MATH 262 - Applied Statistics
Lab 4: Online Dating

Cesar O. Aguilar

A research center conducted surveys pertaining to how relationships are affected by the internet. A portion of the surveys focused on online dating. The data is contained in the file `online_dating_data.csv`. The data includes the responses from 2,252 people age 18 and over. You will find five columns of data in the file. The data was collected during telephone interviews, by landline (1,125) and cell phone (1,127). The data includes the following variables:

- **age** - the age of the interviewee, where 97 = 97 or older, 98 = don't know, and 99 = refused to answer.

- **gender** - the gender of the interviewee, where 1 = male, 2 = female.

- **onlineDating** - response to the question: "Have YOU, personally, ever used an online dating site?", where 1 = yes, 2 = no, 8 = don't know, 9 = refused to answer. If blank, assume the question did not apply to the interviewee.

- **metOnline** - response to the question: "Is your spouse/current partner someone you first met ONLINE or someone you first met OFFLINE?", where 1 = met online, 2 = met offline, 8 = don't know, 9 = refused to answer. If blank, assume the question did not apply to the interviewee.

- **desperate** - response to the question:"Do you agree or disagree with this statement: People who use online dating sites are desperate", where 1 = agree, 2 = disagree, 8 = don't know, 9 = refused to answer.

- **goodWay** - response to the question: "Do you agree or disagree with this statement: Online dating is a good way to meet people", where 1 = agree, 2 = disagree, 8 = don't know, 9 = refused to answer.

Use jamovi to answer the following questions:

1. You will first apply filters to remove all rows that contain missing values for the **age** and **onlineDating** variables. To do this, use a filter to remove all rows where the **age** is greater than 97. These rows correspond to people whose age is unknown. In the main menu, select the **Data** tab, then **Filters** and then input the following filter:

$$\text{age} < 98$$

2. Add another filter to remove all rows where the value to the **onlineDating** variable is unknown. See the above description of the variables.

3. Now create a 95% confidence interval for the average age of **people who have used an online dating site** in two ways:

    (a) using the values for the mean and standard deviation computed by jamovi, and

    (b) by selecting the **Confidence interval for Mean** checkbox in the **Statistics** section of a **Descriptive Statistics**.

   Round both answers to one decimal place. Comment on your findings.

4. What is the margin of error for the confidence interval computed by jamovi? Round to one decimal place.

5. Now create a 95% confidence interval for the **proportion** of people who have used an online dating site. Round to two decimal places.

6. Using the sample proportion of people who have used online dating site, how large of a **total** sample is needed so that the margin of error in computing a 95% confidence interval for the **average age of people using online dating sites** is at most 2 years? Round to the nearest integer.