

Fundamental of Database Management System

and SAS Application

Final Assignment

Ban Bank.

Relations:

employee(eno,ename,esex,ephone)

customer(cno,cname,csex,cphone)

cdrate(cdate,crate)

cdtrans(tno,account,cno, aeno,adate,amount,balance,teno,tdate)

Data sources:

1. employee: In the Sql server database files bank.mdf and bank_log.ldf.
2. customer: In the text file customer.txt
3. cdate: In the CSV file cdate.csv
4. cdtrans: In the text file cdtrans.txt

Operating and programming:

Submitting:

1. Paste your operation steps, programs, output report, models and forecasts in a MS Word file named as “学号_姓名_final.docx”. Compress this word file and your final version SQL Server data files (bank.mdf and bank_log.ldf) in to one rar file: “学号_姓名_final.rar”, attach it to an email titled as “DBMSSAS 学号 姓名 final” and send the mail to: dphu@sjtu.edu.cn before 29 Dec, 2015.
2. Print your word file as a hard copy and submit it before 29 Dec, 2015.

1. Attach the bank.mdf to SQL Server to get table employee of database bank. Create a permanent library named bank in SAS using ODBC connecting SQL Server database bank. Write SAS data steps to create data sets of customer, cdate and cdtrans. (Don't insert the transaction records of November 2015 into table cdtrans)

```
/* create data sets of cdate*/
```

```
DATA bank.cdate;
```

```
INFILE 'C:\Users\Administrator\Desktop\SAS\DMBSSAS Final\cdrate.csv'
```

```
DLM =',';
```

```

FIRSTOBS = 3;
INPUT cdate yymmdd10. crate;
FORMAT cdate mmddyy10.;
PROC PRINT DATA = bank.cdate;
RUN;

```

VIEWTABLE: Bank.cdate

	cddate	crate
1	04AUG1987	0.55
2	24JUL1988	0.69
3	24FEB1989	0.47
4	25MAR1990	0.33
5	10DEC1990	0.25
6	21MAR1992	0.48
7	28OCT1992	0.63
8	13JUN1993	0.17
9	08OCT1994	0.32
10	14SEP1995	0.27

```

/*create data sets of customer*/
DATA bank.customer;
INFILE 'C:\Users\Administrator\Desktop\SAS\DMBSSAS Final\customer.txt'
FIRSTOBS = 3;
INPUT CNO:$ CNAME:$ CSEX:$ CPHONE;
PROC PRINT DATA = bank.customer;
RUN;

```

VIEWTABLE: Bank.customer

	CNO	CNAME	CSEX	CPHONE
1	c100001	张毅	男	14843952666
2	c100002	刘书同	女	13821196491
3	c100003	高远海	男	13832038012
4	c100004	宋俞渝	女	15767573595
5	c100005	张丽丽	女	13683090219
6	c100006	徐中华	男	15530179721
7	c100007	左丽梅	女	15760219496
8	c100008	熊达西	男	15218739413
9	c100009	史博	男	.
10	c100011	宗娟慧	女	13201884668
11	c100012	胡明山	男	.

```

/* create data sets of cdtrans*/
DATA bank.cdtrans;
INFILE 'C:\Users\Administrator\Desktop\SAS\DMBSSAS Final\cdtrans.txt'
DLM = '09'X DSD MISSEVER FIRSTOBS = 3;
LENGTH tno $12 account $10;
INPUT tno $ account $ cno $ aeno $ adate yymmdd10.+1 amount balance teno $ tdate yymmdd10.;
IF tdate >20392 THEN DELETE;
PROC PRINT DATA = bank.cdtrans;
RUN;

```

	tno	account	cno	aeno	adate	amount	balance	teno	tdate
1	ct1987121001	cd15100001	c100003	e106006	10205	5019	5019	e106006	10205
2	ct1987121002	cd15100002	c100020	e106011	10205	244	244	e106011	10205
3	ct1987121101	cd15100003	c100016	e106005	10206	691	691	e106005	10206
4	ct1987121102	cd15100004	c100005	e106010	10206	3361	3361	e106010	10206
5	ct1987121201	cd15100005	c100011	e106011	10207	2768	2768	e106011	10207
6	ct1987121202	cd15100006	c100008	e106011	10207	1387	1387	e106011	10207
7	ct1987121203	cd15100007	c100009	e106011	10207	4304	4304	e106011	10207
8	ct1987121301	cd15100008	c100010	e106009	10208	574	574	e106009	10208
9	ct1987121302	cd15100009	c100004	e106003	10208	980	980	e106003	10208
10	ct1987121401	cd15100010	c100005	e106014	10209	1858	1858	e106014	10209
11	ct1987121402	cd15100011	c100001	e106006	10209	6237	6237	e106006	10209
12	ct1987121403	cd15100012	c100010	e106012	10209	1217	1217	e106012	10209
13	ct1987121404	cd15100013	c100005	e106001	10209	2803	2803	e106001	10209

2. Create permanent library bankloc by selecting a path in your hard drive and copy all data sets from library bank to library bankloc. Set bankloc as the default working library and do the following tasks in this library. All data sets in bankloc should be copied to bank when all tasks have done.

```
/*copy*/
libname bankloc 'C:\Users\Administrator\Desktop\SAS\DMBSSAS Final\data';
PROC COPY IN = bank
OUT = bankloc;
RUN;
/*change the default working library*/
OPTION USER = bankloc;
RUN;
```

Result:

```
101 libname bankloc 'C:\Users\Administrator\Desktop\SAS\DMBSSAS Final\data';
```

