READ THE ENTIRE EXAM BEFORE STARTING

All exam files must be uploaded to the Sakai site by 12:00 PM on December 11th, 2019. There will be no extensions to the final exam.

Exams turned in after 12:00 PM on December 11th, 2019 will not be accepted. Thus, students should begin to prepare their exam files for submission well before the 12:00 PM submission deadline to ensure the files are submitted on time.

If the rare event that a student cannot successfully upload the exam to Sakai due to technical difficulties, the exam files should be sent to Matt Psioda via email (as a ZIP file) before 12:00 PM on December 11th, 2019.

Honor Code: *This is a take home exam. Communication with other students or non-students regarding this exam or about related material for the purpose of obtaining help on the exam constitutes a violation of the honor code.*

You may email Matt Psioda for clarification about exam questions but you may not request clarification from other students or non-students. No help will be given for the exam other than clarify the intent of the questions on the exam.

You may use online resources including SAS documentation, papers written about using SAS, textbooks, SAS help forums, and SAS blogs during the exam because these resources are freely available to all students. You may not post questions to programming blogs or email SAS tech support for help. Doing so constitutes an honor code violation.

All violations of the honor code will be reported to the Honor Court and will result in a grade of 0% being assigned for the final exam. By typing your name in the program header in each of the SAS programs that you are required to produce as a part of this exam, you are making a statement that you have abided by the honor code.

Directory Structure Requirements:

- You *must* use the following folder structure for your exam files. Failure to use this directory structure will result in a deduction of 5 points from your final exam grade.
    - ROOT
        - Parent directory for your final exam files
    - ROOT\programs
        - Location where programs for parts 1-3 should be placed
    - ROOT\output
        - Location where output PDF files should be written by your SAS programs for parts 2 and 3
    - ROOT\macros
        - Location where the FREQ.sas macro program should be stored for access by your SAS program for part 3
    - ROOT\qualtrics_data
        - Location where the SIS16.CSV file should be stored once downloaded
    - ROOT\analysis_data
        - Location where the ADSIS dataset should be written by your SAS program from part 1
    - ROOT\logs
        - Location where the SAS logs should be stored. The SAS logs should correspond to complete runs of your three SAS programs.

Each SAS program you write should include a macro variable named ROOT that defines the path (a location on your computer) to the root directory for your final exam files. Macro variables should be created for the paths to the output, macros, qualtrics_data, and analysis_data subfolders of the root directory and these macro variables should be created using the &ROOT macro variable reference so that I can run your programs by only changing the value of the ROOT macro variable. Thus, all SAS libraries and the locations where files are read from and written to by your SAS programs should be determined by using the macro variables you create for the output, macros, qualtrics_data, and analysis_data subfolders of the root directory.

Submission of Final Exam Files:

To submit final exam files, students should ensure all appropriate files are stored in the directory structure above as described above and then ZIP up the *entire* root directory to create a ZIP file named FINAL-PID.ZIP where PID is your student PID number. Only the ZIP file should be uploaded to Sakai. To check that you have created the ZIP file correctly, you should be able to unzip the file in another location and see the ROOT folder and all its subfolders and files contained within the subfolders. Non-adherence to these requirements will result in a deduction of 5 points from your final exam grade.

Part I: Programming the ADSIS dataset. For this part of the exam you will write a SAS program named PART1-PID.SAS (where PID is your PID number) that produces a permanent SAS dataset (written to the ROOT\analysis_data folder) using as input the SIS16.CSV file (stored in the ROOT\qualtrics_data folder). Be sure to write a well-documented, legible SAS program and otherwise use good programming practices.

