Intermediate Econometrics

Lecture 5 - Interpretation LM

Vincenzo Gioia

28 October 2026

Linear Model

Interpretation

- We use the classic `whiteside` data from package `MASS` (gas consumption vs temperature)

```
#install.packages("MASS")
library(MASS)
```

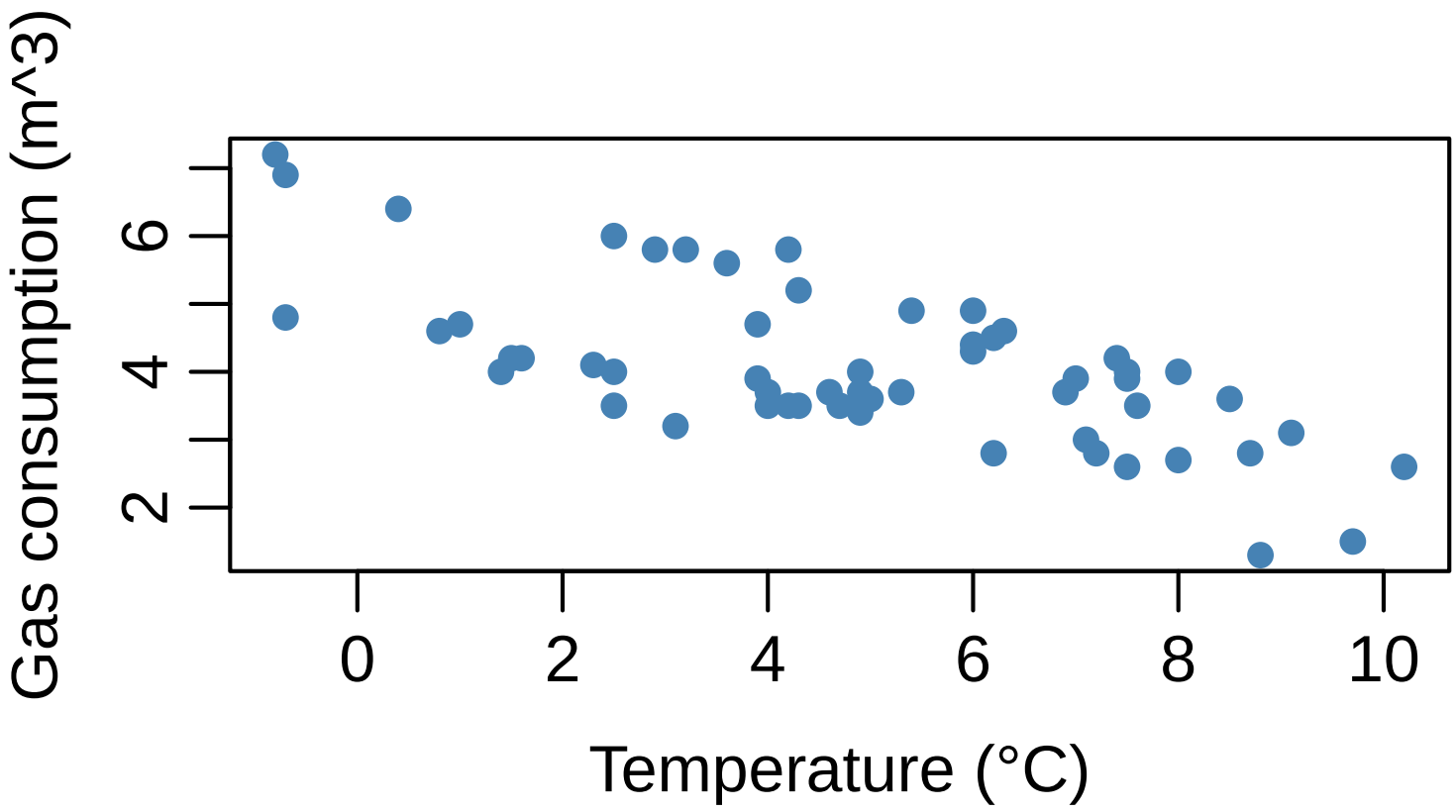
```
## Warning: il pacchetto 'MASS' è stato creato con R versione 4.5.3
```

```
data(whiteside)
```

Linear Model
Interpretation

- The negative association is clear: colder is outside and more gas the house consumes. We want to **quantify** this relationship

