

Statistical Model To Predict The Weight Of Newborns

Enrico Michelon

Introduction

This project concerns the creation of a statistical model to predict the weight of newborns. Our objective is to create a statistical model given the *neonati.csv* dataset that can be extended to the entire population.

Anni.madre	N.gravidanze	Fumatrici	Gestazione	Peso	Lunghezza	Cranio	Tipo.parto	Ospedale	Sesso
26	0	0	42	3380	490	325	Nat	osp3	M
21	2	0	39	3150	490	345	Nat	osp1	F
34	3	0	38	3640	500	375	Nat	osp2	M
28	1	0	41	3690	515	365	Nat	osp2	M
20	0	0	38	3700	480	335	Nat	osp3	F
32	0	0	40	3200	495	340	Nat	osp2	F

Dataset

Studing the first rows of the dataset, we can distinguish 10 variables: Anni.madre, N.gravidanze, Fumatrici, Gestazione, Peso, Lunghezza, Cranio, Tipo.parto, Ospedale and Sesso.

Anni.madre

Anni.madre is a quantitative variable on radio scale. In the dataset we have at least two outliers, which can be found at rows 1152 and 1380, and report an age of 1 and 0, respectively. We decide to compute position measures and standard deviation excluding those rows, and we obtain:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std.dev
13	25	28	28.19	32	46	5.27

N.gravidanze

Fumatrici

Gestazione

Peso

Lunghezza

Cranio

Tipo.parto

Ospedale

Sesso