

A GUIDE TO FACTOR VARIABLES AND MARGINAL EFFECTS IN STATA (FOR ECONOMISTS)

JONATHAN L. GRAVES¹

1. WHAT ARE FACTOR VARIABLES?

1.1. Preamble: Qualitative Variables, Dummies, and Interactions. A common issue faced when doing applied work in econometrics is dealing with *qualitative variables* - that is, variables which only reflect the properties of an object, and don't have any natural ranking. In this example here, imagine we are working with a dataset that relates occupations and wages to happiness, which is measured using a quantitative scale (Y_i with higher meaning happier) and a discrete variable ($D_i \in \{0, 1\}$ which is an indicator for whether the person is happy or not). Then, consider the variable `Occupation` (O_i) which describes the occupation of worker i . In a dataset terms, this variable might look like:

$$O_i = \begin{cases} \text{Baker} & i = 1 \\ \text{Waiter} & i = 2 \\ \text{Hairdresser} & i = 3 \\ \text{Baker} & i = 4 \end{cases}$$

These kinds of variables are problematic, because many common statistics or techniques (like averages) don't make sense when applied to them; for example, what does $\bar{O} = \frac{1}{n} \sum_{i=1}^n O_i$ mean here?

The best solution is to transform a qualitative variable into a *set of dummies* (or indicator) variables corresponding to the possible values of the qualitative variable, and taking on the values 1 (yes) and 0 (no). For example, our variable would become:

$$O_i \implies O_i^B, O_i^W, O_i^H$$

Date: 30-October-2020 (v1.1).

¹Vancouver School of Economics, University of British Columbia.

$$O_i^B = \begin{cases} 1 & O_i = \text{Baker} \\ 0 & \text{Otherwise} \end{cases} \quad O_i^W = \begin{cases} 1 & O_i = \text{Waiter} \\ 0 & \text{Otherwise} \end{cases} \quad O_i^H = \begin{cases} 1 & O_i = \text{Hairdresser} \\ 0 & \text{Otherwise} \end{cases}$$

However, we don't need all of these dummies to capture the same information as O_i . In fact, in many models (like regression) including all three would introduce redundant variables into our model (and result in, for example, multicollinearity). You can see this because if $O_i^B = 0$ and $O_i^W = 0$ then we *must* have $O_i^H = 1$, by definition. In general, a qualitative variable with k unique levels can be represented by a set of $k - 1$ dummies; the excluded level is arbitrary, and is referred to as the *base* or *omitted* level.

A similar issue arises with *interaction terms*, where you want to study the combination of two variables. These can either be qualitative or quantitative; for example, consider O_i above and W_i , the wage (**Wage**) of an individual. You may be interested in a variable like:

$$W_i^H = O_i^H \cdot W_i$$

which is can be interpreted as the wage for hairdressers, or a variable like:

$$W_i^2 = W_i \cdot W_i$$

which is the squared wage. These are very common objects of interest in many models, but they are formed from combinations of other variables. Performing these operations (creating dummies, creating interactions) are common, and need to be carried out in nearly all applied analyses.

1.2. Factor Variables. The natural way to create these terms is to just literally create new variables, which carry out the classification or combinations above. However, there are (at least) two problems with this:

- (1) In most statistical programs, you will lose the association with the original variable(s), which can be confusing or problematic if you lose track of this relationship in subsequent analyses.¹

¹This is particularly problematic when you want to compute marginal effects, but your statistical tool does not understand the relationship between the different variables. For example, if you have a linear regression model like $Y_i = \beta_0 + \beta_1 W_i + \beta_2 X_i + \epsilon_i$ where $X_i = W_i^2$, the correct marginal effect of W_i is $\beta_1 + 2\beta_2 W_i$. However, if your statistical program does not understand the relationship between W_i and X_i , it will only

- (2) It is relatively time consuming, especially if you're doing it by hand, and requires an extra step which is not appealing from an analytical point of view (what if you mess it up?).