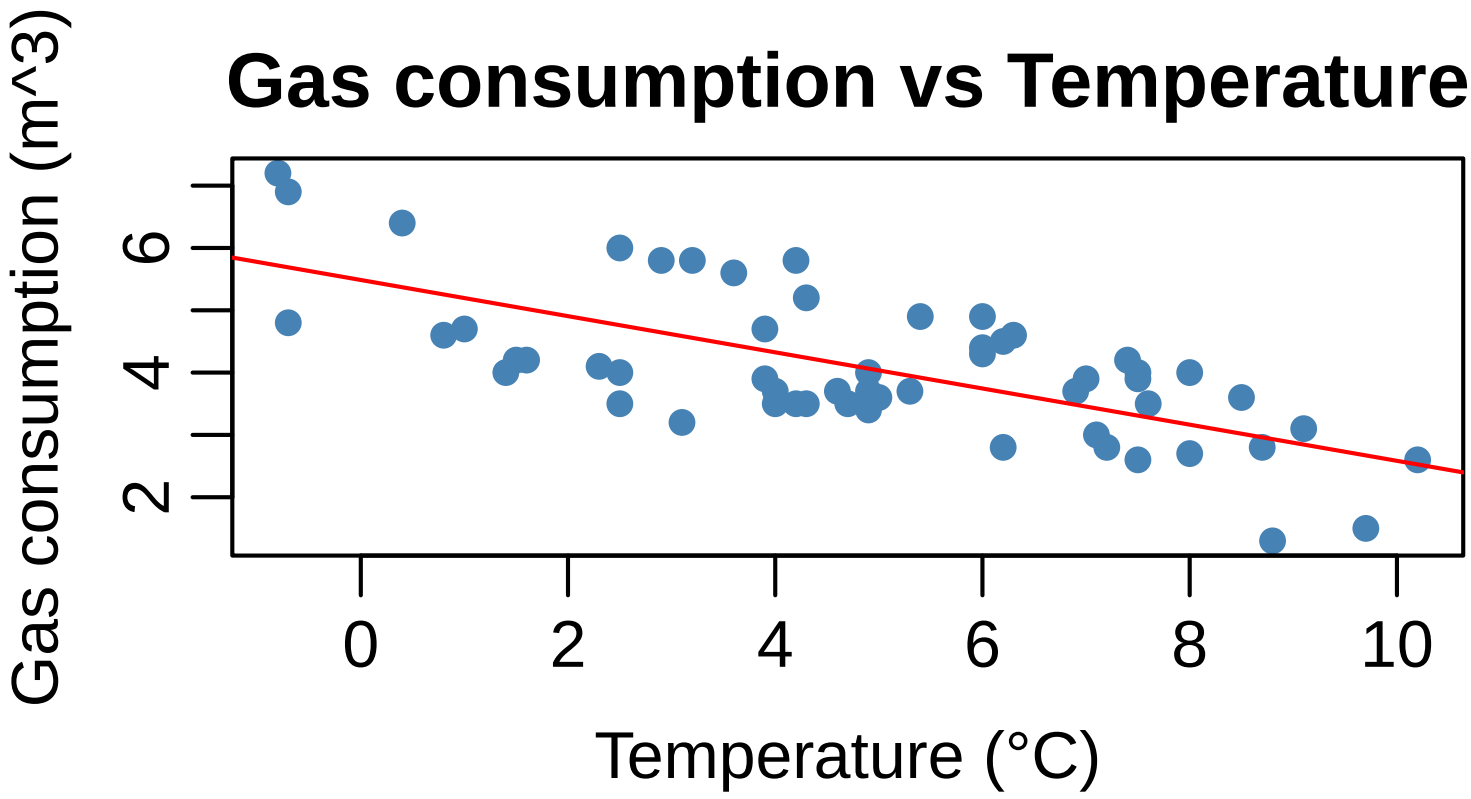
```
plot(whiteside$Temp, whiteside$Gas, pch  = 16,
     col = "steelblue", cex = 0.9,
     xlab = "Temperature (°C)", ylab = "Gas consumption (m^3)")
```



Linear Model

Interpretation

- $\hat{\beta}_0$: estimated average consumption when the temperature is 0
- $\hat{\beta}_1$: estimated average increase/decrease of gas consumption when the temperature is increasing of 1 degree



Linear Model
Interpretation

```
mod1 <- lm(Gas ~ Temp, data = whiteside)
summary(mod1)
```

```
##
## Call:
## lm(formula = Gas ~ Temp, data = whiteside)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6324 -0.7119 -0.2047  0.8187  1.5327
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.4862     0.2357  23.275  < 2e-16 ***
## Temp         -0.2902     0.0422  -6.876 6.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8606 on 54 degrees of freedom
## Multiple R-squared:  0.4668, Adjusted R-squared:  0.457
## F-statistic: 47.28 on 1 and 54 DF,  p-value: 6.545e-09
```

From the output of `mod1`:

- **Intercept ($\hat{\beta}_0 \approx 5.48$)**: estimated mean gas consumption when the external temperature is 0°C. Because 0°C is within the observed range of temperatures, this value is directly interpretable.
- **Slope ($\hat{\beta}_1 \approx -0.29$)**: for each additional degree Celsius of external temperature, the expected gas consumption **decreases by about 0.29 m³**. The negative sign confirms that colder weather drives higher consumption.

