Name: Gavin McRoy

Objectives: The purpose of this project is to introduce the statistical software JMP. Upon successful completion of this project, you will be able to…

- Identify the type of variable in the context of a problem
- Create graphical displays, frequency tables, and relative frequency tables in JMP
- Use the output provided by JMP to answer relevant questions
- Identify the population and the sample in the context of a problem
- Select a random sample using JMP
- Describe the possible effects of using a non-random sample in a study

Delectable Delights is a large consumer food manufacturer selling its products in retail stores nationwide. You have landed your first job after graduation from Clemson in their advertising division. Since you took statistics as a part of your coursework, you are often called upon to perform data analysis for the advertising division, as well as other divisions of the company.

Directions: Answer the following questions using complete sentences as though you were presenting your analysis to the employees of Delectable Delights. Please provide any appropriate output and/or screenshots from JMP. Instructions for creating several types of graphs or tables and statistics can be found on Canvas in the file **JMP Instructions.docx**. Paste your answers and any output into this document.

- (50 points) Delectable Delights would like to feature some of its products in major motion pictures. To assist in selecting possible move projects to approach, your boss, Ray Holtz, would like you to review movie genres to see which is most popular. Since Delectable Delights uses the software JMP, you decide to look into the Hollywood Movies data set that comes with JMP. You can find the file **Hollywood Movies.jmp** on Canvas.

    - Look at the data in the column labeled Genre. What **type** of variable is Genre and what is its **level of measurement**? (5 pts)
        - Response Variable, Qualitative, Nominal

- Using the Genre column in the Hollywood Movies data set, create the following graphical displays in JMP. Copy and paste your tables and graphs into the space provided. You can go to YouTube to see other ways to change the appearance of the tables and graphs you make.

    - **<u>Frequency and Relative Frequency Table</u>** (5 pts)

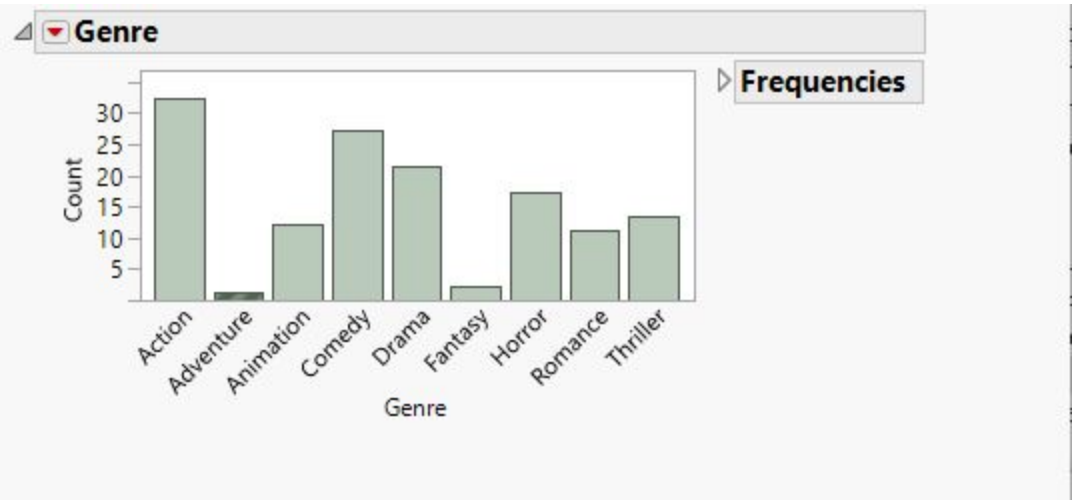
      To create frequency and relative frequency tables, follow the steps on page 2 of the JMP Instructions under **Frequency table, relative frequency table, and histogram**. Change the column titles of the table to Genre, Frequency, and Relative Frequency instead of Level, Count, and Prob. To change a title, simply double-click the title and enter your own. Copy and paste your table in the space below. To copy the table, right click on the table, and select **Copy Table** from the drop-down menu.

