

Exercise: Housing Prices and Location Effects

Context

You are a data analyst working for a real estate firm. The firm wants to understand how the size of a house (in square meters), the location, and the interaction between location and size affect the price of houses in a city.

You collect data on 150 homes sold in the past year.

Variables

- price: Sale price of the house in thousands of USD
- size: Size of the house in square meters
- location: Categorical variable for neighborhood:
 - 'Suburb' (reference group)
 - 'Downtown'
 - 'Waterfront'
- size * location: Interaction between size and location

Regression Output

You run the following model:

$$\text{price}_i = \beta_0 + \beta_1 \cdot \text{size}_i + \beta_2 \cdot \text{Downtown}_i + \beta_3 \cdot \text{Waterfront}_i + \beta_4 \cdot (\text{size}_i \times \text{Downtown}_i) + \beta_5 \cdot (\text{size}_i \times \text{Waterfront}_i) + \varepsilon_i$$

Variable	Estimate	Std. Error	p-value
Intercept	150.0	10.0	<0.001
size	1.20	0.15	<0.001
Downtown	20.0	12.0	0.09
Waterfront	50.0	14.0	<0.001
size × Downtown	0.05	0.19	0.79
size × Waterfront	1.50	0.25	<0.001

$$R^2 = 0.78$$

$$\text{Adjusted } R^2 = 0.76$$

Questions

1. Interpret the coefficient on size.
2. Interpret the coefficient on Waterfront.
3. What is the effect of an additional square meter of size on the price of a house located:
 - a) in the Suburb
 - b) in Downtown
 - c) in Waterfront

4. Is there a statistically significant difference in baseline price (for size 0) between Suburb and Waterfront?
Is the effect of house size significantly different in Downtown compared to Suburb?
5. Suppose you are a buyer: would you expect to pay more per square meter in Downtown or Waterfront? Explain using the model coefficients.
6. 7. Prediction Task:
Consider a house that is 80 square meters in size. What is the predicted price of the house if it is located:
 - a) in the Suburb
 - b) in Downtown
 - c) in WaterfrontShow your calculations using the regression coefficients.
7. 8. Evaluate model fit:
 - a) How much variation in house prices does the model explain?
 - b) Why might adjusted R^2 be lower than R^2 ?