Lesson 04 Multiple Linear Regression Models

Lusine Zilfimian

March 09 (Monday), 2020

Contents

Contents

Contents

Last Lecture ReCap

- How can we estimate the coefficient of linear regression?

Last Lecture ReCap

- How can we estimate the coefficient of linear regression?
- What are TSS, ESS, RSS?

Last Lecture ReCap

- How can we estimate the coefficient of linear regression?
- What are TSS, ESS, RSS?
- What is the distribution of $\varepsilon, y, and\ \hat{\beta}_1^{OLS}$?

Last Lecture ReCap

- How can we estimate the coefficient of linear regression?
- What are TSS, ESS, RSS?
- What is the distribution of $\varepsilon, y, and\ \hat{\beta}_1^{OLS}$?
- How can we estimate the model accuracy?

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

- The total variance of the response variable:

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

- The total variance of the response variable:
-

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

- The total variance of the response variable:
- $$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- The estimated variance of the response variable:

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

- The total variance of the response variable:
- $$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- The estimated variance of the response variable:
- $$ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

- The total variance of the response variable:
- $$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- The estimated variance of the response variable:
- $$ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- $$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$$

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

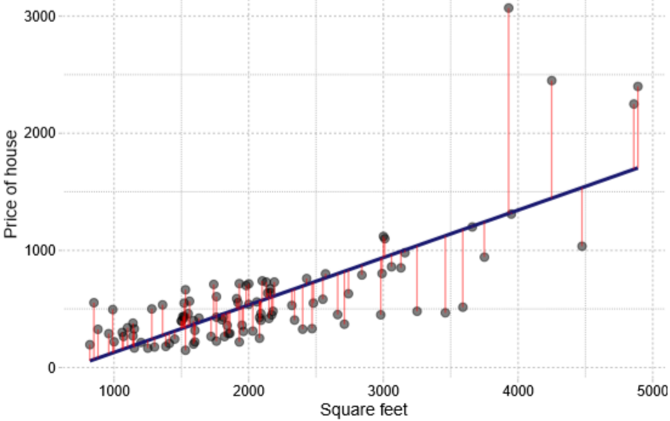
- The total variance of the response variable:
- $$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- The estimated variance of the response variable:
- $$ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- $$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$$
- By definition $0 \leq R^2 \leq 1$

R-squared

R^2 - the proportion of variability in Y that can be explained using X.

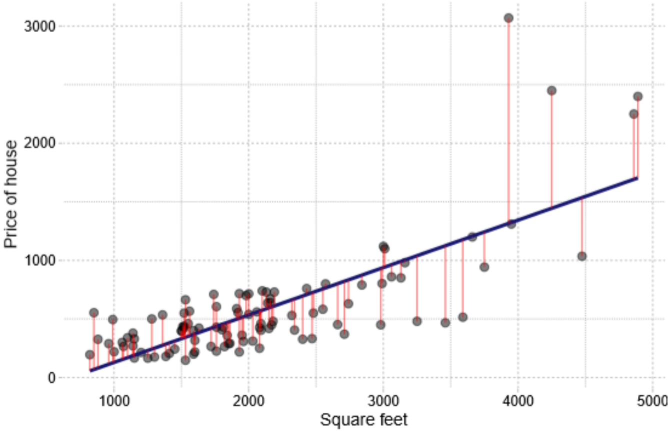
- The total variance of the response variable:
- $$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$
- The estimated variance of the response variable:
- $$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$
- $$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$$
- By definition $0 \leq R^2 \leq 1$
- What is a good R^2 value?

TSS, RSS, ESS



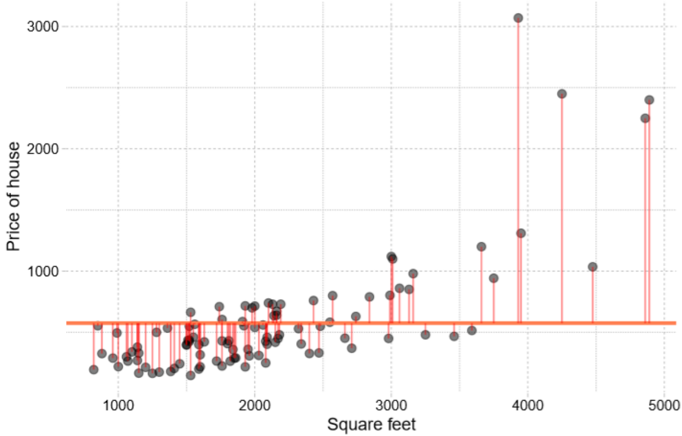
- This is . . .

