

WEEK 5 HOMEWORK

INSTRUCTIONS

- Every learner should submit his/her own homework solutions. However, you are allowed to discuss the homework with each other (in fact, I encourage you to form groups and/or use the forums) – but everyone must submit his/her own solution; you may not copy someone else's solution.
- The homework will be peer-graded. In analytics modeling, there are often lots of different approaches that work well, and I want you to see not just your own, but also others.
- The homework grading scale reflects the fact that the primary purpose of homework is learning:

Rating	Meaning	Point value (out of 100)
4	All correct (perhaps except a few details) <u>with</u> a deeper solution than expected	100
3	Most or all correct	90
2	Not correct, but a reasonable attempt	75
1	Not correct, insufficient effort	50
0	Not submitted	0

Question 8.1

Describe a situation or problem from your job, everyday life, current events, etc., for which a linear regression model would be appropriate. List some (up to 5) predictors that you might use.

Question 8.2

Using crime data from <http://www.statsci.org/data/general/uscrime.txt> (description at <http://www.statsci.org/data/general/uscrime.html>), use regression (a useful R function is `lm` or `glm`) to predict the observed crime rate in a city with the following data:

~~M = 14.0~~
~~So = 0~~
~~Ed = 10.0~~
~~Po1 = 12.0~~
~~Po2 = 15.5~~
~~tF = 0.640~~
~~M.F = 94.0~~
~~Pop = 150~~
~~NW = 1.1~~
~~U1 = 0.120~~
~~U2 = 3.6~~
~~Wealth = 3200~~

Ineq = 20.1
Prob = 0.04
Time = 39.0

Show your model (factors used and their coefficients), the software output, and the quality of fit.

Note that because there are only 47 data points and 15 predictors, you'll probably notice some overfitting. We'll see ways of dealing with this sort of problem later in the course.