What if we have two or more categorical variables in a regression?

Imagine we want to predict median house value (measured in $) with the following predictors:

* Median household income (continuous, measured in $)
* % of individuals with a bachelor’s degree or more (continuous, measured in %)
* Presence of a park (binary, 1=yes, 0=no)
* Police Unit covering the block group (categorical, with units A, B, C and D).
  + We will create 4 dummies and include dummies A, B and C in the regression, so that D is the reference category
* Parking Type (categorical, with values of No parking for most houses, Street parking for most houses, Garage parking for most houses)
  + We will create 3 dummies, and include dummies ‘No parking’ and ‘Street parking’ in the regression, so that ‘Garage parking’ is the reference category.

Here are the results:

|  |  |  |
| --- | --- | --- |
|  | **Beta Estimate** | **p-Value** |
| **Intercept** | 4388 | 0.0004 |
| **Median Household Income** | 2.432 | <0.0001 |
| **% with a Bachelor's Degree or More** | 584.12 | 0.003 |
| **Presence of a Park (1=Yes, 0=No)** | 10834.4 | <0.0001 |
| **Police Unit (D=Reference Category)** |  |  |
| **A** | 12348.23 | <0.0001 |
| **B** | 23.4 | 0.983 |
| **C** | -23480.34 | <0.0001 |
| **Parking Type (Garage =Reference Category)** |  |  |
| **No Parking** | 434.8 | 0.08 |
| **Street Parking** | -15803.43 | 0.004 |

When we hold all other predictors constant:

* As median household income goes up by 1 unit ($1), the average increase in median house value is $2.432.
* As the % with a bachelor’s degree or more goes up by 1 unit (1%), the average increase in median house value is $584.12.
* As the park variable goes up by 1 unit, the average increase in median house value is $10,834.4. Said differently, as we go from a block group that doesn’t have a park (i.e., park = 0) to one that does have a park (park = 1), the average increase in median house value is $10,834.4.
* On average, block groups served by police unit A have median house value that is $12,348.23 higher than block groups served by police unit D.
* Block groups served by police unit aren’t significantly different (p=0.983) from block groups served by police unit D in terms of house value.
* On average, block groups served by police unit C have median house value that is $23,480.34 lower than block groups served by police unit D
* Block groups that have no parking don’t differ significantly (p=0.08) from block groups that have garage parking in terms of median house value.
* On average, median house value is $15,803.43 lower in block groups that have street parking than block groups that have garage parking.
* The intercept is the average median house value\* in (hypothetical\*\*) block groups where all predictors are 0. So, if we have a block group where median household income = 0, % of individuals with at least a bachelor’s degree = 0, presence of parks = 0, police unit dummies A, B, and C are 0, and parking dummies ‘no parking’ and ‘street parking’ are 0, the average median house value in that block group is estimated to be $4,388.
  + Said differently, the intercept ($4,388) is the average median house value in block groups where median household income = 0 and % of individuals with at least a bachelor’s degree = 0, where there are no parks, which is covered by police unit D, and where most houses have garage parking.

\* “Average median house value” may sound confusing. Here, the variable’s name is median house value (across all houses in a single block group). We can take the average of the variable across all block groups, and that will be the average median house value.

\*\* Block groups with such values of the variables may or may not actually exist in the data set.