Data mining with R:

Background:

Fetal well being is a crucial issue during pregnancy as fetal growth restriction (FGR) can lead to adverse perinatal outcomes. The diagnosis of FGR can be performed by means of weight control by ultrasound, usually a percentile below 10%, which is small for gestational age, and is a marker of growth problems, although not in all cases. To control the fetus’s weight during pregnancy, international growth charts are derived. The most used are the Integrowth-21st and the Who model, but there is controversy about its general use. Local charts can be better adjusted to describe the fetus’s growth in local populations.

Methodology:

We propose to build a growth model that adjusts fetus’s weight during pregnancy. Synthetic data will be derived from a retrospective cohort study of births assisted at the Miguel Servet University Hospital (MSUH), between March 2012 and December 2016. The inclusion criteria were as follows: live singleton pregnancies controlled in MSUH from the first trimester of gestation; fetal ultrasound assessment at gestational age of 35 (range 34-36) weeks; and deliveries between 37 and 42 weeks of gestational age with fetuses without stillbirth associated with malformations or chromosomal abnormalities. Of the 19 310 consecutive deliveries assisted in our Hospital in the period studied, the 9 585 cases which fulfilled the specific inclusion criteria such as data availability to estimate percentile weights by standards were considered for the analysis.

Synthetic data will follow a data curation process, exploring outliers and considering data imputation. Descriptive analysis of data will be performed, and a growth model will be built using traditional Hadlock growth model, which is a polynomial model of the logarithm of fetus weight versus gestational age polynomial. In this model, the variation coefficient of Pearson is assumed as constant, and confidence intervals are derived under normality assumption.

In addition, taking into account the non-independence of ultrasound weights, mixed linear model will be derived taken as groups for random effects the fetus. This methodology is similar to Intergrowth-21st where the variance is adjusted depending on gestational age.

Validation:

Both models will be compared on local data. As far as will be possible it should be recommended that growth model will be validated on Spanish and French population. Therefore, tutor and students will be in contact with hospitals to validate the model in their own population, estimating the percentage of SGA derived by the growth model and the international models: Intergrowth21st and Who model.

Expected results will be similar to the figure below



App: For the use of the developed model, a shiny app will be derived and deposited in a free repository on shinyapps.io, similar to [https://quironsaludhurjc.shinyapps.io/appGCQuironMadridHURJC/](https://quironsaludhurjc.shinyapps.io/appGCQuironMadridHURJC/?_ga=2.157391731.831860341.1678714503-1179186838.1675189766)

The proposal covers all steps in data analysis, exploring data, building models, validating them, and providing tools for its use. All analyses are based on data mining with R language programming (The R foundation for Statistical computing, Vienna, Austria).