# Linear Regression Models

## Segment 3 – Other Considerations in the MLRM

## Topic 3 – Interaction between Covariates

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

# Topics



1. Interaction: Basic Ideas
2. Interaction Between Two Continuous Predictors
3. Interaction Between Two Categorical Predictors
4. Interaction Between Continuous & Categorical Predictors

# Interaction: Basic Ideas

# Interaction: Basic Ideas

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.

# Interaction: Basic Ideas

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors;

# Interaction: Basic Ideas

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors; the predictor $newConstruction$ is categorical with two levels: (1) No (2) Yes.

# Interaction: Basic Ideas

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors; the predictor $newConstruction$ is categorical with two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is

# Interaction: Basic Ideas

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors; the predictor $newConstruction$ is categorical with two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes + \hat{\beta}_3 \times livingArea \times newConstructionYes.$

# Interaction: Basic Ideas

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors; the predictor $newConstruction$ is categorical with two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes + \hat{\beta}_3 \times livingArea \times newConstructionYes.$
- Interaction to be considered

# Interaction: Basic Ideas

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors; the predictor $newConstruction$ is categorical with two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes + \hat{\beta}_3 \times livingArea \times newConstructionYes.$
- Interaction to be considered (1) if a particular predictor has a large effect on the response (large coefficient estimate)

# Interaction: Basic Ideas

- An interaction occurs when the effect of a predictor on the response variable depends on the value of another predictor.
- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors; the predictor $newConstruction$ is categorical with two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes + \hat{\beta}_3 \times livingArea \times newConstructionYes.$
- Interaction to be considered (1) if a particular predictor has a large effect on the response (large coefficient estimate) and/or (2) the presence of categorical predictors as coefficients of other predictors may vary across groups.

# Interaction Between Two Continuous Predictors

# Interaction Between Two Continuous Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $rooms$ as the predictors.

# Interaction Between Two Continuous Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $rooms$ as the predictors.
- An MLRM with interaction between $livingArea$ and $rooms$ is

# Interaction Between Two Continuous Predictors

- <u>Example</u>: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $rooms$ as the predictors.
- An MLRM with interaction between $livingArea$ and $rooms$ is
  $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times rooms + \hat{\beta}_3 \times livingArea \times rooms.$

# Interaction Between Two Continuous Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $rooms$ as the predictors.
- An MLRM with interaction between $livingArea$ and $rooms$ is $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times rooms + \hat{\beta}_3 \times livingArea \times rooms.$
- How to interpret the coefficient estimates?

# Interaction Between Two Continuous Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $rooms$ as the predictors.
- An MLRM with interaction between $livingArea$ and $rooms$ is $\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times rooms + \hat{\beta}_3 \times livingArea \times rooms.$
- How to interpret the coefficient estimates? Suppose we increase the living area by 1 unit while keeping the number of rooms fixed.

# Interaction Between Two Continuous Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $rooms$ as the predictors.
- An MLRM with interaction between $livingArea$ and $rooms$ is
$$\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times rooms + \hat{\beta}_3 \times livingArea \times rooms.$$
- How to interpret the coefficient estimates? Suppose we increase the living area by 1 unit while keeping the number of rooms fixed.

$$\underbrace{[\hat{\beta}_0 + \hat{\beta}_1 \times (livingArea + 1) + \hat{\beta}_2 \times rooms + \hat{\beta}_3 \times (livingArea + 1) \times rooms]}_{\widehat{price}_{\text{new}}}$$
$$-\underbrace{[\hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times rooms + \hat{\beta}_3 \times livingArea \times rooms]}_{\widehat{price}_{\text{old}}}$$
$$= \hat{\beta}_1 + \hat{\beta}_3 \times rooms.$$

# Interaction Between Two Categorical Predictors

# Interaction Between Two Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $heating$ and $centralAir$ as the predictors.

# Interaction Between Two Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $heating$ and $centralAir$ as the predictors.
- The predictor $heating$ has three levels: (1) electric (2) hot air (3) hot water/steam.

# Interaction Between Two Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $heating$ and $centralAir$ as the predictors.
- The predictor $heating$ has three levels: (1) electric (2) hot air (3) hot water/steam.
- The predictor $centralAir$ has two levels: (1) No (2) Yes.

# Interaction Between Two Categorical Predictors

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $heating$ and $centralAir$ as the predictors.
- The predictor $heating$ has three levels: (1) electric (2) hot air (3) hot water/steam.
- The predictor $centralAir$ has two levels: (1) No (2) Yes.
- An MLRM with interaction between $heating$ and $centralAir$ is

# Interaction Between Two Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $heating$ and $centralAir$ as the predictors.
- The predictor $heating$ has three levels: (1) electric (2) hot air (3) hot water/steam.
- The predictor $centralAir$ has two levels: (1) No (2) Yes.
- An MLRM with interaction between $heating$ and $centralAir$ is

$$\begin{aligned}
\widehat{price} = {} & \hat{\beta}_0 + \hat{\beta}_1 \times heatinghotair + \hat{\beta}_2 \times heatinghot\ water \\
& + \hat{\beta}_3 \times centralAirYes \\
& + \hat{\beta}_4 \times heatinghotair \times centralAirYes \\
& + \hat{\beta}_5 \times heatinghot\ water/steam \times centralAirYes.
\end{aligned}$$

# Interaction Between Two Categorical Predictors: Continued

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price

$$= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price

$$= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$

- Subtracting

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price
$$= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$
- Subtracting $\Rightarrow \hat{\beta}_3 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among electric-heated houses.

