HW1-ANOVA

(1)

样本被随机抽样且相互独立，各总体服从正态分布，每组方差齐次

(2)

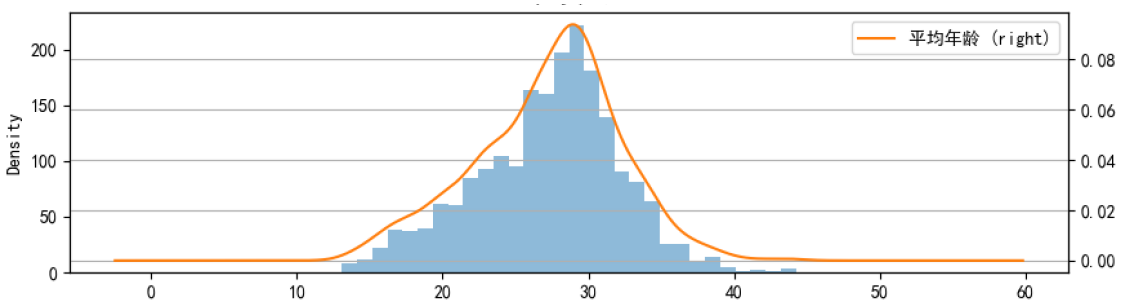
H0：不同类别的成员平均年龄是一样的

H1：至少一个类别的平均年龄是不同的

(3)

a）使用python

绘制经验概率密度函数



使用scipy的几种正态分布检验方式进行测试

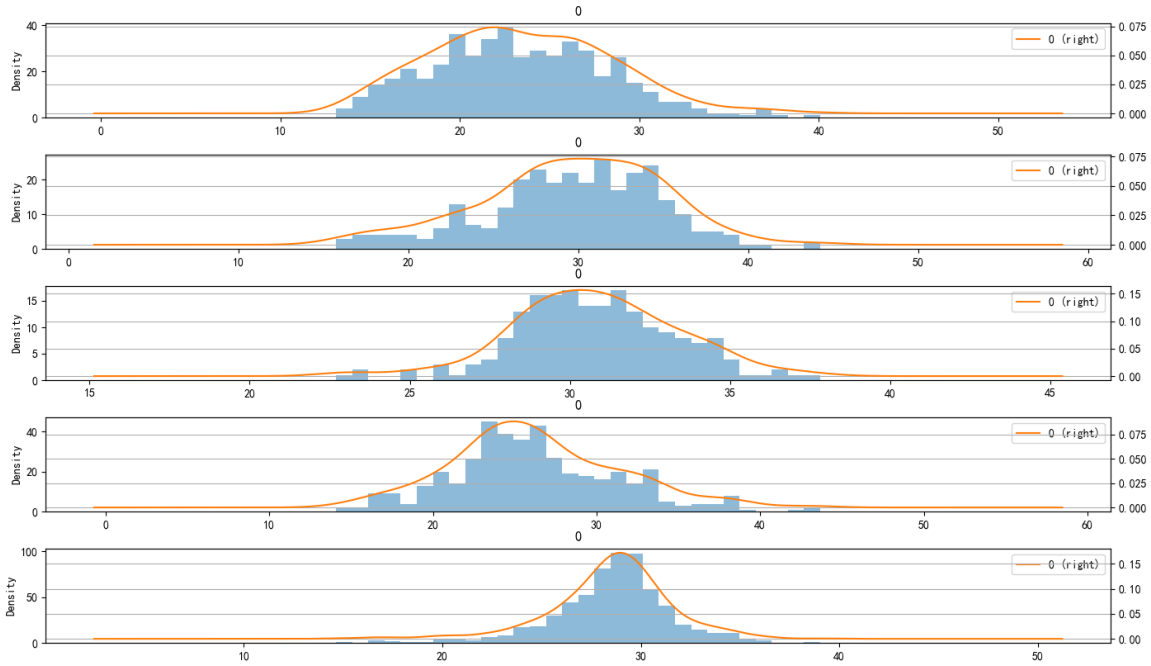
结果如下：

```
D:\Desktop\CV_ML\Courses\大数据分析\HW\HW1-ANOVA>python Ntest_All.py
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.05509327404455733, pvalue=8.012343213416506e-06)
scipy.stats.normaltest统计检验结果：----------------------------------------------------
NormaltestResult(statistic=24.479341544296894, pvalue=4.8348001102946654e-06)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9883749485015869, pvalue=8.787940861121335e-12)
```