TSS, RSS, ESS



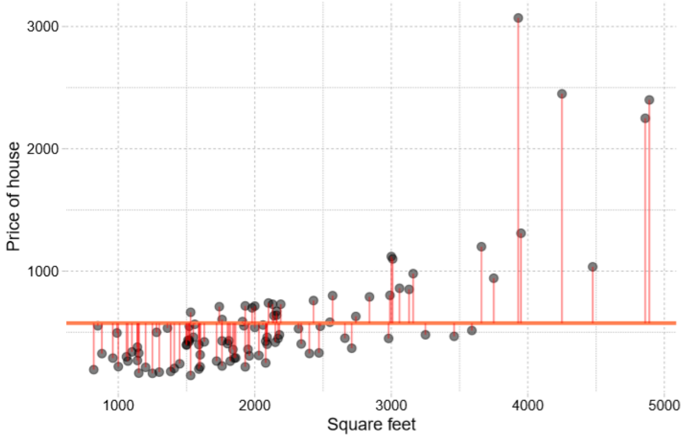
- This is ...
- Residual Sum of Squares

TSS, RSS, ESS



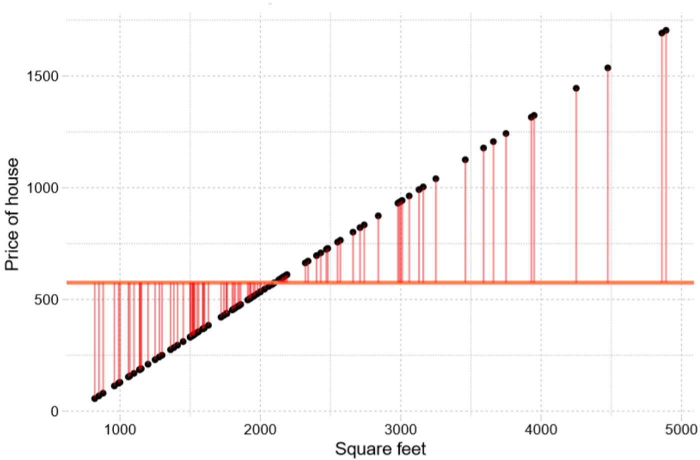
- This is ...

TSS, RSS, ESS



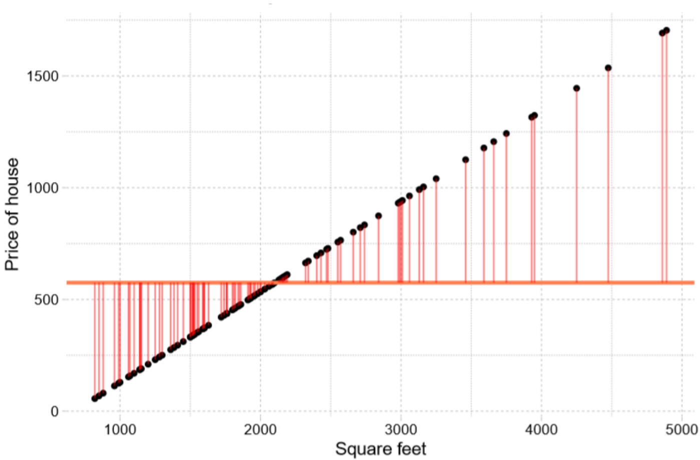
- This is ...
- Total Sum of Squares

TSS, RSS, ESS



- This is . . .

TSS, RSS, ESS



- This is . . .
- Estimated Sum of Squares

Multiple linear regression

Model specification:

- In practice we often have more than one predictor.

