Stata Demo on Prevalence: Part 2

1. **Objectives**
    a. Create an ordinal variable from continuous data
    b. Calculate the prevalence of CHD for different levels of smoking at visit 1

2. **Calculate the prevalence of coronary heart disease (CHD) at visit 1 by categories of cigarettes per day**

    PREVCHD is defined as pre-existing angina pectoris, myocardial infarction (hospitalized, silent or unrecognized), or coronary insufficiency (unstable angina)
    0=Free of disease, 1=Prevalent disease

    Create 4 categories of cigarette packs per day (0,1-20,21-40,≥41). Since the values reflect a particular ordering, it is an ordinal variable.

```
gen packs1=.
replace packs1=0 if (cigpday1==0)
replace packs1=1 if (cigpday1>=1 & cigpday1 <= 20)
replace packs1=2 if (cigpday1>=21 & cigpday1 <= 40)
replace packs1=3 if (cigpday1>=41 & cigpday1<.)
```

    a. Dropdown:
        i. Tables→ Two Way Tables with Measures of Association
        ii. Row variable: packs1
        iii. Column variable: prevchd1
        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”
    b. Command Window Syntax: `tabulate packs1 prevchd1, cell column row`

3. **What could explain the higher prevalence of CHD among non-smokers compared to those who smoke 1 or more cigarettes per day?**
    a. High incidence, Long duration
    b. Cross-sectional data is susceptible to reverse causation
    c. Other common suspects
        i. Bias
        ii. Confounding
        iii. Chance

4. **Conclusions**
    a. Stata can be used to create an ordinal variable based on continuous data.
    b. CHD prevalence was lower among people with higher levels of smoking.
    c. Prevalence is a function of incidence and duration.
    d. In addition to a causal effect of exposure on disease risk, there are several alternative explanations for observing an association between two factors of interest.