**Python实验报告8**

**学号** 117060400230 **姓名**： 何颖岚 **班级：** 应用统计学二班  **指导老师：** 林卫中

**实验名称**： 组合数据类型2

**实验要求：运用组合数据类型进行文本词频统计，运用字典处理复杂的数据信息。**

**实验过程：例：（1）import random**

**def generatesamples1(n:int):**

**birthdays = []**

**days = [31,29,31,30,31,30,31,31,30,31,30,31]**

**for i in range(n):**

**month = randint(1,12)**

**day = randint(1,days[month-1])**

**someday = (month,day)**

**birthdays.append(someday)**

**return birthdays**

**def isRepetitive(ls):**

**set1 = set(ls)**

**if len(set1) == len(ls):**

**return False**

**else:**

**return True**

**birthdays = generateSamples1(50)**

**c=0**

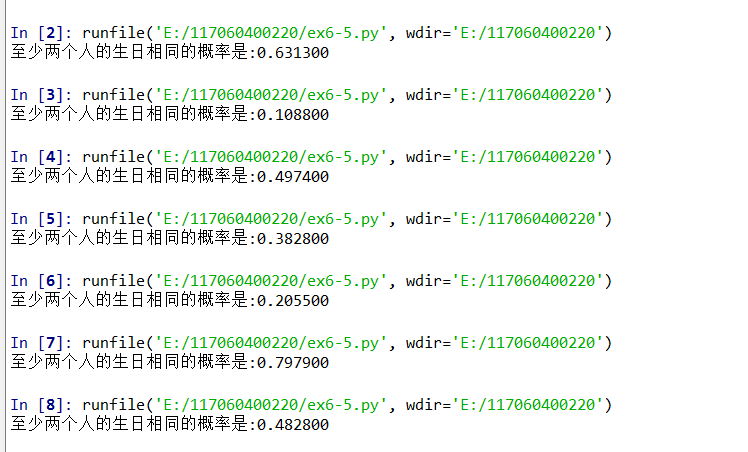
**for i in range(30000):**

**subset = random.sample(birthdays,23)**

**if isRepetitive(subset):**

**c = c+1**

**print("{}个随机样本数量下，23人中相同人数的概率是：{}".format(50,c/30000))**

输出结果：

（2）

**import jieba**

**txt = open("三国演义.txt", "r", encoding='utf-8').read()**

**words = jieba.lcut(txt)**

**counts = {}**

**for word in words:**

**if len(word) == 1:**

**continue**

**else:**

**counts[word] = counts.get(word,0) + 1**

**items = list(counts.items())**

**items.sort(key=lambda x:x[1], reverse=True)**

**for i in range(15):**

**word, count = items[i]**

**print ("{0:<10}{1:>5}".format(word, count))**

****

**实验结果：这周我们更深入的了解了python，在这个过程中我发现自己还没有特别熟悉用python写程序，需要加强动手能力，在这次实验中我了解了字典，列表，还有字典的运用。**