



Swiss Institute of
Bioinformatics



Regression

Isabelle Dupanloup



www.sib.swiss

birthwt {MASS} Risk Factors Associated with Low Infant Birth Weight

The birthwt data frame has 189 rows and 10 columns. The data were collected at Baystate Medical Center, Springfield, Mass during 1986.

This data frame contains the following columns:

- low: indicator of birth weight less than 2.5 kg
- age: mother's age in years
- lwt: mother's weight in pounds at last menstrual period
- race: mother's race (1 = white, 2 = black, 3 = other)
- smoke: smoking status during pregnancy
- ptl: number of previous premature labours
- ht: history of hypertension
- ui: presence of uterine irritability
- ftv: number of physician visits during the first trimester
- bwt: birth weight in grams

Apply several regression models to the birthwt dataset

1. Inspect the dataset
2. Does the mother's age allow us to predict the birth weight below 2.5 kg ?
3. Can you use a linear regression model ? What's wrong with this model ?
4. Which model is better adapted to the current situation ?
5. Check the assumptions of the model
6. Can you better predict birth weight using below 2.5 kg using other variables in the dataset, in addition to mother's age ?