p-value远远小于0.05，有足够的理由认为总体不符合正态分布

b）使用python

从上到下分别画出五个类别的PDF



使用scipy的几种正态分布检验方式对5个类分别进行测试，结果如下：

```
D:\Desktop\CV_ML\Courses\大数据分析\HW\HW1-ANOVA>python Ntest_Classes.py
testing class 1----------------------------------
均值为 23.404277959311987
标准差为 4.923757137352907
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.04509397908812418, pvalue=0.2704200562492669)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=7.959942946031118, pvalue=0.01868617239027858)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.9889171719551086, pvalue=0.0010064239613711834)
testing class 2----------------------------------
均值为 29.618193496517332
标准差为 5.217370879308968
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.04451187431401915, pvalue=0.5763857287698038)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=5.513226640662678, pvalue=0.06350648068044168)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.9866034984588623, pvalue=0.006912337150424719)
testing class 3----------------------------------
均值为 30.79153401603112
标准差为 2.552901286787584
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.04921847792711159, pvalue=0.7103322292131584)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=2.996117937606434, pvalue=0.22356368335514357)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.989406406879425, pvalue=0.15586760640144348)
```

```
testing class 4---------------------------------
均值为 26.255317864947997
标准差为 5.098255869880323
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.07359262797469002, pvalue=0.01901321646759231)
scipy.stats.normaltest统计检验结果：-----------------------------------------------------
NormaltestResult(statistic=15.081227215331731, pvalue=0.000531071652162856)
scipy.stats.shapiro统计检验结果：--------------------------------------------------
ShapiroResult(statistic=0.9838728308677673, pvalue=0.00011372696462785825)
testing class 5---------------------------------
均值为 28.545080703967084
标准差为 3.0189788050760367
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.0815270038579089, pvalue=0.00040480516788924596)
scipy.stats.normaltest统计检验结果：-----------------------------------------------------
NormaltestResult(statistic=90.50233804623075, pvalue=2.226727094510755e-20)
scipy.stats.shapiro统计检验结果：--------------------------------------------------
ShapiroResult(statistic=0.9546676874160767, pvalue=4.545287215183824e-13)
```

从各个结果的p-value看来 类别1，2，3 kstest的p-value大于0.05 满足正态性假设

类别4 5 则不满足正态性假设

类中最大的标准差为5.217 最小的标准差为2.553, 从经验出发不满足方差齐次性

使用levene检验方差齐次性，在统计学角度看不具有方差齐次性

```
print(stats.levene(classList[0].values.reshape(len(classList[0])),
                   classList[1].values.reshape(len(classList[1])),
                   classList[2].values.reshape(len(classList[2])),
                   classList[3].values.reshape(len(classList[3])),
                   classList[4].values.reshape(len(classList[4]))))
```

```
LeveneResult(statistic=61.01927977094263, pvalue=9.677355333795493e-49)
```

c）使用SPSS

使用Q2所答假设进行one-way ANOVA

ANOVA

平均年龄

	平方和	自由度	均方	F	显著性
组间	12782.918	4	3195.730	171.507	.000
组内	37918.618	2035	18.633		
总计	50701.536	2039			

使用python代码得到类似结果

```
F, P = stats.f_oneway(classList[0].values.reshape(len(classList[0])),
                      classList[1].values.reshape(len(classList[1])),
                      classList[2].values.reshape(len(classList[2])),
                      classList[3].values.reshape(len(classList[3])),
                      classList[4].values.reshape(len(classList[4])))
print('F value:', F)
print('P value:', P, '\n')

print(anova_lm(ols('平均年龄~C(群类别)', data).fit()))
```