NOTE: 已成功分配逻辑库引用名 BANKLOC, 如下所示:

引擎: V9

物理名: C:\Users\Administrator\Desktop\SAS\DMBSSAS Final\data

```
102 PROC COPY IN = bank
```

```
103 OUT = bankloc;
```

```
104 RUN;
```

NOTE: 正在复制 BANK.cdtrate 至 BANKLOC.CDRATE (memtype=DATA)。

NOTE: 在跨不同引擎复制时, 没有克隆 BUFSIZE。使用了 BUFSIZE 的系统选项。

NOTE: 从数据集 BANK.cdtrate. 读取了 30 个观测

NOTE: 数据集 BANKLOC.CDRATE 有 30 个观测和 2 个变量。

NOTE: 正在复制 BANK.cdtrans 至 BANKLOC.CDTRANS (memtype=DATA)。

NOTE: 在跨不同引擎复制时, 没有克隆 BUFSIZE。使用了 BUFSIZE 的系统选项。

NOTE: 从数据集 BANK.cdtrans. 读取了 25413 个观测

NOTE: 数据集 BANKLOC.CDTRANS 有 25413 个观测和 9 个变量。

NOTE: 正在复制 BANK.customer 至 BANKLOC.CUSTOMER (memtype=DATA)。

NOTE: 在跨不同引擎复制时, 没有克隆 BUFSIZE。使用了 BUFSIZE 的系统选项。

NOTE: 从数据集 BANK.customer. 读取了 15 个观测

NOTE: 数据集 BANKLOC.CUSTOMER 有 15 个观测和 4 个变量。

NOTE: “PROCEDURE COPY” 所用时间 (总处理时间):

实际时间 0.13 秒

CPU 时间 0.07 秒



- Decompose the relation cdtrans into two relations (account and transaction) because it violates BCNF. Create tables of account and transaction, insert relative values according to records of cdtrans and then delete table cdtrans.

```
/*decompose the relation cdtrans(use SQL language)*/
```

```
PROC SQL;
```

```
CREATE TABLE bankloc.account AS
```

```
SELECT DISTINCT account,cno,aeno,adate
```

```
FROM bankloc.cdtrans;
```

```
CREATE TABLE bankloc.transaction AS
```

```
SELECT account,cno,tno,teno,tdate,amount,balance
```

```
FROM bankloc.cdtrans;
```

```
QUIT;
```

VIEWTABLE: Bankloc.Account				
	account	cno	aeno	adate
1	cd15100001	c100003	e106006	10205
2	cd15100002	c100020	e106011	10205
3	cd15100003	c100016	e106005	10206
4	cd15100004	c100005	e106010	10206
5	cd15100005	c100011	e106011	10207
6	cd15100006	c100008	e106011	10207
7	cd15100007	c100009	e106011	10207
8	cd15100008	c100010	e106009	10208
9	cd15100009	c100004	e106003	10208
10	cd15100010	c100005	e106014	10209
11	cd15100011	c100001	e106006	10209
12	cd15100012	c100010	e106012	10209

VIEWTABLE: Bankloc.Transaction							
	account	cno	tno	teno	tdate	amount	balance
1	cd15100001	c100003	ct1987121001	e106006	10205	5019	5019
2	cd15100002	c100020	ct1987121002	e106011	10205	244	244
3	cd15100003	c100016	ct1987121101	e106005	10206	691	691
4	cd15100004	c100005	ct1987121102	e106010	10206	3361	3361
5	cd15100005	c100011	ct1987121201	e106011	10207	2768	2768
6	cd15100006	c100008	ct1987121202	e106011	10207	1387	1387
7	cd15100007	c100009	ct1987121203	e106011	10207	4304	4304
8	cd15100008	c100010	ct1987121301	e106009	10208	574	574
9	cd15100009	c100004	ct1987121302	e106003	10208	980	980
10	cd15100010	c100005	ct1987121401	e106014	10209	1858	1858
11	cd15100011	c100001	ct1987121402	e106006	10209	6237	6237
12	cd15100012	c100010	ct1987121403	e106012	10209	1217	1217
13	cd15100013	c100005	ct1987121404	e106001	10209	2803	2803
14	cd15100014	c100018	ct1987121501	e106011	10210	5046	5046

- Create data set trans_r with the rate for each transaction record, create a data set named as trans_b with the interest and balance (including interest) for each transaction record and use PROC PRINT to

print transaction records (from year 1987 to year 1988) for account cd1510

```
/*create trans_r*/
OPTIONS USER = bankloc;
DATA trans_r1;
MERGE transaction(IN = intdate)
      cdate(rename = (cdate = tdate) IN = incdate);
BY tdate;
RUN;
DATA trans_r;
SET trans_r1;
RETAIN trate;
IF ^MISSING(crate) THEN trate = crate;
ELSE crate = trate;
DROP trate;
IF tdate = '10077' THEN DELETE;
PROC PRINT DATA = trans_r;
RUN;
```

```
/*create trans_b*/
DATA trans_b1;
SET TRANS_r;
PROC SORT DATA=trans_b1;
BY account;
RUN;
```

```
DATA trans_b2;
MERGE account trans_b1;
BY account;
RUN;
```

```
DATA trans_b3;
SET trans_b2;
IF balance ^= .
THEN interest = 0;
RUN;
```

```
DATA trans_b4;
SET trans_b3;
RETAIN tbalance;
IF ^missing(balance) THEN DO;
tbalance = balance;
tinterest = interest;
lag_tbalance = tbalance;
lag_crate = crate;
dif_tdate = 0;
END;
lag_tbalance = lag0(tbalance);
lag_crate = lag0(crate);
```

```

dif_tdate = dif(tdate);
IF balance = . THEN DO;
tinterest = lag_tbalance*lag_crate*dif_tdate/365/100; /*calculate interest rate*/
tbalance = lag_tbalance+tinterest + amount;
END;
RUN;

PROC SQL;
CREATE TABLE trans_b AS
SELECT cno,tno,account,amount,tbalance,teno,tdate,tinterest
FROM trans_b4;
QUIT;

PROC DATASETS LIBRARY = bankloc;
MODIFY trans_b;
RENAME tbalance=balance tinterest=interest;
QUIT;

