Step 1. Start with our raw fastq datasets.

We can use HBP or HiC-Pro directly to pre-processing our fastq datasets. If we use HBP, we should use following command to generate enzyme site file at first:

generate_enzyme_file(enzyme="DpnII",enzymesite="GATC",chrom_file="chrom_dm3.sizes",enzymedir="annotation",enzymeoverhangs5=0,genomeName="dm3",resolution=5)

```
> generate_enzyme_file(enzyme="DpnII",enzymesite="GATC",chrom_file="chrom_dm3.sizes",enzymedir="annotation",enzymeoverhangs5=0,genomeName="dm3",resolution=5)
Get restriction fragments for chr2L ...
Get restriction fragments for chr2R ...
Get restriction fragments for chr3L ...
Get restriction fragments for chr3R ...
Get restriction fragments for chr4 ...
Get restriction fragments for chrX ...
Get restriction fragments for chrU ...
Get restriction fragments for chrM ...
Get restriction fragments for chr2LHet ...
Get restriction fragments for chr2RHet ...
Get restriction fragments for chr3LHet ...
Get restriction fragments for chr3RHet ...
Get restriction fragments for chrXHet ...
Get restriction fragments for chrYHet ...
Get restriction fragments for chrUextra ...
There were 14 warnings (use warnings() to see them)
```

After that, we will get the enzyme sites file in the annotation dir. This file is just look like following picture. Then, we can use the following command to processing raw dataset.

run_hicpro(hicpro_path = "HiC-Pro",inputfile = "rawdata",configfile = "config-hicpro.txt",outdir = "demoout")

DpnII_resfrag_dm3.bed

```
1  chr2L 0 79  HIC_chr2L_1 0 +
2  chr2L 79  228 HIC_chr2L_2 0 +
3  chr2L 228 369 HIC_chr2L_3 0 +
4  chr2L 369 537 HIC_chr2L_4 0 +
5  chr2L 537 686 HIC_chr2L_5 0 +
6  chr2L 686 827 HIC_chr2L_6 0 +
7  chr2L 827 995 HIC_chr2L_7 0 +
8  chr2L 995 1144  HIC_chr2L_8 0 +
9  chr2L 1144  1285  HIC_chr2L_9 0 +
10 chr2L 1285  1453  HIC_chr2L_10  0 +
11 chr2L 1453  1602  HIC_chr2L_11  0 +
12 chr2L 1602  1743  HIC_chr2L_12  0 +
13 chr2L 1743  1911  HIC_chr2L_13  0 +
14 chr2L 1911  2060  HIC_chr2L_14  0 +
15 chr2L 2060  2201  HIC_chr2L_15  0 +
16 chr2L 2201  2369  HIC_chr2L_16  0 +
17 chr2L 2369  2518  HIC_chr2L_17  0 +
18 chr2L 2518  2659  HIC_chr2L_18  0 +
19 chr2L 2659  2827  HIC_chr2L_19  0 +
20 chr2L 2827  2976  HIC_chr2L_20  0 +
21 chr2L 2976  3117  HIC_chr2L_21  0 +
```

Step 2. Generate interaction frequency matrix.

We can get the interaction file from dir “demoout\hic_results\matrix\SRR389756_split\iced\5000\SRR389756_split_5000_iced.matrix”, and the index file from dir “demoout\hic_results\matrix\SRR389756_split\raw\5000\SRR389756_split_5000_abs.bed”. And then copy these file to the work dir, and use following command to generate matrix:

