


# Stat 401 B – Lecture 22



## Multiple Regression

- Response,  $Y$  (numerical)
- Explanatory variables,  $X_1, X_2, \dots, X_k$  (numerical)
- New explanatory variables can be created from existing explanatory variables.

1

---

---

---


---

---

---

---

---



## Home gas consumption

- Weekly gas consumption for a home in England.
- Average outside temperature.
- There are 26 weeks before insulation was added and 18 weeks after adding insulation.

2

---

---

---


---

---

---

---

---



## Home gas consumption

- Response: Gas, gas consumption in 1000's of cubic feet.
- Explanatory: Temp, average outside temperature in  $^{\circ}\text{C}$ .

3

---

---

---

---

---

---

---

---

# Stat 401 B – Lecture 22

## Home gas consumption

- Response: Gas, gas consumption in 1000's of cubic feet.
- Explanatory: Insul, a dummy or indicator variable
  - Insul = 0, before insulation was added
  - Insul = 1, after insulation was added

4

---

---

---

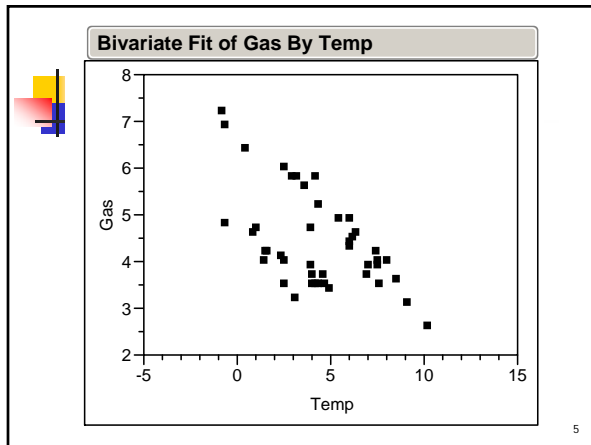
---

---

---

---

---



5

---

---

---

---

---

---

---

---

## General Trend

- As outside temperature increases, gas consumption goes down.
- There is something funny about the plot.

6

---

---

---

---

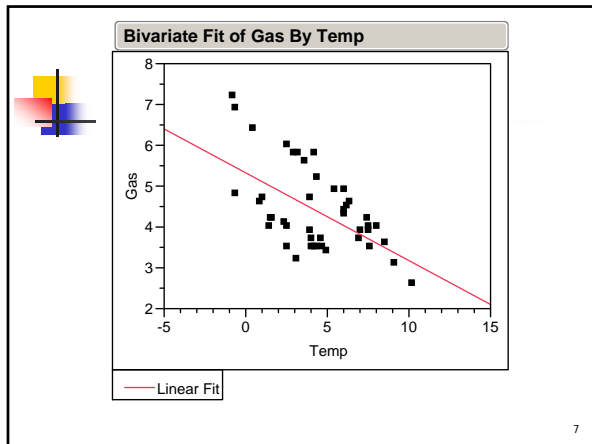
---

---

---

---

# Stat 401 B – Lecture 22



---

---

---

---

---

---

---

---

## Simple Linear Regression

- Predicted Gas =  $5.33 - 0.216 \cdot \text{Temp}$
- $R^2 = 0.328$ , 32.8% of the variation in gas consumption can be explained by the linear relationship with outside temperature.

---

---

---

---

---

---

---

---

## Simple Linear Regression

- Temperature is statistically significant.
  - $t = -4.53$ , P-value  $< 0.0001$
- RMSE = 0.853

---

---

---

---

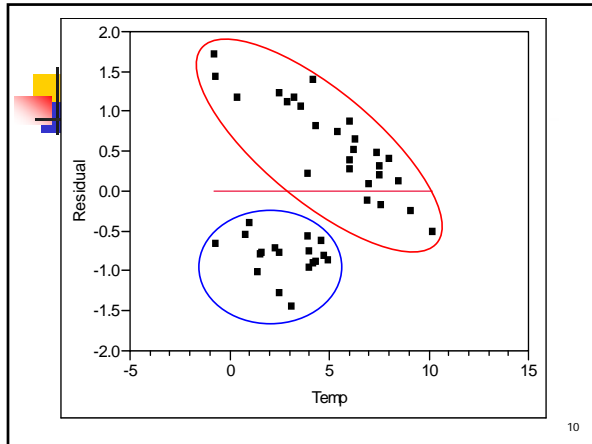
---

---

---

---

# Stat 401 B – Lecture 22



---

---

---

---

---

---

---

---

## Plot of Residuals

- The plot of residuals versus temperature does not appear to be a random scatter.
- There appears to be two groups of values.

