

STAT 540: Test 2

1. Refer to the attached data set Q1, listing names, year of Ph.D., faculty rank, and salary (30 points).
 - (a) Write an INPUT statement to read in the variables FullName, PhDYear, Rank, and Salary.
 - (b) Grad students: write the INPUT statement to read in the variables FirstName, LastName, PhDYear, Rank, Salary.
 - (c) Suppose you had read in FirstName and LastName. Write a SAS statement that creates the variable FullName from these two variables.
 - (d) What format might you use to print Salary?
2. Assume that the data in Q1 has been stored in SAS data set WORK.Faculty (40 points).
 - (a) Write a subsetting IF statement to create a new data set Q2A containing only those records with PhdYear either before 1820 or after 1860.
 - (b) Repeat the same exercise using a WHERE command.
 - (c) Now create a data set Q2C from Faculty with a new variable Period that labels pre-1820 records as "Georgian", 1820-1860 as "Romantic", and post-1860 as "Victorian". Include a LENGTH statement so that Period is not truncated.
 - (d) Write statements to sort Q2C by Rank and compute summary statistics exclusively for Salary for each rank.
 - (e) Write statements you could add to your code in (c) to record the minimum salary in data set Q2C.
 - (f) Grad students. Use an IN clause to create a data set Q2f from Q2C that includes only the Georgian and Romantic periods.
3. Consider the SAS Commands for Q3, which includes analyte concentrations (Ba-Barium, Cu-Copper, Sb-Antimony, Pb-Lead) for a groundwater monitoring well. Concentrations below detection limit were listed as MDL (Below Minimum Detection Limit); any variable with MDL is initially read into GWWell as a character variable. (20 points)
 - (a) Use ARRAY to change MDL to a "" (missing value for character variable) for variables Cu, Sb and Pb.
 - (b) Add a statement to the ARRAY loop to convert the character variables Cu, Sb, and Pb to numeric variables.
 - (c) Grad students. Suppose Ba (Barium) was recorded in parts per million (ppm). How would you print the following label for Barium: "Barium (ppm)"?

4. Consider SAS Program Q4 (10 points).

- (a) The output data set contains 2 lines and 10 variables. What does the output data set look like?

Data set Q1

```
/*--5---10---15---20---25---30---35---40---45*/
```

```
Elizabeth Bennett  1813 Full  120000  
Jane Eyre          1847 Asst   82000  
Bathsheba Everdene 1874 Asst   95000  
Tess Derbyfield   1891 Full  110000  
Emma Woodhouse     1815 Assoc  92000  
Fanny Price        1814 Full  125000  
Eustacia Vye       1878 Asst   82500  
Catherine Earnshaw 1847 Assoc  93000
```

SAS Commands Q3

```
data GWwells;  
input  Year Period $ Ba Cu $ Sb $ Pb $;  
datalines;  
2011 Q1 33.4 MDL MDL MDL  
2011 Q2 45.6 1.2 MDL MDL  
2011 Q3 63.6 1.1 MDL .006  
2011 Q4 36.3 1.5 MDL .007  
2012 Q1 27.6 MDL MDL MDL  
2012 Q2 26.5 1.1 MDL .009  
2012 Q3 18.4 MDL MDL MDL  
2012 Q4 40.3 MDL MDL .006  
;
```

SAS Program Q4

```
data prog1;  
input x1-x5 @@;  
array x x1-x5; array sumx sum1-sum5;  
do over x;  
sumx+x;  
output;  
end;  
datalines;  
0.5 1.2 0.9 0.1 0.1 0.0 0.1 -1.3 0.4 -1.2 0.5 0.9  
;  
run;  
proc print data=prog1; run;
```

i