Simulation of microarray data

Dataset number 4

Assumptions

- We are going to simulate in a simple way microarray data for WT and KO mice.

- Microarray data are supposedly normally distributed.

- Furthermore, we will suppose that there are 10'000 genes

- Among those 100 are taken with a difference of one in mean between WT and KO mice.

- Take 10 WT mice and 10 KO mice.

Exercise

- Generate a dataset modeling that problem.

- Explore and describe the dataset.

- Plot it!

- Check if the data is normally distributed as expected.

- Find out how many significant genes there are, compare them to the 100 true significant genes.

- Change the number of mice to 10 WT and 20 KO and repeat the exercise. Change the difference in means and redo the exercise. How were the results affected ?

Bonus exercise

- Bonus: Simulate data such that on the x axis you plot the variance and the y axis the percentage of true positives found (significant genes that are true significant genes) and the percentage of true negatives (non significant genes that are non significant).