```

VIEWTABLE: Bankloc.Trans_r

	account	cno	tno	teno	tdate	amount	balance	crate
1	cd15100001	c100003	ct1987121001	e106006	10DEC1987	5019	5019	0.55
2	cd15100002	c100020	ct1987121002	e106011	10DEC1987	244	244	0.55
3	cd15100003	c100016	ct1987121101	e106005	11DEC1987	691	691	0.55
4	cd15100004	c100005	ct1987121102	e106010	11DEC1987	3361	3361	0.55
5	cd15100005	c100011	ct1987121201	e106011	12DEC1987	2768	2768	0.55
6	cd15100006	c100008	ct1987121202	e106011	12DEC1987	1387	1387	0.55
7	cd15100007	c100009	ct1987121203	e106011	12DEC1987	4304	4304	0.55
8	cd15100008	c100010	ct1987121301	e106009	13DEC1987	574	574	0.55
9	cd15100009	c100004	ct1987121302	e106003	13DEC1987	980	980	0.55
10	cd15100010	c100005	ct1987121401	e106014	14DEC1987	1858	1858	0.55
11	cd15100011	c100001	ct1987121402	e106006	14DEC1987	6237	6237	0.55
12	cd15100012	c100010	ct1987121403	e106012	14DEC1987	1217	1217	0.55
13	cd15100013	c100005	ct1987121404	e106001	14DEC1987	2803	2803	0.55

VIEWTABLE: Bankloc.Trans_b

	cno	tno	account	amount	balance	teno	tdate	interest
1	c100003	ct1987121001	cd15100001	5019	5019	e106006	10DEC1987	0
2	c100003	ct1988021801	cd15100001	-1227	3797.2940137	e106015	18FEB1988	5.2940136986
3	c100003	ct1988030302	cd15100001	-484	3314.0950867	e106013	03MAR1988	0.8010729837
4	c100003	ct1988032302	cd15100001	4528	7843.0938551	e106004	23MAR1988	0.9987683823
5	c100003	ct1988053101	cd15100001	1799	9650.2485239	e106008	31MAY1988	8.1546688164
6	c100003	ct1988060301	cd15100001	850	10500.684768	e106014	03JUN1988	0.4362441114
7	c100003	ct1988061603	cd15100001	7163	17665.741751	e106002	16JUN1988	2.0569834546
8	c100003	ct1988062202	cd15100001	-7153	10514.338928	e106001	22JUN1988	1.5971768515
9	c100003	ct1988072101	cd15100001	2121	12639.93355	e106010	21JUL1988	4.5946220795
10	c100003	ct1988073103	cd15100001	4412	17054.323017	e106003	31JUL1988	2.3894668903
11	c100003	ct1988091302	cd15100001	-5000	12068.508476	e106004	13SEP1988	14.185458816
12	c100003	ct1988092302	cd15100001	586	12656.78992	e106003	23SEP1988	2.281444068
13	c100003	ct1988101202	cd15100001	-5335	7326.3359615	e106011	12OCT1988	4.5460415302
14	c100003	ct1988102201	cd15100001	-5846	1481.7209401	e106001	22OCT1988	1.384978579

/*print transaction records (from year 1987 to year 1988) for account cd1510*/

```

PROC PRINT DATA = trans_b;
WHERE account = 'cd15100001' AND 1987 <= year(tdate) <= 1988;
RUN;

```


输出 - print b.lst

SAS 系统 2015年12月27日 星期日 下午04时02分47秒 789

Obs	cno	tno	account	amount	balance	teno	tdate	interest
1	c100003	ct1987121001	cd15100001	5019	5019.00	e106006	10DEC1987	0.0000
2	c100003	ct1988021801	cd15100001	-1227	3797.29	e106015	18FEB1988	5.2940
3	c100003	ct1988030302	cd15100001	-484	3314.10	e106013	03MAR1988	0.8011
4	c100003	ct1988032302	cd15100001	4528	7843.09	e106004	23MAR1988	0.9988
5	c100003	ct1988053101	cd15100001	1799	9650.25	e106008	31MAY1988	8.1547
6	c100003	ct1988060301	cd15100001	850	10500.68	e106014	03JUN1988	0.4362
7	c100003	ct1988061603	cd15100001	7163	17665.74	e106002	16JUN1988	2.0570
8	c100003	ct1988062202	cd15100001	-7153	10514.34	e106001	22JUN1988	1.5972
9	c100003	ct1988072101	cd15100001	2121	12639.93	e106010	21JUL1988	4.5946
10	c100003	ct1988073103	cd15100001	4412	17054.32	e106003	31JUL1988	2.3895
11	c100003	ct1988091302	cd15100001	-5000	12068.51	e106004	13SEP1988	14.1855
12	c100003	ct1988092302	cd15100001	586	12656.79	e106003	23SEP1988	2.2814
13	c100003	ct1988101202	cd15100001	-5335	7326.34	e106011	12OCT1988	4.5460
14	c100003	ct1988102201	cd15100001	-5846	1481.72	e106001	22OCT1988	1.3850
15	c100003	ct1988102803	cd15100001	4266	5747.89	e106002	28OCT1988	0.1681
16	c100003	ct1988111803	cd15100001	-994	4756.17	e106003	18NOV1988	2.2818
17	c100003	ct1988120803	cd15100001	5836	10593.97	e106002	08DEC1988	1.7982