Linear Model

Interpretation

In math terms:

$$Gas_i = \mathbb{E}[Gas_i | Temp_i] + \varepsilon_i = \beta_0 + \beta_1 Temp_i + \varepsilon_i$$

- **Intercept**:

$$\mathbb{E}[Gas_i | Temp_i = 0] = \beta_0$$

- **Slope**:

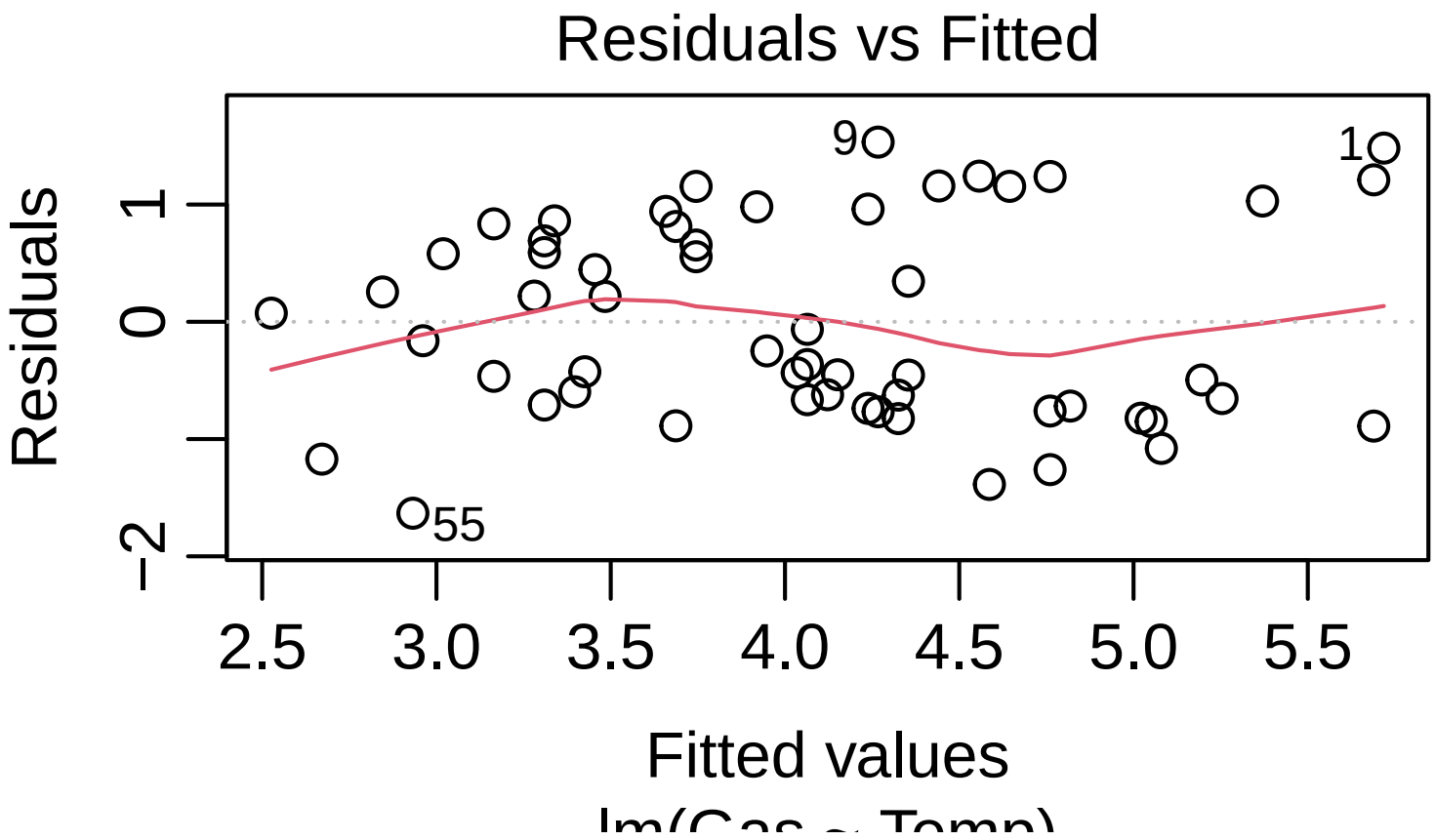
$$\mathbb{E}[Gas_i | Temp_i = c+1] - \mathbb{E}[Gas_i | Temp_i = c] = \beta_0 + \beta_1 \times (c+1) - \beta_0 - \beta_1 \times c = \beta_1$$

Linear Model

Interpretation

- Two distinct clusters of residuals are visible — a clear signal that a relevant variable has been omitted

