## **Topics 1: Age and the level of COVID prevention**

The variable exploring:

* DEMO\_age (numbers | 14, 24, 56 )
* COVID\_prevention\_distancing (categories | not at all, somewhat, very close)
  + - * \_masks
      * \_hand\_washing
      * \_redice\_people
      * \_avoid\_trips
      * \_household

WHY CHOOSING THESE VARIABLES

* Because I want to discuss Influence of different generations' ideologies on behavior

We are estimating the mean COVID-19 prevention score for each age group to determine if age affects prevention behaviors in the general population and the method we are using is **Null Hypothesis Testing** because it allows us to statistically determine if there is a significant relationship between age and COVID prevention behaviors.,where we assume that age has no significant effect on how people prevent COVID. To digitize how much people value COVID-19 prevention, we will transform the category data into numbers (not at all = 0 somewhat = 1 very close = 2), which we can then compare with people of different ages.

We split the age into groups, with 10 years as a group, from 10 to 60. We could use a **Histogram** to collect variance; observing the shape of the histogram would make it easier to determine if the null hypothesis is within the confidence level.

By computing the p-value, the Null hypothesis would be TRUE if the p-value is greater than the significant level. This means that each age group has no difference in age in how they prepare for COVID. Then, we could determine whether the null hypothesis is TRUE or FALSE using the p-value. The test could lead to TWO possible results:

**1) Age has a significant effect to how people prevent COVID**

**2) Age has no significant effect to how people prevent COVID**

Applying the skills I learned from STA130 allows me to solve the real-life problem of the relationship between age and COVID prevention behavior. This topic helps us understand variables better and comprehend the concept of "Hypothesis." In addition, it gives me the chance to study meaningful topics in the real world.

## **Topics 2: Provincial And Wellness Life Satisfaction**

The variable exploring:

* GEO\_province (categories | Ontario, Alberta, Quebec )
* WELLNESS\_life\_satisfaction (numbers | from 1~10 )

WHY CHOOSING THESE VARIABLES

* Because I want to discuss Influence of different generations' ideologies on behavior

We are estimating the average wellness life satisfaction score within each province to assess regional differences in life satisfaction across the population. and the method we use is**Bootstrapping** to estimate the confidence intervals for wellness scores in each province, assessing variability without assuming normality. First, we isolate the dataset based on different provinces. Within the provinces, we repeatedly randomly select several WELLNESS\_life\_satisfaction (reparation based on the length of the dataset), forming a new dataset.

We can create a **box plot for each province**. We chose the box plot because it shows the Q1, Q3, median, and outlier for each wellness, giving a very convincing interval of range.

Once we have applied the Box-Plot method to all provinces, we are left with a series of box plots that represent the wellness ranges for each province. By arranging and aligning them in a vertical order, we create a visually engaging comparison. This method not only provides valuable insights but also makes the comparison process more interesting and engaging for you.

**1) Distinct Differences in Wellness Life Satisfaction by Province**

**2) No Distinct Differences in Wellness Life Satisfaction by Province**

Applying the skills I learned from STA130 helps me understand variability across groups and apply "Bootstrapping" to estimate confidence intervals. Exploring the real-life relationship between province and wellness life satisfaction

## 

## **Topics 3: Relationship Status and Social Connection**

The variable exploring:

* DEMO\_relationship\_status
* (categories | single and not dating, In a relationship)
* CONNECTION\_activities\_talked\_family\_p3m
  + - * + \_ job
        + \_hobbies
  + (categories | not in the past three months, less than monthly, monthly, a few times a month, a few times a week, weekly, daily or all most daily)

WHY CHOOSING THESE VARIABLES

* Because I want to discuss does marriage close people off or expose them to more

We are estimating the mean social connection score for each relationship status group to understand if relationship status impacts connection levels in the population. and the methodwe use is **Two-Sample Group Comparisons** because they best compare the connection level and relationship statuses. We need data to compare, so we transform the categories in CONNECTION\_activities into numbers from 0 to 6: not in the past three months = 0, less than monthly = 1, and so on.

We split the score into groups, with 10 points as a gap. Then, for each relationship status, we add the scores for family, job, and hobbies. Then, we put the score in the corresponding score range. After iterating, we can create a **Bar Chart,** and by combining two Bar charts, we can get a stacked Bar Chart. We use a stacked bar chart because it can clearly show the score of two categories**.** The x-axis would be the score at the connection level, and the y-axis would be the count for that score range. The test could lead to TWO possible results:

1. **Relationship Status Differ Connection Levels**
2. **Relationship Status Does Not Differ in Connection Levels**

Applying the skills I learned from STA130 provides an opportunity to study meaningful aspects of human connection in the real world. It allows me to investigate the real-life relationship between relationship status and social connection.