5. Output a report to display: customer name, total number of accounts, total interest, total balance, account, total number of account transactions, total account interest and account balance.

```

/*use 'merge by' to merge customer into trans_b*/
DATA trans_c1;
SET trans_b;
KEEP cno account tdate interest balance;
RUN;
PROC SORT DATA = trans_c1;
BY cno account;
RUN;
DATA trans_c2;
MERGE trans_c1 customer;
BY cno;
RUN;

```

VIEWTABLE: Bankloc.Trans_c2

	cno	account	balance	tdate	interest	CNAME	CSEX	CPHONE
1	c100001	cd15100011	6237	14DEC1987	0	张毅	男	14843952666
2	c100001	cd15100011	3136.3157507	28DEC1987	1.3157506849	张毅	男	14843952666
3	c100001	cd15100011	9468.1191631	14JAN1988	0.8034123909	张毅	男	14843952666
4	c100001	cd15100011	17607.119163	14JAN1988	0	张毅	男	14843952666
5	c100001	cd15100011	16186.282607	10FEB1988	7.1634443718	张毅	男	14843952666
6	c100001	cd15100011	24953.477733	19FEB1988	2.1951259974	张毅	男	14843952666
7	c100001	cd15100011	20848.502005	14MAR1988	9.0242713995	张毅	男	14843952666
8	c100001	cd15100011	18971.130316	16MAR1988	0.6283110193	张毅	男	14843952666
9	c100001	cd15100011	8714.9889793	26MAR1988	2.8586634723	张毅	男	14843952666
10	c100001	cd15100011	6693.8274839	09APR1988	1.8385045244	张毅	男	14843952666
11	c100001	cd15100011	4737.8274839	09APR1988	0	张毅	男	14843952666
12	c100001	cd15100011	13834.255322	29APR1988	1.4278384198	张毅	男	14843952666
13	c100001	cd15100011	3923.2160885	18MAY1988	3.9607662498	张毅	男	14843952666
14	c100001	cd15100011	8229.8122463	14JUN1988	1.5961577785	张毅	男	14843952666

```

/*calculate total number of accounts, total interest, total balance, account, total number
of account transactions, total account interest and account balance*/

```

```

PROC SQL;
CREATE TABLE trans_c3 AS
SELECT cname,account,interest,balance,
       sum(trans_c2.interest) AS sum_interest,sum(trans_c2.balance) AS sum_balance
FROM trans_c2
GROUP BY cname;
QUIT;

```

VIEWTABLE: Bankloc.Trans_c3

	CNAME	account	interest	balance	sum_interest	sum_balance
1		cd15100019	55.988791881	173307.74305	105690.33215	508115336.84
2		cd15100008	1.4243561293	21730.945177	105690.33215	508115336.84
3		cd15100012	2.6313913312	72301.588656	105690.33215	508115336.84
4		cd15100008	2.1656994794	59495.037667	105690.33215	508115336.84
5		cd15100034	0.423859999	3774.2632179	105690.33215	508115336.84
6		cd15100012	0.4785612859	60860.36047	105690.33215	508115336.84
7		cd15100012	0.8444239788	66408.651757	105690.33215	508115336.84
8		cd15100008	9.155953186	40078.454569	105690.33215	508115336.84
9		cd15100008	17.820023532	66934.39406	105690.33215	508115336.84
10		cd15100008	0.7019972246	41225.061299	105690.33215	508115336.84
11		cd15100019	3.6556798281	26576.690142	105690.33215	508115336.84
12		cd15100008	39.004946111	180763.58067	105690.33215	508115336.84
13		cd15100034	5.1974770602	37102.028325	105690.33215	508115336.84
14		cd15100008	0.3320528236	5246.2960827	105690.33215	508115336.84

```

/*print results using 'proc report'*/
PROC REPORT DATA = bankloc.trans_c3;
COLUMN cname sum_interest sum_balance account interest,SUM balance,SUM;
DEFINE cname/GROUP;
DEFINE sum_interest/GROUP;
DEFINE sum_balance/GROUP;
DEFINE account/GROUP;
RUN;