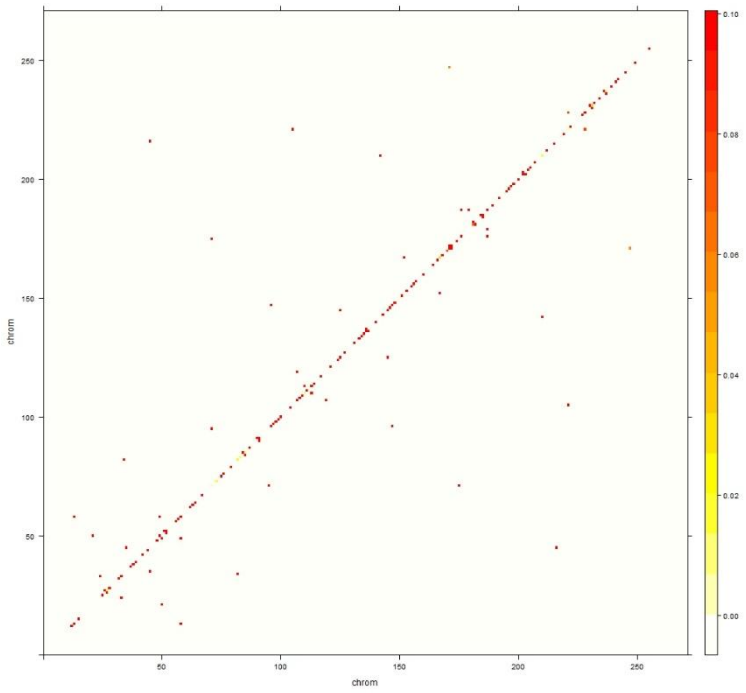
generate_matrix(all_hic_file = "SRR389756_split_5000_iced.matrix",all_bed_file = "SRR389756_split_5000_abs.bed",matrix="dm3_5k",resolution = 5,chrom_file = "chrom_dm3.sizes")

```
> generate_matrix(all_hic_file = "SRR389756_split_5000_iced.matrix",all_bed_file = "SRR389756_split_500
0_abs.bed",matrix="dm3_5k",resolution = 5,chrom_file = "chrom_dm3.sizes")
[1] "Thu Jun 29 11:13:57 2017 chr2L finish!"
[1] "Thu Jun 29 11:15:10 2017 chr2R finish!"
[1] "Thu Jun 29 11:16:47 2017 chr3L finish!"
[1] "Thu Jun 29 11:18:52 2017 chr3R finish!"
[1] "Thu Jun 29 11:18:53 2017 chr4 finish!"
[1] "Thu Jun 29 11:20:14 2017 chrX finish!"
[1] "Thu Jun 29 11:20:30 2017 chrU finish!"
[1] "hic data 0 finish"
[1] "Thu Jun 29 11:20:31 2017 chrM finish!"
[1] "Thu Jun 29 11:20:31 2017 chr2LHet finish!"
[1] "Thu Jun 29 11:20:33 2017 chr2RHet finish!"
[1] "Thu Jun 29 11:20:35 2017 chr3LHet finish!"
[1] "Thu Jun 29 11:20:36 2017 chr3RHet finish!"
[1] "Thu Jun 29 11:20:37 2017 chrXHet finish!"
[1] "Thu Jun 29 11:20:37 2017 chrYHet finish!"
[1] "Thu Jun 29 11:22:55 2017 chrUextra finish!"
```

The results are just like this picture:

chr2L.matrix	2017/6/29 11:11	MATRIX 文件	41,438 KB
chr2L_heatmap.jpeg	2017/6/29 11:11	JPEG 图像	73 KB
chr2LHet.matrix	2017/6/29 11:20	MATRIX 文件	11 KB
chr2LHet_heatmap.jpeg	2017/6/29 11:20	JPEG 图像	69 KB
chr2R.matrix	2017/6/29 11:14	MATRIX 文件	35,008 KB
chr2R_heatmap.jpeg	2017/6/29 11:14	JPEG 图像	78 KB
chr2RHet.matrix	2017/6/29 11:20	MATRIX 文件	848 KB
chr2RHet_heatmap.jpeg	2017/6/29 11:20	JPEG 图像	77 KB
chr3L.matrix	2017/6/29 11:15	MATRIX 文件	47,129 KB
chr3L_heatmap.jpeg	2017/6/29 11:15	JPEG 图像	73 KB
chr3LHet.matrix	2017/6/29 11:20	MATRIX 文件	514 KB
chr3LHet_heatmap.jpeg	2017/6/29 11:20	JPEG 图像	80 KB
chr3R.matrix	2017/6/29 11:16	MATRIX 文件	60,937 KB
chr3R_heatmap.jpeg	2017/6/29 11:17	JPEG 图像	76 KB
chr3RHet.matrix	2017/6/29 11:20	MATRIX 文件	499 KB

