**Problem 1 Extracting Characters Based on Position (2 Points)**

The data set **orion.newcompetitors** has data on competing retail stores that recently opened near existing Orion Star locations.

**orion.newcompetitors**

Postal\_

ID City Code

AU15301W PERTH 6002

AU12217E SYDNEY 2000

CA 150 Toronto M5V 3C6

CA 238 Edmonton T5T 2B2

US 356NC charlotte 28203

US1013CO denver 80201

US 12CA San diego 92139

* 1. Orion Star would like a data set containing only the small retail stores from these observations.
* Create a new variable, **Country**, that contains the first two characters of **ID**.
* Create a new variable, **Store\_Code**, that contains the other characters from the value in **ID**. Left-justify the value so that there are no leading blanks.
* The first character in the value of **Store\_Code** indicates the size of the store, and **1** is the code for a small retail store.
* Write a program to output only the small retail store observations.

Hint: You might need to use a SUBSTR functions as part of a subsetting IF statement

* Make sure that the **City** values appear in proper case (as displayed below).
  1. Print your results with an appropriate title.

Show these columns only once: **Store\_Code**, **Country**, **City**, and **Postal\_Code**.

PROC PRINT output (5 Total Observations)

New Small-Store Competitors

Store\_ Postal\_

Code Country City Code

15301W AU Perth 6002

12217E AU Sydney 2000

150 CA Toronto M5V 3C6

1013CO US Denver 80201

12CA US San Diego 92139

**Problem 2 Converting U.S. Postal Codes to State Names (2 Points)**

The data set **orion.contacts** contains a list of contacts for the U.S. charities that Orion Star donates to.

Partial **orion.contacts**

ID Title Name Address1 Address2

AQI Ms. Farr,Sue 15 Harvey Rd. Macon, GA 31298

CCI Dr. Cox,Kay B. 163 McNeil Pl. Kern, CA 93280

CNI Mr. Mason,Ron 442 Glen Ave. Miami, FL 33054

CS Ms. Ruth,G. H. 2491 Brady St. Munger, MI 48747

CU Prof. Florentino,Helen-Ashe H. PO Box 2253 Washington, DC 20018

1. Create a new data set named **states** that includes the variables **ID** and **Name** as well as a new variable named **Location** that shows the full name in proper case for the state that the contact is based in.

Hint: **Address2** is 24 characters long and the last item in **Address2** is always the ZIP code. ZIPNAMEL is a function that you like to use.

1. Print your results.

**Problem 3 Searching Character Values and Explicit Output (4 Points)**

* The data set **orion.employee\_donations** contains information about charity contributions from Orion Star employees.
* Each employee is allowed to list either one or two charities, which are shown in the **Recipients** variable.

Partial **orion.employee\_donations** (124 Total Observations, 7 Total Variables)

Employee\_ID Recipients

120265 Mitleid International 90%, Save the Baby Animals 10%

120267 Disaster Assist, Inc. 80%, Cancer Cures, Inc. 20%

120269 Cancer Cures, Inc. 10%, Cuidadores Ltd. 90%

120270 AquaMissions International 10%, Child Survivors 90%

120271 Cuidadores Ltd. 80%, Mitleid International 20%

* 1. Use explicit output to create a data set named **work.split**.
* The data set has one observation for each combination of employee and charity to which he donated.
* Some employees made two contributions. Therefore, they have two observations in the output data set. These employees contain a % character in the value of **Recipients**.

**Hint:** Store the position where the % character is found in a variable named **PctLoc**. This can make subsequent coding easier.

* Create a variable named **Charity** with the name and percent contribution of the appropriate charity.
* Read only the first 10 observations from **orion.employee\_donations** to test your program.
  1. Modify the program to read the entire **orion.employee\_donations** data set.
* Print only the columns **Employee\_ID** and **Charity**.
* Add an appropriate title.

Partial PROC PRINT Output (212 Total Observations)

Charity Contributions for each Employee

Employee\_ID Charity

120265 Mitleid International 90%

120265 Save the Baby Animals 10%

120267 Disaster Assist, Inc. 80%

120267 Cancer Cures, Inc. 20%

120269 Cancer Cures, Inc. 10%

**Problem 4 Calculating Statistics and Rounding (4 Points)**

The data set **orion.orders\_midyear** contains an observation for each customer, with the total retail value of the customer’s monthly orders for the first half of the year.

