 92500  
  
 Var. name obs. mean median s.d. min. max.   
1 edad\_meses 92500 20.81 21 8.69 6 35   
2 talla\_edad 92470 -0.63 -0.67 0.93 -5.19 5.88   
3 talla\_edad\_oms 92500 -0.78 -0.82 0.98 -5.82 5.97   
4 peso\_edad 92470 -0.31 -0.4 1.1 -4.99 5.72   
5 peso\_edad\_oms 92500 -0.02 -0.07 1.02 -5.36 4.93   
6 peso\_talla 92485 0.16 0.09 1.01 -3.97 5.9   
7 peso\_talla\_oms 92500 0.52 0.48 1 -4.98 4.98   
8 DBM\_anemia\_stunting\_factor   
  
[[2]]  
  
No. of observations = 8967  
  
 Var. name obs. mean median s.d. min. max.   
1 edad\_meses 8967 19.71 19 7.99 6 35   
2 talla\_edad 8963 -2.33 -2.21 0.52 -5.81 -1.41   
3 talla\_edad\_oms 8967 -2.59 -2.44 0.54 -6 -2.01   
4 peso\_edad 8963 -1.72 -1.73 0.71 -4.76 1.11   
5 peso\_edad\_oms 8967 -1.42 -1.39 0.78 -5.25 1.33   
6 peso\_talla 8963 -0.35 -0.39 0.87 -3.86 3.53   
7 peso\_talla\_oms 8967 -0.07 -0.06 0.93 -4.88 3.52   
8 DBM\_anemia\_stunting\_factor   
  
[[3]]  
  
No. of observations = 175  
  
 Var. name obs. mean median s.d. min. max.   
1 edad\_meses 175 22.41 25 9.35 6 35   
2 talla\_edad 82 -0.23 0.46 2.31 -6 2.18   
3 talla\_edad\_oms 1 -0.18 -0.18 <NA> -0.18 -0.18   
4 peso\_edad 82 2.79 4.3 3.59 -5.72 5.9   
5 peso\_edad\_oms 1 3.1 3.1 <NA> 3.1 3.1   
6 peso\_talla 88 3.39 4.81 3.03 -3.97 5.94   
7 peso\_talla\_oms 1 4.55 4.55 <NA> 4.55 4.55   
8 DBM\_anemia\_stunting\_factor

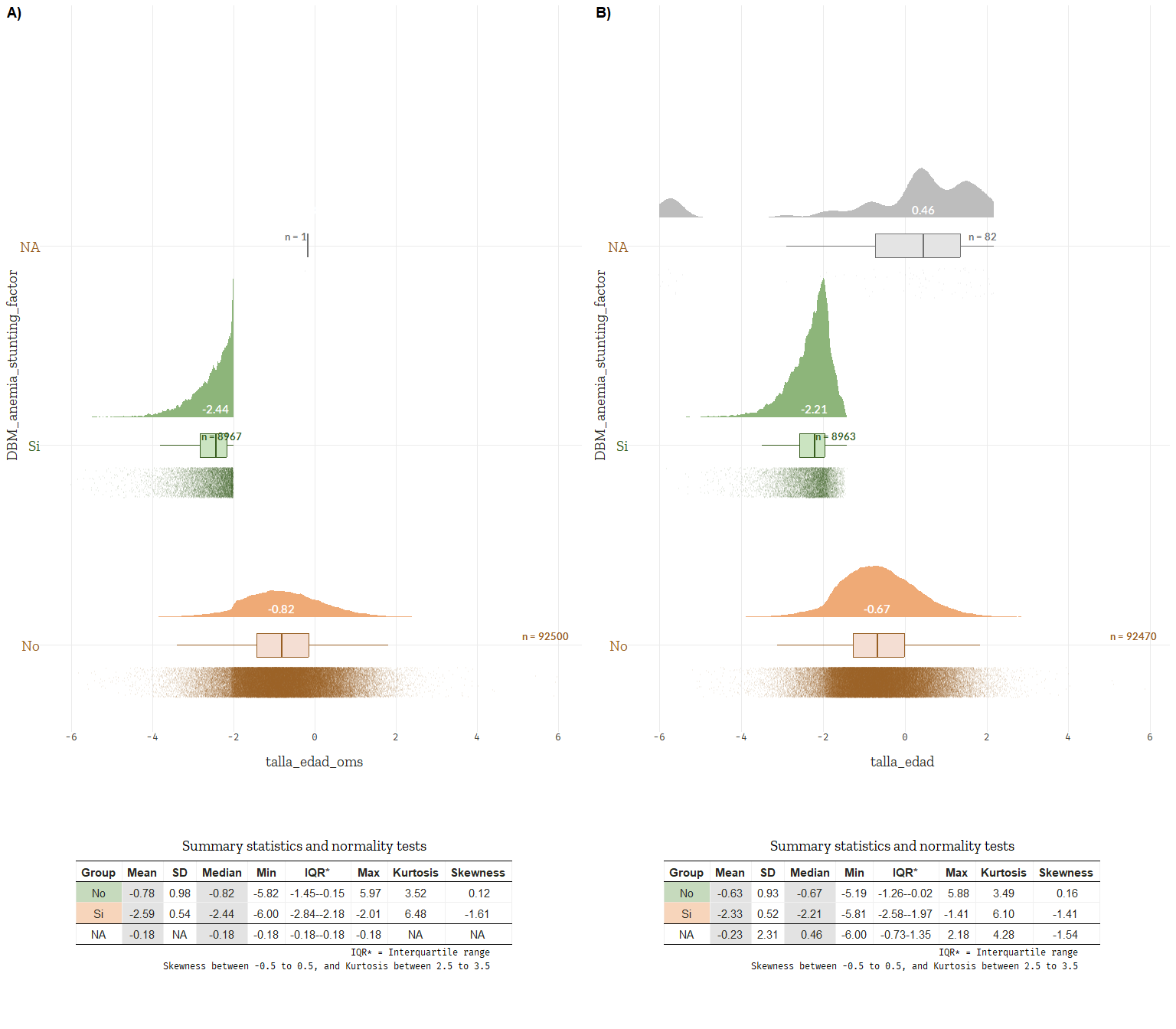
#### Data visualization: Raincloud plots and grouped summary statistics