```

PROC 报表

SAS 系统 2015年12月27日 星期日 下午04时02分47秒 1

CNAME	sum_inter est	sum_balan ce	account	interest SUM	balance SUM
陈代飞	13845.565	65697770	cd15100021	13845.565	65697770
			cd15100001	10233.913	50060607
			cd15100027	10677.382	50804512
			cd15100036	12064.424	64594446
洪存智	27903.255	143654341	cd15100040	11690.228	62017375
			cd15100003	9087.8852	46778449
			cd15100029	8641.7143	43476219
			cd15100038	10173.656	53399673
胡明山	36176.526	170317556	cd15100024	18444.159	86317366
			cd15100028	17732.367	84000191
			cd15100043	19648.697	98468661
			cd15100007	20244.692	104141473
刘书同 史博	19648.697	98468661	cd15100015	19815.232	103287670
			cd15100030	19430.438	98618917
			cd15100031	18550.501	93300503
			cd15100009	6481.301	34398129
宋俞渝	19154.856	94459701	cd15100017	6426.2865	31030094
			cd15100020	6247.2682	29031478
			cd15100014	12489.301	60286090
			cd15100006	19162.581	93714984
魏朝美 熊达西	37275.33	184732697	cd15100016	18112.749	91017713
			cd15100022	28342.317	144702499
			cd15100032	28807.231	156016585
			cd15100002	7262.8365	34769731
薛家珏	14439.678	66020231	cd15100039	7176.8419	31250499
			cd15100004	6799.8807	30217803
张丽丽	20531.27	95356583			

6. Output the statistical moments of variable amount in transaction and draw the histogram of amount

with a fitting normal distribution.

```
PROC UNIVARIATE DATA = transaction;  
VAR amount;  
HISTOGRAM amount / NORMAL (color=red l=21) KERNEL(color=blue l=21);  
RUN;  
QUIT;
```

SAS 系统		2015年12月27日	
UNIVARIATE PROCEDURE			
变量: amount (amount)			
矩			
N	25413	权重总和	25413
均值	491.035612	观测总和	12478688
标准差	6028.86585	方差	36347223.5
偏度	0.25527754	峰度	0.26854241
未校正平方和	9.29783E11	校正平方和	9.23656E11
变异系数	1227.78587	标准误差均值	37.818792

位置和可变性的基本测度:

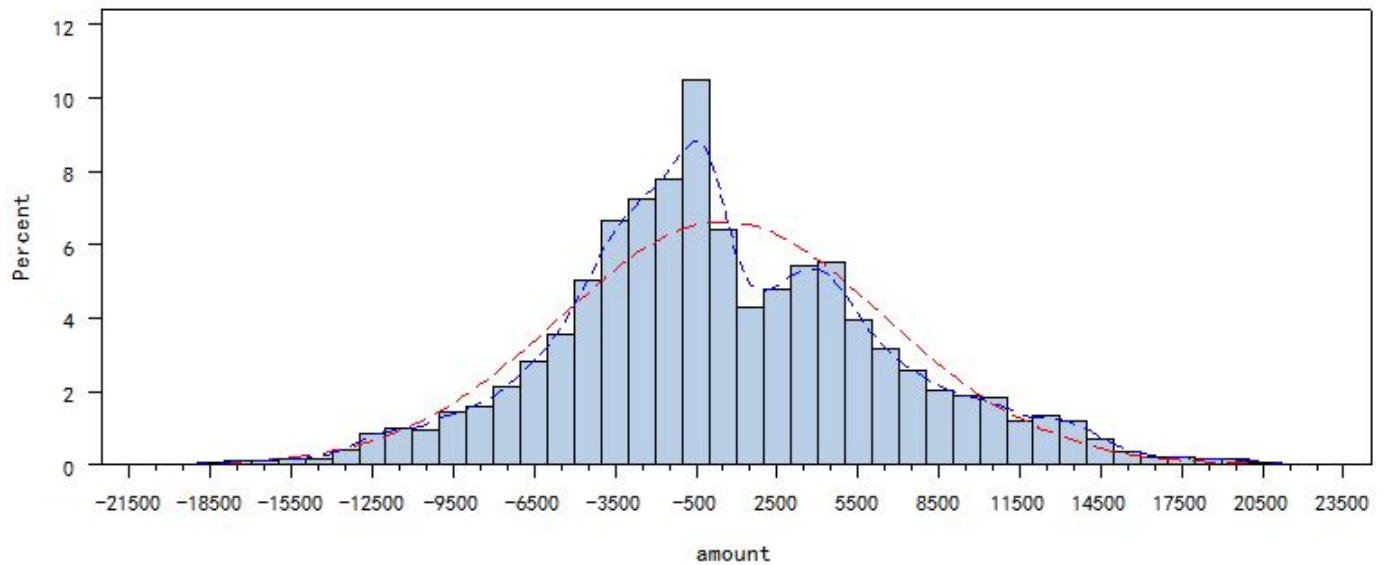
基本统计测度			
位置		变异性	
均值	491.036	标准差	6029
中位数	-164.000	方差	36347223
众数	-62.000	极差	40256
		四分位极差	7544

位置检验: Mu0=0			
检验	--统计量--	-----P 值-----	
Student t	t 12.98391	Pr > t	<.0001
符号	M -652	Pr >= M	<.0001
符号秩	S 8883779	Pr >= S	<.0001

分位数 (定义 5)		
分位数	估计值	
100% 最大值	21383	
99%	15371	
95%	11458	
90%	8612	
75% Q3	4269	
50% 中位数	-164	
25% Q1	-3275	
10%	-6619	
5%	-9213	
1%	-13003	
0% 最小值	-18873	

极值观测			
-----最小值-----		-----最大值-----	
值	观测	值	观测
-18873	7825	20931	22879
-18713	3944	21192	22382
-18615	5556	21297	24104
-18522	2657	21325	23803
-18475	7991	21383	24783

GRAPH 1 BANKLOC.GSEG.UNIVAR (正态分布拟合结果)



