Module 14: Modeling interactions

(The attached PDF file has better formatting.)

*Homework assignment: rating territories and mileage*

This homework assignment continues the exercise in Module 13.

An insurer examines claim frequencies for 15 territories: 5 urban, 5 suburban, and 5 rural. The insurer also has the average miles driven per car in each territory (in thousands). If you live outside the U.S., replace mileage with thousands of kilometers.

| Urban | | | Sub-urban | | | Rural | | |
|---|---|---|---|---|---|---|---|---|
| Territory | Mileage | Claim Freq'y | Territory | Mileage | Claim Freq'y | Territory | Mileage | Claim Freq'y |
| 1 | 5 | 8.45% | 6 | 20 | 6.99% | 11 | 10 | 3.83% |
| 2 | 10 | 10.90% | 7 | 40 | 12.94% | 12 | 20 | 5.06% |
| 3 | 15 | 13.45% | 8 | 60 | 19.01% | 13 | 30 | 6.00% |
| 4 | 20 | 16.04% | 9 | 80 | 25.06% | 14 | 40 | 6.94% |
| 5 | 25 | 18.49% | 10 | 100 | 31.11% | 15 | 50 | 7.92% |

A. How many dummy variables does this regression use?
B. What are the values of the dummy variables for urban, sub-urban, and rural? Assume rural is the base territory, with dummy variables equal to zero.
C. Write the regression equation with all interactions. You should have six terms.
D. Use Excel or other statistical software to run the regression. What are the values of the six regression parameters?

The claim frequencies are chosen so that the standard error of the regression is small. The observed values are very close to the fitted values, so you can tell if your solution is right.


*Jacob:* Do we have separate dummy variables for each territory?

*Rachel:* This homework assignment replicates the scenario in the textbook. It deals with three regions: urban, sub-urban, and rural, not 15 separate territories. The territories within each region just differ by average distance driven.

The solution has three intercepts and three slopes, giving six regression parameters. Look at the slopes first. The exercise says that the stochasticity of the observed values is small. The claim frequency increases about 2.5 percentage points for each five units of mileage in the urban region, about 6 percentage points for each 20 units of mileage in the suburban region, and about 1 percentage point for each 10 units of mileage in the rural region. The intercepts (where mileage = 0) also differ by region; they are about 6 percentage points in

the urban region, about 1 percentage point in the suburban region, and about 3 percentage points in the rural region.

Casual observation shows the formulas: claim frequency is 6% + 5% × mileage for urban; 3% + 1% × mileage for rural; 1% + 3% × mileage for suburban. The homework assignment has you solve for the precise figure using Excel (or R or SAS or Mathlab). Rural is the base, so the rural intercept and slope applies to urban and suburban as well. But urban and suburban (the two dummy variables) has different intercepts and slopes. Excel shows the differences are additions or subtractions to the intercept and slopes: +3 and –2 for the intercepts and +4 and +2 for the slopes.

For the homework assignment, set up the equations and solve them in Excel. The answers differ from the round numbers above by small amounts, and the $p$ values are all significant at the 0.1% level.