Genre	Frequency	Relative Frequency
Action	32	0.23529
Adventure	1	0.00735
Animation	12	0.08824
Comedy	27	0.19853
Drama	21	0.15441
Fantasy	2	0.01471
Horror	17	0.12500
Romance	11	0.08088
Thriller	13	0.09559
Total	136	1

    - **<u>Bar Chart</u>** (5 pts)

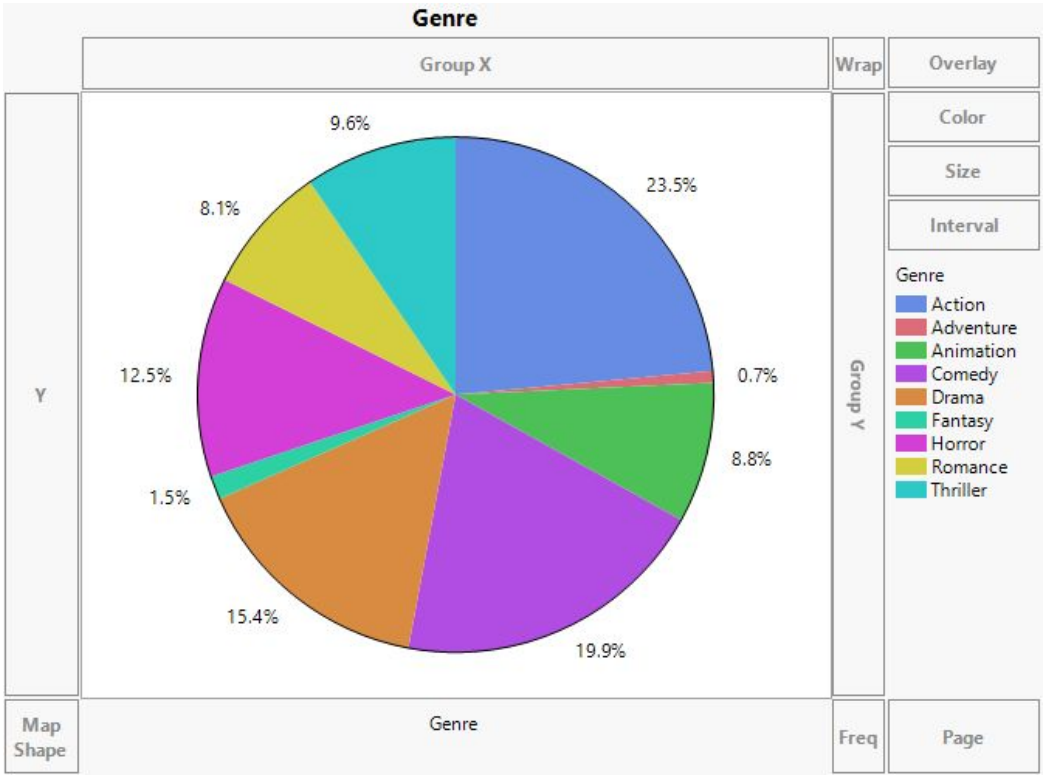
      To create a bar chart, click the red triangle to the left of the column heading (in this case it is Genre). Select **Display Options >> Axes on Left**. Click the red triangle again and select **Histogram Options >> Separate Bars**. Once more click the red triangle and select **Histogram Options >> Count Axis** or **Prob Axis**. This will place either the frequency or relative frequency axis on the bar chart. You can change the title of the vertical axis by double clicking on the default title. Lastly, right click on the movie genre titles and select **Add Axis Label**. If you would like to change the color of the bars, right click in the white space around the bars and select **Histogram Color**. Copy and paste the bar chart. To copy the bar chart, right click on the white space around the bars and **select Edit >> Copy Graph**. (There are instructions on

page 6 on how to use the Graph Builder feature to make a bar graph. JMP often has more than one way to create pictures of your data.)



