Exam 2000 1

We have obeserved the following variables for $n = 100$ apartments in Oslo:

- *Pris*: The price of the apartment

- *Kvm*: The area of the apartment (in sq. meters)

- *Rom*: The number of rooms

- *Leie*: The monthly expenses/rent

- *Balkong*: Whether it has a balcony

- *Garasje*: Wheter it has a garage

- x: The apartment's location in the east/west direction

- y: The apartment's location in the north/south direction

We use linear regression to analyze the data, with *Pris* is the outcome variable

a)

As a first analysis we do the regression against area and the number of rooms as both separately and combined. The printouts give the output from the regressions and descriptive statistics. Give a description and an intuitive explanation of the phenomenon you observe.

b)

The results from a regression analysis with all covariates included is given. Explain what each of the estimated parameters tells you about have the price depends on the covariate. You don't need to look at the significance here.

c)

Find which covariates have a significant effect. Discuss pros and cons for removing insignificant variables from the model. Compare with the next analysis where some covariates have been excluded from the model:

d)

Give a short summary of how residual plots can be used to check if the model assumptions are correct. Use these methods on the given plots which are created from the reduced model.

e)

For the full model we find a cross-validated R^2 of 0.873. For the reduced model, the cross-validated R^2 is 0.882. Compare with the R^2 from the printouts. Explain what these terms mean and discuss the difference between them. How can you use cross-validated R^2 to compare the full and reduced model?