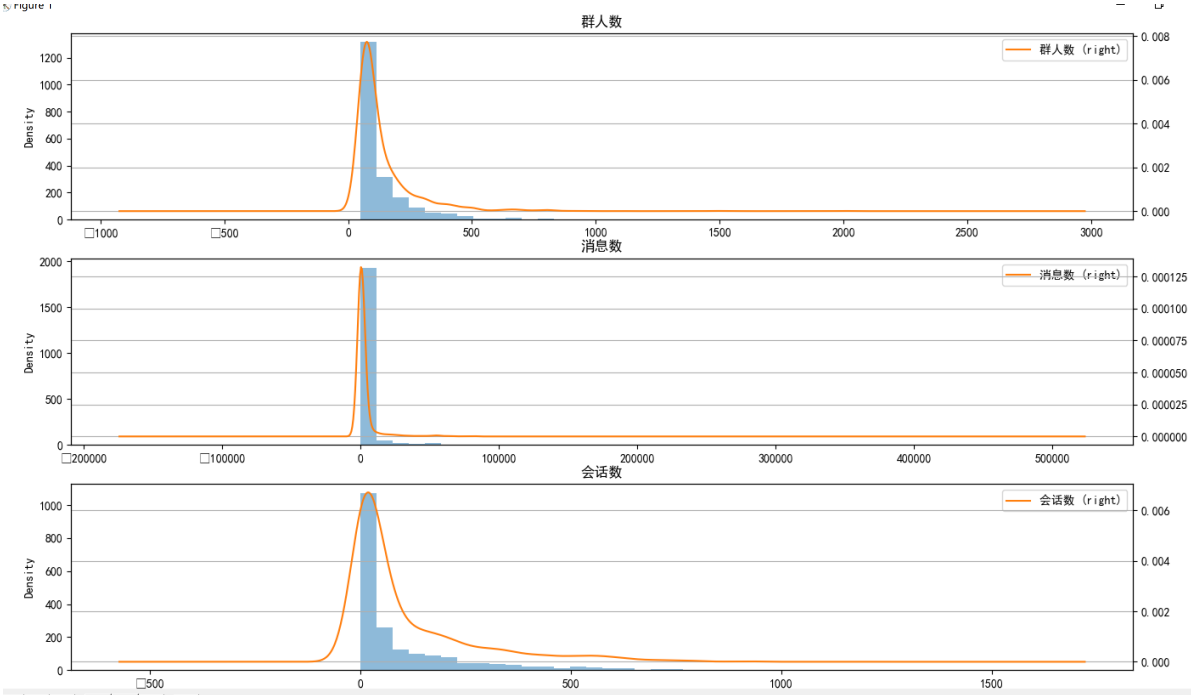
```
F value: 171.50703270711966
P value: 1.0820916064752822e-126

              df        sum_sq      mean_sq           F         PR(>F)
C(群类别)      4.0  12782.918190  3195.729547  171.507033  1.082092e-126
Residual  2035.0  37918.617834    18.633227         NaN           NaN
```

根据表格结果p-value远远小于0.05，拒绝假设H0 得到结论：不同群类别的成员平均年龄有明显差别

(4)

选择群人数、消息数和会话数

分别画PDF



使用python分别验证三列是否符合问题一假设

群人数：

```
testing col 3---------------------------------
testing class 1---------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.2888599351616917, pvalue=2.971638642171804e-36)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=628.219899752222, pvalue=3.835147379021206e-137)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.4973258972167969, pvalue=8.994981495650514e-35)
testing class 2---------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.20214852314506532, pvalue=3.182507480074423e-11)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=121.65423347163187, pvalue=3.82929523696188e-27)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.7712509036064148, pvalue=3.6642075189833225e-20)
testing class 3---------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.17880398444438306, pvalue=5.927628083473053e-06)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=84.01481018977978, pvalue=5.707103756727989e-19)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.7823090553283691, pvalue=8.633045031106147e-16)
```

```
testing class 4---------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.2697173220230211, pvalue=8.4956293745782e-28)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=430.6883551214936, pvalue=3.000626741534548e-94)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.552823543548584, pvalue=1.7453438018961164e-31)
testing class 5---------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.2375320452193162, pvalue=5.253006438848999e-32)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=568.0835163631314, pvalue=4.387648128466781e-124)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.617548942565918, pvalue=3.3403104038383513e-35)
LeveneResult(statistic=23.086127582035257, pvalue=1.1069620118393965e-18)
```

