

## Week 1

1. In Ordinary Least Squares we are trying to minimize:
  - a. The Median Absolute Error
  - b. The Mean Absolute Error
  - c. The Sum of Squared Error
  - d. The Root Mean Squared Error
2. The following formula is which of the following:

$$\sum_i (\bar{y} - \hat{y}_i)^2$$

- a. Sum of Squared Total
- b. Sum of Squared Errors
- c. Sum of Squared Regression
- d. Total Sum of Squares

## Week 2

1. An indicator variable is sometimes referred to as a:
  - a. Hyperparameter
  - b. Dummy variable
  - c. Interaction variable
  - d. Continuous predictor variable
2. Which of the following is an example of an Interaction Term?
  - a. Height\*Gender
  - b.  $Height^2$
  - c. Height - Gender
  - d. Height + Gender
3. If your dataset has a column that contained one of three states for each observation ('New York', 'California', 'Hawaii'), what would be the best way to code them for a regression model with one-hot encoding (not using a factor variable in R)?

a.

Obs	States
1	'New York'
2	'California'
3	'Hawaii'

b.

Obs	States
1	1
2	2
3	3

c.

Obs	'New York'	'California'
1	1	0
2	0	1
3	0	0

d.

Obs	'New York'	'California'	'Hawaii'
1	1	0	0
2	0	1	0
3	0	0	1

4. If our regression model is as follows:

$\text{Nursing\_Home\_Cost\_Per\_Day} = B_0 + B_1 * \text{Age} + B_2 * \text{'New York'} + B_3 * \text{'California'}$

And the base case is 'Hawaii'

After training the model the coefficients are as follows

$B_0 = 200$

$B_1 = 1.5$

$B_2 = -15$

$B_3 = -5$

How much more would a person in California expect to pay per day vs a person in New York of the same age?

- \$5
  - \$-5
  - \$10
  - \$-15
5. From the above scenario, what would a 70-year-old that lives in Hawaii expect to pay per week?
- \$290
  - \$300
  - \$305
  - \$2,135
6. In a linear regression model with one qualitative (categorical) predicting variable with 4 values, we need to include 4 dummy variables.
- True
  - False

	Estimate	S.E.	t Value	Pr> t
Intercept	-6.12	4.72	-1.30	0.20
Salary	.002	.00009	25	<.001
AgeMid	-4.81	6.39	-0.75	0.45
AgeOld	23.28	6.72	3.46	<.001

7. •  $AmountSpent = b_0 + b_1*Salary + b_2*AgeMid + b_3*AgeOld$
- Based on the following regression model summary (Note: the base case is Age = Young), what is the Amount Spent by a Middle-aged customer if his/her salary is 10000?
- 20 – 6.12
  - 20 – 6.12 - 4.81
  - 20 – 6.12 + 23.28
  - 20
8. An interaction term is used to model how the synergies between multiple variables impact the response variable.
- True
  - False

## Answers

### Week1

Question 1: c The Sum of Squared Error (page 18 slide 1)

Question 2: c Sum of Squared Regression (page 18 slide 2)

### Week2

Question 1: b. Dummy variable (page 9 slide 2)

Question 2: a. Height\*Weight (page 18 slide 1)

Question 3: c (page 24 slide 2)

Question 4: c (Page 27 slide 2)

Question 5: d (Page 27 slide 2) \$305\*7

Question 6: b (page 20 slide 2)

Question 7: b (Lesson 2 / Video 4 / Slides 1 – 4)

Question 8: a (Lesson 2 / Video 4 / Slides 6 – 9)

