Problem Set 2

Due in class on Wednesday, February 1st

For this Problem Set, you will be downloading and evaluating a dataset on ICPSR. You will need to register for an account in order to access the data. I do not recommend waiting until the last minute to get the data just in case you run into any problems.

Getting the Data:

Go to <https://www.icpsr.umich.edu/web/pages/>

Click on find data 🡪 find data 🡪 topics 🡪 social institutions and behavior

On the left-hand side click Restriction type 🡪 public use

# Click on “Classical Music Consumer Segmentation Study 2002.” Here you will find a description of the data, codebooks, questionnaires, and other information. If you click on Data and Documentation you can preview the codebooks. Click on these previews to start to get used to how the data is coded and structured. Read the description as well.

# Download the data by clicking on “download.” I typically have better luck downloading data in a non-SAS file format and then importing it into SAS. Don’t ask me why. It is up to you which file format you want. The advantage of Excel is you can always look at the data in Excel.

# Suppose you are hired as a consultant for the National Endowment for the Arts to evaluate the importance of classical music in Americans’ lives. For this homework, you will simply be discussing descriptive statistics of the data.

# Questions on description of the data

# Briefly describe the purpose of the study, including who conducted it and when. You may do this in one paragraph.

# Briefly describe the three surveys. Which of these most accurately represents the overall US population? You may do this in list form.

# Consider the National Survey. How could this survey be potentially biased?

# Consider question Q1C7. Take the mean of the data. Note that there are 13 missing values that you’ll have to drop. You may do so with this coding. Note that I’m saving the data under a new name, classical1 and had originally renamed it to classical.

data classical1; set classical;

IF Q1C7 <100 THEN OUTPUT;

 IF Q1C7 > 100 THEN DELETE;

# Calculate the 95% confidence interval for the sample mean. Interpret this.

# Recall that the purpose of your work is to evaluate the importance of classical music in Americans’ lives. Look through the codebook for the National Survey and choose three variables that you think are helpful in answering this question. List the variables here and describe how they’re coded. Note that you’ll have to choose three variables in addition to question Q3E, Interest in Learning More about Classical Music.

# For the three variables you selected AND question Q3E, calculate descriptive statistics by running proc freq

tables (list the variables you want here)

run;

# Make sure you consider the fact that there may be missing values. You do not have to drop these observations, but you may choose to do so. If you don’t drop them, you should at least discuss or report them.

# Use the dataset to answer the question, “how important is classical music in people’s lives?” Write up a 1-2 page analysis and include a clear table of values supporting your answer. You must turn in your SAS coding as well. Note that you may do this by using just three or four variables and simply looking at means or frequencies. You do NOT need to do work beyond this. (In other words, if you spend 5 hours on this, it is too long.)