Multiple linear regression

Model specification:

- In practice we often have more than one predictor.
- $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + ... + \beta_k x_{ik} + \varepsilon_i$$

Multiple linear regression

Model specification:

- In practice we often have more than one predictor.
- $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + ... + \beta_k x_{ik} + \varepsilon_i$$
- $\mathbb{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \mathbb{X} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

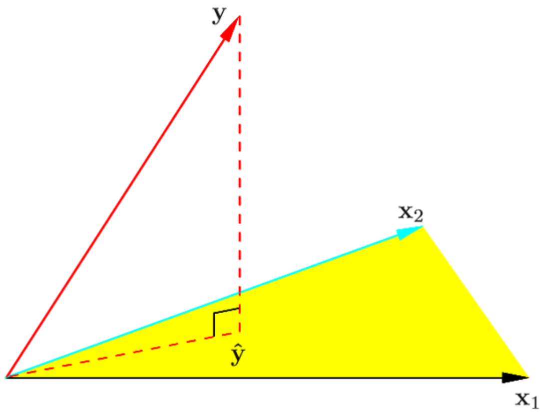
Multiple linear regression

Model specification:

- In practice we often have more than one predictor.
- $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + ... + \beta_k x_{ik} + \varepsilon_i$$
- $\mathbb{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \mathbb{X} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$
- $$\mathbb{Y} = \mathbb{X}\beta + \varepsilon$$

Geometry of least squares regression with two predictors

Source: T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*



Minimizing the RSS

- $$e^T e \to min$$

Minimizing the RSS

- $$e^T e \to min$$
- $$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$

Minimizing the RSS

- $$e^T e \to min$$
- $$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$
- $$\mathbb{E}(\hat{\beta}_{OLS}) = \beta$$

Minimizing the RSS

- $$e^T e \to min$$
- $$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$
- $$\mathbb{E}(\hat{\beta}_{OLS}) = \beta$$
- $$var(\hat{\beta}_{OLS}) = \sigma^2 (X^T X)^{-1}$$

Minimizing the RSS

- $$e^T e \rightarrow min$$
- $$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$
- $$\mathbb{E}(\hat{\beta}_{OLS}) = \beta$$
- $$var(\hat{\beta}_{OLS}) = \sigma^2 (X^T X)^{-1}$$
- $$\hat{\sigma^2} = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Minimizing the RSS

- $$e^T e \to min$$
- $$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$
- $$\mathbb{E}(\hat{\beta}_{OLS}) = \beta$$
- $$var(\hat{\beta}_{OLS}) = \sigma^2 (X^T X)^{-1}$$
- $$\hat{\sigma^2} = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
- **Task**: Derive all formulas by yourself.

Comparison of the models: Adjusted R^2

- The number of predictors $\uparrow \Rightarrow R^2 \uparrow$, even if those variables are only weakly associated with the response.

```
hd <- read.csv("housing.csv")[1:100,]
model1 <- lm(price ~ sqft_living + condition, data = hd)
summary(model1)$r.squared
```

```
## [1] 0.4847802
```

```
set.seed(2708)
model2 <- lm(price ~ sqft_living + condition +
    runif(dim(hd)[1]), hd)
summary(model2)$r.squared
```

```
## [1] 0.5225291
```

Comparison of the models: Adjusted R^2

- The number of predictors $\uparrow \Rightarrow R^2 \uparrow$, even if those variables are only weakly associated with the response.

```
hd <- read.csv("housing.csv")[1:100,]
model1 <- lm(price ~ sqft_living + condition, data = hd)
summary(model1)$r.squared
```

```
## [1] 0.4847802
```

```
set.seed(2708)
model2 <- lm(price ~ sqft_living + condition +
    runif(dim(hd)[1]), hd)
summary(model2)$r.squared
```

```
## [1] 0.5225291
```

- $R^2 \uparrow \not\Rightarrow$ the quality of model $\uparrow$

Comparison of the models: Adjusted R^2

- $R^2_{adj} = 1 - \frac{\frac{ESS}{n-k}}{\frac{TSS}{n-1}}$

Comparison of the models: Adjusted R^2

- $R^2_{adj} = 1 - \frac{\frac{ESS}{n-k}}{\frac{TSS}{n-1}}$
- The adjusted R^2 will only increase on adding a variable to the model if the addition of the variable reduces the residual mean squares.