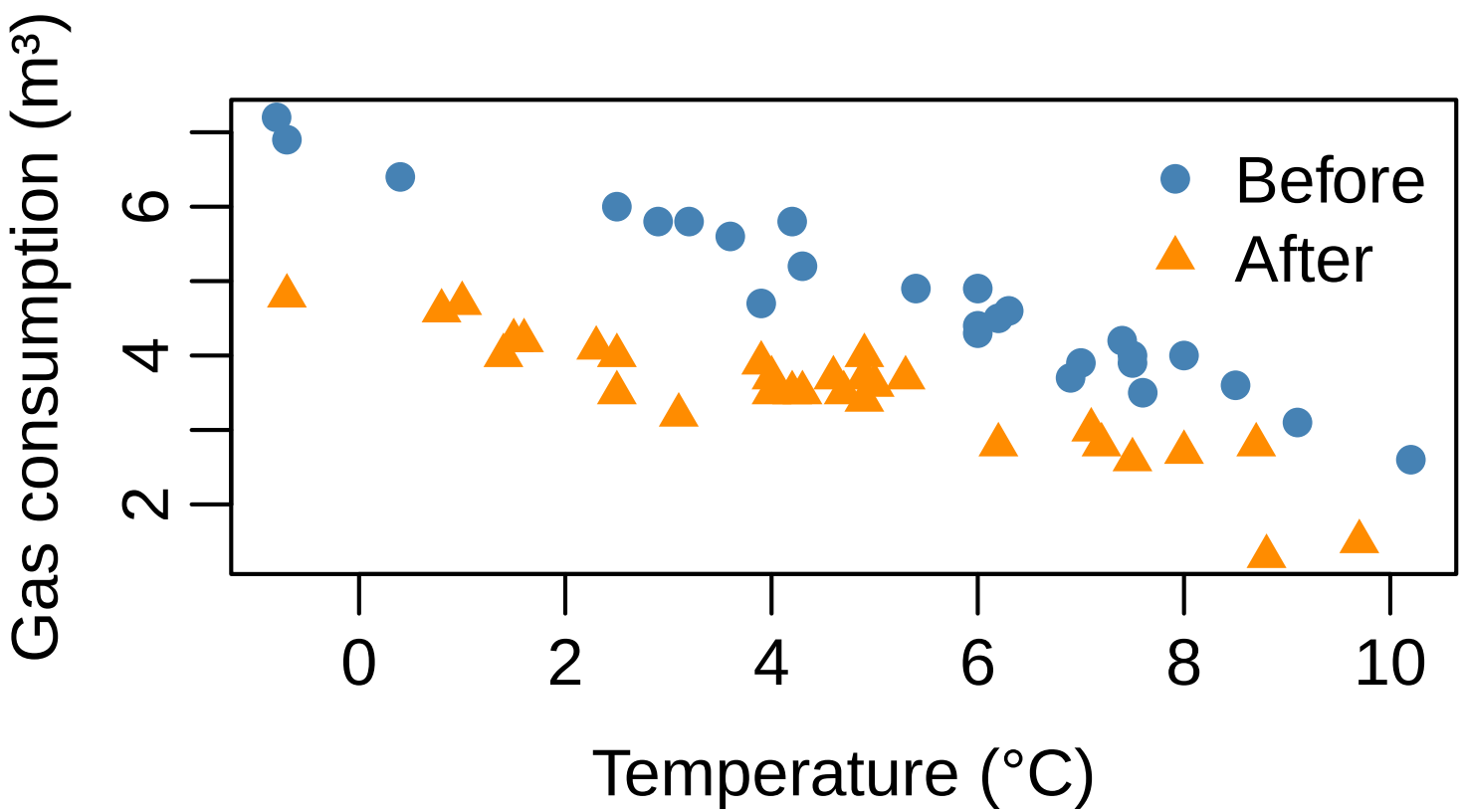
```
plot(mod1, which = 1)
```



Linear Model
Interpretation

- Colouring the scatter plot by the `Insul` variable reveals why: At any given temperature, gas consumption is systematically higher *before* the insulation works than *after*. The two groups clearly follow different trends — a strong signal that `Insul` must be included in the model.

```
plot(whiteside$Temp, whiteside$Gas,
     pch  = c(16, 17)[as.numeric(whiteside$Insul)],
     col  = c("steelblue", "darkorange")[as.numeric(whiteside$Insul)],
     xlab = "Temperature (°C)", ylab = "Gas consumption (m³)")
legend("topright", legend = c("Before", "After"), pch    = c(16, 17),
       col    = c("steelblue", "darkorange"), bty    = "n")
```



Linear Model

Interpretation

$$d_i = \begin{cases} 0 & \text{if } Insul_i = \text{"Before"} \\ 1 & \text{if } Insul_i = \text{"After"} \end{cases}$$

The model then becomes:

$$Gas_i = \beta_0 + \beta_1 \cdot Temp_i + \beta_2 \cdot d_i + \varepsilon_i$$

with

$$\mathbb{E}(Gas_i \mid Temp_i,\ d_i) = \beta_0 + \beta_1 \cdot Temp_i + \beta_2 \cdot d_i$$

which gives **two parallel lines** sharing the same slope β_1:

$$\mathbb{E}(Gas_i \mid Temp_i) = \begin{cases} \beta_0 + \beta_1 \cdot Temp_i & \text{Before } (d = 0) \\ (\beta_0 + \beta_2) + \beta_1 \cdot Temp_i & \text{After } (d = 1) \end{cases}$$

Linear Model

Interpretation

Thus,

- Intercept: $\mathbb{E}(Gas_i | Temp_i = 0, d_i = 0) = \beta_0$ (estimated average consumption before the insulation when the temperature was 0 degree celsius)

- Coefficient of Temp:

$$\mathbb{E}(Gas_i | Temp_i = c + 1, d_i = k) - \mathbb{E}(Gas_i | Temp_i = c, d_i = k) =$$

$$\beta_0 + \beta_1 \times (c + 1) + \beta_2 \times k - \beta_0 - \beta_1 \times c - \beta_2 \times k = \beta_1$$