El test de Shapiro-Wilk evalúa la normalidad de un conjunto de datos, pero en muestras grandes pierde utilidad porque detecta desviaciones mínimas como significativas. En la práctica, se recomienda aplicarlo en tamaños de muestra de hasta 5,000 observaciones, ya que más allá de este límite tiende a rechazar la hipótesis de normalidad incluso cuando la distribución es aproximadamente normal. Por otro lado, el test de Anderson-Darling es más sensible a la forma de las colas de la distribución, pero su cálculo es computacionalmente costoso debido al ordenamiento de los datos y la evaluación de una suma ponderada de diferencias. Con muestras masivas, la normalidad suele asumirse debido al Teorema del Límite Central, salvo en presencia de fuerte asimetría o colas pesadas. Para complementar estas pruebas, es recomendable usar histogramas, QQ-plots y métricas como skewness y kurtosis.

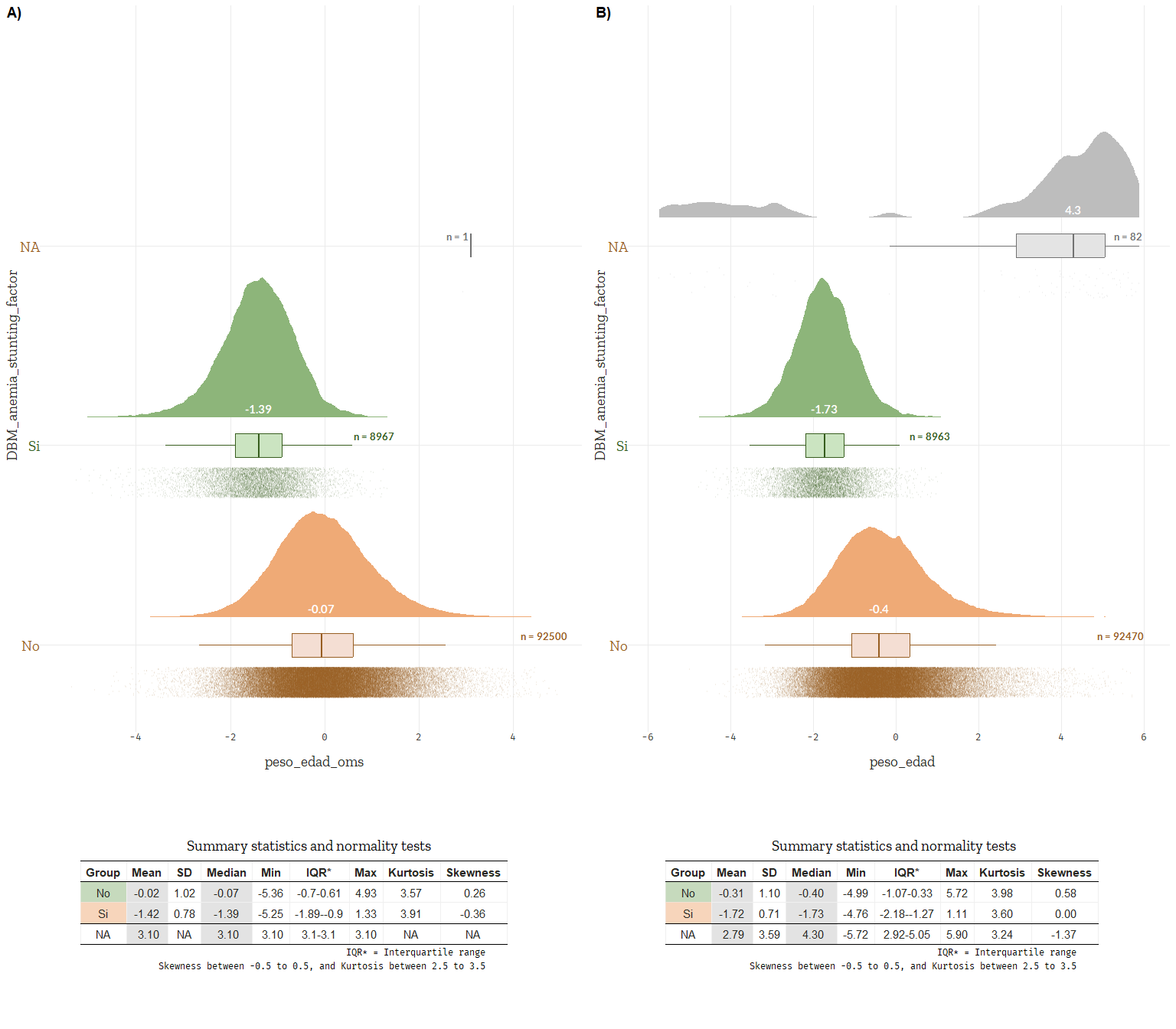
group\_stat\_table\_plot(numerical\_vars, "edad\_meses", "DBM\_anemia\_stunting\_factor")



