

HSCI 416 ○ Lab 2

# **Data manipulation I: IF, THEN, ELSE + creating new variables**

# Today's objectives

Manipulating data: slides & code

1. Subsetting data
2. Cleaning data
3. Creating new variables

# If... then...else statements

- Simple:

*if some condition(s) hold true then a consequence*

- More complex:

*if some condition(s) hold true then a consequence;*

*else some other consequence;*

*condition*

*consequence*

```
/* Using employment status to define weekly hours worked */  
/* Recall: See slides for info on the dataset */  
data samhsa_wh; set samhsa_edu;  
IF EMPLOY = 1 THEN W_H = 40;  
    ELSE IF EMPLOY = 2 THEN W_H = 25;  
    ELSE W_H = 0;  
run;
```

*some other consequence*

# If... then...else statements (cont'd)

- To execute more than one statement within the *then* or *else* clause, use a *do-end* block:

The diagram illustrates the structure of an SQL IF-THEN-ELSE statement. It features a code block with several annotations: 'condition' points to the 'IF EMPLOY = 1' line; 'consequence; do-end block' points to the 'THEN' clause and its associated 'DO' block; 'another condition' points to the 'ELSE IF EMPLOY = 2' line; and 'some other consequence' points to the 'THEN' clause of the second IF block. The code itself is as follows:

```
/* Using employment status and education level to define hourly wage */
/* Calculating weekly income as the product of weekly hours X hourly wage */
data samhsa_wage; set samhsa_wh;
IF EMPLOY = 1 THEN DO;
    if EDU = 'High' THEN H_SA = 40; /* create hourly wage variable according to employment and edu
    else if EDU = 'Mid' THEN H_SA = 30;
    else if EDU = 'Low' THEN H_SA = 20;
    WI = H_SA * W_H; /* weekly income (WI) = hourly salary (H_SA) * working hours (W_H) */
    END;
ELSE IF EMPLOY = 2 THEN DO;
    if EDU = 'High' THEN H_SA = 35;
    else if EDU = 'Mid' THEN H_SA = 25;
    else if EDU = 'Low' THEN H_SA = 15;
    WI = H_SA * W_H;
    END;
ELSE DO; H_SA = 0; WI = 0; END;
run;
```

# Supporting info for today's code

<b>Education level</b>	<i>EDU_L</i>	
<i>EDU_L</i>	<i>EDUC</i>	Category #
High	13-15	4
	16 OR MORE	5
Middle	12	3
Low	9-11	1
	8 YEARS OR LESS	2
<b>Working hours/week</b>	<i>W_H</i>	
<i>W_H</i>	<i>EMPLOY</i>	Category #
40	Full time	1
25	Part time	2
0	UNEMPLOYED	3
	NOT IN LABOR FORCE	4
<b>Hourly salary</b>	<i>H_SA</i>	
EMPLOY status	EDUC level	<i>H_SA</i>
Full time (1)	High (4-5)	\$40
	Middle (3)	\$30
	Low (1-2)	\$20
Part time (2)	High (4-5)	\$35
	Middle (3)	\$25
	Low (1-2)	\$15
Unemployed or not in labor force (3-4)	High (4-5)	\$0
	Middle (3)	\$0
	Low (1-2)	\$0
<b>Income level</b>	<i>INC_L</i>	
<i>INC_L</i>	Weekly income ( <i>WI</i> )	
H	>1000	
M	[500, 1000]	
L	<500	

# Supporting info for today's code

- See provided “SAS\_” files for more this and more!

## Operators in Expressions

Operator	Meaning
= or eq	equal to
^= or ne	not equal to
> or gt	greater than
< or lt	less than
>= or ge	greater than or equal to
<= or le	less than or equal to
AND or &	and, both. If both expressions are true, then the compound expression is true.
OR or	or, either. If neither expression is true, then the compound expression is true.
0	false
1	true
.	false