Question 11.1

Using the crime data set uscrime.txt from Questions 8.2, 9.1, and 10.1, build a regression model using:

1. Stepwise regression

2. Lasso
3. Elastic net

For Parts 2 and 3, remember to scale the data first – otherwise, the regression coefficients will be on different scales and the constraint won't have the desired effect.

Answer 11.1

Part 1 - Using Stepwise regression

	Atributes	Comments	AIC
Stepwise regression	M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob + Time	Using all factors	514.65
Stepwise regression	M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob + Time	Removed "So"	512.65
Stepwise regression	M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob	Removed "Time"	511.01
Stepwise regression	M + Ed + Po1 + Po2 + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob	Removed "LF"	509.37
Stepwise regression	M + Ed + Po1 + Po2 + M.F + Pop + U1 + U2 + Wealth + Ineq + Prob	Removed "NW"	507.77
Stepwise regression	M + Ed + Po1 + M.F + Pop + U1 + U2 + Wealth + Ineq + Prob	Removed "Po2"	506.33
Stepwise regression	M + Ed + Po1 + M.F + U1 + U2 + Wealth + Ineq + Prob	Removed "Pop"	505.07
Stepwise regression	M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob	Removed "Wealth"	503.93

	Atributes	Comments	AIC
Lasso	M + So + Ed + Po1 + LF + M.F + NW + U1 + U2 + Ineq + Prob	Removed “Po2”, “Pop”, “Wealth” and “Time” based on Lasso model output	644.92
Elasticnet	M + So + Ed + Po1 + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob	Removed “Po2”, “LF” and “Time” based on Elastic net model output	645.43

Model Quality using Adjusted R squared and Cross Validated R squared values.

	Factors	Adjusted R squared	Cross Validated R squared
Stepwise Regression	M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob	0.74	0.67
Lasso	M + So + Ed + Po1 + LF + M.F + NW + U1 + U2 + Ineq + Prob	0.723	0.596
Elasticnet	M + So + Ed + Po1 + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob	0.7255	0.590

Comparing the models based on AIC and Cross Validated R squared values, Step wise regression seems to be giving the best factors to be used.

Question 12.1

Describe a situation or problem from your job, everyday life, current events, etc., for which a design of experiments approach would be appropriate.

Answer 12.1

One of the real situation where I will be using Design of experiments approach is to find the best ticket price for flights during summer holidays. Various experiment factors which seem to be equally important are - day when I book, how early/late I buy, when I want to fly (span of summer holidays being approx. 60 days its flexible). This could be combined with additional costs like get flights, hotel, car separately, or package it.

Question 12.2

To determine the value of 10 different yes/no features to the market value of a house (large yard, solar roof, etc.), a real estate agent plans to survey 50 potential buyers, showing a fictitious house with different combinations of features. To reduce the survey size, the agent wants to show just 16 fictitious houses. Use R's FrF2 function (in the FrF2 package) to find a fractional factorial design for this experiment: what set of features should each of the 16 fictitious houses have? Note: the output of FrF2 is "1" (include) or "-1" (don't include) for each feature.

Answer 12.2

Using R FrF2, following is the table which can be used for choosing 16 houses with different features. Output of "1" means include the feature in house, and output of "-1" means don't include feature in house.

House-Feature table-1

House	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7	Feature8	Feature9	Feature10
1	-1	-1	1	-1	1	-1	-1	1	1	-1
2	1	-1	1	-1	-1	1	-1	-1	1	1
3	-1	-1	-1	1	1	1	1	-1	1	-1
4	1	1	-1	1	1	-1	-1	1	-1	-1
5	-1	1	-1	-1	-1	1	-1	1	1	-1
6	1	-1	-1	1	-1	-1	1	1	1	1
7	-1	1	-1	1	-1	1	-1	-1	-1	1
8	-1	1	1	1	-1	-1	1	-1	1	-1
9	1	-1	1	1	-1	1	-1	1	-1	-1
10	-1	-1	-1	-1	1	1	1	1	-1	1
11	1	-1	-1	-1	-1	-1	1	-1	-1	-1
12	-1	1	1	-1	-1	-1	1	1	-1	1
13	-1	-1	1	1	1	-1	-1	-1	-1	1
14	1	1	1	-1	1	1	1	-1	-1	-1
15	1	1	-1	-1	1	-1	-1	-1	1	1
16	1	1	1	1	1	1	1	1	1	1

Question 13.1

For each of the following distributions, give an example of data that you would expect to follow this distribution (besides the examples already discussed in class).

a. Binomial
   b. Geometric c. Poisson
   d. Exponential e. Weibull

Answer 13.1

a. Binomial - If you purchase 100 lottery tickets, chances of winning the lottery might follow binomial distribution.

b. Geometric - Candy crush saga game has lots of levels. To get to the next level, player has to successfully complete the current level. The number of unsuccessful attempts for each level, before successful completion might follow Geometric distribution.

c. Poisson - In my company there is internal Call center to get technical assistance. For example - password reset, account locked, remote connectivity and so on. Number of calls received in call center tomorrow could follow Poisson distribution.

d. Exponential - Further extending the example of Poisson. Time between the calls received (inter arrival time) might follow Exponential distribution.

e. **Weibull** - My 2 boys have iPhones, and they tend to damage screen or something else very often. Time to failure or requirement for maintenance on their phones could follow Weibull distribution. Possibly when they become more responsible, I would expect $k>1$.