In STATA, this is handled using *factor variables* and *factor variable notation*. When using this notation, STATA automatically creates dummies (and interaction terms) for qualitative variables or continuous variables. This is relatively easy to use, and relies on a series of prefixes and operators:

Prefix/Operator	Use	Example
i.	Qualitative	i.Occupation
c.	Quantitative	c.Wage
#	Interaction	c.Wage#i.Occupation
##	Factorial	c.Wage##i.Occupation

The prefixes tell STATA what type of variable it is dealing with; the `i.` prefix tells STATA that the associated variable is qualitative. When used in a command, it automatically performs the work of creating a sequence of dummies. Similarly, the `c.` prefix tells STATA that associated variable is quantitative in nature.

Important Note: all variables in STATA must be numerical in order to use factor variable notation. This means that qualitative variables must be *encoded* using numbers, rather than text (strings). If your data is not already encoded properly, you can use the `encode` command in STATA to easily generate appropriate variables for use with factor variables. For example, if `Occupation` is a string (as above), in order to use factor variables, you need to make a new variable which encodes it numerically:

```
mean i.Occupation //will crash
```

```
encode Occupation, generate(Occupation_n)
```

```
mean i.Occupation_n //will run
```

In what follows, let's assume `Occupation` is encoded as 1 = Baker, 2 = Waiter, 3 = Hairdresser.

report β_1 . A major mistake! Even if you manually correct for this, you'll need to combine regression coefficients and compute the correct standard errors, which is a lot of extra work.

When you want to create interactions of variables, STATA uses the interaction and factorial operators to create interaction terms for the factor variables on either side. The interaction operator `#` creates a new variable which is the product of the two variables - including decomposing qualitative variables into dummies, then creating interaction terms. The factorial operator is really just shorthand for three operations: it represents including both the direct and interaction terms. For example, `c.Wage##i.Occupation` is equivalent to `c.Wage i.Occupation c.Wage#i.Occupation` in STATA's factor variable notation. We can use this notation to easily handle many common situations.

Example 1. (Including a Qualitative Variable as a Set of Dummies)

Suppose we wanted to include O_i in a linear regression of Y_i (happiness, stored in data as `Y`) as a set of dummies $O_i \implies O_i^B, O_i^W, O_i^H$. We can do this using the command:

```
regress Y i.Occupation
```

which represents the model:

$$Y_i = \beta_0 + \beta_1 O_i^W + \beta_2 O_i^H + \epsilon_i$$

Example 2. (Including a Quadratic Term)

Suppose we wanted to include W_i in a quadratic linear regression of Y_i (stored in data as `Y`). We can do this using either the command:

```
regress Y c.Wage c.Wage#c.Wage
```

or the command:

```
regress Y c.Wage##c.Wage
```

both of which represent the model:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 W_i^2 + \epsilon_i$$

Example 3. (Including an Interaction Term)

Suppose we wanted to include W_i interacted with O_i in a linear regression of Y_i (stored in data as `Y`). We can do this using either the command:

```
regress Y c.Wage i.Occupation i.Occupation#c.Wage
```

or the command:

```
regress Y i.Occupation##c.Wage
```

both of which represent the model:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \beta_4 O_i^W \cdot W_i + \beta_5 O_i^H \cdot W_i + \epsilon_i$$

You can see the advantage of using the factorial notation when you have a larger number of variables in a model: it improves readability.

1.3. Base Levels and Referencing Levels. In the preceding section, we introduced factorial variable notation, and looked at some examples. However, you may have noticed when we created our dummy variables, we always represented O_i as O_i^W, O_i^H . However, as we outlined in Section 1.1, this is only one of three possible representations of this variable using dummies: we could have equivalently chosen O_i^B, O_i^H or O_i^W, O_i^H by selecting a different base level. While from an econometric point of view there is no difference between these representations, sometimes we may want to change the base level. For example, we may have a natural “comparison” or “reference level” such as “no schooling” or “unemployed” which will make interpretation easy.

