

ST301-1995-1

We look at the relationship between fuel consumption and the weight, engine displacement and car type for cars. Fuel, weight and disp are continuous variables, while Type can take 5 values: Small, Sporty, Compact, Medium, Large and Van.

a)

Comment the plots and tables below, and discuss them in the context of regression analysis.

b)

A linear regression with variables Weight Disp and Type, and Fuel as response gives this result:

Why is there 5 parameters for Type? Discuss the two last columns in the table.

c)

In the updated model, $(Disp)^2$ is included as a variable. Why do we get a bigger R^2 here? Discuss the interpretation of R^2 . Does it look reasonable to include $(Disp)^2$ in the model?

d)

The table gives a cross-validated estimate for R^2 for different models. Explain what is meant by a cross-validated estimate for R^2 and why this is a more reasonable measure than the normal R^2 . Which model would you choose?

e)

The plots show residual plots. Explain what we can use the different plots for and give an interpretation of the plots for the model.

f)

Give a point estimate for a car with weight 2745kg, engine displacement 124 and of type 'Compact'. What problems arise when determining a the uncertainty.

ST301-2000, 1

For 100 apartments in Oslo the following is registered:

- Price (pris)
- Area of apartment (kvm)
- Number of rooms (rom)
- Monthly rent (leie)
- Balcony (yes=1, no=0)
- Garage (yes=1, no=0)
- x-coordinate (in km, east)
- y-coordinate (in km, north)

a)

Give a description and intuitive explanation of the phenomenon from these regression analyses. (kvm, rom, kvm+rom)

b)

Explain what each of the regression parameters tells you about how the price depends on the covariates.

c)

Which covariates have a significant effect? Discuss advantages/disadvantages of removing insignificant covariates from the model. Compare with the analysis below where some covariates are removed.

d)

Give a short summary of how residual plots can be used to check if the model-assumptions are valid. Use the method on the plots.

e)

For the full model we find cross-validated $R^2 = 0.873$. For the reduced model $R^2 = 0.882$. Compare with R^2 from the read-outs. Explain the terms and the differences between them. How can we use cross-validated R^2 to compare the full and the reduced model.