Stata Demo on Effect Measure Modification and Standardization

1. Objectives

a. Review concepts of
   i. Confounding vs. effect measure modification
   ii. Pooling vs. standardization
b. Conduct a test of homogeneity to determine whether age modifies the association between smoking at visit 1 and the risk of cardiovascular disease.
c. Calculate and interpret a summary estimate using standardization
   i. Total population
   ii. Unexposed individuals
   iii. Exposed individuals

2. Confounding vs. Effect Measure Modification

a. *Confounding*
   i. Incorrect estimates due to the impact of a third factor that is associated with the exposure and a risk factor for the outcome independent of exposure; arises due to non-exchangeability (noncomparability) between the exposed and unexposed
   ii. Results in an invalid estimate; we would like to remove confounding
   iii. Not scale-dependent- if the ratio measure (e.g. rate ratio, risk ratio) is confounded, so is a difference measure (e.g. rate difference, risk difference)
   iv. We can present stratum-specific estimates, standardized estimates or pooled estimate
b. *Effect measure modification*
   i. A third factor that modifies the strength of the association between the exposure-outcome association
   ii. Provides useful information that we would like to highlight and describe in our findings
   iii. Scale dependent: if the strength of the association between exposure and outcome varies for different subgroups (effect measure modification), it can be seen on one scale or on both scales, e.g. the rate ratios for two subgroups may be different but the rate differences for the two subgroups may be similar.
   iv. We can present stratum-specific estimates or standardized estimates. A pooled summary estimate (e.g. Mantel-Haenszel adjusted estimate) is not appropriate.

3. Pooling vs. Standardization

a. *Pooling:* a method for adjusting for confounding when differences between strata are due to sampling variability
b. *Standardization:* a method to compare two populations with different distributions of a stratification factor(s) that confounds and/or modifies an exposure-disease association

4. Test for Effect Measure Modification (EMM)

a. Test of homogeneity
   i. H_0: stratum-specific estimates are homogenous (no EMM)
   ii. H_A: At least one stratum estimate is different from the others (EMM)
   iii. Degrees of freedom= # strata – 1
b. Large p-value: Do not reject null
   i. Insufficient evidence of effect measure modification
   ii. Report stratum-specific estimates or calculate Mantel-Haenszel summary measure
c. Small p-value: Reject null
   i. Effect modification is present
   ii. Report stratum-specific estimates or standardized measure

5. Conduct a test of homogeneity to determine whether age (age4cat) modifies the association between smoking at visit 1 (cursmoke1) and the risk of death.

First, create the 4 categories for age that we have used in previous sessions with the following code:

```
gen age4cat=.
replace age4cat=0 if (age1<=40)
replace age4cat=1 if (age1>40 & age1 <= 50)
replace age4cat=2 if (age1>50 & age1 <= 60)
replace age4cat=3 if (age1>60 & age1<.)
```

To see the 2x2 tables of smoking by death for each age category:

a. Dropdown:
   i. Statistics→ Summaries, tables, and tests→ Two way tables with measures of association
   ii. Main tab
      1. Row variable: death
      2. Column variable: cursmoke1
   iii. By/if/in tab
      1. Stratify on variables: age4cat
   iv. Submit
b. Command Window Syntax: `by age4cat, sort : tabulate death cursmoke1`

Age ≤ 40

		Death		
		Yes	No	Total
Smoker	Yes	67	385	452
	No	25	277	302
	Total	92	662	754

50 < Age < 60

		Death		
		Yes	No	Total
Smoker	Yes	286	281	567
	No	312	500	812
	Total	598	781	1379

40 < Age < 50

		Death		
		Yes	No	Total
Smoker	Yes	266	689	955
	No	110	574	684
	Total	376	1263	1639

Age > 60

		Death		
		Yes	No	Total
Smoker	Yes	169	38	207
	No	315	140	455
	Total	484	178	662

Now, we can look at the stratum specific estimates, crude overall estimate, Mantel-Haenszel estimate and test of homogeneity.

c. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variable: age4cat
      2. Within-stratum weights: Use Mantel-Haenszel weights (default)
   iv. Submit

d. Command Window Syntax: `cs death cursmoke1, by(age4cat)`

```
          age4cat |       RR       [95% Conf. Interval]   M-H Weight
------------------+--------------------------------------------------
                0 |    1.790619     1.158273   2.768188     14.98674
                1 |    1.731975     1.418997   2.113985     64.09396
                2 |    1.312757     1.165099   1.479129     128.2843
                3 |    1.179281     1.078835   1.289079     98.49698
------------------+--------------------------------------------------
            Crude |     1.06826     .9858212   1.157592
    M-H combined |    1.381036     1.279253   1.490917
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(3) =   19.107  Pr>chi2 = 0.0003
```

6. Calculate the association between smoking at visit 1 (cursmoke1) and the risk of death after adjusting for age (age4cat) by standardizing to the total population under study.