从检验结果看各组别都不满足正态性假设，同时也不具备方差齐次性

群人数不满足单因素ANOVA假设

消息数

```
testing col 4--------------------------------
testing class 1--------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.38583043252143356, pvalue=2.3926378606412736e-65)
scipy.stats.normaltest统计检验结果：----------------------------------------------------
NormaltestResult(statistic=789.8472456823433, pvalue=3.06796094552648e-172)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.2946195602416992, pvalue=2.8732461933254984e-39)
testing class 2--------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.3927434762644086, pvalue=3.2820366184333056e-42)
scipy.stats.normaltest统计检验结果：----------------------------------------------------
NormaltestResult(statistic=357.5988946872945, pvalue=2.2304200029955127e-78)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.28002768754959106, pvalue=1.6499238708141627e-32)
testing class 3--------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.366903315054822, pvalue=3.6362011554605315e-24)
scipy.stats.normaltest统计检验结果：----------------------------------------------------
NormaltestResult(statistic=245.48549494799863, pvalue=4.937442811500019e-54)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.3699495792388916, pvalue=1.2110820647058389e-25)
```

```
testing class 4--------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.4141265045680377, pvalue=1.7277705875031156e-66)
scipy.stats.normaltest统计检验结果：----------------------------------------------------
NormaltestResult(statistic=695.0763742922114, pvalue=1.1643487245694757e-151)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.20251929759979248, pvalue=6.566477597333771e-39)
testing class 5--------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.4102027038389371, pvalue=3.7990773307979597e-97)
scipy.stats.normaltest统计检验结果：----------------------------------------------------
NormaltestResult(statistic=1104.5971499662776, pvalue=1.3796741325236286e-240)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.19262337684631348, pvalue=1.401298464324817e-45)
LeveneResult(statistic=17.355146155365478, pvalue=5.1948143755841847e-14)
```

从检验结果看各组别都不满足正态性假设，同时也不具备方差齐次性

消息数不满足单因素ANOVA假设

会话数：

```
testing col 11---------------------------------
testing class 1---------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.2309825352605927, pvalue=3.471070981483411e-23)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=122.17426114250313, pvalue=2.9525433149639207e-27)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.7506336569786072, pvalue=2.2404396969095158e-26)
testing class 2---------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.24062768525734257, pvalue=8.983858210183508e-16)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=204.3229997414737, pvalue=4.283752226644693e-45)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.6932970285415649, pvalue=5.0453899511809434e-23)
testing class 3---------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.24591805878036666, pvalue=6.338596341426485e-11)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=85.10285415084782, pvalue=3.312455282050774e-19)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.7087076902389526, pvalue=3.094697324929834e-18)
```

```
testing class 4---------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.2976559327323658, pvalue=7.042241827969836e-34)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=316.0614139074955, pvalue=2.3341888958759127e-69)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.5594570636749268, pvalue=2.647298618941661e-31)
testing class 5---------------------------------
scipy.stats.kstest统计检验结果：----------------------------------------------------
KstestResult(statistic=0.22924865095564212, pvalue=8.104973339459487e-30)
scipy.stats.normaltest统计检验结果：------------------------------------------------
NormaltestResult(statistic=261.1985960975389, pvalue=1.911813469144232e-57)
scipy.stats.shapiro统计检验结果：---------------------------------------------------
ShapiroResult(statistic=0.7298407554626465, pvalue=8.302614400987805e-31)
LeveneResult(statistic=24.866641023403535, pvalue=3.9569195407590673e-20)
```