7. Draw the block chart of quarterly total interest for the three customers (刘书同, 宗娟慧, 蒋开源).

```
/*use function qtr() to select the quarters among the dates*/
DATA trans_d1;
SET trans_c2;
RUN;
PROC SORT DATA = trans_d1;
BY cname tdate;
RUN;
DATA trans_d2;
SET trans_d1;
quarter = qtr(tdate);
IF cname in ('刘书同', '宗娟慧', '蒋开源');
RUN;
```

	cno	account	balance	tdate	interest	CNAME	CSEX	CPHONE	quarter
1	c100002	cd15100043	7053	27DEC1987	0	刘书同	女	13821196491	4
2	c100002	cd15100043	17114.594171	11JAN1988	1.5941712329	刘书同	女	13821196491	1
3	c100002	cd15100043	27734.236212	29JAN1988	4.6420406108	刘书同	女	13821196491	1
4	c100002	cd15100043	16842.072038	31JAN1988	0.8358262968	刘书同	女	13821196491	1
5	c100002	cd15100043	30133.446931	04MAR1988	8.374893356	刘书同	女	13821196491	1
6	c100002	cd15100043	27865.961198	08MAY1988	29.514266515	刘书同	女	13821196491	2
7	c100002	cd15100043	14622.64079	12MAY1988	1.6795921818	刘书同	女	13821196491	2
8	c100002	cd15100043	9566.064543	23MAY1988	2.4237527885	刘书同	女	13821196491	2
9	c100002	cd15100043	22101.812344	18JUN1988	3.7478006292	刘书同	女	13821196491	2
10	c100002	cd15100043	16369.627803	14AUG1988	23.815459712	刘书同	女	13821196491	3
11	c100002	cd15100043	24823.24671	16AUG1988	0.6189064759	刘书同	女	13821196491	3
12	c100002	cd15100043	18370.709779	05OCT1988	23.463068808	刘书同	女	13821196491	4
13	c100002	cd15100043	11732.601055	14NOV1988	13.891276435	刘书同	女	13821196491	4
14	c100002	cd15100043	8585.9318213	20NOV1988	1.3307662567	刘书同	女	13821196491	4

```
/*draw graphics*/
PROC CHART DATA = bankloc.trans_d2;
BLOCK quarter /GROUP = cname
SUMVAR = interest TYPE = SUM DISCRETE;
```

RUN ;

按 CNAME 分组并由 quarter 划分的 interest 的总和

	1	2	3	4
CNAME	宗娟慧	宗娟慧	宗娟慧	宗娟慧
	1478.653	1537.914	1284.247	1710.492
	4579.9110383	5151.3992037	4710.9517781	5206.4354133

8. Draw the 3D scatter plot to display the sum of amount operated by each employee for each customer.

```

/* use first.var and last.var to accumulate relative variables*/
DATA trans_e1;
SET transaction;
KEEP cno teno amount;
RUN;

PROC SORT DATA = trans_e1;
BY cno teno;
RUN;

DATA trans_e2;
SET trans_e1;
BY cno teno;
IF first.teno = 1
THEN sum_of_amount = 0;
IF first.teno = 0
THEN sum_of_amount + amount;
IF last.teno;
DROP amount;
RUN;

```

VIEWTABLE: Bankloc.Trans_e2			
	cno	teno	sum_of_amount
1	c100001	e106001	110768
2	c100001	e106002	55719
3	c100001	e106003	85955
4	c100001	e106004	66588
5	c100001	e106005	127848
6	c100001	e106006	141067
7	c100001	e106007	60631
8	c100001	e106008	136088
9	c100001	e106009	74979
10	c100001	e106010	119707
11	c100001	e106011	18951
12	c100001	e106012	-54939
13	c100001	e106013	196234
14	c100001	e106014	147490

```
/*transfer string type into numeric type*/
```

```
PROC SQL;
```

```
CREATE TABLE trans_e3 AS
```

```
SELECT *,Substr(cno,6,2) AS n_c1,Substr(teno,6,2) as n_tel
```

```
FROM trans_e2;
```

```
QUIT;
```

```
DATA trans_e;
```

```
SET bankloc.trans_e3;
```

```
    n_c=input(n_c1,best12.);
```

```
    n_te=input(n_tel,best12.);
```

```
RUN;
```