- **<u>Pie Chart</u>** (5 pts)

  Follow the steps on page 5 of the JMP Instructions to make your pie chart. To add either counts or percentages to your pie chart, use the **Label** dropdown on the bottom left of the Graph Builder window. Copy and paste your chart here. To copy the pie chart, right click on the chart, go to **Edit >> Copy Graph.** This should both copy the graph and the legend.

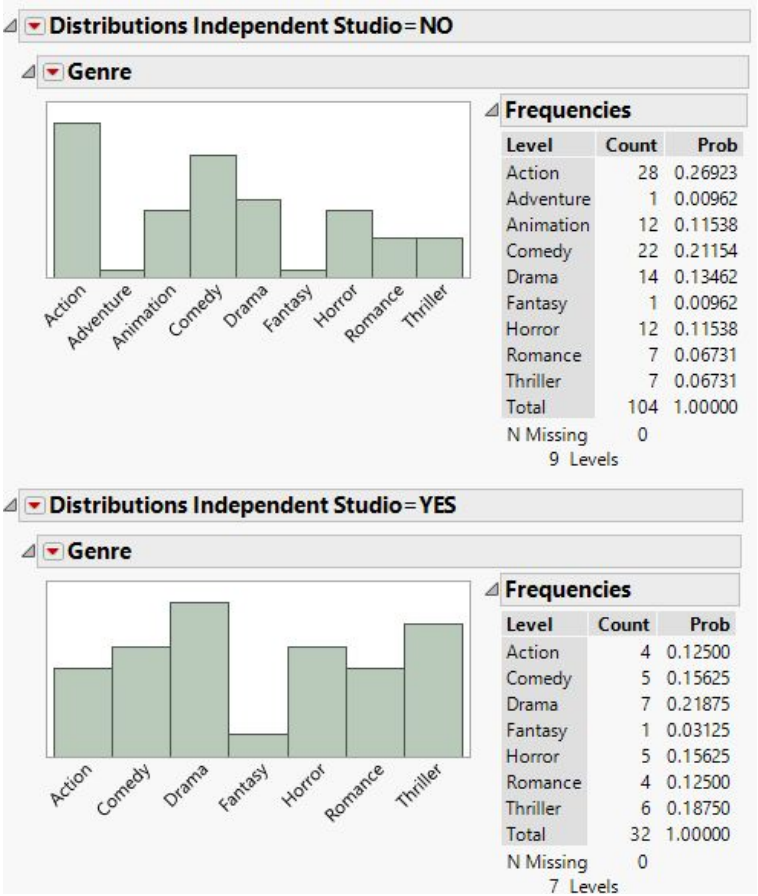


- C. You need to make a short report to Ray on the Genre that is most popular. **Which** of the tables/graphs that you made in Part (b) would you like to use? **Why** did you select that table/graph? (There is no one right answer to this question. You need to select the type of graph that you believe quickly conveys the count or proportion or percentage of each genre). (5 pts)
  - I would use the pie graph since it displays very quickly which genre is the most popular.

- D. Using the graph or table you chose in Part (c), **report** to Ray the most popular genre. **Explain** your answer based on the graph or table you selected. (5 pts)
  - According to the graph Action is the most popular with 23.5

- E. Let's separate the movie Genres based on whether they are produced by an independent studio or one of the major studios. (10 points)

  Redo the to the JMP analysis from Part (b) i., but this time place the Independent Studio column in the **By** block (see JMP instructions pages 2 and 3 under **Frequency table, relative frequency table, and histogram)**. Copy and Paste the JMP output in the space below. Either take a screenshot of the output or copy and paste the graphs and tables separately. You do not need to change any titles on the tables or separate the bars in the graph.



- Suppose that Delectable Delights wants to advertise with a Major studio. Does your answer in Part (d) change? **Support** your answer using the JMP output from Part (e). (5 points)
  - No it does not since Action is still the most popular genre

- Suppose that Delectable Delights wants to advertise with an Independent studio. Does your answer in Part (d) change? **Support** your answer using the JMP output from Part (e). (5 points)
  - Advertising with an independent studio changes which genres are most popular to drama.

