

SCIO

December 8, 2020

Data preparation

The raw data consists on two dataframes, SCIO and DXA. The former contains the results of all the SCIO measurements, plus additional feature information for each measurement. The latter contains the results from the DXA measurements. In detail:

SCIO columns:

- folio: woman identifier.
- mama: breast identifier (right, left, left2).
- ubicacion: location of SCIO shot (3pm, 6pm, 9pm, 12pm, pezon).
- feature columns: BMI, weight, height, fitzpatrick color, nipple radius, bra size, age.
- spectrum columns: 331 SCIO normalized measurements for 331 different spectrum values.

DXA columns:

- folio: woman identifier.
- mama: breast identifier (right, left, left2).
- DXA density: DXA density measurement.

After merging both tables we get

- 197 folios.
- 3 breast per folio.
- 5 locations per breast.
- 3 spectrum measurements per location

Variability within left breast measurements

We examine now how much does the variability of the SCIO measurements can be attributed to differences in the breast locations, and how much is due the imprecisions of the instrument. The following histogram illustrates the percentage variation between the two measurements of the left breast for each folio, and the percentage variation of a folio with respect to the average.

More in concrete:

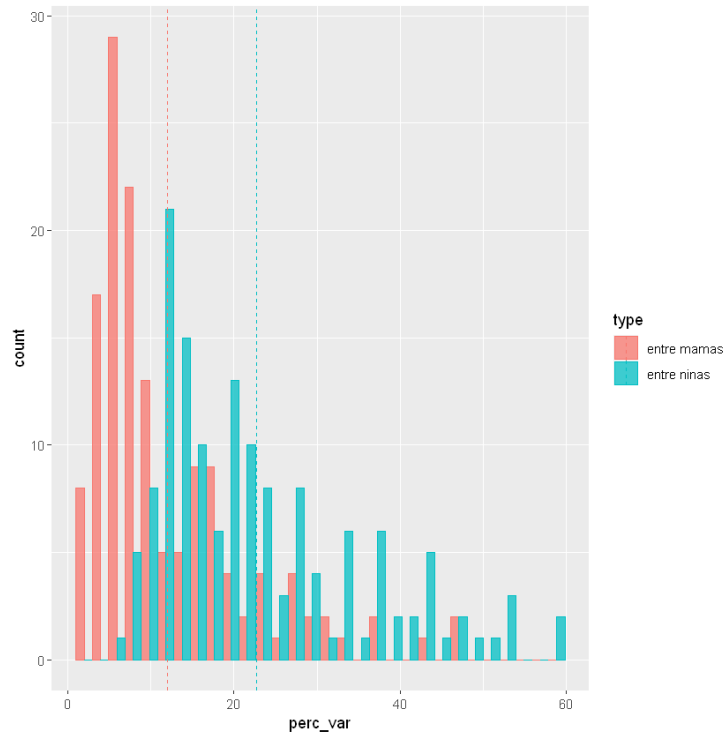


Figure 1: .

Around 30% of folios have a left breast variation that is indistinguishable from the variation between folios. From now on, we drop these folios.

1 Analysis

The following is the distribution of spectrum values per breast by BMI of folio:

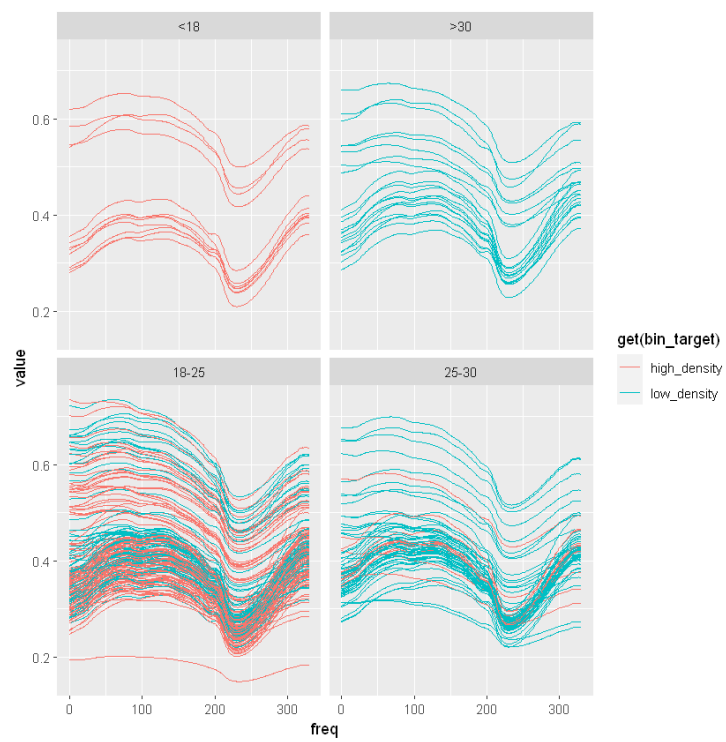


Figure 2: .