There are many equivalent models when we use dummies
STA9750

Suppose we have the following variables:

1. Score: teaching evaluation score
2. Beauty: numerical/quantitative score, assessment of "beauty" by students
3. Male: 1 if subject is male, 0 otherwise
4. Female: 1 if subject if female, 0 otherwise

You can code up models in different ways, which are all equivalent.

Suppose you run the regression and you get an estimated regression mean:

Score = 1 + 2 beauty + 3 male

Then, this is equivalent to running a model with female, as a variable, and you'd get

Score = 4 + 2 beauty - 3 female

You could even do away with the intercept altogether, and fit a model with female, male, and beauty:

Score = 4 male + 1 female + 2 beauty

These models assume that the difference between men and women can be modeled with a "shift" that doesn't depend upon the value of beauty (parallel lines). They all say the same thing:

Score = 4 + 2 beauty, for males
Score = 1 + 2 beauty, for females

If you try to add an intercept ***and*** both female and male, SAS is going to complain. Why? Because the intercept is equal to "female + male." Intuitively, if you put in a variable that is a perfect function of the ones that are already in the model, you can't get anything new out of it, so you shouldn't add it. Techincally, finding the regression coefficients involves inverting a matrix, and it becomes rank-deficient if you don't have linearly independent variables.

You can do this with more complicated models, which may have interactions.

For example, if you have a model which is

Score = 1 + 3 male + 2 beauty + 1 male*beauty

It is equivalent to:

Score = 1 + 3 male + 3 male*beauty + 2 female*beauty

These models are equivalent to other models that have different intercepts and slopes for males and females [that's what's important: if the intercepts and/or slopes are the same]. They all say the same thing:

Score = 4 + 3 beauty, for male
Score = 1 + 2 beauty, for female