

SAS Training 2019

Exercise 2

- (a) Using a data-step, create a SAS dataset called wtsbp reading in the following variables from the tomhss SAS dataset: ptid, group, age, sex, marital, wtbl, wt12, sbpbl, and sbp12. (Use a SET and KEEP statement).
- (b) Compute two new variables in the data step; change in weight and change in systolic blood pressure after 12-months. Name the variables wtdif and sbpdif. The formula for wtdif is:

`wtdif = wt12 - wtbl`; Define sbpdif in a similar manner.

Note: by computing the differences in this way negative values will indicate a *decrease* from baseline and positive values will indicate an *increase* from baseline.

Also compute a new variable to indicate whether the participant was taking active treatment or placebo.

```
if group = 6 then active = 2; else active = 1;
```

- (c) Run **proc means** on all numeric variables to help verify the new variables where defined correctly. *What was the average weight change and SBP change for all participants?*
- (d) Run **proc means** on the variable sbpdif separately by active and placebo groups. *What was the average change in SBP in the active groups; in the placebo group?* Test if the change in SBP differed significantly between the active and placebo groups using **proc ttest**.
- (e) Run **proc univariate for the variable wtdif**. *What are the 25th, 50th, and 75th percentiles of weight change.*
- (f) Use a procedure to display the cross-tabulation of participants marital status and gender. *What is the percentage of men that are currently married? What are the percentage of women that are currently married?* Note: The values of marital status are 1=never married, 2=separated, 3=divorced, 4=widowed, 5=married. Test if the distribution of marital status differs significantly between men and women. Use the **chisq** option on the table statement.
- (g) Using **proc sgplot** display side-by-side boxplots of weight change by gender.

2. In a new SAS data-step read the following variables from the tomhss SAS dataset: ptid, age, sex, and income creating a SAS dataset called ageincome.
 - (a) Using if/then/else statements compute a new variable with 5 categories for age (45-49, 50-54, 55-59, 60-64, 65-69).
 - (b) Using the IN function compute a new variable with two categories for reported income: making \$40,000 or more (values 5-8) versus making less than \$40,000 (values 1-4).
 - (c) Using **proc freq** display the distribution of age and income categories. *What percentage of participants are 65-69? What percentage make \$40,000 or more?*
 - (d) Using **proc freq** display the percentage of participants making \$40,000 or more by gender.