September 14, 2025

10.6First generate the variable BMI, where BMI equals weight in kg divided by height in meters squared (BMI = weight/(height^2)). Construct a categorical variable for BMI that considers the commonly used categories: i) underweight, BMI below 18.5; ii) normal weight, BMI larger or equal to 18.5 and lower than 25; iii) overweight, BMI larger or equal to 25 and lower than 30; iv) obese, BMI of 30 or higher. Compute and report the prevalence of overweight and obesity by ethnic group (black vs non-black). What differences do you observe?

We begin by generating the bmi variable using the gen command. Note that, since Body Mass Index is calculated as weight in kilograms divided by height in meters squared[1], we must convert the height variable from centimeters to meters before the calculation:

stata gen height$_m = height/100 genbmi = weight/(height_m^2)$

[1]World Health Organization. (2000). *Obesity: preventing and managing the global epidemic.* WHO Technical Report Series, 894. Geneva: World Health Organization.

We can then construct bmi$_c at, our categorical BMI variable$:

stata gen bmi$_c at$ = $.replace bmi_c at$ = $1 if bmi < 18.5 ! missing(bmi) replace bmi_c at$ = $2 if bmi >= 18.5 bmi < 25 ! missing(bmi) replace bmi_c at = 3 if bmi >= 25 bmi < 30 ! missing(bmi) replace$
$4 if bmi >= 30 ! missing(bmi)$

By adding !missing(bmi) to our if statement, we ensure that only individuals with a reported BMI get a categorical BMI value. This is due to the fact that Stata codes missing values as larger than any nonmissing values[2].

aCompute and report the prevalence of overweight and obesity by ethnic group (black vs non-black). What differences do you observe?

In order to compute the prevalence of overweight and obesity by ethnic group, two dummy variables can be created. The first, overweight, includes all overweight and obese individuals. The second, obese, includes only obese individuals.

stata gen overweight = (bmi$_c at >= 3) if ! missing(bmi_c at) genobese = (bmi_c at ==$
$4) if ! missing(bmi_c at)$

In order to facilitate comparison, tables can be generated with the following commands:

stata tab black overweight, row missing tab black obese, row missing

The option row allows us to see row percentages, while the option missing includes individuals with missing data.

[2]Gould, W. (n.d.). *Stata — FAQ: Logical expressions and missing values.* Stata Corp. Retrieved September 14, 2025, from https://www.stata.com/support/faqs/data-management/logical-expressions-and-missing-values/