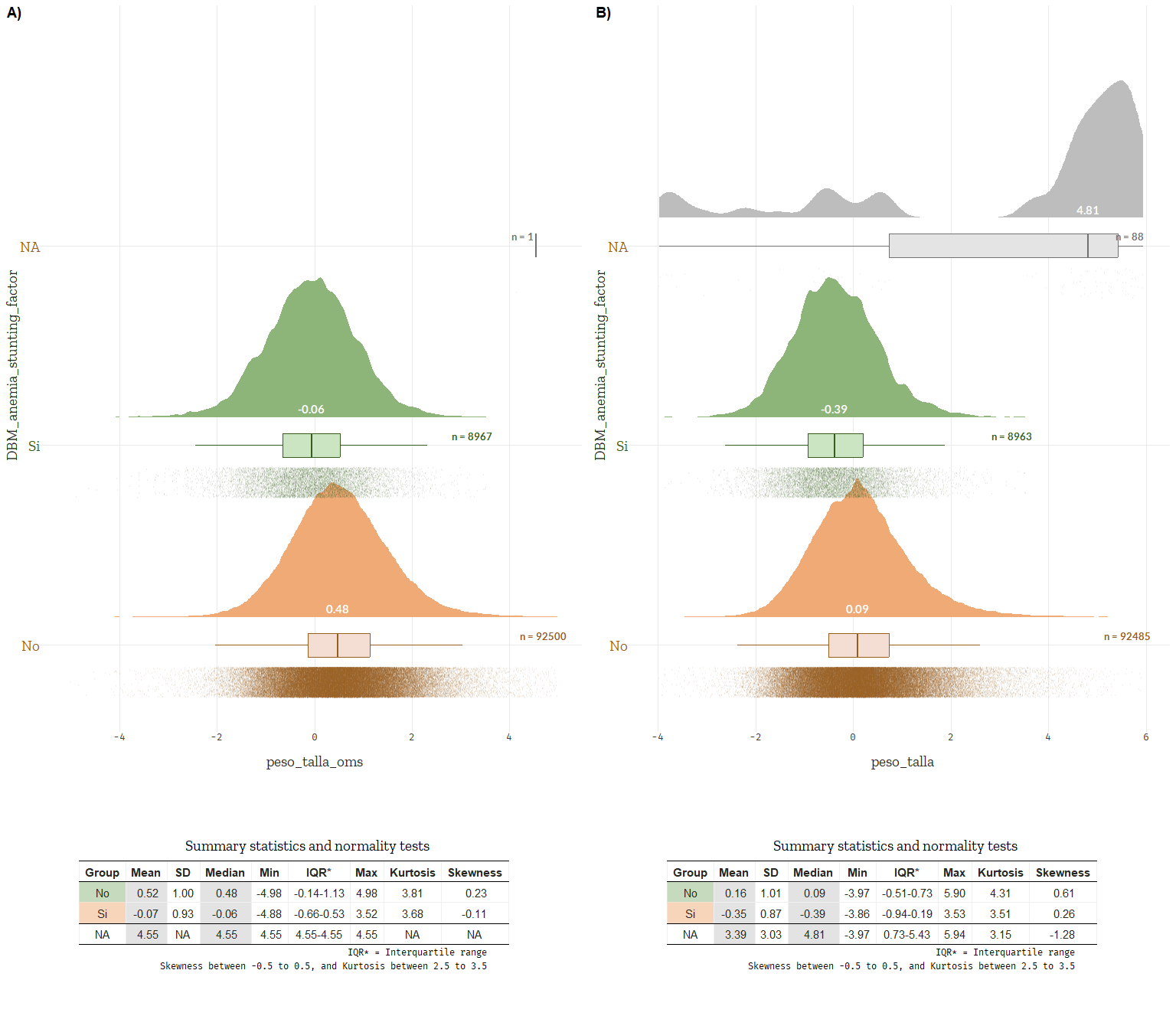
a <- group\_stat\_table\_plot(numerical\_vars, "talla\_edad\_oms", "DBM\_anemia\_stunting\_factor")  
b <- group\_stat\_table\_plot(numerical\_vars, "talla\_edad", "DBM\_anemia\_stunting\_factor")  
  