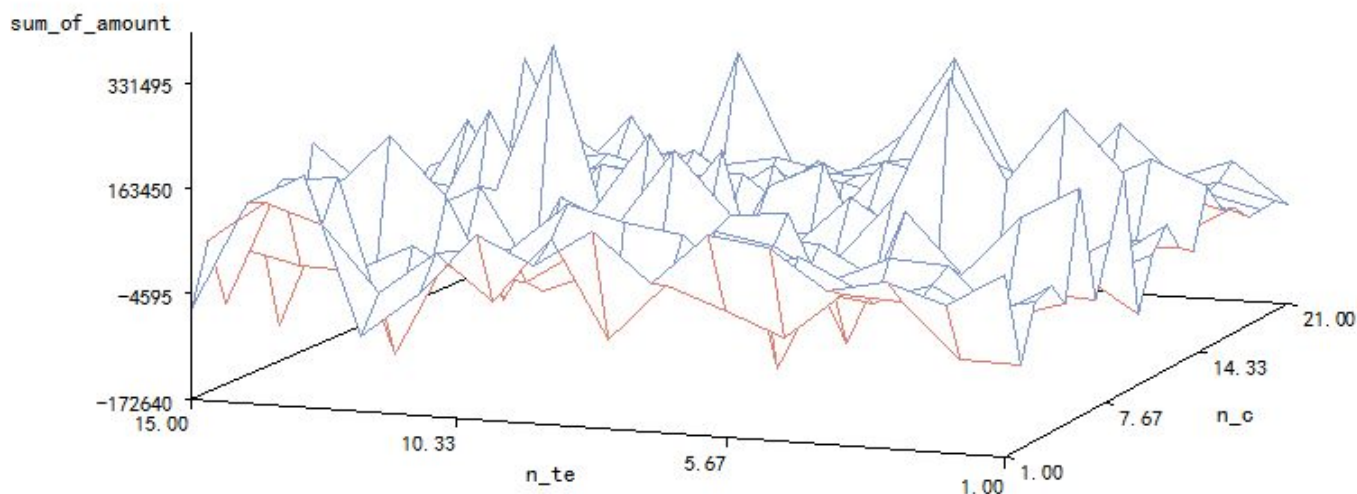
VIEWTABLE: Bankloc.Trans_e							
	cno	teno	sum_of_amount	n_c1	n_tel	n_c	n_te
1	c100001	e106001	110768	01	01	1	1
2	c100001	e106002	55719	01	02	1	2
3	c100001	e106003	85955	01	03	1	3
4	c100001	e106004	66588	01	04	1	4
5	c100001	e106005	127848	01	05	1	5
6	c100001	e106006	141067	01	06	1	6
7	c100001	e106007	60631	01	07	1	7
8	c100001	e106008	136088	01	08	1	8
9	c100001	e106009	74979	01	09	1	9
10	c100001	e106010	119707	01	10	1	10
11	c100001	e106011	18951	01	11	1	11
12	c100001	e106012	-54939	01	12	1	12
13	c100001	e106013	196234	01	13	1	13
14	c100001	e106014	147490	01	14	1	14

```
/*draw 3D scatter plot*/
```

```
PROC G3D DATA = trans_e;
```

```
PLOT n_te * n_c = sum_of_amount;
```

RUN;



9. Create data set named as balanceyearly including variables of year and balance in bank for each year (1988 to 2014) and draw the line plot graph. Build a nonlinear regression model by using PROC NLIN and forecast balance from year 2015 to year 2020.

```
/*use function year( ) , first.var and last.var to calculate variables which we need and
then create a data set called balance*/
```

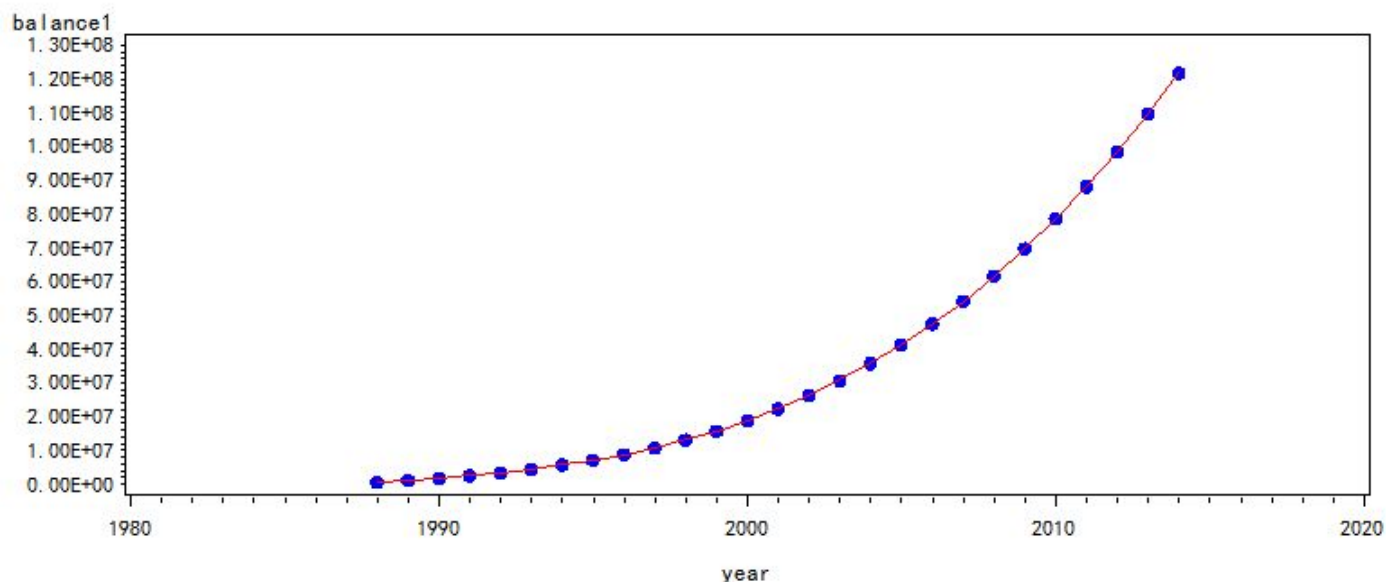
```
DATA trans_f1;
SET trans_c2;
    year = year(tdate);
RUN;
PROC SORT DATA = trans_f1;
BY account year;
RUN;
DATA trans_f2;
SET trans_f1;
BY account year;
IF last.year;
RUN;
PROC SORT DATA = trans_f2;
BY year;
RUN;
DATA balance;
SET trans_f2;
BY year;
RETAIN balancel 0;
IF first.year = 0
THEN balancel + balance;
IF last.year;
IF year ^= 1987 AND year ^= 2015;
KEEP year balancel;
RUN;
```



```

/*draw the line plot graph*/
SYMBOL VALUE = dot CV = blue
      INTERPOL = join CI = red;
PROC GPLOT DATA = balance;
PLOT balance1 * year;
RUN;