Comparison of the models: Adjusted R^2

- $R^2_{adj} = 1 - \frac{\frac{ESS}{n-k}}{\frac{TSS}{n-1}}$
- The adjusted R^2 will only increase on adding a variable to the model if the addition of the variable reduces the residual mean squares.
- **Task:** Show that $R^2_{adj} = 1 - (1 - R^2)\dfrac{n-1}{n-k}$

Comparison of the models: Adjusted R^2

- $R^2_{adj} = 1 - \frac{\frac{ESS}{n-k}}{\frac{TSS}{n-1}}$
- The adjusted R^2 will only increase on adding a variable to the model if the addition of the variable reduces the residual mean squares.
- **Task:** Show that $R^2_{adj} = 1 - (1 - R^2)\dfrac{n-1}{n-k}$
- **Task:** Prove that $R^2 \geq R^2_{adj}, k > 1$

F test

- The hypotheses are

F test

- The hypotheses are
- $$H_0 : \beta_1 = \beta_2 = ... = \beta_n = 0, \; H_\alpha : \exists \; \beta_i \neq 0$$

F test

- The hypotheses are
- $$H_0 : \beta_1 = \beta_2 = ... = \beta_n = 0,\ H_\alpha : \exists\ \beta_i \neq 0$$
- $$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} \sim F_{k-1,n-k}$$

F test

- The hypotheses are
- $$H_0 : \beta_1 = \beta_2 = ... = \beta_n = 0,\ H_\alpha : \exists\ \beta_i \neq 0$$
- $$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} \sim F_{k-1,n-k}$$
- **Task:** Prove that for the single linear regression $F = t$.

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters
2. Random sampling

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters
2. Random sampling
3. $rank(\mathbb{X}) = m$ - absence of multicollinearity

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters
2. Random sampling
3. $rank(\mathbb{X}) = m$ - absence of multicollinearity
4. $\mathbb{E}(\varepsilon) = 0$ - expected zero error term

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters
2. Random sampling
3. $rank(\mathbb{X}) = m$ - absence of multicollinearity
4. $\mathbb{E}(\varepsilon) = 0$ - expected zero error term
5. $var(\varepsilon) = \sigma^2 I$ - homoskedasticity of errors

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters
2. Random sampling
3. $rank(\mathbb{X}) = m$ - absence of multicollinearity
4. $\mathbb{E}(\varepsilon) = 0$ - expected zero error term
5. $var(\varepsilon) = \sigma^2 I$ - homoskedasticity of errors
6. $cov(\varepsilon_i, \varepsilon_j) = 0$ - the absence of autocorrelation

The Theorem of Gauss-Markov: OLS is BLUE

If the linear regression model satisfies the following six classical assumptions:

1. True model is linear in parameters
2. Random sampling
3. $rank(\mathbb{X}) = m$ - absence of multicollinearity
4. $\mathbb{E}(\varepsilon) = 0$ - expected zero error term
5. $var(\varepsilon) = \sigma^2 I$ - homoskedasticity of errors
6. $cov(\varepsilon_i, \varepsilon_j) = 0$ - the absence of autocorrelation

- then ordinary least squares regression produces unbiased estimates that have the smallest variance of all possible linear estimations.

OLS is BLUE

1. Unbiased: $\mathbb{E}(\hat{\beta}_{OLS}) = \beta$

OLS is BLUE

1. Unbiased: $\mathbb{E}(\hat{\beta}_{OLS}) = \beta$
2. Minimum Variance: $var(\hat{\beta}_{OLS}) < var(\hat{\beta}_{other})$

OLS is BLUE

1. Unbiased: $\mathbb{E}(\hat{\beta}_{OLS}) = \beta$
2. Minimum Variance: $var(\hat{\beta}_{OLS}) < var(\hat{\beta}_{other})$
3. Consistent: with the increase of the number of observations the estimators converge to the true parameters

OLS is BLUE

1. Unbiased: $\mathbb{E}(\hat{\beta}_{OLS}) = \beta$
2. Minimum Variance: $var(\hat{\beta}_{OLS}) < var(\hat{\beta}_{other})$
3. Consistent: with the increase of the number of observations the estimators converge to the true parameters