ggarrange(a, b, ncol = 2, nrow = 1, labels = c("A)", "B)"))



a <- group\_stat\_table\_plot(numerical\_vars, "peso\_edad\_oms", "DBM\_anemia\_stunting\_factor")  
b <- group\_stat\_table\_plot(numerical\_vars, "peso\_edad", "DBM\_anemia\_stunting\_factor")  
  
ggarrange(a, b, ncol = 2, nrow = 1, labels = c("A)", "B)"))



a <- group\_stat\_table\_plot(numerical\_vars, "peso\_talla\_oms", "DBM\_anemia\_stunting\_factor")  
b <- group\_stat\_table\_plot(numerical\_vars, "peso\_talla", "DBM\_anemia\_stunting\_factor")  
  
ggarrange(a, b, ncol = 2, nrow = 1, labels = c("A)", "B)"))



# Produce outputs

## Design

# Survey design  
data\_dbm\_svy <- srvyr::as\_survey\_design(  
 data\_dbm,  
 ids = cluster,  
 strata = estrato,  
 weights = pesos,  
 nest = TRUE  
)

## Table 1: Sociodemographic characteristics of Peruvian children

This is the univariate analysis

### Unweighted descriptive analysis

# Absolute frequency  
table\_1a <- data\_dbm\_svy |>  
 gtsummary::tbl\_svysummary(  
 include = c(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor,  
 DBM\_anemia\_stunting\_factor  
 ),  
 statistic = list(  
 all\_categorical() ~ "{n\_unweighted}",  
 all\_continuous() ~ "{mean}"  
 ),  
 digits = list(  
 all\_continuous() ~ c(1, 1),   
 all\_categorical() ~ c(0, 1)  
 ),  
 missing = "no"  
 ) |>  
 modify\_header(  
 stat\_0 ~ "\*\*N = {N\_unweighted}\*\*"  
 ) |>  
 bold\_labels()

### Weighted descriptive analysis

Use the show\_header\_names() to see the column names that can be modified.

# Weighted proportion  
table\_1b <- data\_dbm\_svy |>  
 gtsummary::tbl\_svysummary(  
 include = c(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor,  
 DBM\_anemia\_stunting\_factor  
 ),  
 statistic = list(  
 all\_categorical() ~ "{p}",  
 all\_continuous() ~ "{mean} ({sd})"   
 ),  
 digits = list(  
 all\_continuous() ~ c(1, 1),   
 all\_categorical() ~ c(2, 1)),  
 missing = "no"  
 ) |>  
 add\_ci(  
 method = list(  
 all\_categorical() ~ "svyprop.logit",   
 all\_continuous() ~ "svymean"),   
 statistic = list(  
 all\_categorical() ~ "{conf.low} - {conf.high}",   
 all\_continuous() ~ "{conf.low} - {conf.high}"),  
 conf.level = 0.95,  
 style\_fun = list(  
 all\_categorical() ~ purrr::partial(style\_percent, digits = 2))  
 ) |>  
 bold\_labels()  
  
# Merge tables  
table\_1 <- gtsummary::tbl\_merge(  
 tbls = list(table\_1a, table\_1b),  
 tab\_spanner = c("\*\*Absolute frequency\*\*", "\*\*Weighted proportion\*\*")) |>  
 modify\_caption("Table 1. Sociodemographic characteristics of Peruvian children") |>  
 modify\_footnote(  
 stat\_0\_1 ~ "Unweighted N; Mean",  
 stat\_0\_2 ~ "Weighted percentage; Mean (SD)")  
  
