

PCB data validation (before 2004 and after 2004)

Xuelong Wang

2018-09-28

Contents

1	Motivation	1
2	Comparison of the 2000 and 2005 PCB data	1
2.1	sample size	1
2.2	box-plot comparison for each covairates	2

1 Motivation

To get a better model performance evaluation of the proposed method, we need make sure that the sample size n is greater than the combined covaraites(main and interaction) p . Therefore, we want to add more observations to our simuation dataset. But the format of the PCB data has changed a little after 2004 in the nhanes database. So we need to run some basic data validation steps to make sure that we can combine those data from different years together.

2 Comparison of the 2000 and 2005 PCB data

2.1 sample size

PCB_2000 1000 PCB_2005 247

2.2 box-plot comparison for each covairates