- An excerpt from the SIS16.CSV file is show below (data from columns E through L are suppressed):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	ResponseId	Q1	Q2	Q3	.	.	.	.	.	.	.	.	Q12	Q13	Q14	Q15	Q16
2	Response ID	In the past 2 weeks, how difficult was it to dress the top part of your body?	In the past 2 weeks, how difficult was it to bathe yourself?	In the past 2 weeks, how difficult was it to get to the toilet on time?	.	.	.	.	.	.	.	.	In the past 2 weeks, how difficult was it to walk fast?	In the past 2 weeks, how difficult was it to climb one flight of stairs?	In the past 2 weeks, how difficult was it to walk one block?	In the past 2 weeks, how difficult was it to get in and out of a car?	how difficult was it to carry heavy objects (e.g. bag of groceries) with your affected hand?
3	1	4	3	3	.	.	.	.	.	.	.	.	5	3	4	5	3
4	2	5	5	5	.	.	.	.	.	.	.	.	5	5	4	5	5
5	3	2	1	2	.	.	.	.	.	.	.	.	3	3	3	2	2
6	4	1	3	3	.	.	.	.	.	.	.	.	2	3	2	2	2
7	5	2	4	3	.	.	.	.	.	.	.	.	4	3	5	4	3
8	6	4	5	5	.	.	.	.	.	.	.	.	5	5	5	5	5
9	7	2	5	3	.	.	.	.	.	.	.	.	4	4	3	3	2
10	8	1	4	3	.	.	.	.	.	.	.	.	1	2	5	3	2
11	9	5	5	4	.	.	.	.	.	.	.	.	4	4	5	4	4
12	10	3	3	5	.	.	.	.	.	.	.	.	4	4	5	4	4
13	11	5	5	5	.	.	.	.	.	.	.	.	5	5	4	5	5

- The SIS16.CSV file has 17 columns with the first column corresponding to the subject ID number and the next 16 columns corresponding to the 16 items on the Stroke Impact Scale (SIS-16).
    - The SIS-16 is commonly administered to stroke survivors to provide a self-assessment of social and physical functioning status following a defined period of time after their stroke (e.g. 90 days post-hospital discharge).
    - For more information on the SIS-16 instrument, see the following paper (not necessary):

      P.W. Duncan, S.M. Lai, R.K. Bode, S. Perera, J. DeRosa, the GAIN Americas Investigators. Neurology Jan 2003, 60 (2) 291296; DOI:10.1212/01.WNL.0000041493.65665.D6
    - Patient responses to the 16 survey items follow a 5-point Likert scale. For example, the 1st survey item asks "*In the past 2 weeks, how difficult was it to dress the top part of your body?*"

      Possible responses are:
      1 → Could not do at all
      2 → Very difficult
      3→ Somewhat difficult
      4 → A little difficult
      5 → Not difficult at all
    - A patient's responses are scored as integers ranging from 1-5 in the CSV file. **Note that some patients have missing values for 1 or more of the 16 survey items.**
    - The first row of the CSV file corresponds to the variable names for the SIS16.CSV dataset and the second row corresponds to their labels which, for variables Q1-Q16, correspond to the SIS-16 survey items.
    - DO NOT manually modify the CSV file in any way (e.g. in Excel). The CSV file must be left as it is and all programming of the ADSIS dataset must be done in SAS.
- You must write a SAS program that reads in the SIS16.CSV file and creates an analysis dataset named ADSIS which must have **ONLY** the following variables with the column order shown below:

ADSIS Variables & Attributes				
#	**Variable**	**Type**	**Len**	**Label**
1	USUBJID	Char	4	Unique Subject ID
2	QSSEQ	Num	8	Item Sequence Number
3	QSTESTCD	Char	10	Survey Item Code
4	QSTEST	Char	200	Survey Item
5	QSTYP	Char	10	Survey Item Type
6	AVALC	Char	20	Analysis Value (Character)
7	AVAL	Num	8	Analysis Value
8	QSSTAT	Char	50	SIS-16 Score Status
9	QSREASND	Char	50	Reason SIS-16 Score Not Calculated

- The following is a print out of data for two subjects to help you understand the correct ADSIS data structure. The data values do not necessarily match the CSV file you have been given and should not be used to check your dataset.