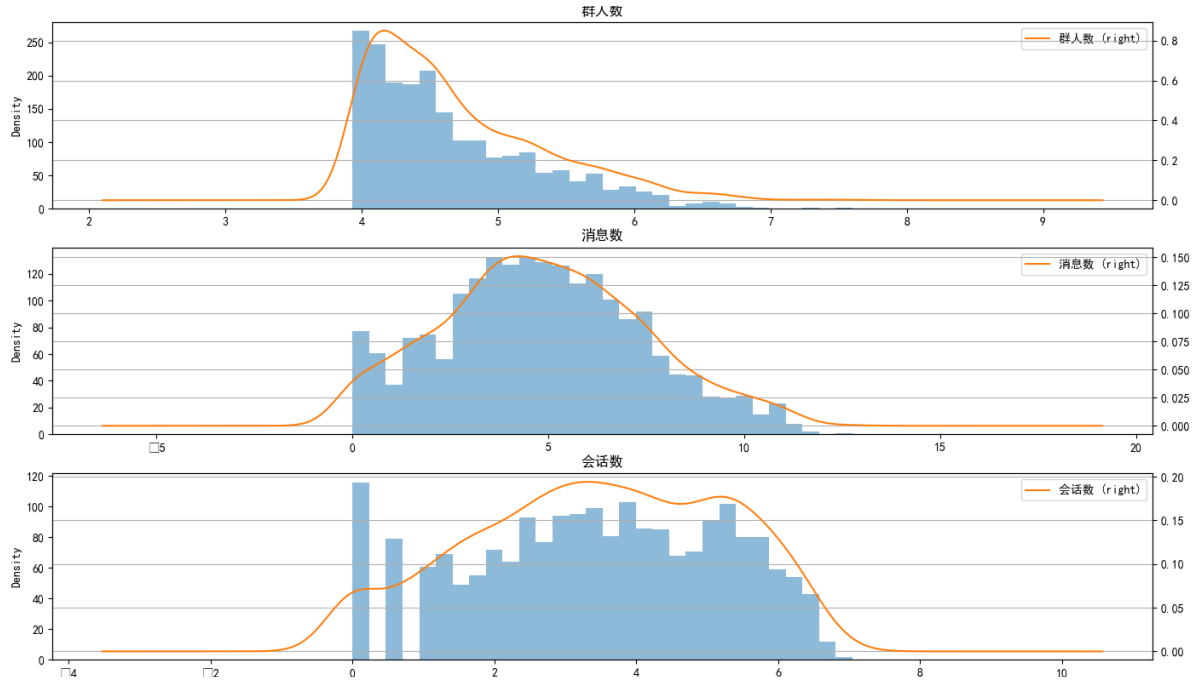
从检验结果看各组别都不满足正态性假设，同时也不具备方差齐次性

会话数不满足单因素ANOVA假设

进行对数转换

```
data["群人数"] = data["群人数"].apply(np.log)
data["消息数"] = data["消息数"].apply(np.log)
data["会话数"] = data["会话数"].apply(np.log)
```