(estimated average increase/decrease of consumption for a unitary increase of the temperature, before ($k = 0$) or after ($k = 1$) the insulation)

- Coefficient of Insulation:

$$\mathbb{E}(Gas_i | Temp_i = c, d_i = 1) - \mathbb{E}(Gas_i | Temp_i = c, d_i = 0) =$$

$$\beta_0 + \beta_1 \times c + \beta_2 \times 1 - \beta_0 - \beta_1 \times c - \beta_2 \times 0 = \beta_2$$

(estimated average difference in gas consumption between after and before insulation, holding fixed/for a given value of temperature)

Linear Model

Interpretation

```r
mod2 <- lm(Gas ~ Temp + Insul, data = whiteside)
summary(mod2)
```

```
##
## Call:
## lm(formula = Gas ~ Temp + Insul, data = whiteside)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74236 -0.22291  0.04338  0.24377  0.74314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.55133    0.11809   55.48   <2e-16 ***
## Temp        -0.33670    0.01776  -18.95   <2e-16 ***
## InsulAfter  -1.56520    0.09705  -16.13   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3574 on 53 degrees of freedom
## Multiple R-squared:  0.9097, Adjusted R-squared:  0.9063
## F-statistic: 267.1 on 2 and 53 DF,  p-value: < 2.2e-16
```

Linear Model

Interpretation

```r
b0 <- coef(mod2)["(Intercept)"]
b1 <- coef(mod2)["Temp"]
b2 <- coef(mod2)["InsulAfter"]

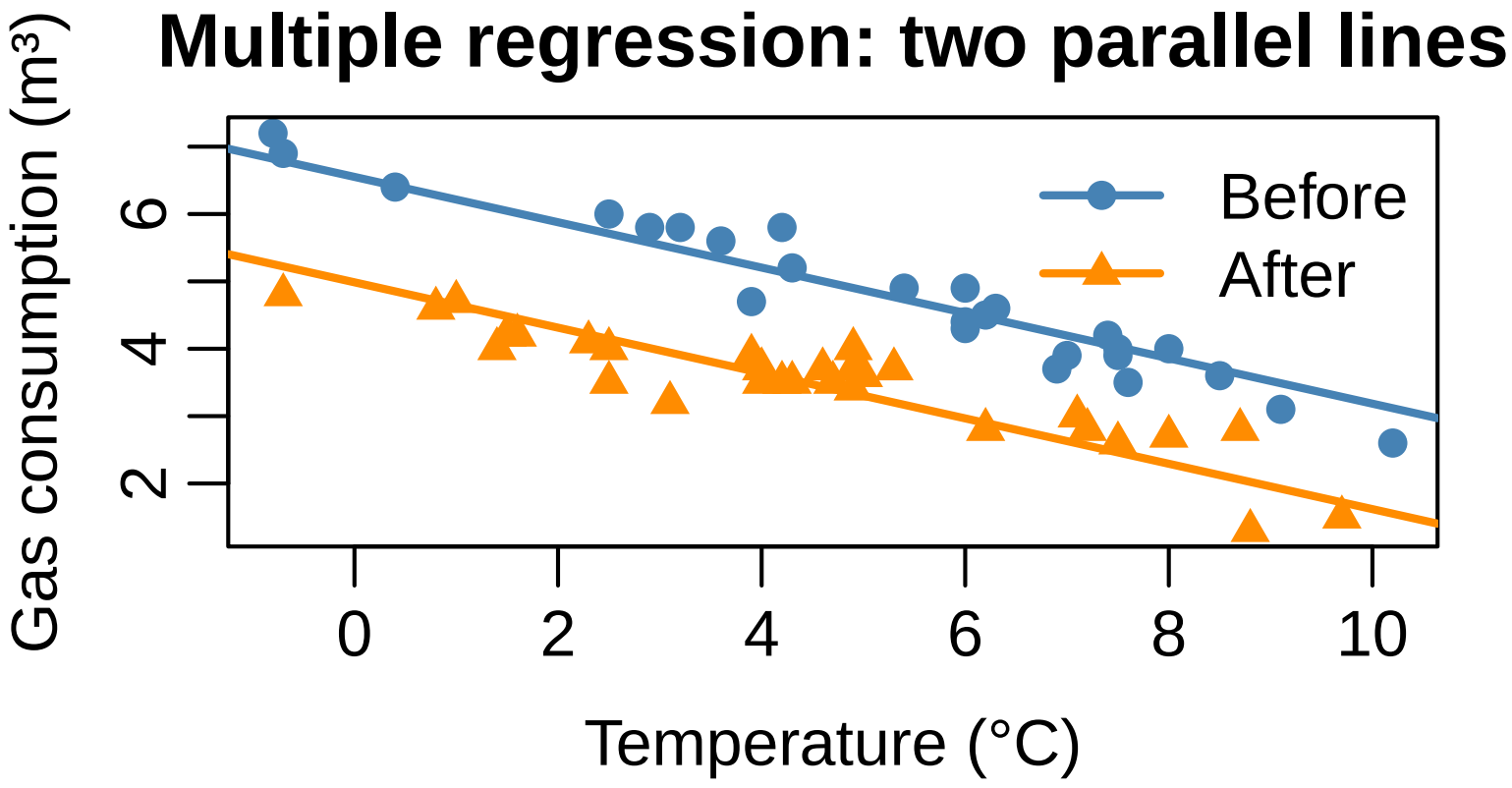
plot(whiteside$Temp, whiteside$Gas,
     pch  = c(16, 17)[as.numeric(whiteside$Insul)],
     col  = c("steelblue", "darkorange")[as.numeric(whiteside$Insul)],
     xlab = "Temperature (°C)",
     ylab = "Gas consumption (m³)",
     main = "Multiple regression: two parallel lines")

abline(b0,      b1, col = "steelblue",   lwd = 2)   # Before
abline(b0 + b2, b1, col = "darkorange",  lwd = 2)   # After

legend("topright",
       legend = c("Before", "After"),
       col    = c("steelblue", "darkorange"),
       pch    = c(16, 17), lwd = 2, bty = "n")
```