In STATA, this is easy to do using a variation on the `i.` command. By default, `i.` chooses the smallest ordered level of the encoding as the base level, which is usually zero or one, in most set-ups. However, you can adjust it by using the following commands instead:

Prefix	Interpretation
<code>ibx.</code>	Use level x as the base
<code>ib(#x).</code>	Use the x -th ordered level as base (e.g. $x = 2$ means second)
<code>ib(first).</code>	Use the smallest level as base (default)
<code>ib(last).</code>	Use the largest level as base
<code>ib(freq).</code>	Use the most frequent level as base
<code>ibn.</code>	Do not use a base; provide all dummies

Example 4. (Including a Qualitative Variable as a Set of Dummies with a Set Base)

Suppose we wanted to include O_i in a linear regression of Y_i (stored in data as `Y`) as a set of dummies $O_i \implies O_i^B, O_i^H$. We can do this using either of the commands:

```
regress Y ib2.Occupation
```

```
regress Y ib(#2).Occupation
```

both of which represent the model:

$$Y_i = \beta_0 + \beta_1 O_i^B + \beta_2 O_i^H + \epsilon_i$$

Example 5. (Including a Qualitative Variable as a Set of Dummies with no Base)

Suppose we wanted to include O_i in a table of summary statistics, but want to make sure we report all our dummies. We can do this using the command:

```
summarize ibn.Occupation
```

After we use a command which uses factor variables, we may want to refer to the dummies or interaction terms in subsequent commands. You can do this using the `ix.` or `io.` prefixes. You can also use these prefixes as an alternative to the `i.` prefixes above to manually create or handle specific dummies (for example, if you don't want an entire set for some reason).

Prefix/Operator	Interpretation
<code>ix.</code> OR <code>x.</code>	Creates (or refers) to the dummy for level x
<code>iox.</code> OR <code>ox.</code>	Creates or refers to all dummies <i>except</i> level x

There are a number of variations on this syntax which STATA will accept, but this is the most basic and useful version of the syntax. This is also

Example 6. (Testing a Coefficient)

Suppose we included O_i in a linear regression of Y_i (stored in data as `Y`) as a set of dummies $O_i \implies O_i^W, O_i^H$, then wanted to test if the value of the coefficient on O_i^H is 0.5. We can do this using the commands:

```
regress Y i.Occupation
```

```
test (_b[i3.Occupation] == 0.5)
```

or

```
test (_b[3.Occupation] == 0.5)
```

Which is equivalent to testing $\beta_2 = 0.5$ in the model $Y_i = \beta_0 + \beta_1 O_i^W + \beta_2 O_i^H + \epsilon_i$

Example 7. (Creating a regression with only selected dummies)

Suppose we wanted to regress on only a specific occupation, like being a baker. We can do this using the command:

```
regress Y i1.Occupation
```

Which is equivalent to the model $Y_i = \beta_0 + \beta_1 O_i^B + \epsilon_i$; notice how we have not included the full set of dummies here.

This ability to reference specific dummies is very useful when we want to study the effects of different variables in a model, using commands like marginal effects - as we discuss in the next section.

2. MARGINS AND MARGINAL EFFECTS

2.1. Preamble: Marginal Effects. One of the most common uses of econometric models is to study the effect one variable has on another. In linear regression models, this is usually pretty simple, due to the structure of the model. Consider the linear regression model:

$$(2.1) \quad Y_i = \beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \epsilon_i$$

In this model, the interpretation of the coefficients is fairly straightforward:

$$\frac{dY_i}{dW_i} = \beta_1 = \frac{dE[Y_i|W_i, O_i^W, O_i^H]}{dW_i}$$

This is just the usual statement about the marginal effects: holding other variables constant, a 1-unit change in W_i leads to a β_1 -unit change in Y_i . For our dummy variables we have a similar statement: holding other variables constant, being a waitress leads to a β_2 -unit change in Y_i (relative to being a baker, our omitted level). We are not going to get into a long-winded review of how we interpret regression coefficients in different models here; we are only interested in what we are trying to do by interpreting coefficients.

This is to study margins: these are both examples of **marginal effects** - one is a continuous margin (W_i) while the other is a discrete margin (O_i) but they both capture the effect an explanatory variable has on the outcome variable. These are the most common types of

marginal effects we study in economics - but there can be other types as well, which we will revisit later.

