SAS Data set for Q1: WORK.Hist_sales

Q1.Using SAS Array technique to replace all the numeric variables' missing values with 0.

SAS Data set for Q2: WORK.Overdue

Q2. Browsing the SAS data set 'WORK.Overdue', the variables 'month1' to 'month12' are all binary variables(1 means 'overdue' and 0 otherwise)for each customer (the customer's ID is given by the column 'id' in the table). Please use SAS data step functions and array to solve the following questions

(1)Total overdue times for each customer in 12 months

(2)The overdue rate (i.e. %) for each customer in 12 months

(3)Create 12 new columns of character type -- 'overdue_1', 'overdue_2',..'overdue_12'. It is required that the value of variable overdue_j (j=1, 2...12) is either 'YES' if month*j*=1 (j=1, 2...12) or 'No' if month*j*=0 (j=1, 2...12).

Q3. You have the following raw data of phone numbers (see the following data step program) and wish to verify that all the numbers are in the following format: (nnn)nnn-nnnn, where n must be a digit 0-9. Extra spaces are permitted, and all numbers could be 15 characters or less. Create a new variable 'valid' and assign the value 'YES" if the phone number is valid otherwise assign the value 'NO'.

```
data phone;
  infile datalines DSD;
  length tel_number $20.;
  input id tel_number $;
datalines;
1,(988)463-4490
2,(241) 343-2233
3,456-5034
4,(123)456-7890
5,(271)SH4-1234
6,(592)2578362
;
run;
```

1, (988)463-4490 (valid)
2, (241) 343-2233 (valid)
3, 456-5034 (invalid)
4, (123)456-7890 (valid)

5, (271)SH4-1234 (invalid)
6, (592)2578362 (invalid)

SAS Data set for Q4: WORK.Exam_grade

Q4. Observing the SAS data set 'WORK.Exam_grade', you can find there are 739 students taking part in the exam in the form of multiple choices. The variables 'anwser1' to 'anwser12' are the answers for each student. The correct answer is ('A' 'B' 'D' 'A' 'C' 'A' 'B' 'B' 'D' 'A' 'B' 'B'). Based on the standard answer and student's answer to grade each student. If the answer is correct then he/she gets 1 point otherwise gets 0 point. Also student can leave answer as missing, and in that case he/she get 0.5 point. Please write SAS data step codes to calculate the grade for each student (hint: use temporary array).

SAS Data set for Q5: WORK.transaction

Q5. Check the SAS data set 'WORK.transaction' below. You can find it contains the columns 'customer_id', 'tran_time' (transaction time), 'amount' (transaction amount), 'last' (indicator of being last transaction for this customer), 'first' (indicator of being first transaction for this customer).

	customer_id	tran_time	amount	last	first
1	136	.	$485	N	Y
2	136	02JUL13:08:41:26	$352	N	N
3	136	04JUL13:08:38:17	$475	N	N
4	136	05JUL13:21:32:05	$2,449	N	N
5	136	06JUL13:08:20:58	$532	N	N
6	136	08JUL13:23:45:31	$248	N	N
7	136	13JUL13:15:40:32	$231	N	N
8	136	16JUL13:14:55:17	$1,289	N	N
9	136	17JUL13:07:46:33	$29	N	N

You now apply the data step functions 'DIFn()' and 'LAGn()' on the table above to obtain the following SAS data set 'WORK.transaction_lag', where the columns 'time_diff_1' to 'time_diff_3' contain the value of time (by hours) lag (or duration) from the last 2, 3, and 4 transaction to the last transaction. The columns 'amount_1' to 'amount_3' contain the value of transaction amount (by dollar) difference from the last 2, 3, and 4 transaction to the last transaction. Finally, the last column 'last_tran_time' stands for the last transaction time for each customer. Please note the value of variable 'customer_id' is unique in the resulting data set.

	customer_id	time_dif_1	time_dif_2	time_dif_3	amount_1	amount_2	amount_3	last_tran_time
1	136	150.05	177.98	216.88	$1,338	$1,034	$183	29JUL13:07:12:42
2	305	8.38	115.44	121.82	$2,206	$16	$1,041	25JUL13:19:22:08
3	325	8.59	61.44	242.23	$287	$1,135	$143	28JUL13:18:39:20
4	401	12.62	45.85	70.74	$874	$710	$76	28JUL13:18:24:31
5	412	11.74	150.1	150.45	$201	$1,355	$1,046	29JUL13:21:43:47
6	568	41	112.28	136.86	$1,212	$248	$457	30JUL13:23:35:32
7	653	44.61	109.3	143.08	$358	$504	$252	28JUL13:15:06:46
8	736	16.16	23.44	46.8	$1,273	$1,057	$400	26JUL13:00:32:58
9	832	32.19	84.78	153.6	$48	$121	$429	30JUL13:15:02:46

SAS Data set for Q6: WORK.dailyprice

Q6. The data set 'work.dailyprice' contains the daily stock prices (open price) and volumes from the biggest stock exchange in US (check exec='AMEX','NASDAQ' and 'NYSE' and stock ticker variable 'symbol'). The columns 'pc_1' to 'pc_100' stand for the daily price today, yesterday, the day before yesterday…until 99 days before today (i.e. pc_100). In the similar way, the variables 'vo_1' to 'vo_100' represent daily volumes in the past 100 days. Please solve the following questions based on the data set.

(a) Create the following new variable 'meanprice50', 'maxprice50', 'minprice50', 'rangeprice50' and 'stdprice50' which are the values of MEAN, MAX, MIN, RANGE and STANDARD DEVIATION of the daily price in the past 50 days. Save these columns the existing columns 'EXEC' and 'SYMBOL' in a new SAS table.

(b) Calculate (and create) new variables 'changerate_price_1', 'changerate_price_2'… 'changerate_price_30' and 'changerate_volume_1', 'changerate_volume_2'… 'changerate_volume_30' based on the values of daily price and volume. These variables are the changing rates from yesterday to today, the day before yesterday to yesterday,...from 31^th day to 30^th day. (For example, changerate_price_2=100*(pc_2-pc_3)/pc_3, changerate_volume_15=100*(vo_15-pc_16)/vo_16).

SAS Data set for Q7: WORK.Ticketinfo

Q7. Checking the data set 'WORK.Ticketinfo', you may find the first column contains the ticket information. For example, if ticket='TO*LOS_200 31111', then the 'LOS' stands for the destination name, '200 31111' indicates the starting time=NOV 11 2003, and the last column represents the birthday of each customer. Please apply SAS data step functions on this data set to generate the following three

new variables (1) ‘destination’ (2) ‘start_time’ and (3)‘age’ (the age of each customer)

SAS Data set for Q8: WORK.Customers

Q8. Use the uniform distribution generation function ‘RANUNI()’ to draw some random samples from the data set ‘WORK.Customers’. It is required to follow the following rules (1) The samples only contain the population with lan_spoken=’E’ (2) The sample size is around 1000.