# View  
table\_1

|  | **Absolute frequency** | **Weighted proportion** | |
| --- | --- | --- | --- |
| **Characteristic** | **N = 101642***1* | **N = 54,167***2* | **95% CI** |
| **Área de residencia** |  |  |  |
| Urbano | 68,487 | 70.97 | 70.22 - 71.71 |
| Rural | 33,155 | 29.03 | 28.29 - 29.78 |
| **Sexo del infante** |  |  |  |
| Masculino | 51,887 | 51.00 | 50.56 - 51.44 |
| Femenino | 49,755 | 49.00 | 48.56 - 49.44 |
| **Nivel educativo de la madre** |  |  |  |
| Sin educación | 1,737 | 1.73 | 1.605 - 1.865 |
| Primaria | 21,083 | 21.16 | 20.65 - 21.68 |
| Secundaria | 64,146 | 64.50 | 63.93 - 65.06 |
| Superior | 12,863 | 12.61 | 12.22 - 13.01 |
| **Edad del infante (MINSA)** |  |  |  |
| 6 a 11 meses | 19,674 | 19.08 | 18.75 - 19.43 |
| 12 a 23 meses | 40,240 | 39.83 | 39.40 - 40.27 |
| 24 a 35 meses | 41,728 | 41.08 | 40.66 - 41.51 |
| **Doble carga de malnutrición** |  |  |  |
| No | 92,500 | 91.84 | 91.57 - 92.10 |
| Si | 8,967 | 8.16 | 7.899 - 8.435 |
| *1*Unweighted N; Mean | | | |
| *2*Weighted percentage; Mean (SD) | | | |
| Abbreviation: CI = Confidence Interval | | | |

### Confirm with dplyr package

data\_dbm |>  
 dplyr::select(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor,  
 DBM\_anemia\_stunting\_factor,  
 DBM\_anemia\_desnutricion\_factor,  
 pesos  
 ) |>  
 tidyr::pivot\_longer(  
 cols = c(area\_residencia\_factor:DBM\_anemia\_desnutricion\_factor),  
 names\_to = "Variable",  
 values\_to = "Valor"  
 ) |>  
 dplyr::group\_by(Variable, Valor) |>  
 dplyr::filter(!is.na(Valor)) |>  
 dplyr::summarise(n = n(), weighted\_prop = sum(pesos)) |>  
 dplyr::mutate(weighted\_prop = round(weighted\_prop / sum(weighted\_prop) \* 100, 2))

## Table 2: Sociodemographic characteristics of Peruvian children by DBM status

This is the bivariate analysis

Chi-squared test with Rao & Scott’s second-order correction and Holm-Bonferroni method to adjust for multiple tests. Use (García-Pérez 2023) to understand the goodness of fit.

### Weighted bivariate analysis

table\_2 <- data\_dbm\_svy |>  
 dplyr::filter(!is.na(DBM\_anemia\_stunting\_factor)) |>  
 gtsummary::tbl\_svysummary(  
 percent = "row",   
 by = DBM\_anemia\_stunting\_factor,  
 include = c(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor  
 ),  
 type = all\_categorical() ~ "categorical",  
 statistic = all\_categorical() ~ "{n\_unweighted} ({p})",  
 digits = all\_categorical() ~ c(0, 2),  
 missing = "no"  
 ) |>  
 add\_ci(  
 method = all\_categorical() ~ "svyprop.logit",   
 statistic = all\_categorical() ~ "{conf.low} - {conf.high}",  
 conf.level = 0.95,  
 style\_fun = list(  
 all\_categorical() ~ purrr::partial(style\_percent, digits = 2))  
 ) |>  
 add\_p(pvalue\_fun = ~style\_pvalue(.x, digits = 3)) |>  
 bold\_labels() |>  
 bold\_p(t = 0.05) |>  
 modify\_caption(  
 "Table 2. Sociodemographic characteristics of Peruvian children by double burden of malnutrition status"  
 )  
  