In linear regressions and some other models, marginal effects can be read directly off using the coefficients in one way or another, sometimes using transformations of the coefficients. However, what happens if you have a model like:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \beta_4 O_i^W \cdot W_i + \beta_5 O_i^H \cdot W_i + \epsilon_i$$

This is still a linear regression (the simplest case) but notice that margins are more complicated:

$$\frac{dY_i}{dW_i} = \beta_1 + \beta_4 O_i^W + \beta_5 O_i^H$$

There are two major challenges with this marginal effect:

- (1) It now depends not just on parameters $(\beta_1, \beta_4, \beta_5)$ but *also* values of the covariates (O_i^W, O_i^H) . This is because our model, while still linear in the coefficients is no longer linear in the variables because of the interactions. Interactions mean different effects as values of related variables change - this means we have a bunch of different marginal effects possible, depending on which occupation we are looking at.
 - To be clear, this is an advantage from a modeling point of view: we may want a model where (in this example) bakers can have a different relationship between Y_i and W_i than that of waiters or hairdressers.
- (2) It also makes it challenging to evaluate this statistically; we have to combine our coefficients in a particular way, and it's not clear that the right way to do that is when there's different possible values for our covariates.

There are two main solutions to this problem:

- (1) Consider the average marginal effect: $E\left[\frac{dY_i}{dW_i}\right]$
- (2) Consider the marginal effect at specific values of the covariates - most frequently at the means.

You will notice that in this particular example, these two yield the same results, but that's not always the case.

For many models, you can analytically compute expressions for these things and estimate them using sample analogues or combinations of coefficients. However, for some models (including many popular ones) this is not possible or is very difficult. Moreover, it's time consuming - this is where using statistical software to compute marginal effects really shines.

2.2. Marginal Effects in STATA: dy/dx . Most of the difficult work in STATA is handled using the `margins` command, which handles all of the different kinds of margins that STATA can compute. This is both a good and a bad thing: it's a fairly complicated command, and not very clearly exactly what you are doing when you use it. Margins is what STATA calls a **postestimation command** which means that you estimate a model, and then you compute the margins after the fact. This is because STATA uses the stored information from the estimation procedure to compute the associated margins. The most basic form of this command looks like:

```
margins, dydx(*)
```

or

```
margins, dydx(*) atmeans
```

The first version computes the average marginal effect for all variables in your model, while the second compute the marginal effect for all variables in your model at the average values of your covariates. You can replace the `*` with a list of variables, if you don't want all of them - but remember to use factor variable notations.

Example 8. (Basic Linear Model)

Suppose we want to compute the average marginal effect of W_i in a model like:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \epsilon_i$$

We can do this by running the following commands:

```
regress Y Wage ib1.Occupation
```

```
margins, dydx(Wage)
```

If we wanted to compute the marginal effect at the means, we would use `margins, dydx(Wages) atmeans` instead. You should verify that (i) both types of marginal effects give the same results and (ii) that they are the same as the analytical margin (the coefficient).

Example 9. (Complicated Linear Model)

This is easy to extend to more complicated models such as:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \beta_4 O_i^W \cdot W_i + \beta_5 O_i^H \cdot W_i + \epsilon_i$$

We can estimate the average marginal effect for wages using:

```
regress Y c.Wages##i1.Occupation
margins, dydx(Wages)
```

If we wanted the average marginal effect for all of the Occupation dummies we could use

```
margins, dydx(Occupation)
```

If we just wanted the margin effect of being a Waiter, we would use

```
margins, dydx(i2.Occupation)
```

Example 10. (Polynomial Terms)

One area where margins really shine are when you are fitting a model with polynomial terms for a variable, such as a quadratic regression (or something similar). Consider the cubic polynomial model:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 W_i^2 + \beta_3 W_i^3 + \epsilon_i$$

The analytical marginal effect of W_i is $\beta_1 + 2\beta_2 W_i + 3\beta_3 W_i^2$, but we can easily compute this using STATA using:

```
regress Y c.Wages c.Wages#c.Wages c.Wages#c.Wages#c.Wage
margins, dydx(Wages)
```

This type of model really comes into its own when we have wholly non-linear models, where analytical representations of the marginal effects are not easy to compute.

