# Problem set 4: Dummy variables
# Data analysis part, ØKA201

Bjørnar Karlsen Kivedal

## Excercise 1

Consider data on an outcome variable $Y_i$ and an indicator variable $D_i$.

### a)

Suppose $\bar{Y} = 30$ for the subsample with $D_i = 1$ and $\bar{Y} = 20$ for the subsample with $D_i = 0$ . Give the fitted model from OLS regression of $Y_i$ on $D_i$ using the full sample.

### b)

Suppose OLS regression using all data yields the fitted model $\hat{Y}_i = 3 + 5D_i$. Give $\bar{Y}$ for the subsample with $D_i = 0$ and $\bar{Y}$ for the subsample with $D_i = 1$.

## Excercise 2

Use the data set `HousePrice.gdt` (the same as in PS3) for this excercise.

### a)

Create three separate dummy variables `Dlotsize_1`, `Dlotsize_1`, `Dlotsize_1` for lotsize equal to 1, 2 and 3 using Gretl.[1]

### b)

Verify that these dummy variables are mutually exclusive.

### c)

Regress price on size and lotsize.

### d)

Regress price on size, `Dlotsize_2` and `Dlotsize_3`. Are `Dlotsize_2` and `Dlotsize_3` jointly statistically significant?

### e)

Which model do you prefer, that in (b) or that in (c)? Explain.

### f)

Redo part (d) with `Dlotsize_1` and `Dlotsize_2` rather than `Dlotsize_2` and `Dlotsize_3`. How have your answers changed?

### g)

Regress price on size, `Dlotsize_1`, `Dlotsize_2` and `Dlotsize_3`. Explain what has happened.

### h)

Regress price on size, `Dlotsize_1`, `Dlotsize_2` and lotsize. Explain what has happened.

[1] Hint to Gretl: Choose the variable `lotsize` and go to `Add > Dummies for discrete variable...` (or right-click the variable and select `dummify...`).