Statistical Data Science (Module 1): Homework 1

Table 1: The Table shows the descriptive statistics of the "Births" dataset, which contains 500 singleton births in London. The **Variable** column in the Table represents the eight columns in a dataset. The number of non-null observations in the dataset is represented by **# of non-null observations.** Each variables minimum and maximum values are represented by the **Min** and **Max** columns. The additional columns represent the aggregate values, such as **Mean**, **Standard Deviation (SD)**, and **percentiles (25, 50, 75).**

Variable	# of non-null observations	Min	Mean	SD	25 percentile	50 percentile (Median)	75 percentile	Max
id	500	1	250.5	144.482	125.75	250.5	375.25	500
bweight	500	628	3136.884	637.451	2862	3188.5	3551.25	4553
lowbw	500	0	0.12	0.325	0	0	0	1
gestwks	490	24.69	38.722	2.314	37.94	39.125	40.09	43.16
preterm	490	0	0.129	0.335	0	0	0	1
matage	500	23	34.028	3.898	31	34	37	43
hyp	500	0	0.144	0.351	0	0	0	1
sex	500	1	1.472	0.5	1	1	2	2

Table 2: The Table represents four columns. First and second column represents the information such as mean birth weight of preterm and term births, respectively. Second and third columns corresponds the difference in means (percentages) and p-value used in **t-test** for comparing the two groups.

	Preterm birth (preterm = 1)	Term birth (pre-term = 0)	Difference	p-value
Mean birth weight	2168.825	3281	-1112.175	2.2e-16
Mean maternal age	33.9	34.079	-0.17487	0.7484
% of male babies	49.2	52.693	-3.48686	0.6292
% of mothers with hypertension	30.15	13.86	17.98074	0.003973

Figure 1: The figure shows the density estimation of the **Birth Weights** based on **Gender** information.