Example 11. (Probit Model) Suppose we want to change our model above, so that we're looking at the relationship of wages and happiness, but now only measuring happiness using our discrete (dummy) variable D_i . Our econometric knowledge tells us that appropriate model would be a probit model, which is specified as:

$$D_i = I(\beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \epsilon_i \geq 0)$$

where $\epsilon_i \sim N(0, 1)$ and I is an indicator function for whether or not the condition is true. We can estimate this in STATA as:

```
probit D c.Wages ib1.Occupation
```

And then compute the marginal effects exactly as before:

```
margins, dydx(Wages)
```

or

```
margins, dydx(Wages) atmeans
```

You should notice in this model there are differences between (i) the average marginal effects, (ii) the marginal effects at the means, and (iii) the coefficients in the model. This highlights the importance of thinking about what you are reporting carefully!

2.2.1. *Varying Marginal Effects (Response Surface Mapping)*. In some situations, particularly when working with highly non-linear models (such as a probit model), the marginal effects can vary quite dramatically at different values of the covariates. Alternatively, your research question may be particularly interested in the marginal effects on certain groups characterized by particular values of covariates - such as the impact on low income individuals, or the elderly. This means that the average marginal effect is probably not interesting; it also means the marginal effect for a typical individual (compute at the means) is also probably not interesting.

We can address these kinds of questions by directly controlling where the marginal is computed at: any possible value is allowed, not just at the means of the covariates. The plot of these marginal effects is sometimes referred to as a “response surface” since it varies (multidimensionally) and can be visualized using graphing tools.

This is done using the `at` option, followed by a series of variables and values:

```
at(variable=(value list))
```

where *value list* is a series of values separated by spaces. For example:

```
at(Wages=(10 20 30))
```

would compute the average marginal effects while fixing wages at three levels (10, 20, 30). This will report three marginal effects - one for each level. If we wanted the marginal effects

at the average values of other covariates, while fixing wages at these level we could use the option:

```
at(Wages=(10,20,30)) atmeans
```

This allows us to compute a wide variety of different marginal effects at different values. Here are some useful tools for specifying specific values:

Command	Meaning	Example
<code>(a(c)b)</code>	The sequence a to b in steps of c	<code>(0(10)30)</code>
<code>(mean) x</code>	The mean of x	<code>(mean) Wages</code>
<code>(median) x</code>	The median of x	<code>(median) Wages</code>
<code>(min) x</code>	The smallest value of x in data	<code>(min) Wages</code>
<code>(max) x</code>	The largest value of x in data	<code>(max) Wages</code>
<code>(pa) x</code>	The a -th percentile of x in the data	<code>(p35) Wages</code>
<code>x = gen(exp)</code>	At the value of x specified by function exp	<code>Wage = gen(2*Wages)</code>

Example 12. (Using `at` in Probit Model)

For example, consider our probit model:

$$D_i = I(\beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \epsilon_i \geq 0)$$

where $\epsilon_i \sim N(0, 1)$ and I is an indicator function for whether or not the condition is true. We can estimate this in STATA as:

```
probit D c.Wages ib1.Occupation
```

We can now directly control where we estimate the average marginal effects at:

```
margins, dydx(Wages) at(Wages = 15)
```

if we wanted the marginal effect at the averages of the other variables, we could use any of:

```
margins, dydx(Wages) at(Wages = 15) atmeans
```

```
margins, dydx(Wages) at(Wages = 15 (mean) Occupation)
```

```
margins, dydx(Wages) at(Wages = 15 (mean) i2.Occupation i3.Occupation)
```

As you can, there are many different but equivalent ways to get the same results.

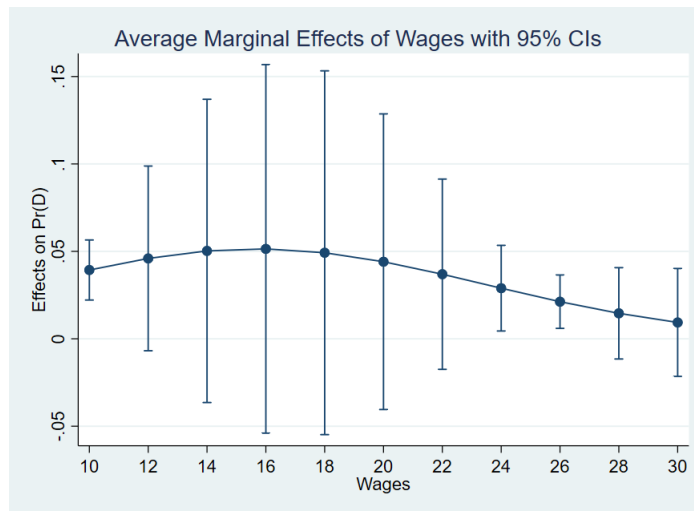
STATA also offers a very useful command to visualize these results. After you plot a series of different margins (or in fact any margin), you can run the `marginsplot` command to generate a visualization of the result. We will not discuss this in detail here, but you can see how this works in the following example.

