

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.

Table 1: Dataset first rows

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 ratio 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. Computing position measures and standard deviation excluding those rows, we obtain:

Table 2: Position measures and standard deviation for Anni.madre

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std.dev
13	25	28	28.19	32	46	5.22

N.gravidanze

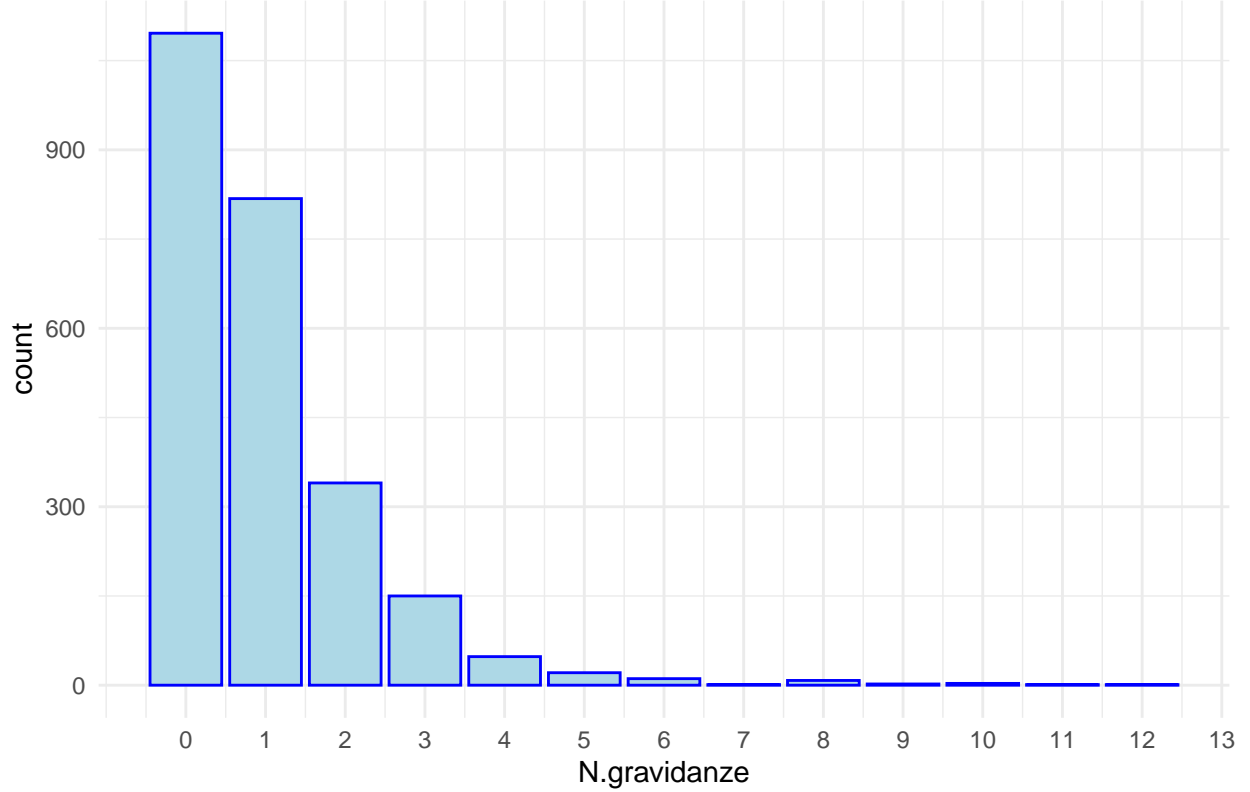
N.gravidanze is a quantitative variable on ratio scale. In Table 3 position measures and standard deviation for the variable are shown. We can see that mean and standard deviation and third interquartile are around 1 (0.98, 1.28 and 1 respectively), while the maximum reaches a value of 12.

Table 3: Position measures and standard deviation for Anni.madre

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std.dev
0	0	1	0.98	1	12	1.28

We can look now at the distribution of *N.gravidanze*. From the graphic we can notice that it is a normal positive skewed distribution, with mean 0.98 and standard deviation of 1.28.

