

Mini-Project 2

Stat 212: Interim 2021

Chamee Vang, Marco Ruiz, Ivana Kramarevsky

1/13/2021

1. Identify Research Question and Background

a. Research Questions:

- 1) Does the education level of the participant affect how active they are through physical activities?
- 2) Do people would have a healthy heart rate overall?

b. Variable Codebook:*

Variable name	Original name	Description	Type	Levels/Encoding
Education Level	Education	Level of education for the participant	categorical	8th Grade, 9-11th Grade, High School, Some college, College Grad
Physical Activity	PhysActive	If the participant is physically active	categorical	Yes, No
Heart Rate	Pulse	The participant's pulse in 1 minute	Numeric	Identifier
College Level	College	If the participant went to college, regardless of graduation status	categorical	College, No college
Heart Health	HealthyHeart	If the participant's heart rate is within the healthy 60-100 BPM range	categorical	Yes, No

c. Literature Review

Accelerometer-Measured Physical Activity and Sedentary Time Differ According to Education Level in Young Adults: [link](#)

Accelerometer was used to measure the physical activities of young adults throughout the week and weekends. The study uses the education level as a criteria to determine the eligibility of the participants. The study's result found that there was no relationship between the education level and the measurements of physical activity. However by comparing the gender of the participants and how physically active they are throughout the week, both male and female participants with similar education levels have similar measurements.

Age-predicted maximal heart rate revisited: [link](#)

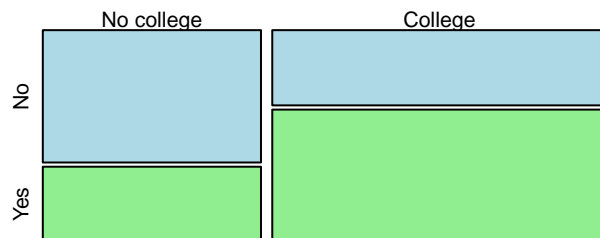
HRMax is a measure of the maximum heart rate by age, and it does not depend on gender nor physical activity. The relationship that they found is modelled with an equation with $r = -0.90$ (high negative correlation): The HRmax for some age is $208 - (0.7 * \text{age})$. For this study, only healthy, non-medicated non-smokers were recruited from ages 18 to 81. In addition, subjects with severe obesity (above 35% bmi) were excluded to avoid confounding variables.

2. Variable Exploration: EDA and CUSS

Question 1: Education Level vs. Physical Activity

```
##  
##           No  Yes  
## No college 1764  988  
## College    1537 2681
```

College and Physical Activity

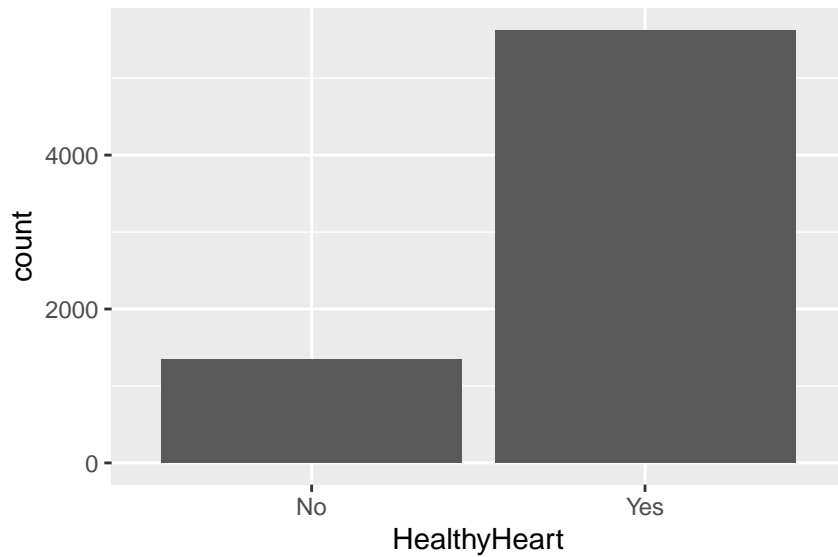


```
##  
##           No      Yes  
## No college 0.6409884 0.3590116  
## College    0.3643907 0.6356093  
## [1] 0.2765977
```

CUSS: Because this is a mosaic plot with two categorical variables, there is no center. The proportional difference between those who are at a college level and those who are not and if they were physically active is about 27.66%. Because of this difference in proportion, it seems like there are more of the participant who has gone to college or finished college are more physically active.

Question 2: Proportion of people with healthy heart rate

```
##  
## No  Yes  
## 1344 5626  
  
##  
## No      Yes  
## 0.1928264 0.8071736
```



CUSS: There is no center because this is a categorical variable with two categories. The “yes” category contains 80.7% of data points. The “no” category contains 19.3% of data points. There is a much larger proportion of data in “yes”, so, visually, it appears that most people have healthy heart rates.

