

工作记录19之pandas基础统计

2017-09-20 20:59:32

```
np.random.seed(1234) d1=pd.Series(2*np.random.normal(size=100)+3) d2=np.random.f(2,4,size=100)
d3=np.random.randint(1,100,size=100) d1.count()#非空元素计算 d1.min()#最小值 d1.max()#最大值 d1.idxmin()#最小
值的位置，类似于R中的which.min函数 d1.idxmax()#最大值的位置，类似于R中的which.max函数 d1.quantile(0.1)#10%分位
数 d1.sum()#求和 d1.mean()#均值 d1.median()#中位数 d1.mode()#众数 d1.var()#方差 d1.std()#标准差 d1.mad()#平均
绝对偏差 d1.skew()#偏度 d1.kurt()#峰度 d1.describe()#一次性输出多个描述性统计指标
df = pd.DataFrame(np.array([d1,d2,d3]).T,columns=['x1','x2','x3']) df.head() df.apply(stats) df.describe() df.corr() 关
于相关系数的计算可以调用pearson方法或kendall方法或spearman方法，默认使用pearson方法。 df.corr('spearman') 如果只想
关注某一个变量与其余变量的相关系数的话，可以使用corrwith,如下方只关心x1与其余变量的相关系数: df.corrwith(df['x1']) 数值
型变量间的协方差矩阵 df.cov() source:http://shujuren.org/article/207.html
```