Linear Model

Interpretation

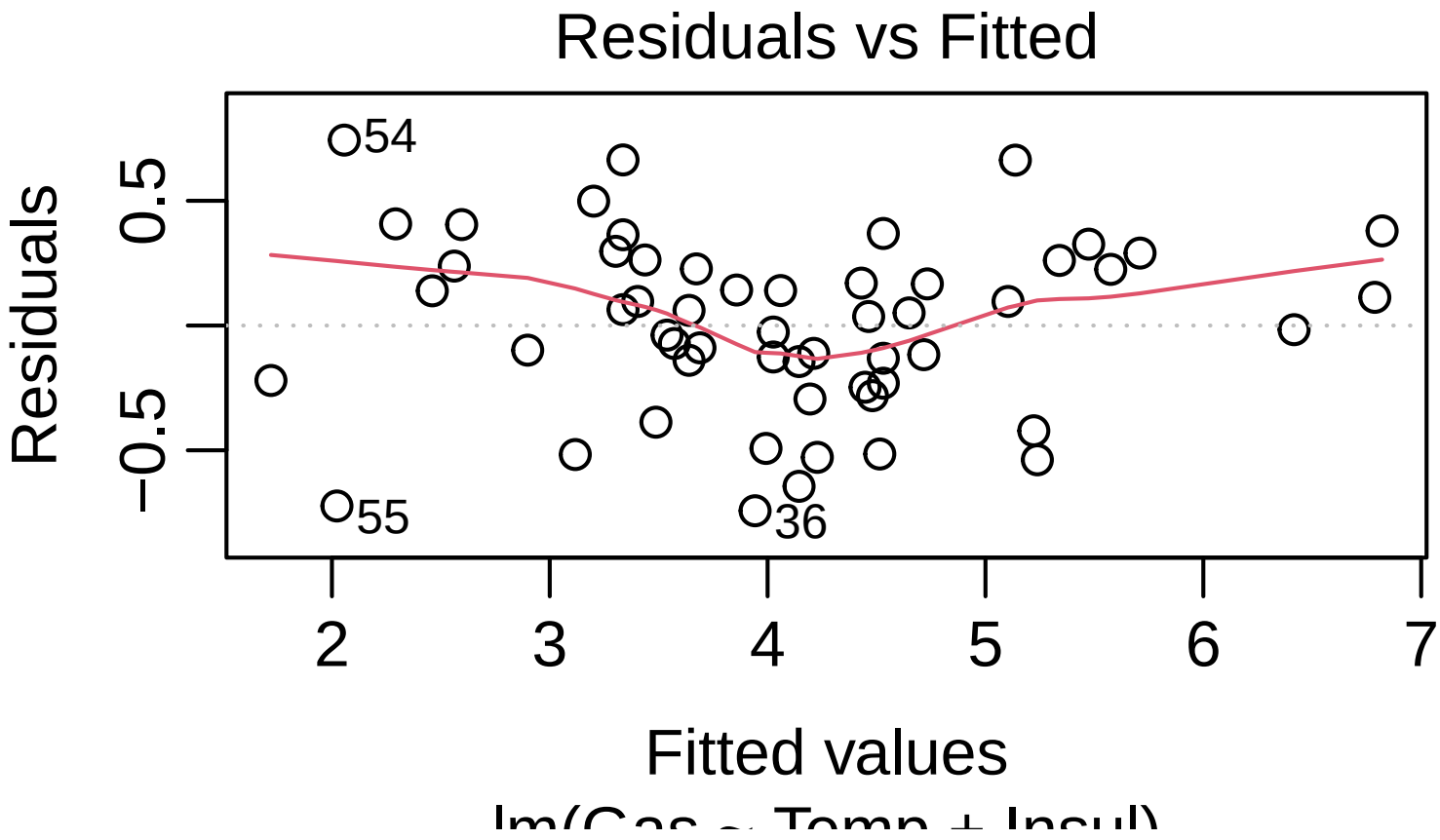


Linear Model

Interpretation

- The two-cluster pattern has disappeared. Residuals now scatter randomly around zero with roughly constant spread — the model assumptions look much more plausible.

```
plot(mod2, which = 1)
```