3. Hypothesis Testing

a. Null and Alternative Hypothesis

Question 1: education level vs. physical activities

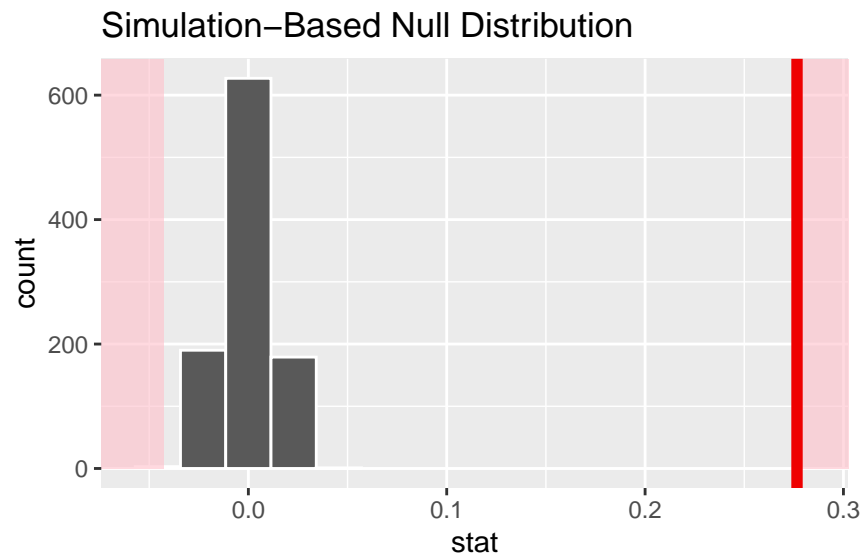
- Null: The level of education and how active a person is are not related.
 - $H_0 : p_{Education} - p_{PhysActive} = 0$
- Alternative: The level of education and how active a person is are in some way related to each other.
 - $H_a : p_{Education} - p_{PhysActive} \neq 0$

Question 2: proportion of people with a healthy heart rate

- Null: Overall, the proportion of people with healthy or unhealthy heart rates is the same.
 - $p_{healthy} = 0.5$
- Alternative: Overall, the proportion of people with healthy to unhealthy heart rates is not the same.
 - $p_{healthy} \neq 0.5$

b. Histogram of Null Distribution*

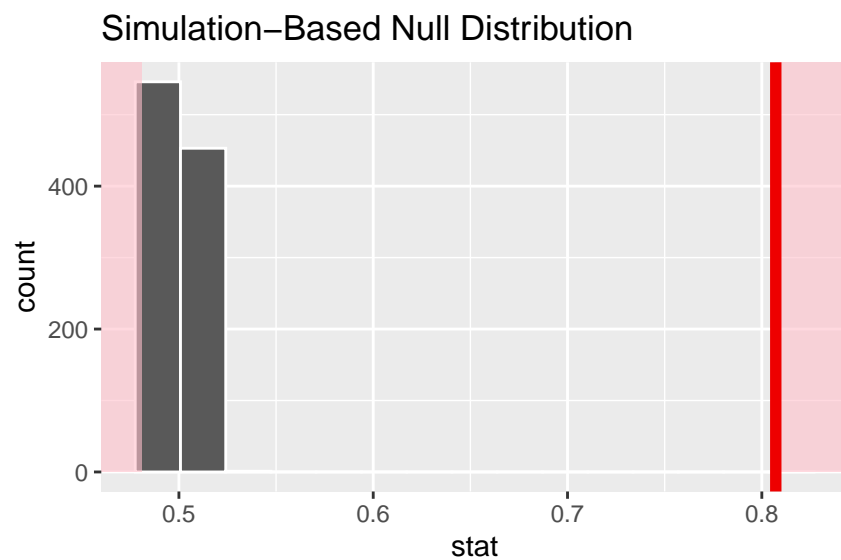
Question 1: Education level vs. Physical activities



```
## # A tibble: 1 x 1
##   p_value
##   <dbl>
## 1      0
```

Question 2: Proportion of people with a healthy heart rate

```
##
##   No  Yes
## 1344 5626
```



```
## # A tibble: 1 x 1
##   p_value
##   <dbl>
## 1      0
```

c. P-Value

Question 1: The p-value is 0. Almost no data in the null distribution is as unusual as our observation.

Question 2: The p-value is 0. There is virtually no data that is above the observed sample value compared to the null distribution using a mean of 0.5.

d. Interpretation

Question 1: There is enough evidence to reject the null hypothesis. It is likely that attending higher education increases the proportion of people that engage in moderate to intense physical activity.

Question 2: There is sufficient evidence to reject the null hypothesis since it is extremely unlikely that the observed data sample is due to chance under the null distribution. This means that the alternative hypothesis is likely: overall, people have healthy heart rates that are within 60 to 100 BPM