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price
$$= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$
- Subtracting $\Rightarrow \hat{\beta}_3 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among electric-heated houses.
- Suppose we have a house that is hot air heated and not centrally air-conditioned;

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price
$$= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$
- Subtracting $\Rightarrow \hat{\beta}_3 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among electric-heated houses.
- Suppose we have a house that is hot air heated and not centrally air-conditioned; the predicted price of the house is $\hat{\beta}_0 + \hat{\beta}_1$.

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price
$$= \hat{\beta}_0 + \hat{\beta}_3 \times \color{magenta}{centralAirYes} = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$
- Subtracting $\Rightarrow \hat{\beta}_3 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among electric-heated houses.
- Suppose we have a house that is hot air heated and not centrally air-conditioned; the predicted price of the house is $\hat{\beta}_0 + \hat{\beta}_1$.
- Now consider another house that is hot air heated and centrally air-conditioned;

# Interaction Between Two Categorical Predictors: Continued

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- For electric-heated houses, predicted price $= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$
- Subtracting $\Rightarrow \hat{\beta}_3 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among electric-heated houses.
- Suppose we have a house that is hot air heated and not centrally air-conditioned; the predicted price of the house is $\hat{\beta}_0 + \hat{\beta}_1$.
- Now consider another house that is hot air heated and centrally air-conditioned; the predicted price of this house is $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_4$.

# Interaction Between Two Categorical Predictors: Continued

- For electric-heated houses, predicted price

$$= \hat{\beta}_0 + \hat{\beta}_3 \times centralAirYes = \begin{cases} \hat{\beta}_0 & \text{if not centrally air-conditioned} \\ \hat{\beta}_0 + \hat{\beta}_3 & \text{if centrally air-conditioned.} \end{cases}$$

- Subtracting $\Rightarrow \hat{\beta}_3 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among electric-heated houses.
- Suppose we have a house that is hot air heated and not centrally air-conditioned; the predicted price of the house is $\hat{\beta}_0 + \hat{\beta}_1$.
- Now consider another house that is hot air heated and centrally air-conditioned; the predicted price of this house is $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_4$.
- Subtracting the two results above, we see that $\hat{\beta}_3 + \hat{\beta}_4 =$ difference between average prices of centrally air-conditioned and not centrally air-conditioned houses among hot air-heated houses.

# Interaction Between Continuous & Categorical Predictors

# Interaction Between Continuous & Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors.

# Interaction Between Continuous & Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors.
- The predictor $newConstruction$ has two levels: (1) No (2) Yes.

# Interaction Between Continuous & Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors.
- The predictor $newConstruction$ has two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is

# Interaction Between Continuous & Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors.
- The predictor $newConstruction$ has two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is

$$\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes$$
$$+ \hat{\beta}_3 \times livingArea \times newConstructionYes.$$

# Interaction Between Continuous & Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors.
- The predictor $newConstruction$ has two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is

$$\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes$$
$$+ \hat{\beta}_3 \times livingArea \times newConstructionYes.$$

- The predicted house price

$$\widehat{price} = \begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \times livingArea & \text{if old house,} \\ \left(\hat{\beta}_0 + \hat{\beta}_2\right) + \left(\hat{\beta}_1 + \hat{\beta}_3\right) \times livingArea & \text{if new house.} \end{cases}$$

# Interaction Between Continuous & Categorical Predictors

- Example: suppose that we consider the *SaratogaHouses* dataset with $price$ as the response and $livingArea$ and $newConstruction$ as the predictors.
- The predictor $newConstruction$ has two levels: (1) No (2) Yes.
- An MLRM with interaction between $livingArea$ and $newConstruction$ is

$$\widehat{price} = \hat{\beta}_0 + \hat{\beta}_1 \times livingArea + \hat{\beta}_2 \times newConstructionYes + \hat{\beta}_3 \times livingArea \times newConstructionYes.$$

- The predicted house price

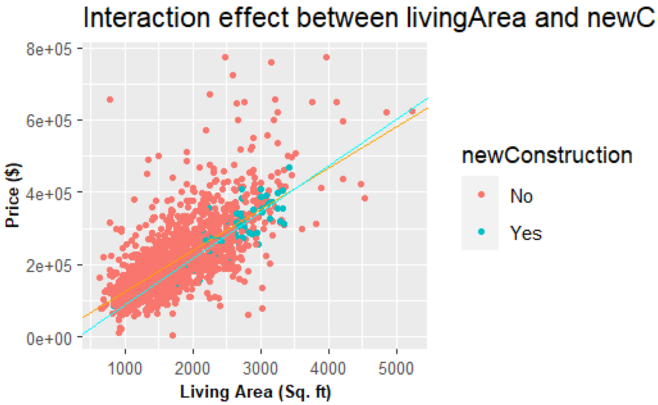
$$\widehat{price} = \begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \times livingArea & \text{if old house,} \\ \left(\hat{\beta}_0 + \hat{\beta}_2\right) + \left(\hat{\beta}_1 + \hat{\beta}_3\right) \times livingArea & \text{if new house.} \end{cases}$$

- Note the differences in both intercept and slope for new houses.

# Interaction Between Continuous & Categorical Predictors: Continued

# Interaction Between Continuous & Categorical Predictors: Continued

The scatter plot indicates that a higher slope is needed for new houses:



# Summary

# Summary

- Describe the effects of interaction between predictors.

# Summary

- Describe the effects of interaction between predictors.
- Construct linear regression models with interacting predictors