Example 13. (Using the `Marginsplot` command)

Suppose we use our probit model from before, and want to visualize the average margins for all the wages between \$10 and \$30 dollars, in steps of \$2. We can do this as follows:

```
probit D c.Wages ib1.Occupation
margins, dydx(Wages) at(Wages = (10(2)30))
```

Which looks produce a table like:



2.3. Other Kinds of Margins and Marginal Effects.

2.3.1. *Elasticities.* In economics, when we talk about marginal effects we are almost always referring to the preceding term. However, those are not the only types you may encounter. The second most common are related items, called elasticities or semi-elasticities. In many models (for example, log-log or log-level), elasticities are interpretable naturally as the coefficients - but you can also compute them directly. This can be useful in several situations: (i) when you want to compare models which are reporting elasticities, (ii) when elasticity is the object of interest (for example, in a macroeconomic model of labour supply), or (iii) when the units and sizes of your values make the interpretations of elasticities more natural (for

example, if a 1-unit change is too large or too small to be relevant economically). The three elasticities in STATA are:

- Elasticity of Y with respect to X ($\frac{dy}{dx} \frac{x}{y}$): implemented as `eyex(*)`, with interpretation of “the % change in Y for a 1% change in X ” (where 1 = 100%)
- (Inverse) Semi-elasticity of X with respect to Y ($\frac{dy}{dx} x$): implemented as `dyex(*)`, with interpretation of “the unit change in Y for a 1% change in X ”
- Semi-elasticity of Y with respect to X ($\frac{dy}{dx} \frac{1}{y}$): implemented as `eydx(*)`, with interpretation of “the % change in Y for a 1-unit change in X ”.
- You can compare these to the marginal effects command ($\frac{dy}{dx}$) we saw earlier, which was implemented as `dydx(*)` and had the interpretation of “the unit change in Y for a 1-unit change in X ”

Implementing these into your `margins` command is straightforward: simply replace `dydx(*)` with the appropriate implementation above. Everything we said above applies to these options!

Example 14. (Computing elasticities)

Suppose we have our probit model from earlier:

```
probit D c.Wages ib1.Occupation
```

We can average marginal effects using:

```
margins, dydx(Wages)
```

or the average elasticity using

```
margins, eyex(Wages)
```

If we wanted the elasticity at the means, we could use

```
margins, eyex(Wages) atmeans
```

2.3.2. *Margins.* Another exercise we may want to perform is to calculate hypothetical (counterfactual) predictions of the outcome variables, while changing specific covariates. For instance, we are very familiar in a regression model with computing the fitted values of a model:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 W_i + \hat{\beta}_2 O_i^W + \hat{\beta}_3 O_i^H$$

where $\hat{\beta}_k$ is the estimate value of the coefficient from your model. These are computed at the actual values of the explanatory variables (W_i, O_i^W, O_i^H) but you can calculate predicted values for any hypothetical set of values. This is what **predictive margins** do in STATA; they calculate the average hypothetical predicted value setting some variables for every observation to a specified level.

For example, we could compute the predictive margin of O_i^W by setting $O_i^W = 1$ for all i (and the other associated dummies to zero):

$$\hat{M}(O_i^W = 1) = \hat{\beta}_0 + \hat{\beta}_1 W_i + \hat{\beta}_2 0 + \hat{\beta}_3 1$$

This is why factor variables are essential to margins in STATA - it is necessary to understand the underlying relationship between O_i^W and O_i^H .

For continuous variables, you can also compute predictive margins, but it is necessary to specify at which level they will be held. For factor variables, this is not necessary since it's unambiguous: you can only set one factor equal to one at a time. However, this is not true for continuous variables since they can take on many levels. You specify this using the `at` option, just as we did when varying marginal effects.