# View  
table\_2

| **Characteristic** | **No** N = 49,650*1* | **95% CI** | **Si** N = 4,413*1* | **95% CI** | **p-value***2* |
| --- | --- | --- | --- | --- | --- |
| **Área de residencia** |  |  |  |  | **<0.001** |
| Urbano | 64,578 (95.07) | 72.72 - 74.16 | 3,766 (4.93) | 41.01 - 44.63 |  |
| Rural | 27,922 (83.93) | 25.84 - 27.28 | 5,201 (16.07) | 55.37 - 58.99 |  |
| **Sexo del infante** |  |  |  |  | **<0.001** |
| Masculino | 46,374 (90.51) | 49.78 - 50.70 | 5,406 (9.49) | 57.93 - 60.64 |  |
| Femenino | 46,126 (93.22) | 49.30 - 50.22 | 3,561 (6.78) | 39.36 - 42.07 |  |
| **Nivel educativo de la madre** |  |  |  |  | **<0.001** |
| Sin educación | 1,352 (77.02) | 1.339 - 1.570 | 382 (22.98) | 4.208 - 5.615 |  |
| Primaria | 17,484 (82.98) | 18.64 - 19.62 | 3,567 (17.02) | 42.61 - 45.63 |  |
| Secundaria | 59,523 (93.91) | 65.40 - 66.51 | 4,518 (6.09) | 46.50 - 49.61 |  |
| Superior | 12,500 (98.07) | 13.05 - 13.89 | 334 (1.93) | 2.544 - 3.472 |  |
| **Edad del infante (MINSA)** |  |  |  |  | **<0.001** |
| 6 a 11 meses | 18,047 (92.42) | 18.85 - 19.56 | 1,592 (7.58) | 16.71 - 18.76 |  |
| 12 a 23 meses | 35,686 (89.68) | 38.47 - 39.38 | 4,507 (10.32) | 49.03 - 51.79 |  |
| 24 a 35 meses | 38,767 (93.66) | 41.43 - 42.32 | 2,868 (6.34) | 30.61 - 33.17 |  |
| *1*n (unweighted) (%) | | | | | |
| *2*Pearson's X^2: Rao & Scott adjustment | | | | | |
| Abbreviation: CI = Confidence Interval | | | | | |

### Confirm with dplyr package

# By row  
data\_dbm |>  
 dplyr::select(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor,  
 DBM\_anemia\_stunting\_factor,  
 pesos  
 ) |>  
 tidyr::pivot\_longer(  
 cols = c(area\_residencia\_factor:edad\_meses\_minsa\_cat\_factor),  
 names\_to = "Variable",  
 values\_to = "Valor"  
 ) |>  
 dplyr::filter(!is.na(DBM\_anemia\_stunting\_factor)) |>  
 dplyr::group\_by(Variable, Valor, DBM\_anemia\_stunting\_factor) |>  
 dplyr::summarise(n = n(), weighted\_prop = sum(pesos), .groups = "drop\_last") |>  
 dplyr::mutate(weighted\_prop = round(weighted\_prop / sum(weighted\_prop) \* 100, 2)) |>  
 dplyr::ungroup()  
  
# By column  
data\_dbm |>  
 dplyr::select(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor,  
 DBM\_anemia\_stunting\_factor,  
 pesos  
 ) |>  
 tidyr::pivot\_longer(  
 cols = c(area\_residencia\_factor:edad\_meses\_minsa\_cat\_factor),  
 names\_to = "Variable",  
 values\_to = "Valor"  
 ) |>  
 dplyr::filter(!is.na(DBM\_anemia\_stunting\_factor)) |>  
 dplyr::group\_by(DBM\_anemia\_stunting\_factor, Variable, Valor) |>  
 dplyr::summarise(n = n(), weighted\_prop = sum(pesos), .groups = "drop\_last") |>  
 dplyr::mutate(weighted\_prop = round(weighted\_prop / sum(weighted\_prop) \* 100, 2)) |>  
 dplyr::ungroup()

## Table 3: Binomial logistic regression analysis

Let’s use the svyglm() function from the survey package to fit a generalised linear model to data obtained from a complex survey design, incorporating inverse-probability weighting and design-based standard errors that account for stratification, clustering, and sample weights. This function returns *model-robust* standard errors. Specifically, these are Horvitz–Thompson-type standard errors, which are a generalisation of the sandwich estimators used in robust regression. In particular, a quasi-Poisson svyglm will return correct standard errors for relative risk regression models. This is a common approach in epidemiological studies to account for the complex survey design, including stratification, clustering, and sample weights.