Linear Model

Interpretation

- Of course we can evaluate different thinks, e.g.

$$\mathbb{E}(Gas_i | Temp_i = c_1, d_i = 1) - \mathbb{E}(Gas_i | Temp_i = c_2, d_i = 0) =$$
$$= \beta_0 + \beta_1 \times c_1 + \beta_2 - \beta_0 - \beta_1 \times c_2 - \beta_2 \times 0 = \beta_1(c_1 - c_2) - \beta_2$$

(estimated averaged difference in gas consumption between a house after the isolation with c_1 degrees of temperature and a house before the isolation with c_2 degrees of temperature)

```
predict(mod2, newdata = data.frame(Temp = 10, Insul = "After"))-
  predict(mod2, newdata = data.frame(Temp = 5, Insul = "Before"))
```

```
##         1
## -3.248689
```

Linear Model

Interpretation

- Let's fit the model including the interaction term: in our mind we can suspect that the effect of isolation is different for different levels of temperatures (I guess that lower the temperatures and greater is the effect of isolation)

```
mod3 <- lm(Gas ~ Temp*Insul, data = whiteside)
summary(mod3)
```

```
##
## Call:
## lm(formula = Gas ~ Temp * Insul, data = whiteside)
##
## 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 ***
## InsulAfter       -2.12998    0.18009 -11.827 2.32e-16 ***
## Temp:InsulAfter   0.11530    0.03211   3.591 0.000731 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.323 on 52 degrees of freedom
## Multiple R-squared:  0.9277, Adjusted R-squared:  0.9235
## F-statistic: 222.3 on 3 and 52 DF,  p-value: < 2.2e-16
```

Linear Model

Interpretation

Mathematically speaking

$$Gas_i = \beta_0 + \beta_1 \cdot Temp_i + \beta_2 \cdot d_i + \beta_3 \cdot Temp_i \cdot d_i + \varepsilon_i$$

- Intercept:
$$\mathbb{E}(Gas_i | Temp_i = 0, d_i = 0) = \beta_0$$
- Now, β_1 and β_2 are simply called the main effects, but to say something more the coefficient β_3 must be considered
- For instance, given a level of temperature (c_1) what is the effect of insulation?

$$\mathbb{E}(Gas_i | Temp_i = c_1, d_i = 1) - \mathbb{E}(Gas_i | Temp_i = c_1, d_i = 0) =$$

$$\beta_0 + \beta_1 \cdot c_1 + \beta_2 \cdot 1 + \beta_3 \cdot c_1 \cdot 1 - \beta_0 - \beta_1 \cdot c_1 = \beta_2 + \beta_3 c_1$$

(expected difference in gas consumption between after and before isolation is changing with temperature)

- For instance, given the isolation ($d_i = k$) what is the effect of a unitary increase of temperature?

$$\mathbb{E}(Gas_i | Temp_i = c_1 + 1, d_i = k) - \mathbb{E}(Gas_i | Temp_i = c_1, d_i = k) =$$

$$\beta_0 + \beta_1 \cdot (c_1 + 1) + \beta_2 \cdot k + \beta_3 \cdot (c_1 + 1) \cdot k - \beta_0 - \beta_1 \cdot c_1 - \beta_2 \cdot k - \beta_3 \cdot c_1 \cdot k =$$
$$= \beta_1 + \beta_3 k$$

(expected difference in gas consumption for a unit increase of temperature is different before or after the insulation)

Linear Model

Interpretation

```
mod3 <- lm(Gas ~ Temp*Insul, data = whiteside)
summary(mod3)
```

```
##
## Call:
## lm(formula = Gas ~ Temp * Insul, data = whiteside)
##
## 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 ***
## InsulAfter       -2.12998    0.18009 -11.827 2.32e-16 ***
## Temp:InsulAfter   0.11530    0.03211   3.591 0.000731 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.323 on 52 degrees of freedom
## Multiple R-squared:  0.9277, Adjusted R-squared:  0.9235
## F-statistic: 222.3 on 3 and 52 DF,  p-value: < 2.2e-16
```