11

---

---

---

---

---

---

---

---

## How can we do better?

- If the two groups in the residual plot are associated with data from the un-insulated and insulated house, adding the dummy (indicator) variable *Insul* can explain more of the variation in gas consumption.

12

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 22

 **Multiple Regression Model**

- Predicted Gas =  $6.72 - 0.368 \cdot \text{Temp} - 1.79 \cdot \text{Insul}$
- $R^2 = 0.919$ , 91.9% of the variation in gas consumption can be explained by the multiple regression model with outside temperature and the dummy variable.

13

---

---

---


---

---

---

---

---

 **Interpretation**

- For an un-insulated house ( $\text{Insul} = 0$ ) when the average outside temperature is  $0^\circ\text{C}$ , the predicted amount of gas used is 6.72 (1000 cubic feet).

14

---

---

---


---

---

---

---

---

 **Interpretation**

- Holding  $\text{Insul}$  constant, gas consumption drops, on average, 368 cubic feet for every  $1^\circ\text{C}$  increase in average outside temperature.

15

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 22

 **Interpretation**

- This model is a no interaction model because the relationship between gas and temperature has the same slope for both the un-insulated and insulated house.

16

---

---

---


---

---

---

---

---

 **Interpretation**

- Holding outside temperature constant, changing Insul from 0 to 1 (going from an un-insulated house to an insulated house), the amount of gas used drops 1.79 (1000 cubic feet), on average.

17

---

---

---


---

---

---

---

---

 **Interpretation**

- The dummy variable changes the level of the gas usage on a day when the outside temperature is 0 °C.

18

---

---

---

---

---

---

---

---

# Stat 401 B – Lecture 22

## Two regression lines

- Before insulation
  - Predicted Gas =  $6.72 - 0.368 \cdot \text{Temp}$
- After insulation
  - Predicted Gas =  $4.93 - 0.368 \cdot \text{Temp}$

19

---

---

---

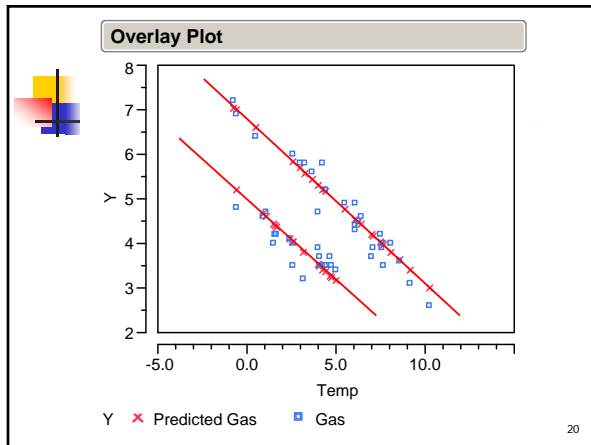
---

---

---

---

---



---

---

---

---

---

---

---

---

## Statistical Significance

- Model Utility
  - $F = 233.48$ ,  $P\text{-value} < 0.0001$
- The model with Temp and Insul is useful. The P-value for the test of model utility is very small.
- $RMSE = 0.299$

21

---

---

---

---


---

---

---

---

# Stat 401 B – Lecture 22

 **Statistical Significance**

- Temp
  - $t = -19.47$ , P-value  $< 0.0001$
- Because the P-value is small, Temp adds significantly to the model with Insul.

22

---

---

---


---

---

---

---

---

 **Statistical Significance**

- Insul
  - $t = -17.33$ , P-value  $< 0.0001$
- Because the P-value is small, Insul adds significantly to the

23

---

---

---

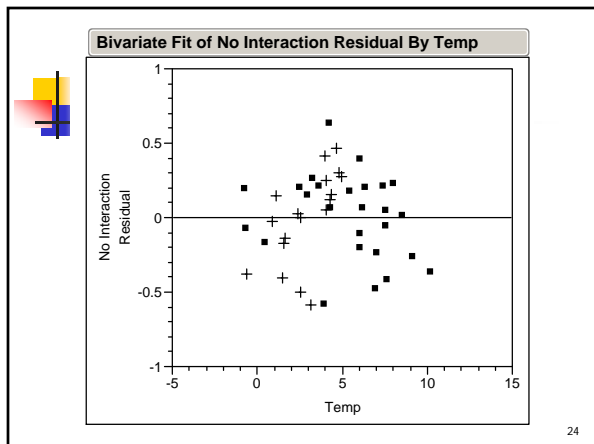
---

---

---

---

---



---

---

---

---

---

---

---

---