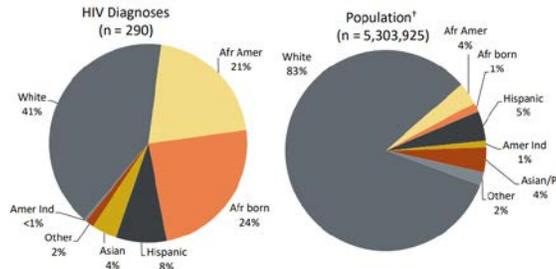


For the first couple problems on this take-home portion we will work with data from the STD, HIV and Hepatitis C 2016 Report published by the Minnesota Department of Health.

Report Source: <http://www.health.state.mn.us/divs/idepc/diseases/hiv/stats/2016/webinar2016.pdf>

Consider the information provided on Slide #49 (and the data presented on Slide #50).

HIV Diagnoses\* in Year 2016 and General Population in Minnesota by Race/Ethnicity



Next, consider the first talking points for the presenter on Slide #49.

Statement: “When you look at the racial/ethnic distribution of new HIV diagnoses for both genders compared to the racial/ethnic distribution of the general population of Minnesota, it is apparent that disparities continue.

**Presenter**  
 When you look at the racial/ ethnic distribution of new HIV diagnoses for both genders compared to the racial/ ethnic distribution of the general population of Minnesota, it is apparent that disparities continue to exist among persons of color.

- Using Handout #6, provide a statistical evaluation of the statement the Presenter is supposed to make on this slide. The observed values must be counts (not percentages) in Handout #6. (5 pts)

Research Question:

P-value: \_\_\_\_\_

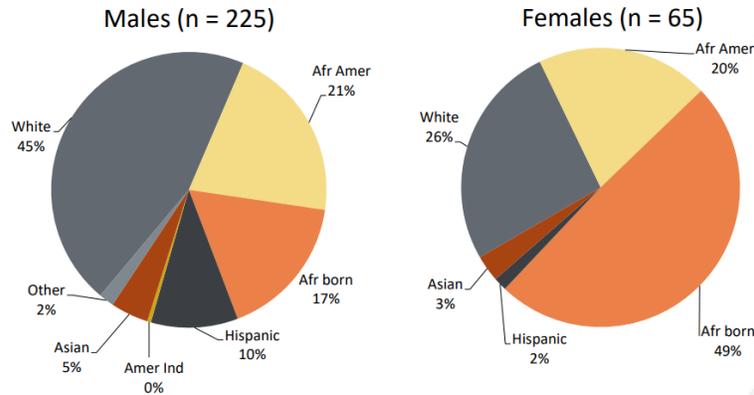
Conclusion:

- JMP automatically computes the Pearson ChiSquare value which is computed using the following formula. Which race contributes the most to this calculation? What does this mean for this race? Discuss. (3 pts)

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Next, consider the information provided on Slide #51.

### HIV Diagnoses\* Diagnosed in Year 2016 by Gender and Race/Ethnicity



Once again, consider the talking points for the presenter on Slide #51.

Statement: “There are differences in the racial/ethnic distribution by gender.”

**Presenter**  
There are differences in the racial/ethnic distribution by gender.

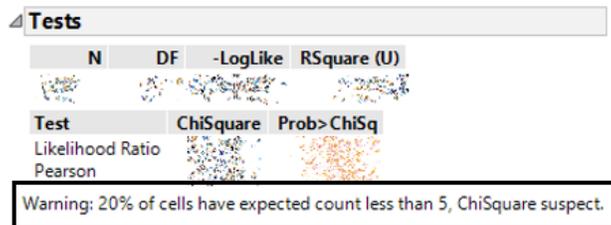
- Using Handout #8, provide a statistical evaluation of the statement the Presenter is supposed to make on this slide. Again, counts must be used for the analyses in Handout #8. (5 pts)

Research Question:

P-value: \_\_\_\_\_

Conclusion:

- When I run my analysis in JMP, JMP returned the following message: “Warning: 20% of cells have expected count less than 5, ChiSquare suspect.”



- Which cells have expected cell counts less than 5? What percent of the cells have expected cell counts less than 5? (2 pts)
- As a practicing statistician, should I be worried about this warning? Use whatever resources you'd like, e.g. other teachers, Google, etc., to answer this question. You should cite your source. (3 pts)

The popularity of bike share programs continues to grow. CitiBike is one company that runs the bike share program in the City of New York. They publish their data each month and we will use their data to answer the following questions.



Source: <https://www.citibikenyc.com/system-data>

Data Download: <https://s3.amazonaws.com/tripdata/index.html>

Download the most recent data, JC-201707, which is near the bottom of the data download page. The file being downloaded is a \*.zip file. WSU laptops come with software to unzip files of this type. Unzip this file so that the JC-201709-citibike-tripdata.csv can be opened into JMP.

■ JC-201709-citibike-tripdata.csv.zip	Oct 3rd 2017, 10:52:49 am	667 KB	ZIP file
---------------------------------------	---------------------------	--------	----------

Consider the following columns in this dataset.

- tripduration (duration of trip in seconds)
- usertype (Customer = 24 hour pass or 3-day pass -- temporary type member; Subscriber = Annual pass – more permanent type member)
- Year of Birth
- Gender (Zero = unknown, 1 = male, 2=female)

Answer the following questions.

1. Investigate the relationship between usertype (predictor) and Gender (response). You should relabel 1-to-male and 2-to-female before plotting. Write a brief data story about what you discovered here. (3 pts)
2. Create a new variable, say Trip Length, as follows

$$\text{Trip Length} = \begin{cases} \text{"Short Trip"}, & \text{if } \text{tripduration} < 15 \text{ minutes} \\ \text{"Long Trip"}, & \text{if } \text{tripduration} \geq 15 \text{ minutes} \end{cases}$$

Investigate the relationship between usertype (predictor) and Trip Length (response). Write a brief data story about what you discovered here. (3 pts)

3. Use the Day of Week function under the Data Time functions in JMP to rip off the day of week from stoptime. Next, use the recode feature in JMP (under Cols menu) to create a new variable, say DayType, where you label each rental as either weekday or weekend.

Investigate the relationship between usertype (predictor) and DayType (response). Write a brief data story about what you discovered here. (3 pts)

4. Consider once again the relationship investigated above between usertype (predictor) and Trip Length (response). Does DayType have an influence this relationship? Discuss. (3 pts)