```



```

/*Build a nonlinear regression model */
PROC NLIN DATA = balance ;
PARMS a0 = 0 a1 = 1 a2 = 1;
MODEL balance1 = a0 + a1 * year + a2 * year * year;
RUN;

```



The NLIN Procedure
 Dependent Variable balance1
 Method: Gauss-Newton

Iterative Phase				
Iter	a0	a1	a2	Sum of Squares
0	0	1.0000	1.0000	6.321E16
1	9.095E11	-9.133E8	229295	1.113E14

NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	1
R	3.103E-6
PPC(a0)	5.124E-7
RPC(a0)	9.095E17
Object	0.998239
Objective	1.113E14
Observations Read	27
Observations Used	27
Observations Missing	0

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	2	3.486E16	1.743E16	3757.91	<.0001
Error	24	1.113E14	4.639E12		
Corrected Total	26	3.497E16			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
a0	9.095E11	3.065E10	8.462E11	9.727E11
a1	-9.133E8	30633235	-9.765E8	-8.501E8
a2	229295	7654.5	213497	245093

Approximate Correlation Matrix			
	a0	a1	a2
a0	1.0000000	-0.9999985	0.9999940
a1	-0.9999985	1.0000000	-0.9999985
a2	0.9999940	-0.9999985	1.0000000

10. Create data set name ecsamount with variables of row number, employee name, customer name, sum of saving amount. Build a regression model by using PROC GLM and analyze the parameters of model.

```
/* select amount into saving_amount when amount > 0*/
```

```
DATA ecsamount_g1;
```

```
SET cdtrans;
```

```
saving_amount = amount;
```

```
IF amount > 0;
```

```
KEEP aeno cno saving_amount;
```

```
RUN;
```

```
DATA ecsamount;
```

```
SET ecsamount_g1;
```

```
row_number=_N_;
```

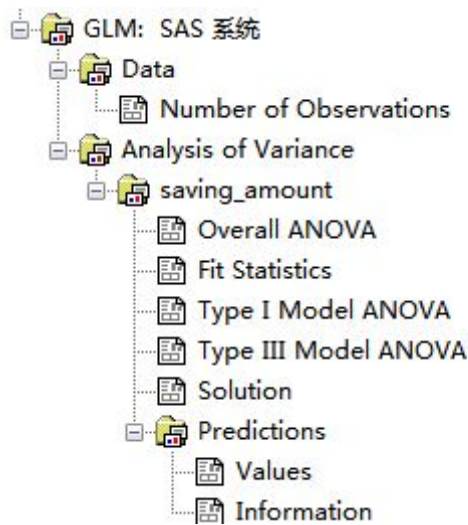
```
RUN;
```

```
/*Build a regression model by using PROC GLM*/
```

```
PROC GLM DATA = ecsamount;
```

```
MODEL saving_amount = row_number / CLM;
```

```
RUN;  
QUIT;
```



Number of observations:

The GLM Procedure

Number of Observations Read	12053
Number of Observations Used	12053

Overall ANOVA:

The GLM Procedure

Dependent Variable: saving_amount

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	35012990.134	35012990.134	2.02	0.1556
Error	12051	209205421366	17360005.092		
Corrected Total	12052	209240434357			

Fit statistics:

R-Square	Coeff Var	Root MSE	saving_amount Mean
0.000167	76.83627	4166.534	5422.613

Type I model ANOVA:

Source	DF	Type I SS	Mean Square	F Value	Pr > F
row_number	1	35012990.13	35012990.13	2.02	0.1556

Type III model ANOVA:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
row_number	1	35012990.13	35012990.13	2.02	0.1556

Solution:

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	5329.252681	75.90744370	70.21	<.0001
row_number	0.015490	0.01090744	1.42	0.1556

Predictions

Values:

The GLM Procedure					
Observation	Observed	Predicted	Residual	95% Confidence Limits for Mean Predicted Value	
1	5019.00000	5329.26817	-310.26817	5180.49589	5478.04046
2	244.00000	5329.28366	-5085.28366	5180.52989	5478.03743
3	691.00000	5329.29915	-4638.29915	5180.56390	5478.03441
4	3361.00000	5329.31464	-1968.31464	5180.59790	5478.03138
5	2768.00000	5329.33013	-2561.33013	5180.63191	5478.02836
6	1387.00000	5329.34562	-3942.34562	5180.66591	5478.02534
7	4304.00000	5329.36111	-1025.36111	5180.69991	5478.02232
8	574.00000	5329.37660	-4755.37660	5180.73391	5478.01930
9	980.00000	5329.39209	-4349.39209	5180.76791	5478.01628
10	1858.00000	5329.40758	-3471.40758	5180.80191	5478.01326
11	6237.00000	5329.42308	907.57692	5180.83591	5478.01024
12	1217.00000	5329.43857	-4112.43857	5180.86991	5478.00723
13	2803.00000	5329.45406	-2526.45406	5180.90390	5478.00421
14	5046.00000	5329.46955	-283.46955	5180.93790	5478.00119
15	9039.00000	5329.48504	3709.51496	5180.97190	5477.99818
16	40000.00000	5329.50053	46709.99947	5181.00590	5477.99516

Information:

Sum of Residuals	4.7966751E-9
Sum of Squared Residuals	209205421366
Sum of Squared Residuals - Error SS	-0.000366211
PRESS Statistic	209275162693
First Order Autocorrelation	0.000683392
Durbin-Watson D	1.9986318261