Author: Meghan Ng

Title: How to Date & Related Trends

Summary of research questions:

1. Do women and men in general share the same thoughts on what they look for in the other gender? For example, do both women and men value intelligence the most in potential partners? My hypothesis: no, they do not share the same thoughts. Participants were asked to allocate 100 points to various attributes of people based on what they valued the most. For each gender, I plan to average the points in each attribute and compare the averages for each gender. This is to see if women and men value the attributes in the other gender in the same way.
2. Is speed dating effective based on when comparing what participants look for in their potential partner with the actual attributes that their preferred dates hold? My hypothesis: what participants look for in their potential partner is accurately reflective of the actual attributes that their preferred dates hold. Participants were asked to allocate 100 points to various attributes of people as well as themselves based on what they value the most. I will use this information to compare the accuracy of their hopes versus what they chose. This is to see whether speed dating (such little time!) is effective in helping participants find the partners they prefer.
3. Do people generally prefer partners who have the same career paths as they do? My hypothesis: yes, people generally prefer partners who have the same career paths as they do. Participants were asked to write down their career path and I will compare their own career path with those of their preferred dates. This is to see whether participants who share the same career paths find each other more compatible.
4. Do people who go out (not necessarily on dates) more often get more matches than those who don't? My hypothesis: yes, people who go out (not necessarily on dates) more often get more matches than those who don't. Participants ranked themselves on a scale of 1-10 whether they go out often or not. I will then compare their rankings to the number of matches they had. This is to see whether spending time with other people will enhance their chances of finding matches during speed dating events.

Motivation and background:

Too often, the same questions and thoughts run through the minds of women and men after a breakup or being ghosted by a first date: "Is there something wrong with me? Men/women are so frustrating!" This dataset can be used to debunk myths about the other gender and allow people to better understand their own habits, interests, what they look for in a partner, and what a potential partner could look for in them. Hopefully by learning about these findings, people who are looking for a relationship can do so with the confidence in knowing that they are many steps closer to finding "the one."

Dataset:

URL to dataset (needs a simple sign up for Kaggle to download), located at the bottom of the page in the "Files" section: https://www.kaggle.com/annavictoria/speed-dating-experiment

**If it cannot be accessed, please let me know and I will send the dataset via email!

This dataset was compiled by Sheen Iyengar and Ray Fisman, professors of the Columbia Business School. The experimental speed dating events that produced the data were conducted from October 16, 2002 to April 7, 2004. Participants were asked to fill multiple questionnaires before, during, and after the events to describe themselves (demographics, personal values, habits), what they hoped to see in other participants, and their overall experience with the event. The answers to these questionnaires as well as information about their matches are included in the dataset.

Methodology (algorithm or analysis):

1. Do women and men in general share the same thoughts on what they look for in the other gender?

   In the .csv file, ignore the rows where the column "wave" has the integers 6 through 9. These rows contain information where participants were asked to rank various attributes of people based on what they valued the most on a scale of 1 to 10. For consistency, the focus will be on participants who were asked to allocate 100 points to these various attributes, where the column "wave" has integers 1 through 5 and 10 through 21. This will be followed in questions below in order to stay consistent with this question.

   For each gender, average (i.e., (a + b + c) / 3, etc.) the points in each attribute, then compare the averages for each gender. The resulting graph will be a bar plot where the x-axis consists of the attributes and the y-axis will be floats. Each attribute will have two bars: one for the women's average and the other for the men's average.

2. Is speed dating effective based on when comparing what participants look for in their potential partner with the actual attributes that their preferred dates hold

   Before the event started, participants allocated 100 points to various attributes of people based on what they valued the most (basically, they ranked the attributes). They also ranked themselves. At the end of the event, participants chose certain people they believed they had good connections with.

   For each participant, find out how they ranked the attributes they look for in others (with the exception of the attribute, "has shared interests/hobbies"). Then, for each participant they chose at the end of the night, find out how each of these "chosen ones"

ranked themselves. As an example, there are attributes a, b, and c. If the participant (the chooser) ranked these attributes from most important to least important as “b, c, a”, and one of the participants they chose ranked themselves as “a, b, c”, the chooser will have an accuracy of 0. This means that the participant got none of the positions of the attributes correct. If the chosen participant ranked themselves “b, a, c”, then the chooser will have an accuracy of 1, and so on. The accuracy will be calculated for each participant chosen, then averaged based on the number of participants chosen. The resulting graph will plot a histogram of the accuracy per participant. The x-axis will be the accuracy averages, and the y-axis will be the frequencies of these values.

3. Do people generally prefer partners who have the same career paths as they do?

   Participants wrote down their intended careers.

   If a choosing participant chooses a participant with the same career path as themselves, the chooser gets an accuracy of 1. If not, an accuracy of 0. The accuracy will be calculated for each participant chosen, then averaged based on the number of participants chosen. The resulting graph will be a histogram of the accuracy per participant. The x-axis will be the accuracy averages, and the y-axis will be the frequencies of these values.

4. Do people who go out (not necessarily on dates) more often get more matches than those who don’t?

   Participants ranked themselves (on a scale of 1 to 7, with more information related to this scale in the metadata) on how often they go out (not necessarily on dates).

   For each participant who ranked themselves as a 1, the number of matches these participants had will be summed up, and so on for each number on the scale. The resulting graph will be a histogram that shows the number of matches per ranking. The x-axis will be the scale and the y-axis will be integers.

Work Plan:

1. Pseudocode the entire program (7 hours)
2. Map out the functions required to implement the program (3 hours)
3. Write test cases for each function that requires them (2 hours)
4. Create report (3 hours)
5. Create presentation (2 hours)

Questions:

I’m a little stumped on how I should answer the questions once I have the results. I will be producing histograms, not scatterplots, and therefore (I think) will not be using correlations to justify my answers. Overall, how can I use histogram results to answer my questions?