Partial Listing of **orion.orders\_midyear** (24 Total Observations)

Customer\_ID month1 month2 month3 month4 month5 month6

5 213.1 . 478.0 525.80 394.35 191.79

10 188.1 414.09 2876.9 3164.59 2373.44 169.29

11 78.2 . . . . 70.38

12 135.6 . 117.6 129.36 97.02 122.04

18 . . 29.4 32.34 24.26 .

24 93.0 265.80 . . . 83.70

27 310.7 782.90 . . . 279.63

31 1484.3 293.30 . . . 1335.87

34 642.5 . 86.3 94.93 71.20 578.25

1. Create a data set named **work.sale\_stats** with three new variables for all months in which the customer placed an order.

* The variable **MonthAvg** should contain the average.
* The variable **MonthMax** should contain the maximum.
* The variable **MonthSum** should contain the sum of values.
* Round **MonthAvg** to the nearest integer.

1. Print the variables **Customer\_ID**, **MonthAvg**, **MonthMax**, and **MonthSum**. Add   
   an appropriate title.

Partial PROC PRINT Output (24 Total Observations)

Statistics on Months in which the Customer Placed an Order

Month Month Month

Customer\_ID Avg Max Sum

5 361 525.80 1803.04

10 1531 3164.59 9186.41

11 74 78.20 148.58

12 120 135.60 601.62

1. 29 32.34 86.00

**Problem 5 Calculating Statistics for Missing, Median, and Highest Values (4 Points)**

The data set **orion.orders\_midyear** contains an observation for each customer, with   
the total retail value of the customer’s monthly orders for the first half of the year.

Partial **orion.orders\_midyear** (24 Total Observations)

Customer\_ID month1 month2 month3 month4 month5 month6

5 213.1 . 478.0 525.80 394.35 191.79

10 188.1 414.09 2876.9 3164.59 2373.44 169.29

11 78.2 . . . . 70.38

12 135.6 . 117.6 129.36 97.02 122.04

18 . . 29.4 32.34 24.26 .

* 1. Orion Star wants to look at information about the median order and the top two months’ orders,   
     but only for frequent customers.
* Create a data set named **work.freqcustomers** that contains the requested statistics.
* **Frequent customers are defined to be those who placed an order in at least five of the six months.**
  1. Print your results with an appropriate title.

Partial PROC PRINT Output (9 Total Observations)

Month Statistics on Frequent Customers

Month\_

Month\_ Month\_ 2nd

Customer\_ID month1 month2 month3 month4 month5 month6 Median Highest Highest

5 213.10 . 478.0 525.80 394.35 191.790 394.35 525.80 478.00

10 188.10 414.09 2876.9 3164.59 2373.44 169.290 1393.77 3164.59 2876.90

12 135.60 . 117.6 129.36 97.02 122.040 122.04 135.60 129.36

34 642.50 . 86.3 94.93 71.20 578.250 94.93 642.50 578.25

1. 134.00 119.20 313.0 344.30 258.23 120.600 196.11 344.30 313.00

**Problem 6 Changing a Variable’s Data Type (4 Points)**

The data set **orion.US\_newhire** contains information about newly hired employees.

Partial **orion.US\_newhire**

ID Telephone Birthday

120-012-40-4928 5467887 05DEC1972

120-012-83-3816 6888321 03MAY1969

120-341-44-0781 9418123 23NOV1976

120-423-01-7721 7839191 28JUN1971

Partial PROC CONTENTS Output of **orion.US\_newhire**

Variables in Creation Order

# Variable Type Len

1 ID Char 15

2 Telephone Num 8

3 Birthday Char 9

1. Create a new data set from **orion.US\_newhire**.

* Name the new data set **US\_converted**.
* Remove the embedded hyphens in **ID**.
* Convert **ID** to a numeric value.
* Convert **Telephone** to character and place a – (hyphen or dash) between the third and fourth digits.
* Convert **Birthday** to a SAS date value.

1. Print **US\_converted** with an appropriate title and use PROC CONTENTS to check the variables types.

Partial **US\_converted** (10 Total Observations)

US New Hires

ID Telephone Birthday

120012404928 546-7887 4722

120012833816 688-8321 3410

120341440781 941-8123 6171

120423017721 783-9191 4196

Partial PROC CONTENTS Output

Variables in Creation Order

# Variable Type Len

1 ID Num 8

2 Telephone Char 8

3 Birthday Num 8