### Unadjusted analysis

# Univariate regression models  
table\_3a <- data\_dbm\_svy |>  
 gtsummary::tbl\_uvregression(  
 method = survey::svyglm,  
 method.args = list(family = quasipoisson(link = "log")),  
 y = DBM\_anemia\_stunting, # Tener cuidado  
 include = c(  
 area\_residencia\_factor,  
 sexo\_child\_factor,  
 education\_factor,  
 edad\_meses\_minsa\_cat\_factor  
 ),  
 hide\_n = TRUE,  
 exponentiate = TRUE,  
 conf.int = TRUE,  
 conf.level = 0.95,  
 pvalue\_fun = label\_style\_pvalue(digits = 1),  
 estimate\_fun = ~ style\_number(.x, digits = 2)  
 ) |>  
 bold\_labels()

### Adjusted analysis

# Multivariable analysis  
table\_3b <-   
 survey::svyglm(  
 DBM\_anemia\_stunting ~   
 area\_residencia\_factor +  
 sexo\_child\_factor +  
 education\_factor +  
 edad\_meses\_minsa\_cat\_factor,  
 design = data\_dbm\_svy,  
 family = quasipoisson(link = "log")) |>  
 gtsummary::tbl\_regression(  
 exponentiate = TRUE,   
 conf.int = TRUE,  
 conf.level = 0.95,  
 tidy\_fun = broom.helpers::tidy\_parameters,  
 pvalue\_fun = label\_style\_pvalue(digits = 1),  
 estimate\_fun = ~ style\_number(.x, digits = 2)  
 ) |>  
 add\_vif() |>  
 bold\_labels() |>  
 bold\_p(t = 0.05)  
  
# Merge tables  
table\_3 <- gtsummary::tbl\_merge(  
 tbls = list(table\_3a, table\_3b),  
 tab\_spanner = c("\*\*Unadjusted Analysis\*\*", "\*\*Adjusted Analysis\*\*"))  
  
# View  
table\_3

|  | **Unadjusted Analysis** | | | **Adjusted Analysis** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **IRR** | **95% CI** | **p-value** | **IRR** | **95% CI** | **p-value** | **GVIF** | **Adjusted GVIF***1* |
| **Área de residencia** |  |  |  |  |  |  | 1.3 | 1.2 |
| Urbano | — | — |  | — | — |  |  |  |
| Rural | 3.26 | 3.06, 3.48 | <0.001 | 2.19 | 2.04, 2.35 | **<0.001** |  |  |
| **Sexo del infante** |  |  |  |  |  |  | 1.0 | 1.0 |
| Masculino | — | — |  | — | — |  |  |  |
| Femenino | 0.71 | 0.68, 0.75 | <0.001 | 0.71 | 0.67, 0.75 | **<0.001** |  |  |
| **Nivel educativo de la madre** |  |  |  |  |  |  | 1.4 | 1.1 |
| Sin educación | — | — |  | — | — |  |  |  |
| Primaria | 0.74 | 0.65, 0.84 | <0.001 | 0.81 | 0.72, 0.92 | **0.001** |  |  |
| Secundaria | 0.26 | 0.23, 0.30 | <0.001 | 0.40 | 0.35, 0.46 | **<0.001** |  |  |
| Superior | 0.08 | 0.07, 0.10 | <0.001 | 0.15 | 0.12, 0.19 | **<0.001** |  |  |
| **Edad del infante (MINSA)** |  |  |  |  |  |  | 1.0 | 1.0 |
| 6 a 11 meses | — | — |  | — | — |  |  |  |
| 12 a 23 meses | 1.36 | 1.27, 1.46 | <0.001 | 1.35 | 1.26, 1.44 | **<0.001** |  |  |
| 24 a 35 meses | 0.84 | 0.77, 0.90 | <0.001 | 0.82 | 0.76, 0.88 | **<0.001** |  |  |
| *1*GVIF^[1/(2\*df)] | | | | | | | | |
| Abbreviations: CI = Confidence Interval, GVIF = Generalized Variance Inflation Factor, IRR = Incidence Rate Ratio | | | | | | | | |

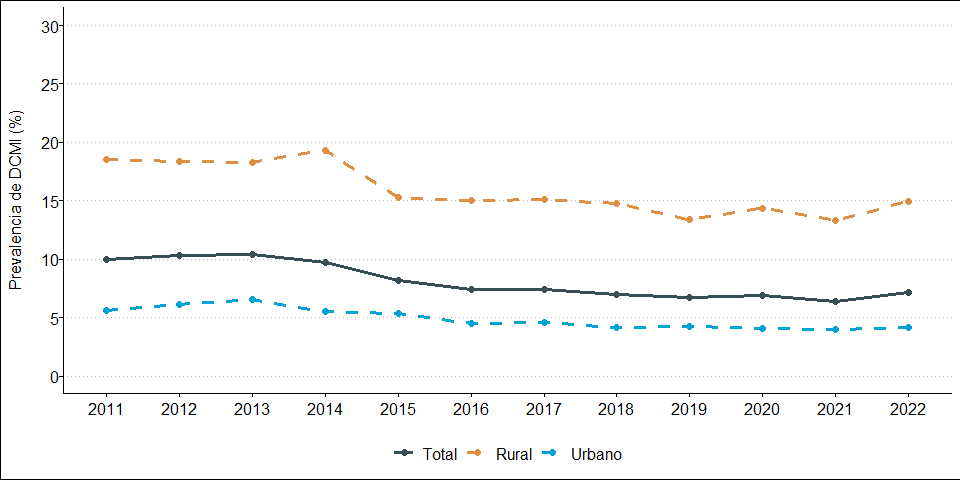
## Figure 1

survey::svytable(  
 ~ DBM\_anemia\_stunting\_factor + area\_residencia\_factor + year, data\_dbm\_svy, na.rm = TRUE) |>  
 prop.table(margin = 1) \* 100