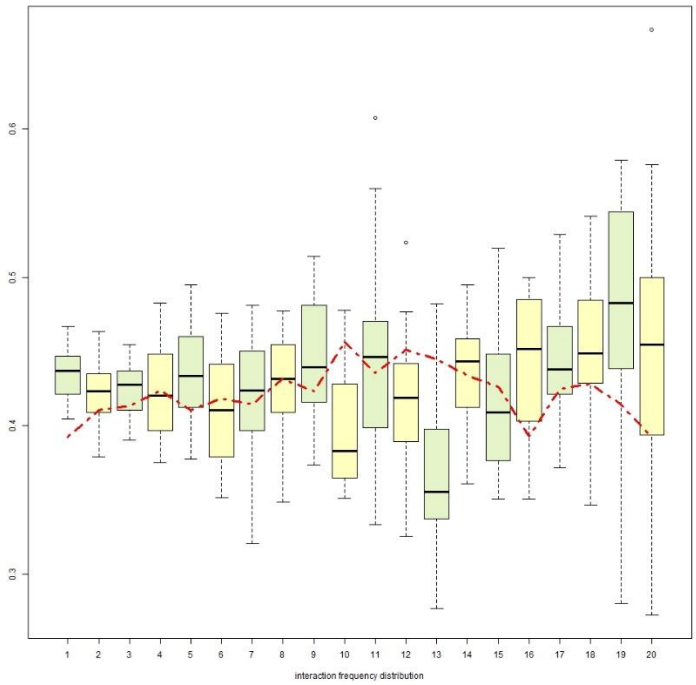
HBP generate interaction frequency matrix file and heatmap at here.



Step 3. Interaction frequency distribution analysis

Then we can use the command to make Interaction frequency distribution analysis:
if_distribution_analysis(all_hic_file = "SRR389756_split_5000_iced.matrix",all_bed_file = "SRR389756_split_5000_abs.bed",bedFile = "dm3_mars.bed",inter_chromfile = NULL,groupNum = 20,random_analysis = TRUE,threshold_percent = 0.005,if_bin_number = 20,matrix_dir = "dm3_5k",slide_window = TRUE)
The results are looking like following picture.

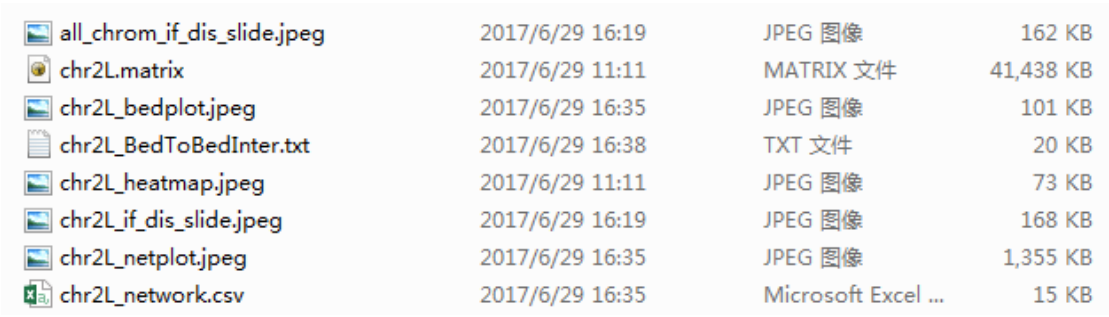
all_chrom_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	162 KB
chr2L.matrix	2017/6/29 11:11	MATRIX 文件	41,438 KB
chr2L_heatmap.jpeg	2017/6/29 11:11	JPEG 图像	73 KB
chr2L_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	168 KB
chr2LHet.matrix	2017/6/29 11:20	MATRIX 文件	11 KB
chr2LHet_heatmap.jpeg	2017/6/29 11:20	JPEG 图像	69 KB
chr2LHet_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	63 KB
chr2R.matrix	2017/6/29 11:14	MATRIX 文件	35,008 KB
chr2R_heatmap.jpeg	2017/6/29 11:14	JPEG 图像	78 KB
chr2R_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	179 KB
chr2RHet.matrix	2017/6/29 11:20	MATRIX 文件	848 KB
chr2RHet_heatmap.jpeg	2017/6/29 11:20	JPEG 图像	77 KB
chr2RHet_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	63 KB
chr3L.matrix	2017/6/29 11:15	MATRIX 文件	47,129 KB
chr3L_heatmap.jpeg	2017/6/29 11:15	JPEG 图像	73 KB
chr3L_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	168 KB