- (50 points) In this question, we will explore a possible effect of using a self-selected sample (which is a type of convenience sample) rather than a random sample. Take a look at the list of the 200 top grossing movies of 2019. The list can be found on Canvas in the file **Top Movies 2019.pdf**.

  - In the context of this problem, what is the **population**? (5 points)

    -All movies released in 2019

  - What is the **sample**? (5 points)

    -The top 200 movies released in 2019

  - What is the **variable**? (5 points)

    -Top grossing movies of 2019

  - What **type** of variable is this and what is its **level of measurement**? (5 points)

    -Quantitative, Interval

  - Select a sample of 10 movies that you saw (or wanted to see) in theaters and write the titles of those 10 movies in the table below, along with the amount they grossed in 2019. This is your **self-selected sample**. Notice that the listing of movies gives the gross income rounded to the nearest million, so that a movie listed as earning $191.5 really grossed $191,500,000. The order in which you write the movies in the table below does not matter. (5 points)

Movie Title	Gross Income (Millions)
Avengers Endgame	**858.4**
Captain Marvel	**426.8**
Spiderman Far From Home	**390.5**
Joker	**333.8**
Jumanji The Next Level	**192.1**
Fast and Furious Hobbs and Shaw	**173.8**
John Wick Chapter 3	**171.0**
Shazam!	**140.4**
Knives Out	**115.7**
Rocketman	**96.4**

- Compute the **average** gross income for the sample of movies you selected. Show how you computed the average and include units in your answer. If you forgot how to calculate an average, simply add up the 10 gross income values and divide by 10. (5 points)

  Average Gross Income = 289.89 (millions)

- Is the average you calculated in Part (f) a **parameter** or a **statistic**? **How** do you know? (5 points)

  -Statistic: Because it is a fact about a sample

- Now you will select a random sample of 10 movies. Open a new blank data table in JMP. Select **Rows >> Add Rows** and enter **10**. Highlight the rows you just created. Next select **Cols >> Formula**. From the list of options on the left, select **Random >> Random Integer.**

  Double click the blue box and enter **200**. This will then generate 10 random integers between 1 and 200. Click **OK**.

Record the random numbers in the ID column in the table below. If you have a repeated number, generate another random integer so that there are no duplicates in the ID column. To generate an additional random integer in JMP, simply follow the steps above to add an additional row and JMP will generate another number. Select the 10 movies from the population of 200 movies that correspond to these ID numbers. Record the Movie Title and Gross Income of these 10 movies in the table below. (5 points)

ID	Movie Title	Gross Income (Millions)
67	**Queen & Slim**	**40.7**
16	**The Secret Life of Pets 2**	**158.3**
41	**Angel Has Fallen**	**69.0**
28	**The Lego Movie 2: The Second Part**	**105.8**
24	**Glass**	**111.0**
2	**The Lion King**	**543.6**
38	**Yesterday**	**73.3**
89	**The Art of Racing in the Rain**	**26.4**
91	**Uncut Gems**	**25.3**
105	**The Peanut Butter Falcon**	**20.5**

- Compute the **average** gross income for the random sample of movies. Show how you computed the average and include units in your answer. (5 points)
  - Average Gross Income = 117.39

- We often use statistics calculated from samples to estimate the true parameter value of a population. In this case, we are considering the average gross income for the population of 200 movies, which is $55.1 million.
  - Which of your samples came **closest** to the true average? (2 points)
    - The second estimate came the closest

- Regardless of your answer in Part (j) i. above, when using the self-selected sample, are you likely to **overestimate** or **under-estimate** the average gross? **Why**? (3 points)

- <u>You are likely to overestimate because you have a higher probability of seeing a higher advertised movie that grossed within the top 40 of all films that year compared to other less advertised and budgeted films.</u>