, , year = 2011  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.455339 2.399531  
 Si 3.684294 6.136656  
  
, , year = 2012  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.514613 2.522524  
 Si 4.073952 6.390512  
  
, , year = 2013  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.195718 2.235498  
 Si 4.113262 5.637148  
  
, , year = 2014  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.969982 2.189891  
 Si 3.973886 5.898926  
  
, , year = 2015  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 8.066857 2.833684  
 Si 5.163464 5.772780  
  
, , year = 2016  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 7.602093 2.536162  
 Si 4.070066 5.040778  
  
, , year = 2017  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 6.847330 2.283521  
 Si 3.687789 4.565701  
  
, , year = 2018  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 6.304803 2.038696  
 Si 3.108673 3.986733  
  
, , year = 2019  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.962950 2.023758  
 Si 3.012383 3.517331  
  
, , year = 2020  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.570782 1.900696  
 Si 2.688902 3.590252  
  
, , year = 2021  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.675751 1.812595  
 Si 2.637862 3.129686  
  
, , year = 2022  
  
 area\_residencia\_factor  
DBM\_anemia\_stunting\_factor Urbano Rural  
 No 5.277882 1.779344  
 Si 2.592352 3.526612

# Prevalencia por año y área  
tabla\_area <- svyby(  
 ~ DBM\_anemia\_stunting\_factor,  
 ~ year + area\_residencia\_factor,  
 design = data\_dbm\_svy,  
 svymean,  
 na.rm = TRUE) |>  
 as\_tibble() |>  
 mutate(  
 area = area\_residencia\_factor,  
 prevalencia = DBM\_anemia\_stunting\_factorSi \* 100) |>  
 select(year, area, prevalencia)  
  
# Prevalencia total por año  
tabla\_total <- svyby(  
 ~ DBM\_anemia\_stunting\_factor,  
 ~ year,  
 design = data\_dbm\_svy,  
 svymean,  
 na.rm = TRUE) |>  
 as\_tibble() |>  
 mutate(  
 area = "Total",   
 prevalencia = DBM\_anemia\_stunting\_factorSi \* 100) |>  
 select(year, area, prevalencia)  
  
# Combinar en una sola tabla  
tabla\_final <- bind\_rows(tabla\_area, tabla\_total) |>  
 arrange(year, factor(area, levels = c("Total", "Rural", "Urbano")))  
  
tabla\_final <- tabla\_final |>  
 mutate(area = factor(area, levels = c("Total", "Rural", "Urbano")))

F1 <- ggplot(  
 data = tabla\_final,   
 aes(  
 x = year,  
 y = prevalencia,  
 group = area,   
 colour = area,   
 linetype = area)) +  
 geom\_line(linewidth = 1.2) +  
 geom\_point(size = 2) +  
 labs(y = "Prevalencia de DCMI (%)") +  
 scale\_y\_continuous(limits = c(0, 30), breaks = seq(0, 100, by = 5)) +  
 ggthemes::theme\_clean() +  
 ggsci::scale\_color\_jama() +  
 scale\_linetype\_manual(  
 values = c("Total" = "solid", "Rural" = "dashed", "Urbano" = "dashed")) +  
 theme(  
 legend.position = "bottom",   
 legend.title = element\_blank(),  
 legend.text = element\_text(size = 12, color = "black", family = "Arial"),  
 legend.background = element\_blank(),  
 axis.title.x = element\_blank(),  
 axis.text = element\_text(size = 12, color = "black", family = "Arial"),  
 axis.title = element\_text(size = 12, colour = "black", family = "Arial"),  
 axis.text.y = element\_text(margin = margin(l = 10)))  
  
# View  
F1



# Save outputs

# Save figure 1 (PNG)  
ggsave(  
 F1,   
 filename = here("outputs", "prevalence\_plot.png"),   
 width = 10,   
 height = 5,   
 dpi = 500)  
  
# Save figure 1 (JPEG)  
ggsave(  
 F1,   
 filename = here("outputs", "prevalence\_plot.jpeg"),   
 width = 10,  
 height = 5,  
 dpi = 500)