Stata Demo on Prevalence: Part I

1. **Objectives**
   a. Calculate the prevalence of smoking in the Framingham Data Set and interpret the results
   b. Restrict an analysis to non-missing data
   c. Create a 2x2 table to examine changes in self-reported smoking status between visit 1 and visit 2

2. **Calculate the proportion of people at each visit that report current smoking.**
   In this data set, current smoking status us coded as 0=not current smoker, 1=current smoker
   a. Dropdown:
       i. Statistics→ Summaries, tables and tests→Tables→ Multiple One Way Tables
       ii. cursmoke1 cursmoke2 cursmoke3 or cursmoke*
       iii. Check box for "treat missing values like other values"
       iv. Submit
   b. Command Window Syntax: `tab1 cursmoke1 cursmoke2 cursmoke3, missing`

3. **Calculate the proportion of people at each visit that report current smoking among those with data on smoking status at that visit.**
   a. Dropdown:
       i. Statistics→ Summaries, tables and tests→Tables→ Multiple One Way Tables
       ii. cursmoke1 cursmoke2 cursmoke3 or cursmoke*
       iii. Submit
   b. Command Window Syntax: `tab1 cursmoke1 cursmoke2 cursmoke3`

4. **Calculate the proportion of people at each visit that report current smoking among those with data on smoking status at all 3 visits.**
   A missing datum counts as infinity when making comparisons. Therefore, if the value is not missing, it is considered less than infinity.

   We can create a variable for no missing smoking status at any examination cycle:
   `gen cursmokenotmiss = cursmoke1<. & cursmoke2<. & cursmoke3<.`

   If all 3 cursmoke variables are less than infinity (i.e. not missing), then cursmokenotmiss equals true which is recorded as 1, but otherwise cursmokenotmiss equals false which is recorded as 0.

   a. Dropdown:
       i. Statistics→ Summaries, tables and tests→Tables→ Multiple One Way Tables
       ii. Categorical variables: `cursmoke1 cursmoke2 cursmoke3` or `cursmoke*`
       iii. Check box for "treat missing values like other values"
       iv. Under tab for by/if/in Restrict observations: `cursmokenotmiss==1`
       v. Submit
   b. Command Window Syntax: `tab1 cursmoke1 cursmoke2 cursmoke3 if cursmokenotmiss==1, missing`

5. **What could explain the declining prevalence of smoking?**
   a. Over time, the prevalence of smoking is declining in the population
   b. Current smokers have a shorter life
   c. Several smokers choose not to participate in the 2nd and 3rd visits

6. **Calculate the change in smoking prevalence between the 1st and 2nd visit.**
   a. Dropdown:
      i. Tables→ Two Way Tables with Measures of Association
      ii. Row variable: cursmoke1
      iii. Column variable: cursmoke2
      iv. Check boxes for
         1. "treat missing values like other values"
         2. Cell contents→"Within-column relative frequencies"
         3. Cell contents→"Within-row relative frequencies"
         4. Cell contents→"Relative frequencies"
      v. Submit
   b. Command Window Syntax: `tabulate cursmoke1 cursmoke2, cell column miss row`

7. **Calculate the change in smoking prevalence between the 1st and 2nd visit among those with data on smoking status at both visits.**
   a. Dropdown:
      i. Tables→ Two Way Tables with Measures of Association
      ii. Row variable: cursmoke1
      iii. Column variable: cursmoke2
      iv. Check boxes for
         1. Do not check box for "treat missing values like other values"
         2. Cell contents→"Within-column relative frequencies"
         3. Cell contents→"Within-row relative frequencies"
         4. Cell contents→"Relative frequencies"
      v. Submit
   b. Command Window Syntax: `tabulate cursmoke1 cursmoke2, cell column row`

8. **Conclusions**
   a. Smoking prevalence declined over time
      i. Smokers are quitting
      ii. Smokers have a shorter life
      iii. Smokers are less likely to participate
   b. Stata can be used to
      i. Restrict an analysis to non-missing data
      ii. Create a 2 x 2 table to cross-classify two nominal variables