从PDF图看出正态性有所改善



测试结果如下

```
testing col 3----------------------------------
testing class 1----------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.1385260590974159, pvalue=1.4537144422388028e-08)
scipy.stats.normaltest统计检验结果：--------------------------------------------------
NormaltestResult(statistic=114.47854529020555, pvalue=1.38452162068625e-25)
scipy.stats.shapiro统计检验结果：--------------------------------------------------
ShapiroResult(statistic=0.8738664984703064, pvalue=2.213164897373843e-19)
testing class 2----------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.13832442028782915, pvalue=1.8083708290154666e-05)
scipy.stats.normaltest统计检验结果：--------------------------------------------------
NormaltestResult(statistic=34.716809663527116, pvalue=2.8929475972361947e-08)
scipy.stats.shapiro统计检验结果：--------------------------------------------------
ShapiroResult(statistic=0.9294371008872986, pvalue=9.961873687069911e-11)
testing class 3----------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.1122112438318057, pvalue=0.013220374362912369)
scipy.stats.normaltest统计检验结果：--------------------------------------------------
NormaltestResult(statistic=14.521877678263348, pvalue=0.0007024482400382735)
scipy.stats.shapiro统计检验结果：--------------------------------------------------
ShapiroResult(statistic=0.9458172917366028, pvalue=9.536948937238776e-07)
testing class 4----------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.16118001463726733, pvalue=4.099982260239173e-10)
scipy.stats.normaltest统计检验结果：--------------------------------------------------
NormaltestResult(statistic=166.9845434247539, pvalue=5.492462668297059e-37)
scipy.stats.shapiro统计检验结果：--------------------------------------------------
ShapiroResult(statistic=0.819685697555542, pvalue=1.7436496772865077e-21)
```

```
testing class 5---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.09526225780643666, pvalue=1.818426811737231e-05)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=79.65128950846281, pvalue=5.0575739897515686e-18)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9342203140258789, pvalue=4.2054280564713634e-16)
LeveneResult(statistic=35.30989016931306, pvalue=1.4929114658768898e-28)
F value: 59.73150864257041
P value: 9.493707242830655e-48

testing col 4---------------------------------
testing class 1---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.046785036613765985, pvalue=0.2325906013675254)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=56.76896783219424, pvalue=4.707336705665444e-13)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9782127141952515, pvalue=1.2339296517893672e-06)
testing class 2---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.06256989470472127, pvalue=0.18293013308628203)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=13.30655851527809, pvalue=0.0012897856236455197)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9710918664932251, pvalue=9.89702130027581e-06)
testing class 3---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.060301393480815646, pvalue=0.4564509114587786)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=2.843443720605013, pvalue=0.24129817723842242)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9862481951713562, pvalue=0.05355570837855339)
```

```
testing class 4---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.04568756530855708, pvalue=0.32762241841720063)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=9.996055568292558, pvalue=0.006751248797676029)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9829906821250916, pvalue=6.851421494502574e-05)
testing class 5---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.021001532855179306, pvalue=0.9364955411475273)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=0.2432271001582241, pvalue=0.8854905001237201)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9959505200386047, pvalue=0.1010657474398613)
LeveneResult(statistic=28.814557168557386, pvalue=2.510112212224113e-23)
F value: 20.011754697728104
P value: 3.530671426877148e-16

testing col 11---------------------------------
testing class 1---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.07247319093686488, pvalue=0.011759021688606406)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=112.04211544375038, pvalue=4.6812707599506e-25)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9545251131057739, pvalue=4.678083961873192e-11)
testing class 2---------------------------------
scipy.stats.kstest统计检验结果：--------------------------------------------------
KstestResult(statistic=0.09575532581827761, pvalue=0.007601056180461768)
scipy.stats.normaltest统计检验结果：----------------------------------------------
NormaltestResult(statistic=45.05673777149828, pvalue=1.6445750855457745e-10)
scipy.stats.shapiro统计检验结果：-------------------------------------------------
ShapiroResult(statistic=0.9331091642379761, pvalue=2.2376359098963405e-10)
```

```
testing class 3--------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.0550976511548158, pvalue=0.5721375520343143)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=3.4636233483183663, pvalue=0.17696351914808223)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.9781529903411865, pvalue=0.003733379067853093)
testing class 4--------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.05946083186025758, pvalue=0.09511445465470936)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=22.5317811353402, pvalue=1.2802237883525106e-05)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.9736391305923462, pvalue=5.902439284000138e-07)
testing class 5--------------------------------
scipy.stats.kstest统计检验结果：-----------------------------------------------------
KstestResult(statistic=0.05202903480652332, pvalue=0.062015369991711866)
scipy.stats.normaltest统计检验结果：-------------------------------------------------
NormaltestResult(statistic=24.114156831821568, pvalue=5.803331458517489e-06)
scipy.stats.shapiro统计检验结果：----------------------------------------------------
ShapiroResult(statistic=0.9762263894081116, pvalue=1.2336465182727352e-08)
LeveneResult(statistic=11.856914013664593, pvalue=1.5949172505297402e-09)
F value: 36.20654087706869
P value: 2.860119121302016e-29
```

很多组的正态性得到显著改善，但仍存在不满足正态性假设的组别以及方差仍不具备齐次性

(5)

a)

