

## **Kolmogorov–Smirnov test**

'k-s\_data.csv' file provides grades of two independent groups.

The following tutorial asks whether the grades of group A distribute similarly to the grades of group B.

1. State the hypothesis for testing the grades distribute similarly or not.
2. load data and present the distributions. (you can use simple histogram function). What is your intuitive guess regarding the hypothesis?
3. construct the cdf vectors of the two groups in a new table

For this, span the x axis between below the minimal grade and 100 in bins as you find fit.

Next, for each bin, and each group, calculate the percent to entries below or equal to the cutoff value.

4. calculate the difference between groupA cdf and groupB cdf (the vector D).
5. find the maximal difference between the cdf functions of the two groups.
6. present the cdfs and the D vector in a plot. (use "par(new=TRUE)" to overly plots into one figure).
7. Conclude in 5% significance level if the grades distribute similarly in the two groups.