- $$\varepsilon \sim \mathbb{N}(0, \sigma^2) \;\Rightarrow\; \hat{\beta}_{OLS} \sim \mathbb{N}(\beta, \sigma^2(X^T X)^{-1})$$

OLS is BLUE

1. Unbiased: $\mathbb{E}(\hat{\beta}_{OLS}) = \beta$
2. Minimum Variance: $var(\hat{\beta}_{OLS}) < var(\hat{\beta}_{other})$
3. Consistent: with the increase of the number of observations the estimators converge to the true parameters

- $$\varepsilon \sim \mathbb{N}(0, \sigma^2) \;\Rightarrow\; \hat{\beta}_{OLS} \sim \mathbb{N}(\beta, \sigma^2(X^TX)^{-1})$$
- We can apply statistical tests to them

Interpretations

Continuous predictors

Regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant.

- $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k$$

Be attentive. Gayane's question is appropriate.

Interpretations

Continuous predictors

Regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant.

- $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k$$

Be attentive. Gayane's question is appropriate.

- Holding $x_2, ..., x_k$ constant, 1 unit $\uparrow x_1 \Rightarrow \ \hat{\beta}_1 \ unit \ \uparrow y$

Interpretations

Continuous predictors

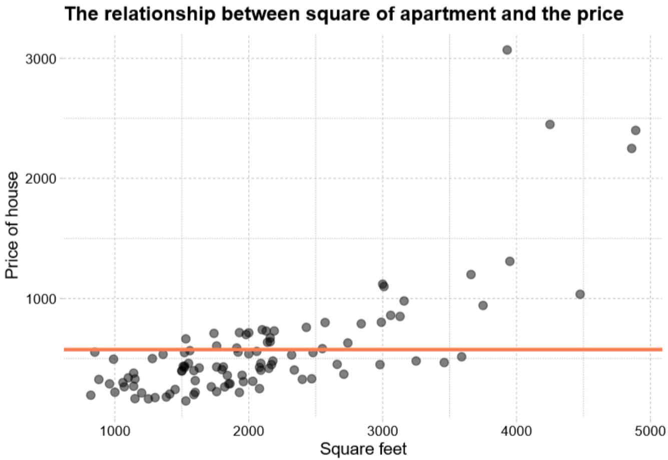
Regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant.

- $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k$$

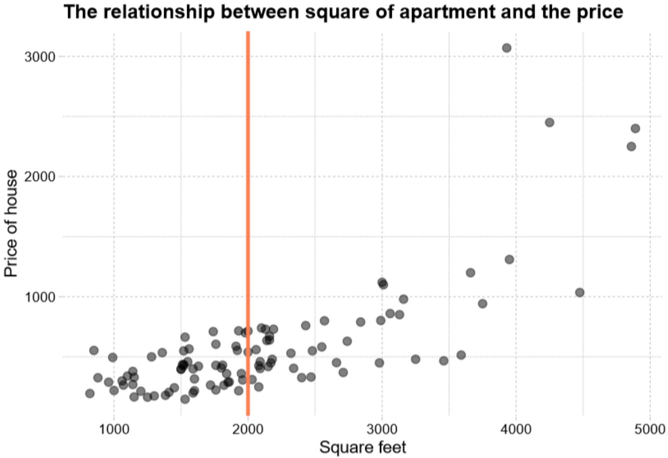
Be attentive. Gayane's question is appropriate.

- Holding $x_2, ..., x_k$ constant, 1 unit $\uparrow x_1 \Rightarrow \ \hat{\beta}_1 \ unit \ \uparrow y$
- **Intercept** - $\hat{\beta}_0$ is the average value of y when $x = 0$.