Continuous Density Plot



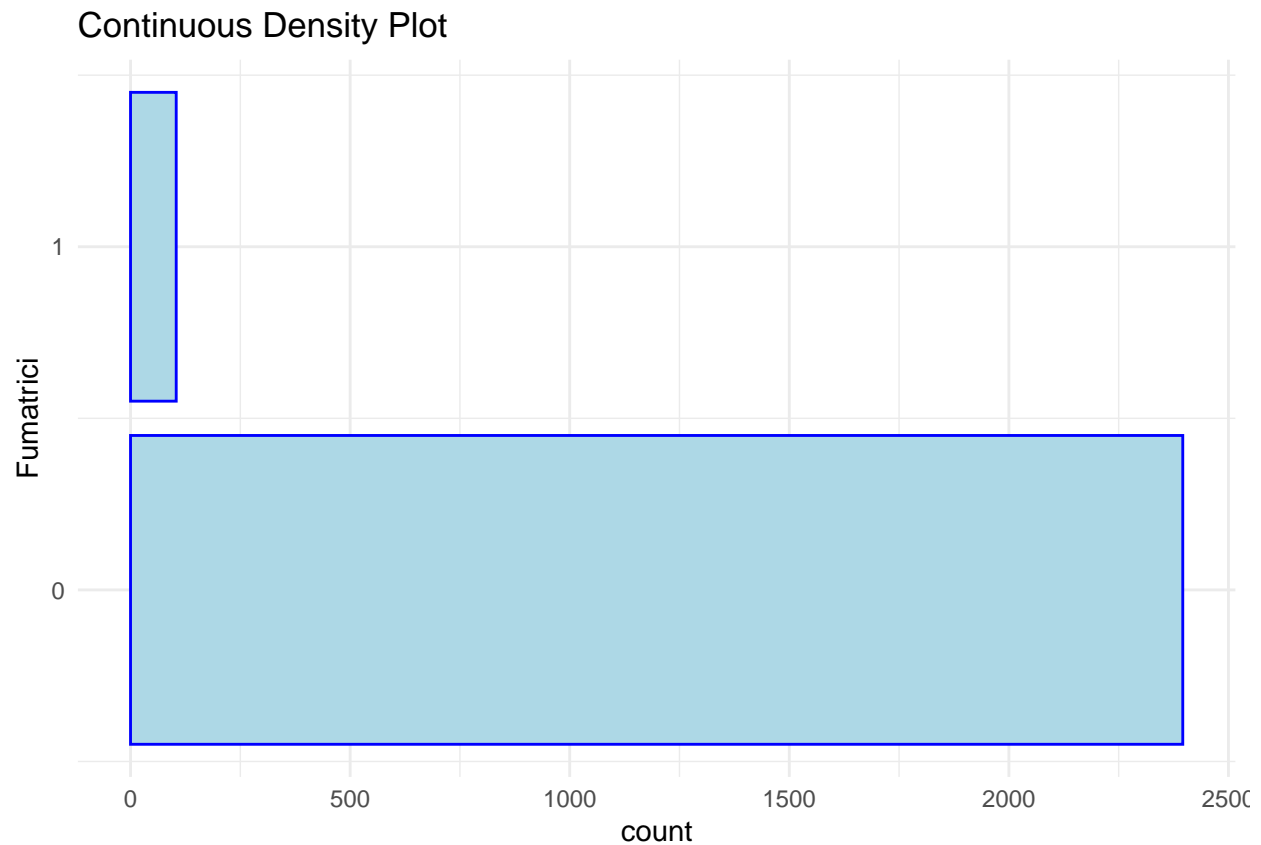
Fumatrici

Fumatrici is a qualitative encoded variable on nominal scale, with values 0 and 1. Value 0 means that the mother is not a smoker, while mothers with “Fumatrici” value of 1 means she is a smoker. In Table 4 we can observe measures of position and standard deviation.

Table 4: Position measures and standard deviation for Fumatrici

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std.dev
0	0	0	0.04	0	1	0.2

Media of 0.04, means that the 60% of the mothers are not smokers. We can also observe it from the next figure.



Gestazione

Gestazione is a quantitative variable on ratio scale, with position and standard deviation measures that can be seen in Table 5.

Table 5: Position measures and standard deviation for Gestazione

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std.dev
25	38	39	38.98	40	43	1.87

Peso

Lunghezza

Cranio

Tipo.parto

Ospedale

Sesso