Step 4. Interaction network topological analysis
At here we can use the following command to make interaction network topological analysis:
network_analysis(bedFile = "dm3_mars.bed",matrix_dir = "dm3_5k",resolution = 5)

```
Console G:/R/HBP/demo/
> network_analysis(bedFile = "dm3_mars.bed",matrix_dir = "dm3_5k",resolution = 5)
[1] "dm3_5k/chr2L.matrix"
[1] "chr2L"
Mapped Fragments: 1152
[1] "interacion bin 1000 finish"
[1] "interacion bin 2000 finish"
[1] "interacion bin 3000 finish"
[1] "interacion bin 4000 finish"
[1] 0.9735668
[1] "plot chr2Lbed picture"
[1] "dm3_5k/chr2LHet.matrix"
[1] "dm3_5k/chr2R.matrix"
[1] "chr2R"
Mapped Fragments: 1117
[1] "interacion bin 1000 finish"
[1] "interacion bin 2000 finish"
[1] "interacion bin 3000 finish"
[1] "interacion bin 4000 finish"
[1] 0.9703837
```

After this, we can get some plot and list at the dir “dm3_5k”:



all_chrom_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	162 KB
chr2L.matrix	2017/6/29 11:11	MATRIX 文件	41,438 KB
chr2L_bedplot.jpeg	2017/6/29 16:35	JPEG 图像	101 KB
chr2L_BedToBedInter.txt	2017/6/29 16:38	TXT 文件	20 KB
chr2L_heatmap.jpeg	2017/6/29 11:11	JPEG 图像	73 KB
chr2L_if_dis_slide.jpeg	2017/6/29 16:19	JPEG 图像	168 KB
chr2L_netplot.jpeg	2017/6/29 16:35	JPEG 图像	1,355 KB
chr2L_network.csv	2017/6/29 16:35	Microsoft Excel ...	15 KB

The chr2L_bedplot.jpeg is the Hi-C heatmap with specific sites. The chr2L_BedToBedInter.txt is a list contains data about the interaction, which is generated by HBP and used when processing datasets. The chr2L_netplot.jpeg is topological network plot of this chromosome. If the range is too big, this picture maybe not clearly to investigate, and can be optimize by adjusting parameters. And the chr2L_network.csv is the list of nodes in this network. This list contains degree, closeness, betweenness, local cluster coefficient, eigenvector centrality and cluster membership information of these nodes.

Step 5. Visualization of interactions and tracks.

We can use following command to plot circos picture of this network:

circos_plot(bedFile = "dm3_mars.bed", wig_dir = "dm3wig",matrix_dir = "dm3_5k", outputpdf = FALSE,resolution = 5)

```
> circos_plot(bedFile = "dm3_mars.bed", wig_dir = "dm3wig",matrix_dir = "dm3_5k", outputpdf = FALSE,res
olution = 5)
[1] "generate genome frame"
[1] "chr2L frame generate finished!"
[1] "chr2LHet frame generate finished!"
[1] "chr2R frame generate finished!"
[1] "chr2RHet frame generate finished!"
[1] "chr3L frame generate finished!"
[1] "chr3LHet frame generate finished!"
[1] "chr3R frame generate finished!"
[1] "chr3RHet frame generate finished!"
[1] "chr4 frame generate finished!"
[1] "chrM frame generate finished!"
[1] "chrU frame generate finished!"
[1] "chrUextra frame generate finished!"
```

The wig file is stored in the wig_dir:

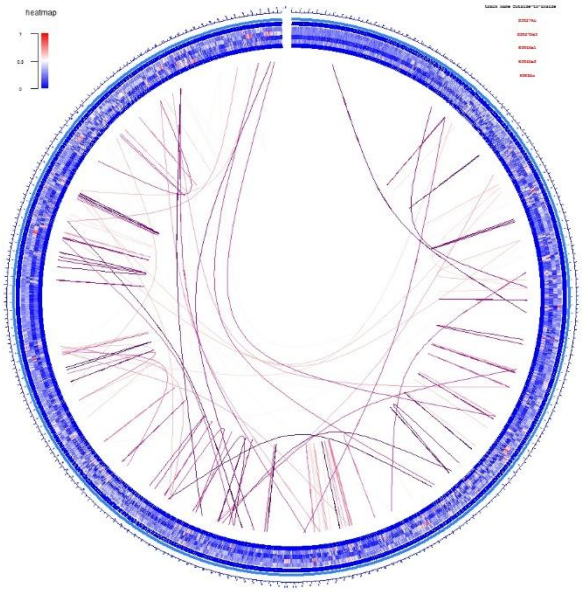
```
G:\R\HBP\demo\dm3wig>ls
H3K27Ac.wig  H3K27Me3.wig  H3K4Me1.wig  H3K4Me3.wig  H3K9Ac.wig

G:\R\HBP\demo\dm3wig>
```