Intercept-only model



Mean change in the response variable



Predictors with only two levels

- Dummy variables

Predictors with only two levels

- Dummy variables
- $x_i = \begin{cases} 1, & \textit{if } i^{th} \textit{ observation has the property}, \\ 0, & \textit{if } i^{th} \textit{ observation does not have the property} \end{cases}$

Predictors with only two levels

- Dummy variables
- $x_i = \begin{cases} 1, & \textit{if } i^{th} \textit{ observation has the property}, \\ 0, & \textit{if } i^{th} \textit{ observation does not have the property} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$

Predictors with only two levels

- Dummy variables
- $x_i = \begin{cases} 1, & \textit{if } i^{th} \textit{ observation has the property}, \\ 0, & \textit{if } i^{th} \textit{ observation does not have the property} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$
- β_0 - the average of y among observations which do not have the property

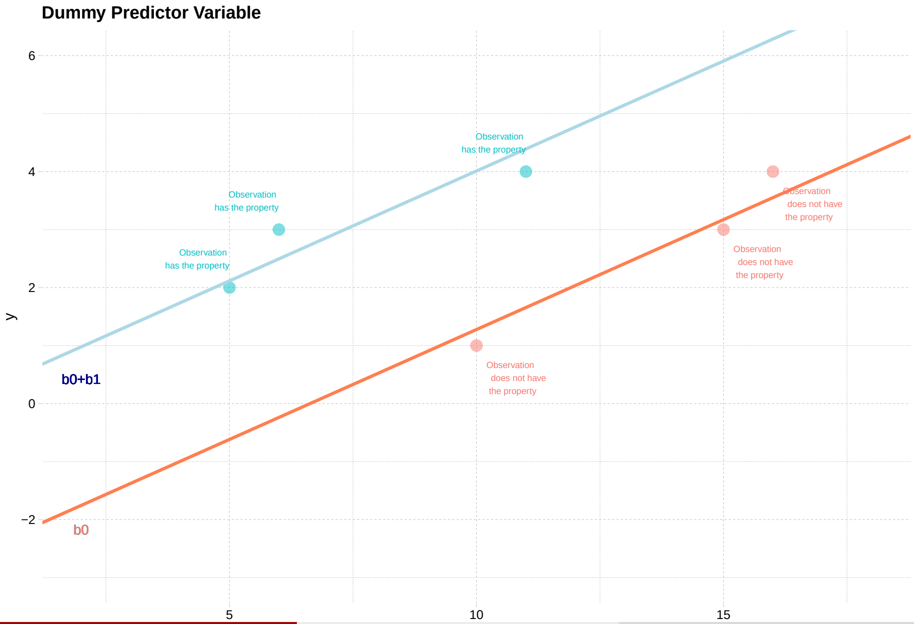
Predictors with only two levels

- Dummy variables
- $x_i = \begin{cases} 1, & \textit{if } i^{th} \textit{ observation has the property}, \\ 0, & \textit{if } i^{th} \textit{ observation does not have the property} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$
- β_0 - the average of y among observations which do not have the property
- $\beta_0 + \beta_1$ - the average of y among observations which have the property

Predictors with only two levels

- Dummy variables
- $x_i = \begin{cases} 1, & \textit{if } i^{th} \textit{ observation has the property}, \\ 0, & \textit{if } i^{th} \textit{ observation does not have the property} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$
- β_0 - the average of y among observations which do not have the property
- $\beta_0 + \beta_1$ - the average of y among observations which have the property
- What about another way of coding?

Predictors with only two levels



Predictors with more that two levels

- Dummy variables

Predictors with more that two levels

- Dummy variables
- $x_{i1} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from AUA}, \\ 0, & \textit{if } i^{th} \textit{ student is not from AUA} \end{cases}$

Predictors with more that two levels

- Dummy variables
- $x_{i1} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from AUA}, \\ 0, & \textit{if } i^{th} \textit{ student is not from AUA} \end{cases}$
- $x_{i2} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from YSU}, \\ 0, & \textit{if } i^{th} \textit{ student is not from YSU} \end{cases}$

Predictors with more that two levels

- Dummy variables
- $x_{i1} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from AUA}, \\ 0, & \textit{if } i^{th} \textit{ student is not from AUA} \end{cases}$
- $x_{i2} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from YSU}, \\ 0, & \textit{if } i^{th} \textit{ student is not from YSU} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \beta_2 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$