使用正态分布图直观判断正态分布的特质，而不是检验的方法。检验方法比较严格，现实数据满足钟形曲线特征即可

进行数据转换，将数据分布转换为正态分布

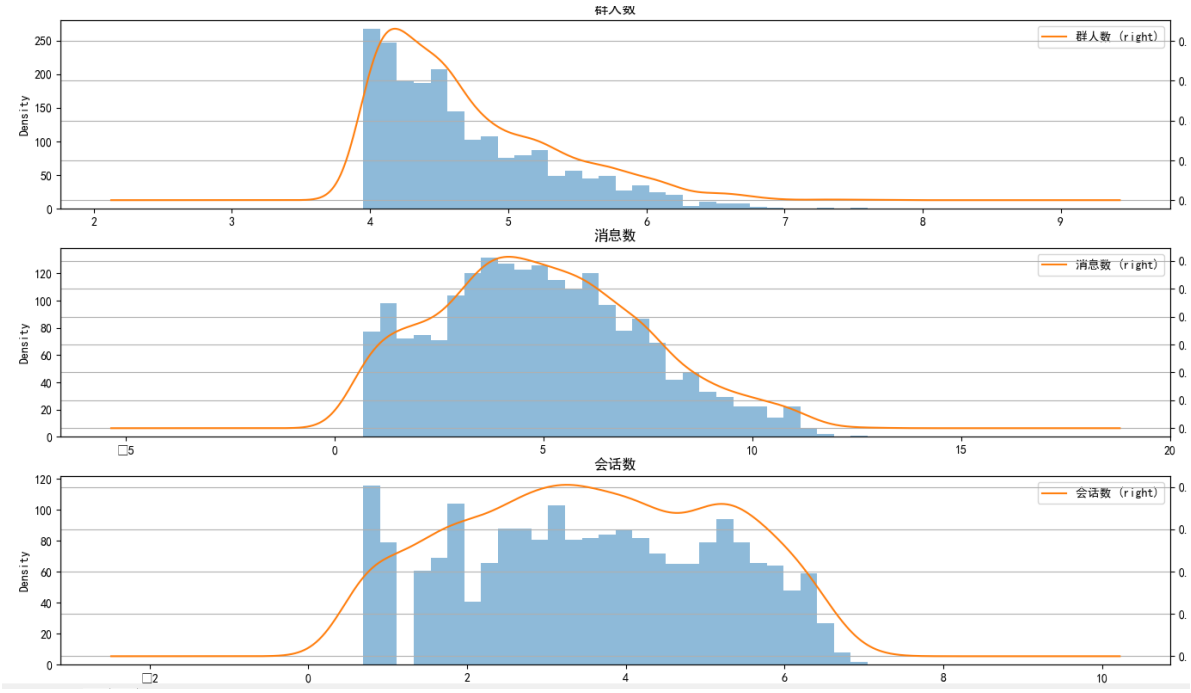
- 取对数或开根号等压缩处理
- BOX-COX转换等

使用不需要正态性假设的非参数检验等方法，如kruskal-wallis H检验

不满足方差齐次性时，可使用Welch检验或者Brown-Forsythe检验

b)

对各组使用log变换 + box-cox变换改善正态性后



原总体的正态性有所改善，但仍不具备明显正态特征

直接进行单因素ANOVA结果为并进行事后LSD检验

群人数：

```
F value: 512044.72163549793
P value: 0.0
```

多重比较

因变量: 群人数

LSD

(I) 群类别	(J) 群类别	平均值差值 (I-J)	标准 错误	显著性	95% 置信区间	
					下限	上限
1	2	-.271301355*	.0433765619	.000	-.356368450	-.186234261
	3	-.116707528*	.0499786065	.020	-.214722092	-.018692963
	4	.294737312*	.0392414637	.000	.2177796845	.3716949396
	5	-.207786055*	.0356193428	.000	-.277640231	-.137931879
2	1	.271301355*	.0433765619	.000	.1862342610	.3563684495
	3	.154593828*	.0542166373	.004	.0482679317	.2609197233
	4	.566038667*	.0445137635	.000	.4787413725	.6533359621
	5	.0635153003	.0413560182	.125	-.017589244	.1446198450
3	1	.116707528*	.0499786065	.020	.0186929632	.2147220923
	2	-.154593828*	.0542166373	.004	-.260919723	-.048267932
	4	.411444840*	.0509687171	.000	.3114885392	.5114011404
	5	-.091078527	.0482354147	.059	-.185674465	.0035174110
4	1	-.294737312*	.0392414637	.000	-.371694940	-.217779684
	2	-.566038667*	.0445137635	.000	-.653335962	-.478741372
	3	-.411444840*	.0509687171	.000	-.511401140	-.311488539
	5	-.502523367*	.0369957646	.000	-.575076886	-.429969848
5	1	.207786055*	.0356193428	.000	.1379318788	.2776402311
	2	-.063515300	.0413560182	.125	-.144619845	.0175892443
	3	.0910785272	.0482354147	.059	-.003517411	.1856744653
	4	.502523367*	.0369957646	.000	.4299698482	.5750768857

