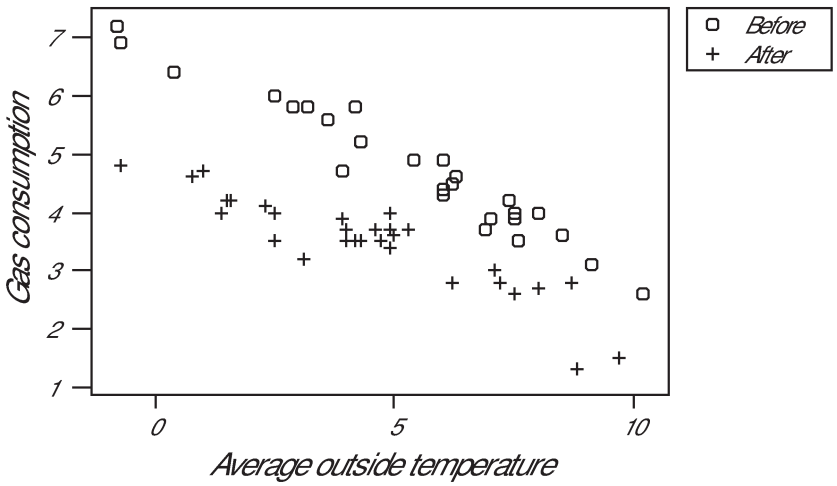
UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM: Example — Cavity-wall insulation **Semester 1, 2022–2023**

Numerical Example 1.9 in the lecture notes concerned the relationship between the weekly gas consumption (in thousands of cubic feet) and the average outside temperature (in °C) for a house which had gas-fired central heating. These two variables were recorded for 26 weeks before and 30 weeks after cavity-wall insulation was installed. The house thermostat was set to 20°C throughout. The data are given in the file Insulate.txt along with a column indicating whether they were recorded before or after installation.

A plot of the data (given again below) shows that, as expected, gas consumption is lower when the outside temperature is increased and when insulation is installed. We can use regression analysis to quantify these two effects, and also compare the two regression lines.



We first define a vector called 'indicator' taking values 0 and 1 for observations made before and after installation, and a vector 'ind_temp' equal to zero for the 'before' observations and to the outside temperature for those made afterwards. Including these as explanatory variables allows different slopes and intercepts to be fitted for the two lines.

```
Insulate.dat <- read.table("Insulate.txt",header=T)
attach(Insulate.dat)
indicator <- 1*(Before.after=="After")  # 0 for "Before" 1 for "After"
temp <- Average_outside_temperature
ind_temp <- indicator*temp
fit1 <- lm(Gas_consumption ~ temp + indicator + ind_temp)
summary(fit1)
```

```
Call:
lm(formula = Gas_consumption ~ temp + indicator + ind_temp)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97802 -0.18011  0.03757  0.20930  0.63803

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.85383    0.13596  50.409  < 2e-16 ***
temp        -0.39324    0.02249 -17.487  < 2e-16 ***
indicator   -2.12998    0.18009 -11.827 2.32e-16 ***
ind_temp     0.11530    0.03211   3.591 0.000731 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimated intercepts and slopes for the weeks before installation are

$$\widehat{\alpha}_1 = 6.854\,, \quad \widehat{\beta}_1 = -0.393\,,$$

and those for the weeks afterwards are

$$\widehat{\alpha}_2 = 6.854 - 2.130 = 4.724\,, \quad \widehat{\beta}_2 = -0.393 + 0.115 = -0.278\,.$$

The large value (3.59) of the t statistic for '`ind_temp`' shows that there is strong evidence that the true slopes of the two regression lines differ: the effect of an increase in temperature on consumption is reduced after the insulation is installed.

If we had found little or no evidence of a difference between the true slopes, we might have wanted to fit regression lines with a common slope. To do this, we could regress gas consumption on the variables '`temp`' and '`indicator`' only. The t statistic for '`indicator`' would then show whether the difference between the estimated intercepts was statistically significant under the assumption of a common slope.

Note that we ought to be sceptical about whether these data satisfy the assumptions of linear regression: there are several apparent outliers, and successive weekly gas consumption values may be correlated.