And these wig files are downloaded from UCSC:

H3K4Me1.wig

```
chr2L 5110 5145 1.21343350926097
chr2L 5210 5245 0.357175764549395
chr2L 5310 5345 -1.33594854414335
chr2L 5410 5445 -3.40173370133267
chr2L 5510 5545 -5.11492778000223
chr2L 5610 5645 -5.91298426985645
chr2L 5710 5745 -5.70009960020549
chr2L 5810 5845 -4.74063295661428
chr2L 5910 5945 -3.49139369748715
chr2L 6010 6045 -2.45439555963301
chr2L 6110 6145 -1.64413308655459
chr2L 6210 6245 -1.18248381928938
chr2L 6310 6345 -0.775649440517197
chr2L 6410 6445 -0.448845242771466
chr2L 6510 6545 -0.0501316593295282
chr2L 6610 6645 0.217184751970445
chr2L 6710 6745 0.125806907069577
chr2L 6810 6845 -0.319838038094358
chr2L 6910 6945 -0.896141304122238
chr2L 7010 7045 -1.36077197600532
chr2L 7110 7145 -1.48599391641456
chr2L 7210 7245 -1.19679252195848
```

According to this step, we can get a picture named “*_ circos.jpeg”:



If the range is too big, this picture maybe cannot be see clearly in jpeg format, so we can output pdf format to make it clearly by change the parameter “outputpdf”. And the color of lines and other elements in this picture can be optimized by adjusting other parameters.

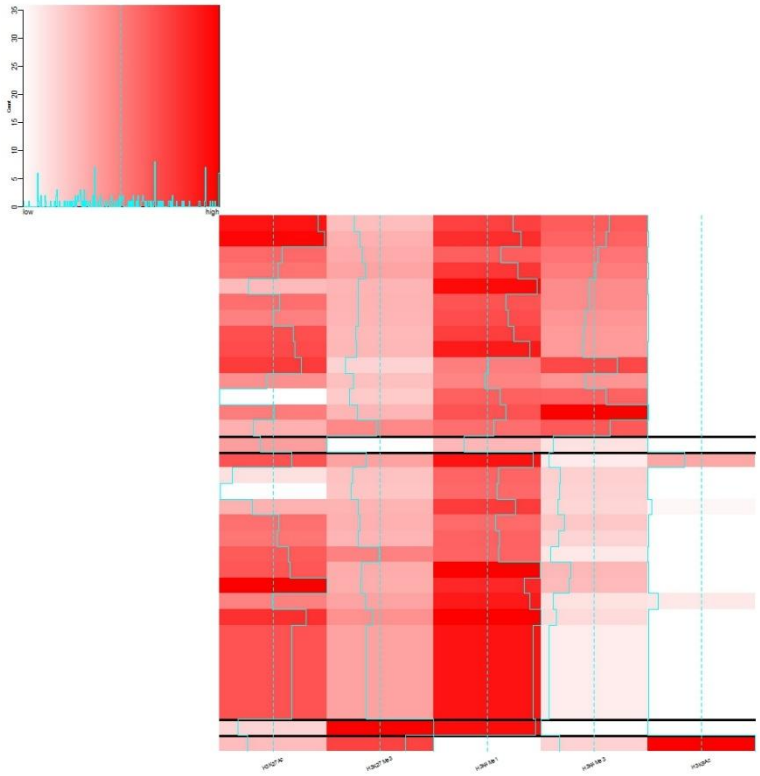
Step 6. Statistical significance tests.