Linear Model

Interpretation

```r
b0 <- coef(mod3)["(Intercept)"]
b1 <- coef(mod3)["Temp"]
b2 <- coef(mod3)["InsulAfter"]
b3 <- coef(mod3)["Temp:InsulAfte"]

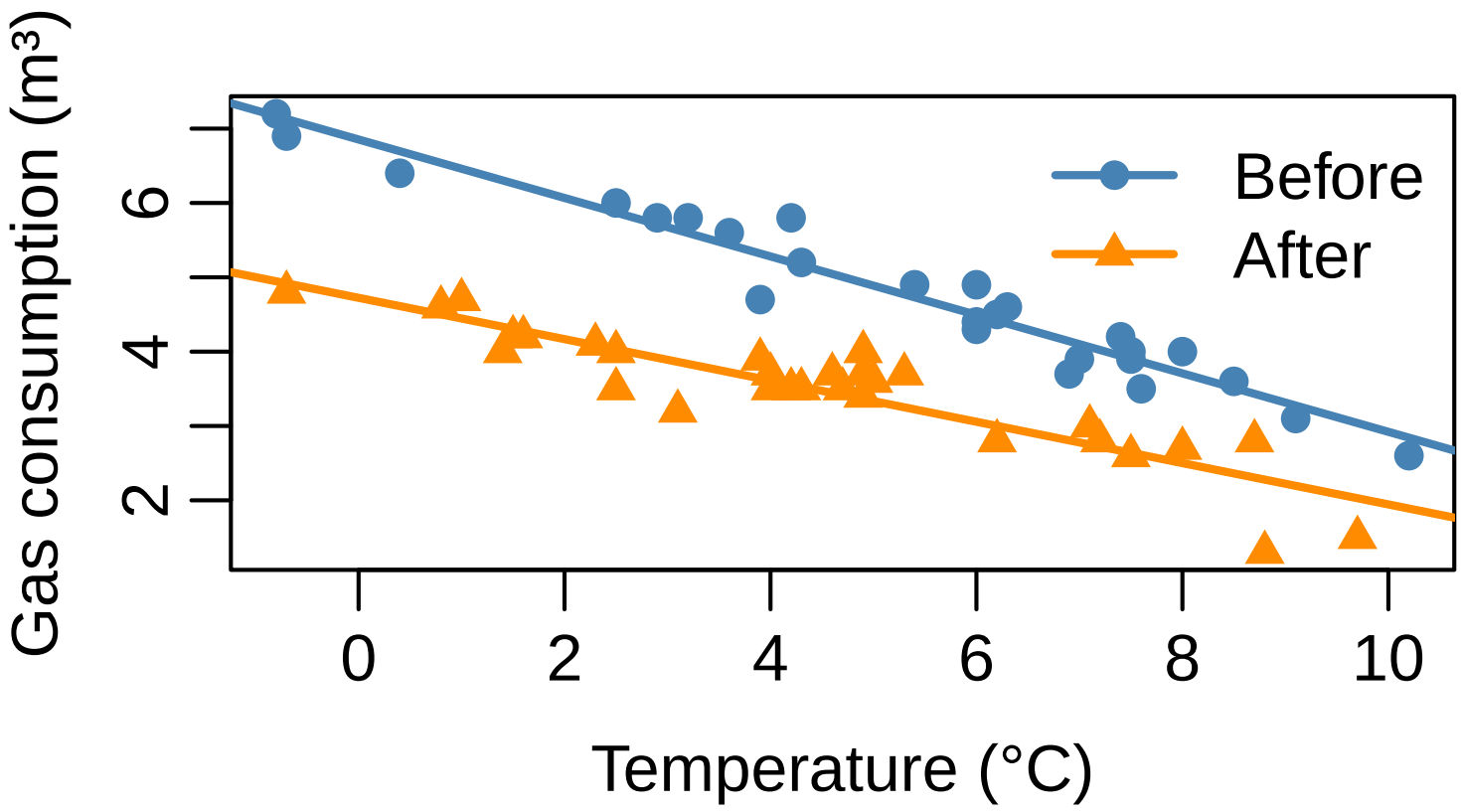
plot(whiteside$Temp, whiteside$Gas,
     pch  = c(16, 17)[as.numeric(whiteside$Insul)],
     col  = c("steelblue", "darkorange")[as.numeric(whiteside$Insul)],
     xlab = "Temperature (°C)", ylab = "Gas consumption (m³)")

abline(b0,        b1, col = "steelblue",   lwd = 2)   # Before
abline(b0 + b2,   b1 + b3, col = "darkorange",  lwd = 2)   # After

legend("topright",
       legend = c("Before", "After"),
       col    = c("steelblue", "darkorange"),
       pch    = c(16, 17), lwd = 2, bty = "n")
```

Linear Model

Interpretation



Linear Model

Interpretation

- Here we can see that the insulation effect is lower for higher temperature, as it is estimated by $\hat{\beta}_2 + \hat{\beta}_3 \mathit{Temp} = -2.12998 + 0.1153 \cdot \mathit{Temp}$

```
predict(mod3, newdata = data.frame(Temp = 2, Insul = "After")) -
  predict(mod3, newdata = data.frame(Temp = 2, Insul = "Before"))
```

```
##         1
## -1.89937
```

```
predict(mod3, newdata = data.frame(Temp = 10, Insul = "After")) -
  predict(mod3, newdata = data.frame(Temp = 10, Insul = "Before"))
```

```
##           1
## -0.9769393
```

- Also we can see that unitary change in temperature is more pronounced before the isolation than after, as it is estimated by $\hat{\beta}_1 + \hat{\beta}_3 d = -0.3932 + 0.1153 \cdot d$

```
predict(mod3, newdata = data.frame(Temp = 3, Insul = "After")) -
  predict(mod3, newdata = data.frame(Temp = 2, Insul = "After"))
```

```
##         1
## -0.277935
```

```
predict(mod3, newdata = data.frame(Temp = 3, Insul = "Before")) -
  predict(mod3, newdata = data.frame(Temp = 2, Insul = "Before"))
```

```
##           1
## -0.3932388
```

Linear Model

Interpretation

- The model assumptions look much more plausible

```
plot(mod3, which = 1)
```