In order to standardize to the total population in Stata, we need to tell Stata that each category of age should be weighted equally, so we create a new variable equal to the proportion of the population in that age category:

```
table age4cat
```

```
----------------------
  age4cat |      Freq.
----------+-----------
        0 |         754
        1 |       1,639
        2 |       1,379
        3 |         662
----------------------
```

754/4434=	0.1700496
1639/4434=	0.3696437
1379/4434=	0.3110059
662/4434=	0.1493009

```
gen all=.

replace all= 754/4434 if (age4cat==0)
replace all= 1639/4434 if (age4cat==1)
replace all= 1379/4434 if (age4cat==2)
replace all= 662/4434 if (age4cat==3)
```

THIS SECTION HAS BEEN REVISED:

PLEASE USE THIS CODE AND NOT THE CODE PRESENTED IN THE VIDEO!

a. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variables: age4cat
      2. User-specified variable: all
   iv. Submit
b. Command Window Syntax: `cs death cursmoke1, by(age4cat) standard(all)`

```
         age4cat |       RR       [95% Conf. Interval]       Weight
-----------------+------------------------------------------------
               0 |    1.790619     1.158273   2.768188     .1700496
               1 |    1.731975     1.418997   2.113985     .3696437
               2 |    1.312757     1.165099   1.479129     .3110059
               3 |    1.179281     1.078835   1.289079     .1493009
-----------------+------------------------------------------------
           Crude |     1.06826     .9858212   1.157592
    Standardized |    1.372987     1.275679   1.477717
```

In a population with the age distribution of the **total population**, the risk of death is 1.37 times greater among smokers than among the nonsmokers.

7. Calculate the association between smoking at visit 1 (cursmoke1) the risk of death after adjusting for age (age4cat) by standardizing to the unexposed population (not current smokers at visit 1).

a. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variables: age4cat
      2. Within-stratum weights: Use external
         estandard: external weights are the total number of unexposed
   iv. Submit
b. Command Window Syntax: `cs death cursmoke1, by(age4cat) estandard`

```
 age4cat |       RR       [95% Conf. Interval]       Weight
-----------------+------------------------------------------------
               0 |    1.790619     1.158273   2.768188          302
               1 |    1.731975     1.418997   2.113985          684
               2 |    1.312757     1.165099   1.479129          812
               3 |    1.179281     1.078835   1.289079          455
-----------------+------------------------------------------------
           Crude |     1.06826     .9858212   1.157592
 E. Standardized |    1.333775      1.24473    1.42919
```

In a population with the age distribution of the **non-smokers**, the risk of death is 1.33 times greater among smokers than among nonsmokers.

8. Calculate the association between smoking at visit 1 (cursmoke1) the risk of death after adjusting for age (age4cat) by standardizing to the exposed population (current smokers at visit 1).

a. Dropdown:
   i. Statistics→ Epidemiology and Related→ Tables for Epidemiologists→ Cohort study risk-ratio etc.
   ii. Main tab
      1. Case variable: death
      2. Exposed variable: cursmoke1
   iii. Options tab
      1. Stratify on variables: age4cat
      2. Within-stratum weights: Use internal standard: internal weights are the total number of exposed
   iv. Submit

b. Command Window Syntax: `cs death cursmoke1, by(age4cat) istandard`

age4cat	RR	[95% Conf.	Interval]	Weight
0	1.790619	1.158273	2.768188	452
1	1.731975	1.418997	2.113985	955
2	1.312757	1.165099	1.479129	567
3	1.179281	1.078835	1.289079	207
Crude	1.06826	.9858212	1.157592	
I. Standardized	1.4271	1.3129	1.551233	

In a population with the age distribution of the **smokers**, the risk of death among smokers is 1.43 times greater than among the nonsmokers.

9. Conclusions

a. Confounding and effect measure modification both involve a third factor, but they are separate concepts- a factor may or may not be a confounder and it may or may not modify the association between the exposure and outcome.
b. Both pooled estimates and standardized estimates adjust for confounding to the degree of stratification by that factor, but pooling is not appropriate in the presence of effect measure modification.
c. In the Framingham Heart study, self-reported smoking at visit 1 is associated with a higher risk of death, especially among younger participants.
d. The standardized estimates for smoking status at visit 1 and the risk of death are adjusted for differences in the age distribution between smokers and non-smokers. These standardized estimates and reflect the association between smoking and death in a population with the age distribution of the group to which we standardize.