USUBJID	QSSEQ	QSTESTCD	QSTEST	QSTYP	AVALC	AVAL	QSSTAT	QSREASND
0003	1	ITEM01	In the past 2 weeks, how difficult was it to dress the top part of your body?		Very difficult	2		
0003	2	ITEM02	In the past 2 weeks, how difficult was it to bathe yourself?		Could not do at all	1		
0003	3	ITEM03	In the past 2 weeks, how difficult was it to get to the toilet on time?		Very difficult	2		
0003	4	ITEM04	In the past 2 weeks, how difficult was it to control your bladder (not have an accident)?		Somewhat difficult	3		
0003	5	ITEM05	In the past 2 weeks, how difficult was it to control your bowels (not have an accident)?		Very difficult	2		
0003	6	ITEM06	In the past 2 weeks, how difficult was it to stand without losing balance?		A little difficult	4		
0003	7	ITEM07	In the past 2 weeks, how difficult was it to go shopping?		Could not do at all	1		
0003	8	ITEM08	In the past 2 weeks, how difficult was it to do heavy household chores (e.g. vacuum, laundry or yard work)?		Very difficult	2		
0003	9	ITEM09	In the past 2 weeks, how difficult was it to stay sitting without losing your balance?		Could not do at all	1		
0003	10	ITEM10	In the past 2 weeks, how difficult was it to walk without losing your balance?		Somewhat difficult	3		
0003	11	ITEM11	In the past 2 weeks, how difficult was it to move from a bed to a chair?		Somewhat difficult	3		
0003	12	ITEM12	In the past 2 weeks, how difficult was it to walk fast?		Somewhat difficult	3		
0003	13	ITEM13	In the past 2 weeks, how difficult was it to climb one flight of stairs?		Somewhat difficult	3		
0003	14	ITEM14	In the past 2 weeks, how difficult was it to walk one block?		Very difficult	2		
0003	15	ITEM15	In the past 2 weeks, how difficult was it to get in and out of a car?		Very difficult	2		
0003	16	ITEM16	In the past 2 weeks, how difficult was it to carry heavy objects (e.g. bag of groceries) with your affected hand?			.		
0003	17	SIS16	Stroke Impact Scale 16 Score	DERIVED	31.67	31.67		
0105	1	ITEM01	In the past 2 weeks, how difficult was it to dress the top part of your body?		Not difficult at all	5		
0105	2	ITEM02	In the past 2 weeks, how difficult was it to bathe yourself?		Not difficult at all	5		
0105	3	ITEM03	In the past 2 weeks, how difficult was it to get to the toilet on time?			.		
0105	4	ITEM04	In the past 2 weeks, how difficult was it to control your bladder (not have an accident)?			.		
0105	5	ITEM05	In the past 2 weeks, how difficult was it to control your bowels (not have an accident)?			.		
0105	6	ITEM06	In the past 2 weeks, how difficult was it to stand without losing balance?			.		
0105	7	ITEM07	In the past 2 weeks, how difficult was it to go shopping?			.		
0105	8	ITEM08	In the past 2 weeks, how difficult was it to do heavy household chores (e.g. vacuum, laundry or yard work)?			.		
0105	9	ITEM09	In the past 2 weeks, how difficult was it to stay sitting without losing your balance?			.		
0105	10	ITEM10	In the past 2 weeks, how difficult was it to walk without losing your balance?			.		
0105	11	ITEM11	In the past 2 weeks, how difficult was it to move from a bed to a chair?			.		
0105	12	ITEM12	In the past 2 weeks, how difficult was it to walk fast?			.		
0105	13	ITEM13	In the past 2 weeks, how difficult was it to climb one flight of stairs?			.		
0105	14	ITEM14	In the past 2 weeks, how difficult was it to walk one block?			.		
0105	15	ITEM15	In the past 2 weeks, how difficult was it to get in and out of a car?			.		
0105	16	ITEM16	In the past 2 weeks, how difficult was it to carry heavy objects (e.g. bag of groceries) with your affected hand?			.		
0105	17	SIS16	Stroke Impact Scale 16 Score	DERIVED		.	NOT CALCULATED	Only 2 Items Answered

- The ADSIS dataset that you program should have variable names, types, lengths, labels, and column ordering that matches the given ADSIS specifications **exactly**.

- The ADSIS dataset should have 17 observations for each observation in the SIS16.CSV file with the first 16 corresponding to the SIS-16 survey items and the 17th corresponding to a computed SIS-16 summary score.

- USUBJID → 4 character subject ID variable created from the Response ID column in the CSV file. For example, if Response ID equal 1 then USUBJID will equal 0001. **Hint**: zero-padding can be done with a SAS-supplied format and using the PUT function.