*. 平均值差值的显著性水平为 0.05。

消息数:

```
F value: 40.906940547521536
P value: 5.1270281277475835e-33
```

多重比较

因变量: 消息数

LSD

(I) 群类别	(J) 群类别	平均值差值 (I-J)	标准 错误	显著性	95% 置信区间	
					下限	上限
1	2	.582927864*	.1774088173	.001	.2350060393	.9308496894
	3	.2191424304	.2044109785	.284	-.181734154	.6200190149
	4	1.38603988*	.1604963918	.000	1.071285528	1.700794235
	5	.787675984*	.1456820276	.000	.5019745310	1.073377437
2	1	-.582927864*	.1774088173	.001	-.930849689	-.235006039
	3	-.363785434	.2217443958	.101	-.798655110	.0710842421
	4	.803112017*	.1820599371	.000	.4460687397	1.160155295
	5	.2047481199	.1691448553	.226	-.126966998	.5364632379
3	1	-.219142430	.2044109785	.284	-.620019015	.1817341540
	2	.3637854339	.2217443958	.101	-.071084242	.7986551099
	4	1.16689745*	.2084605008	.000	.7580792255	1.575715677
	5	.568533554*	.1972813772	.004	.1816390473	.9554280603
4	1	-1.38603988*	.1604963918	.000	-1.70079423	-1.07128553
	2	-.803112017*	.1820599371	.000	-1.16015529	-.446068740
	3	-1.16689745*	.2084605008	.000	-1.57571568	-.758079225
	5	-.598363897*	.1513115508	.000	-.895105580	-.301622215
5	1	-.787675984*	.1456820276	.000	-1.07337744	-.501974531
	2	-.204748120	.1691448553	.226	-.536463238	.1269669981
	3	-.568533554*	.1972813772	.004	-.955428060	-.181639047
	4	.598363897*	.1513115508	.000	.3016222150	.8951055797

*. 平均值差值的显著性水平为 0.05。

会话数:

```
F value: 107.27009779932048
P value: 5.085343516808963e-83
```

多重比较

因变量: 会话数

LSD

(I) 群类别	(J) 群类别	平均值差值 (I-J)	标准 错误	显著性	95% 置信区间	
					下限	上限
1	2	.1320226964	.1158149931	.254	-.095105607	.3591510003
	3	.0442020903	.1334423871	.740	-.217495832	.3059000125
	4	1.03622535*	.1047743218	.000	.8307492427	1.241701458
	5	-.063675062	.0951032946	.503	-.250185025	.1228348999
2	1	-.132022696	.1158149931	.254	-.359151000	.0951056075
	3	-.087820606	.1447578878	.544	-.371709701	.1960684885
	4	.904202654*	.1188513101	.000	.6711197365	1.137285572
	5	-.195697759	.1104201615	.076	-.412246094	.0208505768
3	1	-.044202090	.1334423871	.740	-.305900012	.2174958320
	2	.0878206062	.1447578878	.544	-.196068489	.3717097009
	4	.992023260*	.1360859727	.000	.7251409219	1.258905598
	5	-.107877153	.1287880822	.402	-.360447376	.1446930706
4	1	-1.03622535*	.1047743218	.000	-1.24170146	-.830749243
	2	-.904202654*	.1188513101	.000	-1.13728557	-.671119736
	3	-.992023260*	.1360859727	.000	-1.25890560	-.725140922
	5	-1.09990041*	.0987783273	.000	-1.29361759	-.906183232
5	1	.0636750624	.0951032946	.503	-.122834900	.2501850248
	2	.1956977588	.1104201615	.076	-.020850577	.4122460945
	3	.1078771527	.1287880822	.402	-.144693071	.3604473760
	4	1.09990041*	.0987783273	.000	.9061832322	1.293617593

*. 平均值差值的显著性水平为 0.05。

观察p-value，三列的p-value都极小，从统计的角度可以认为群人数、消息数、会话数在不同的群类别中有显著差异

事后比较得出，两两之间群类别和群人数、消息数以及会话数存在显著关系