HBP can use following command to calculate statistical significance:

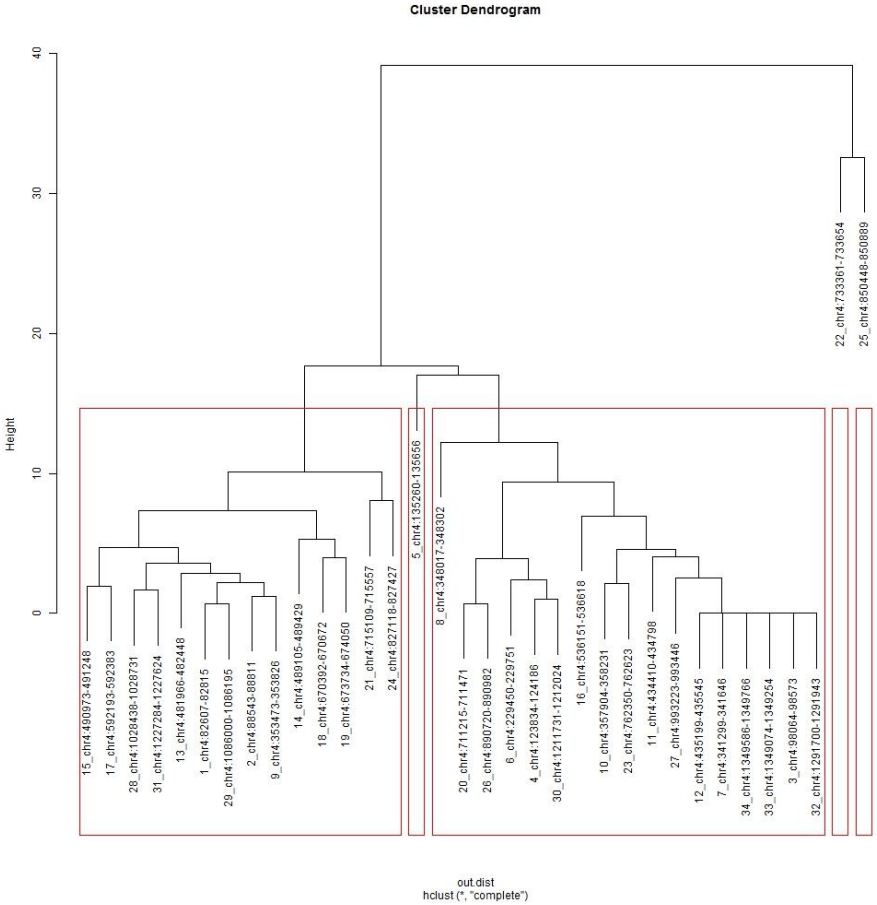
statistical_analysis(bedFile = "dm3_mars.bed",wig_dir = "dm3wig",matrix_dir = "dm3_5k",resolution=5)

```
> statistical_analysis(bedFile = "dm3_mars.bed",wig_dir = "dm3wig",matrix_dir = "dm3_5k",resolution=5,g
roupNum = 20)
[1] "dm3_5k/chr2L.matrix"
[1] "chr2L"
Mapped Fragments: 1152
[1] "Analysis start time: 2017-06-29 22:03:17"
[1] "2017-06-29 22:03:17  start make clusters"
```

And we can get several results file from this step. The chr*_cluster_heatmap.jpeg is heatmap of tracks clusters:



The chr*_cluster_tree.jpeg is the cluster tree of these node:



The chr*_statistic.txt contains the statistical difference of these interactions:

```
the statistic test of interaction frequency between b2b and b2o :

test name : Kruskal-Wallis rank sum test
Kruskal-Wallis chi-squared : 0.147419324621797
Kruskal-Wallis df : 1
Kruskal-Wallis p value : 0.701013844360815

test name : Multiple comparison test after Kruskal-Wallis
significance level : 0.05
observed difference  : 1.34166666666667
critical difference  : 6.85987394589019
exist difference  : FALSE
b2b frequency mean  : 3.6761882
b2o frequency mean  : 3.3607649



the statistic test of interaction number between b2b and o2o :

test name : t-test
numbers of random group : 20
95 percent confidence interval of random group : 11.9241495544692 ~ 14.4758504455308
numbers of b2b : 15
t test p value : 0.00816966706548562
exist difference  : TRUE
```