- QSSEQ → Within-subject observation sequence number that matches the ordering above. **The submitted ADSIS dataset should be ordered by USUBJID and then by QSSEQ.**

- QSTESTCD → Will have values equal to ITEM01 - ITEM16 for the observations corresponding to the 16 survey items and SIS16 for the observations corresponding to the computed SIS-16 score.

- QSTEST → Will have values equal to the SIS-16 survey item descriptions that are found in row #2 of the SIS16.CSV file or "Stroke Impact Scale 16 Score". The survey item descriptions can be manually programmed (i.e., user-written LABEL statements) but the optimal solutions will not do this.

- QSTYP → Will be missing for all observations corresponding to SIS-16 survey items and equal to "DERIVED" for observations corresponding to the computed SIS-16 score.

- AVALC → Text version of the patient's response to the SIS-16 survey item or 2-decmial place character version of the patient's computed SIS-16 score *without extra whitespace (e.g., " 22.5" is not correct even if the value of AVAL is 22.5)*.

- AVAL → Numeric version of the patient's response to the SIS-16 survey item (taken directly from the SIS16.CSV file) or patient's computed SIS-16 score which should be rounded to 2 decimal places using the ROUND function.

    - Computing the SIS-16 Score:

      [1] NMISS = # of survey items not answered by the patient

      [2] MAX_RAW = 5*(16 – NMISS) →maximum possible raw score based on NMISS

      [3] MIN_RAW = 1*(16 – NMISS) → minimum possible raw score based on NMISS

      [4] RAW_SCORE = sum of non-missing survey items for the patient

      [5] The SIS-16 score should only be computed if the patient answered at least 12 of the survey items and should be rounded to 2 decimal places once computed.

      SIS-16 Score = (RAW_SCORE - MIN_RAW) / (MAX_RAW - MIN_RAW) * 100

      The SIS-16 score is the percentage of the maximum possible score for the patient based on the number of survey items the patient answered.

- QSSTAT → Will only be non-missing for the observations corresponding to the computed SIS-16 score and will be equal to "NOT CALCULATED" when the SIS-16 score is not calculated because the patient responded to too few survey items. The variable will be missing otherwise.

- QSREASND → Will only be non-missing when QSSTAT is non-missing and will be equal to "Only X Items Answered" where 0<=X<12 is the number of items answered. Be sure to program the value of X depending on the actual data.

*** Suggestion: Manually spot check ADSIS dataset for several subjects. Choose subjects with problematic data. ***

Part II: Producing a basic graph in SAS. For this part of the exam you will write a SAS program named PART2-PID.SAS (where PID is your PID number, stored in the ROOT\programs folder) that produces a simple histogram and horizontal box plot of the computed SIS-16 scores from Part I. Be sure to write a well-documented, legible SAS program and otherwise use good programming practices. The graphics should adhere to the following specifications:

[1] Both graphs should be written to a single page PDF file named PART2-PID.PDF and stored in the ROOT\output folder. Your SAS program must use ODS statements to create and name the PDF file.

[2] The single page PDF file should have as a title "Distribution of SIS-16 Scores" and there should be no titles in the images themselves nor should there be any x-axis labels.

[3] Each graph must be 6 inches in width and 4 inches in height.

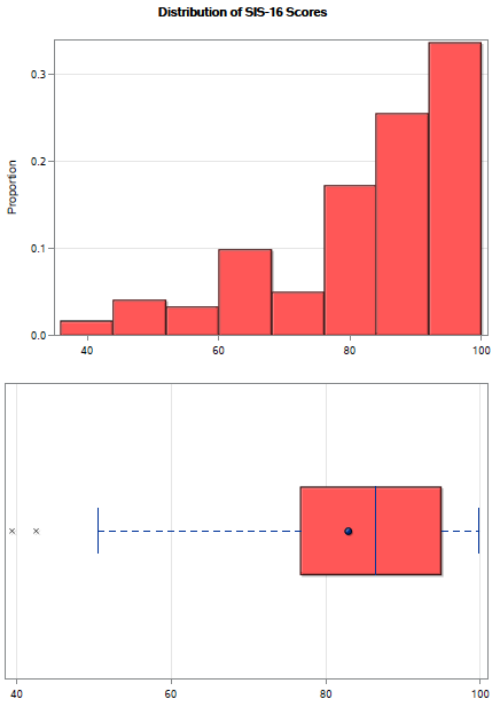
[4] The color filled elements in each graph should be light red and the bars/filled area in the box plot should use the CRISP data skin.

