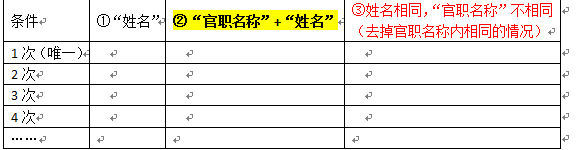
2、官职、姓名重复的情况：  
   
对于这一步需求不是很请求，请详细解释下。



李：这一部分就是统计一下同一“姓名”出现过几次，每种次数的数据量是多少。“官职名称”+“姓名”出现过几次，数据量是多少。而需要统计输出的数据就是每种次数出现的数据量和比例。

例如，

①“姓名”：1980个数据量中，“姓名”为“1次“的共有1000个，说明这1000个人只出现了1次，”姓名“为”2次”有100个，说明其中有100个人出现了2次。

②“官职名称”+“姓名”：1980个数据中， “官职名称”+“姓名”出现2次的数据量有50个，说明有50个人在同一官职名称中任职2次。

③姓名相同，“官职名称”不相同（去掉官职名称内相同的情况）：

问题2：   
“履历不正确”是指“任职时间”-“生年”小于0或者大于90的情况  
   
数据中很多数据是没有”任职时间“和”生年“的，这一步该怎么处理呢？

李：

1、筛选履历是否正确的前提是**具有履历，即“姓名内编码”=1**，加上这个条件应该可以过滤掉一些数据。

原则就是：先用任职时间和出生时间做差，如果条件不具备，再用履历中的时间和任职时间做差。（履历中的任职时间是指“时间1”或“时间2”）

 2、如果两个时间都具备，就按照条件筛选。

 3、无任职时间，有生年的情况：一般没有任职时间的情况是兼任的情况，也就是已经有官职，同时兼任这个官职；或者是任职期间有外调然后又调回来。

 如果这个官员出现过两次以上，可用其他任职时间的进行比较。

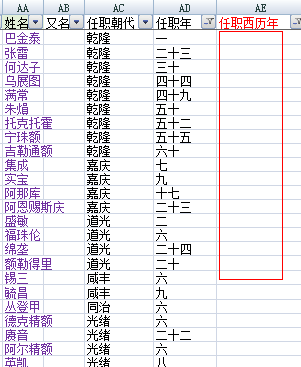
例如，张三在乾隆三十年任职，在乾隆四十-五十年间任职（只有朝代，没有具体年，任职公历年是缺失的），这种情况就可以参照乾隆三十年的数据做减法看是否符合条件。这种情况只是针对姓名出现2次以上的情况。

4、如果姓名只出现过一次，而且任职时间是缺失的，但是有出生日期履历信息中（任职时间1或任职时间2）里有时间的，也可以进行比较。那就需要用任职时间1或2与出生时间比较。（可以选取最早的时间或者最晚的时间，这个没有影响，因为时间差都在0-90之内，对结果不会产生太大影响）

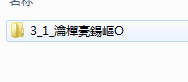
5、如果没有出生时间，只有任职时间，可以用任职时间与履历中的任职时间1或2中的任意时间求差值。

6、如果没有任职时间和出生时间，但是有履历信息，那就只能比较履历中的朝代与任职时间的朝代是否一致，这个朝代最好选取履历信息中任职时间最中间的朝代，因为最早和最晚任职时间的朝代可能出现跨朝代的情况。如果朝代相同则可以认为履历正确。（由于材料所限，朝代相同只能保证大体正确，重名的情况基本不太可能。所以这里可以认为履历正确）

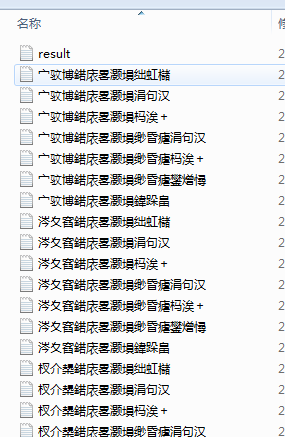
补充：请你再进行筛选一下，有部分数据的“任职时间”空白的情况，只有任职朝代和任职年，即没有换算公历年的情况，这个需要写程序进行补充下。



附件问题：

另外，查看过28日发送的结果附件，解压后的为

打开后为下图，出现了字体乱码的情况，但是 “result”中的字体是正确的。请问是什么情况？



另外，文件中一起下载下来的下图中最后一个文件不知道是什么格式和内容，对统计结果是否有影响呢？