Predictors with more that two levels

- Dummy variables
- $x_{i1} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from AUA}, \\ 0, & \textit{if } i^{th} \textit{ student is not from AUA} \end{cases}$
- $x_{i2} = \begin{cases} 1, & \textit{if } i^{th} \textit{ student is from YSU}, \\ 0, & \textit{if } i^{th} \textit{ student is not from YSU} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \beta_2 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$
- β_0 - the average of y among students from universities apart from YSU and AUA

Predictors with more that two levels

- Dummy variables
- $x_{i1} = \begin{cases} 1, & \text{if } i^{th} \text{ student is from } AUA, \\ 0, & \text{if } i^{th} \text{ student is not from } AUA \end{cases}$
- $x_{i2} = \begin{cases} 1, & \text{if } i^{th} \text{ student is from } YSU, \\ 0, & \text{if } i^{th} \text{ student is not from } YSU \end{cases}$
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \beta_2 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$
- β_0 - the average of y among students from universities apart from YSU and AUA
- $\beta_0 + \beta_1$ - the average of y among students from AUA

Predictors with more that two levels

- Dummy variables
- $x_{i1} = \begin{cases} 1, & \text{if } i^{th} \text{ student is from AUA}, \\ 0, & \text{if } i^{th} \text{ student is not from AUA} \end{cases}$
- $x_{i2} = \begin{cases} 1, & \text{if } i^{th} \text{ student is from YSU}, \\ 0, & \text{if } i^{th} \text{ student is not from YSU} \end{cases}$
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i \\ \beta_0 + \beta_2 + \varepsilon_i \\ \beta_0 + \varepsilon_i \end{cases}$
- β_0 - the average of y among students from universities apart from YSU and AUA
- $\beta_0 + \beta_1$ - the average of y among students from AUA
- $\beta_0 + \beta_2$ - the average of y among students from YSU

Some aspects of Linear Regression

Additivity

- The additive assumption means that the effect of changes in any predictor on the response is independent of the values of the other predictors.

Some aspects of Linear Regression

Additivity

- The additive assumption means that the effect of changes in any predictor on the response is independent of the values of the other predictors.
- Interaction term: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i$

Some aspects of Linear Regression

Additivity

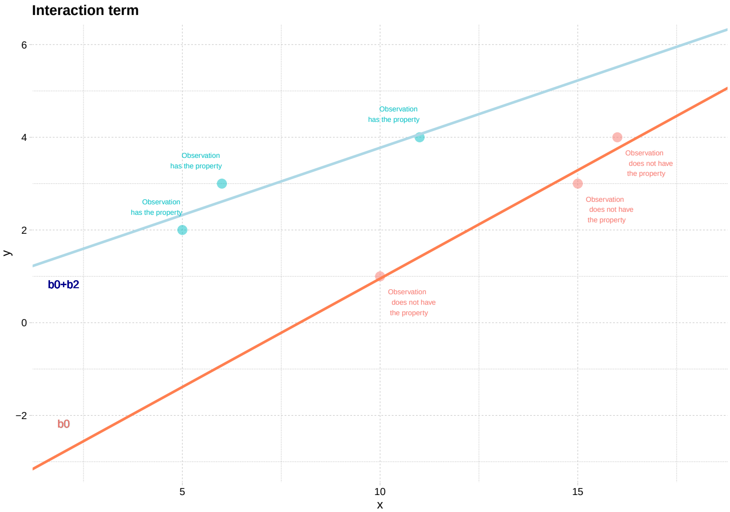
- The additive assumption means that the effect of changes in any predictor on the response is independent of the values of the other predictors.
- Interaction term: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i$
- An interaction between a qualitative variable and a quantitative variable:

Some aspects of Linear Regression

Additivity

- The additive assumption means that the effect of changes in any predictor on the response is independent of the values of the other predictors.
- Interaction term: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \varepsilon_i$
- An interaction between a qualitative variable and a quantitative variable:
- $y_i = \beta_0 + \beta_1 x_{i1} + \begin{cases} \beta_2 + \beta_3 x_{i1} + \varepsilon_i \\ 0 + \varepsilon_i \end{cases}$

- Slope for blue line is $b_1 + b_3$



- Slope for blue line is $b_1 + b_3$
- Slope for red line is b_1