[5] The images should have no outer border (Hint: option on the ODS GRAPHICS statement).

[6] Grid lines should be used for the y-axis in the histogram.

[7] The mean symbol, outlier symbol, and the whiskers in the box plot should be customized as shown.

If the code is correctly written, the single page PDF should closely mirror the following:



NOTE: The data used to produce this figure is simulated and DOES NOT correspond the SIS-16 scores you will derive. Thus, you should not interpret anything about the distribution. Only use these figures to compare the layout and properties of the graphs.

Part III: Producing and using a SAS macro. For this part of the exam you will write a SAS program named PART3-PID.SAS (where PID is your PID number, stored in the ROOT\programs folder) that accesses a macro named %FREQ that is defined in a SAS program named PART3-FREQ-PID.SAS (stored in the ROOT\macros folder). Be sure to write a well-documented, legible SAS program/macro and otherwise use good programming practices.

Contents of the PART3-PID.SAS program:

[1] You must include the following PROC FORMAT step in the PART3-PID.SAS program (prior to any use of the %FREQ macro):

```
proc format;
 value fmtA
  1-3 = 'At Least Somewhat Difficult'
  4   = 'A Little Difficult'
  5   = 'No Difficulty';
run;
```

This PROC FORMAT step defines a format that will be used when your program calls the %FREQ macro.

[2] You must include the PART3-FREQ-PID.SAS program in your PART3-PID.SAS program using a %INCLUDE statement (to gain access to the %FREQ macro that is defined within the PART3-FREQ-PID.SAS program).

[3] The PART3-PID.SAS program should call the %FREQ macro using the following arguments:

```
%freq(cd=ITEM01, fmt=FMTA);
%freq(cd=ITEM02, fmt=NONE);
```

The output produced by the macro should be written to a two-page PDF file named PART3-PID.PDF which must be created in the ROOT\output folder using ODS statements. You will want to test/develop the macro using other values of the macro parameters but these should be the only the two macro calls (that are not commented out) in your submitted program.

Contents of the %FREQ macro defined in the PART3-FREQ-PID.SAS Program:

[1] The macro should have two non-positional parameters, named CD and FMT.

[2] The value of CD should be the word ITEM followed by a two digit representation of the numbers between 1 and 16 (e.g. 01, 02, ..., 16). Your macro should verify that the user provided value is correct for the ITEM macro variable and should abort execution if it is not (providing the user a relevant message for the reason execution was aborted).

[3] The value of the FMT macro variable should be FMTA or NONE and the macro should abort execution if it is not (providing the user a relevant message for the reason execution was aborted).

[4] The macro should perform a one-way frequency analysis using the variable AVAL for the observations from the ADSIS dataset where QSTESTCD is equal to the value of the ITEM macro variable. The one-way frequency analysis should use the format specified by the value of the FMT macro variable. If the value of FMT is NONE, then no format statement should be used in the analysis (i.e., no format).

For example, the macro calls shown below should produce the tables shown thereafter.

```
%freq(cd = ITEM14, fmt=FMTA);
%freq(cd = ITEM02, fmt=NONE);
```

Frequency Analysis of Survey Item 14
In the past 2 weeks, how difficult was it to walk one block?

Analysis Value	Frequency Count	Percent of Total Frequency
At Least Somewhat Difficult	76	60.80
A Little Difficult	30	24.00
No Difficulty	19	15.20

Frequency Analysis of Survey Item 2
In the past 2 weeks, how difficult was it to bathe yourself?

Analysis Value	Frequency Count	Percent of Total Frequency
1	24	19.20
2	24	19.20
3	30	24.00
4	32	25.60
5	15	12.00

Notice the survey item number and the survey item description are used in the title for the output. The SAS macro you create should do this as well. The data values have been simulated in this illustration and **DO NOT** match the correct ADSIS dataset.