tsv-utils之计算指定数值列的数值统计信息: stats

一、tsv-utils stats 介绍

功能描述：

`tsv-utils stats` 计算指定数值列的数值统计信息，包括：最大值，最小值，均值，方差等；

命令行接口：

```
$ tsv-utils  stats

Usage: tsv-utils stats [options] <tsv>
Options:
  -c INT     column number [1].
  -s         skip first line.
```

可选参数：

```
-c  指定   设置要计算的列
-s         跳过第一行
```

二、使用场景实例及其用法

经典使用案例

统计序列文件的长度分布，比如做454扩增子数据分析，序列将序列截断到合适长度，可选择中位数 `median`.

示例演示

示例文件：`454.fna.gz`。

```
$ zcat 454.fna.gz | head -n 2

>EJFW8:00682:05789
GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGGTTGAATCGGAATAACTGGGCGTGAAAGCAGCACGCAGGCG
GTTTTGTTAAGTCAGATGTGGAAATCCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCG
TAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCC
CTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGATACGTCCACGCCGT
AAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGGAG
TACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGATCGAATAACCTT
```

运行命令：

获取长度分布信息：

```
$fastx-utils length 454.fna.gz | head -n 6

EJFW8:00682:05789       434
EJFW8:04202:01162       423
EJFW8:00713:05834       427
0FHCN:04212:06762       425
EJFW8:00683:05878       319
EJFW8:03285:04566       424
```

参数使用**1**：设置`-c`参数，计算第二列的数值统计信息，包括：最大值，最小值，均值，方差等。

```
$ fastx-utils length 454.fna.gz | tsv-utils  stats -c 2 -

#number sum     mean    min     max     std.dev skewness        25%-
percentile  median  75%     2.5%    97.5%   1%      99%     0.5%    99.5%
9999    3272321.000000  327.265 25      501     133.851 -1.14523        269
  415     424     37      430     30      433     28      436
```


Last Update: Friday, August 28, 2020