Natan Alper 2/10/2020

Business Intelligence & Consumer Insights- Professor Kovtun

HW #2

**1 (a)**

Lotsize, bed, bath, rooms, age

**1 (b)**

Yi = a + b1x1 + b2x2 + b3x3 + b4x4 + b5x5 + ei ei = epsilon

Value of Home = a + b1\*(Lotsize)i + b2\*(bed)i + b3\*(bath)i + b4\*(rooms)i + b5\*(age)i + ei

**1 (c)**

(Intercept) Lotsize Bed Bath Rooms Age

97.9262 1.5441 0.8288 25.6778 11.0958 -0.6812

**1 (d)**

Estimated Model 🡪 Value = 97.9262 + 1.5441\*(Lotsize)i + 0.8288\*(bed)i + 25.6778\*(bath)i + 11.0958\*(rooms)i + -0.6812\*(age)i + ei

**1 (e)**

Home 32 is the biggest bargain

Est value = 251.1827

Current value = 195.0

**2**

A screenshot of a cell phone

Description automatically generated

Rooms and Bath

A screenshot of a cell phone

Description automatically generated

Modern bath and modern kitchen

**3 (a)**

Estimated Model 🡪 Value = 79.6399 + 2.0248\*(Lotsize)i + -2.5764\*(bed)i + 30.1494\*(bath)i + 11.8863\*(rooms)i + -0.3057\*(age)i + ei

MSE 🡪 1081.305

CODE

HOMES\_Train <- HOMES[(1:60),]

HOMES\_Test <- HOMES[-(1:60),]

HOMES\_Train

HOMES\_Test

train\_model <- lm(Value ~ Lotsize + Bed + Bath + Rooms + Age, data = HOMES\_Train)

train\_model

preds <- predict(train\_model, newdata=HOMES\_Test[,-1]) ## to come up with predictions we don't include the Value column in the newdata argument!

#MSE

sum((HOMES\_Test$Value-preds)^2)/length(preds)

**3 (b)**

> # 3 (b) # Only checking combos of Lotsize, Bed, Bath and Rooms (not Age)

> # Lotsize only

> model <- lm(Value ~ Lotsize, data = HOMES\_Train)

> predictions <- predict(model, newdata = HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1937.404

>

> # Bed only

> model <- lm(Value ~ Bed, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1820.055

>

> # Bath only

> model <- lm(Value ~ Bath, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1296.692

>

> # Rooms only

> model <- lm(Value ~ Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1177.982

>

> # Lotsize + Bed

> model <- lm(Value ~ Lotsize + Bed, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1770.283

>

> # Lotsize + Bath

> model <- lm(Value ~ Lotsize + Bath, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1354.382

>

> # Lotsize + Rooms

> model <- lm(Value ~ Lotsize + Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1055.35

>

> # Bed + Bath

> model <- lm(Value ~ Bed + Bath, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1240.416

>

> # Bed + Rooms

> model <- lm(Value ~ Bed + Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1207.263

>

> # Bath + Rooms

> model <- lm(Value ~ Bath + Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1091.839

>

> # Lotsize + Bed + Bath

> model <- lm(Value ~ Lotsize + Bed + Bath, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1328.584

>

> # Lotsize + Bed + Rooms

> model <- lm(Value ~ Lotsize + Bed + Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1100.43

>

> # Lotsize + Bath + Rooms

> model <- lm(Value ~ Lotsize + Bath + Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1075.437

>

> # Bed + Bath + Rooms

> model <- lm(Value ~ Bed + Bath + Rooms, data=HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1119.319

>

> # Lotsize + Bed + Bath + Rooms

> model <- lm(Value ~ Lotsize + Bed + Bath + Rooms, data = HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1115.437

> ### Using 4 pred vars out of 5 (5 choose 4)

> # Lotsize + Bed + Bath + Rooms

> model <- lm(Value ~ Lotsize + Bed + Bath + Rooms, data = HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1115.437

>

> # Lotsize + Bed + Bath + Age

> model <- lm(Value ~ Lotsize + Bed + Bath + Age, data = HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1290.98

>

> # Lotsize + Bed + Rooms + Age

> model <- lm(Value ~ Lotsize + Bed + Rooms + Age, data = HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 964.688

>

> # Lotsize + Bath + Rooms + Age

> model <- lm(Value ~ Lotsize + Bath + Rooms + Age, data = HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1043.409

>

> # Bed + Bath + Rooms + Age

> model <- lm(Value ~ Bed + Bath + Rooms + Age, data = HOMES\_Train)

> predictions <- predict(model, newdata=HOMES\_Test[,-1])

> sum((HOMES\_Test$Value-predictions)^2)/length(predictions)

[1] 1062.182

**3 (c)**

# When using every combination of the same 4, out of 5

Lotsize + Rooms

# When using 4 out of 5

Lotsize + Bed + Rooms + Age