The syntax for this in STATA is slightly different from marginal effects:

```
margins variables
```

or

```
margins, at(continuous)
```

where *variables* is a list of factor variables you want to compute predictive margins for and *continuous* are the levels you want predictive margins for continuous variables computed at. The syntax for this is exactly as above, so we will not revisit it here. Notice you can use this not only to compute predictive margins for continuous variables, but also entertain any counterfactual predictive margin with specified explanatory variables you want - essentially, you can combine the two types of commands.

Example 15. (A simple probit model)

Consider our probit model from before:

$$D_i = I(\beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \epsilon_i \geq 0)$$

where $\epsilon_i \sim N(0, 1)$ and I is an indicator function for whether or not the condition is true. We estimate this in STATA as:

```
probit D c.Wages ib1.Occupation
```

Then we can compute our predictive margin:

```
margins Occupation
```

which will report all three possible predictive margins, corresponding to each level of `Occupation` in the data.

If we wanted to compute the margin for `Wages`, we would need to include the `at` option and some levels

```
margins, at(Wages=(10(10)30))
```

This will give use the predictive margin for these three values.

Interactions with predictive margins are a little more complicated, because by their very nature two (or more) variables are changing at the same time. By default, the margins syntax above computes each margin independently even if you specify multiple factor variables which interact. Consider the model:

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 O_i^W + \beta_3 O_i^H + \beta_4 O_i^W \cdot W_i + \beta_5 O_i^H \cdot W_i + \epsilon_i$$

The command `margins Occupation` will only change the specific variables (`Occupation`) and the value of wages. This can be problematic if you want to examine a particular combination of variables (low wages in a particular profession, for example). However, you can do this in two ways, depending on what kind of variables are interacting.

- (1) If one variable is a factor variable and the other is a continuous variable you combine the syntaxes above:


```
margins variables, at(continuous)
```

 where *variables* and *continuous* are as above.
- (2) If both of your variables are factor variables, using an interaction term (as above) in the regular syntax where *variables* goes, using the `#` or `##` operators. These will compute the margins setting both of the interacting variables at the same time.
- (3) If both your variables are continuous variable, use the regular syntax for continuous predictive margins but specify the levels for both of the variables in the `at` option.

Example 16. (Interactions)

Suppose, in this example we have another dummy variable `Dummy` and another continuous variable `Cont` in our dataset, and we have estimated the model:

```
probit D i.Occupation##i.Dummy i.Occupation##c.Wages c.Wages##c.Cont
```

We can specify each different direct and interaction term as follows:

- (1) `margins Occupation Dummy` - yield the direct (non-interacting) predictive margins for both factor variables
- (2) `margins, at(Wages=10)` and `margins, at(Cont=10)` - yields the direct (non-interacting) predictive margins for both continuous variables
- (3) `margins Occupation#Dummy` - yields the interacting predictive margins for the factor variables
- (4) `margins, at(Wages=10 Cont=10)` - yields the interacting predictive margins for the continuous variables at the level 10 for oth.
- (5) `margins Occupation, at(Wages=10)` - yields the interacting predictive margins for `Wages` and `Occupation`

As you can see, there are many version of this command you can develop to examine specific situations and predictive margins.

3. FINAL NOTE

Margins and factor variables are a powerful tool for doing applied statistical analysis and learning how to use them is very rewarding, since it not only simplifies your analysis but opens up many convenient options for counterfactual analysis - beyond just the analysis of the relationship between two variables. If you have any suggestions or comments on this guide (or if you find any errors or ambiguities), please let me know at jonathan.graves@ubc.ca.

Happy coding!

4. REFERENCES AND RESOURCES

This guide is based on my own experiences with STATA and marginal effects, but is also greatly assisted by the following sources, which you will likely also find valuable:

- STATA Corp's documentation for the margins command: <https://www.stata.com/manuals13/rmargins.pdf>

- Richard William's very excellent and helpful articles and notes on the subject: <https://www3.nd.edu/~rwilliam/stats/Margins01.pdf> and <https://www.stata-journal.com/article.html?article=st0260>. You will likely find these very useful if there's something here that was unclear.

Email address: jonathan.graves@ubc.ca