Linear Regression Models

Segment 1 – Simple Linear Regression Model

Topic 6 – Feature Engineering: Transforming Data

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

Topics

1. Centering
2. Standardizing
3. Logarithmic Transformation

Centering

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1:

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values:

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1^{(i)}$.

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1^{(i)}$.
- The intercept estimate $\hat{\beta}_0$ can now be interpreted as approximately the average response value for an average predictor input.

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1^{(i)}$.
- The intercept estimate $\hat{\beta}_0$ can now be interpreted as approximately the average response value for an average predictor input.
- The response values can also be centered:

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1^{(i)}$.
- The intercept estimate $\hat{\beta}_0$ can now be interpreted as approximately the average response value for an average predictor input.
- The response values can also be centered: $\tilde{y}^{(i)} = y^{(i)} - \bar{y}_n$.

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1^{(i)}$.
- The intercept estimate $\hat{\beta}_0$ can now be interpreted as approximately the average response value for an average predictor input.
- The response values can also be centered: $\tilde{y}^{(i)} = y^{(i)} - \bar{y}_n$.
- The resulting SLRM will have zero intercept:

Centering

- Suppose we have a dataset with n samples and build an SLRM to predict the response Y from a single predictor X_1: $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- We can make the coefficient estimates more interpretable by centering the predictor values: $\tilde{x}_1^{(i)} = x_1^{(i)} - \bar{x}_n$.
- The resulting SLRM is $\hat{y}^{(i)} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_1^{(i)}$.
- The intercept estimate $\hat{\beta}_0$ can now be interpreted as approximately the average response value for an average predictor input.
- The response values can also be centered: $\tilde{y}^{(i)} = y^{(i)} - \bar{y}_n$.
- The resulting SLRM will have zero intercept: $\hat{\tilde{y}}^{(i)} = \hat{\beta}_1 \tilde{x}_1^{(i)}$.

Standardizing

Standardizing

- Sometimes, it is also helpful to standardize the predictor:

Standardizing

- Sometimes, it is also helpful to standardize the predictor:
  $\tilde{x}_1^{(i)} = \frac{x_1^{(i)} - \bar{x}_n}{\hat{\sigma}_{x_1}}$,

Standardizing

- Sometimes, it is also helpful to standardize the predictor: $\tilde{x}_1^{(i)} = \frac{x_1^{(i)} - \bar{x}_n}{\hat{\sigma}_{x_1}}$, where $\hat{\sigma}_{x_1}$ is the sample standard deviation of the predictor.

Standardizing

- Sometimes, it is also helpful to standardize the predictor: $\tilde{x}_1^{(i)} = \frac{x_1^{(i)} - \bar{x}_n}{\hat{\sigma}_{x_1}}$, where $\hat{\sigma}_{x_1}$ is the sample standard deviation of the predictor.
- This is helpful typically in the multiple linear regression setup where different scales may be present in the data.

Logarithmic Transformation

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values:

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- This can be seen as a *multiplicative* model by exponentiating:

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- This can be seen as a *multiplicative* model by exponentiating: $\hat{y}^{(i)} = e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 x_1^{(i)}}$.

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- This can be seen as a *multiplicative* model by exponentiating: $\hat{y}^{(i)} = e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 x_1^{(i)}}$.
- What is the interpretation of the estimate $\hat{\beta}_1$ now?

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- This can be seen as a *multiplicative* model by exponentiating: $\hat{y}^{(i)} = e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 x_1^{(i)}}$.
- What is the interpretation of the estimate $\hat{\beta}_1$ now?
- Suppose there is a 1 unit increase in the predictor value x_1:

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- This can be seen as a *multiplicative* model by exponentiating: $\hat{y}^{(i)} = e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 x_1^{(i)}}$.
- What is the interpretation of the estimate $\hat{\beta}_1$ now?
- Suppose there is a 1 unit increase in the predictor value x_1: $\frac{\hat{y}_{\text{new}}}{\hat{y}_{\text{old}}} = e^{\hat{\beta}_1} \approx 1 + \hat{\beta}_1$ for small $\hat{\beta}_1$.

Logarithmic Transformation

- We may have response variables such as height, weight etc., which cannot be predicted to be negative values using an SLRM.
- One way to overcome that issue is to build an SLRM for logarithmically transformed response values: $\log\left(\hat{y}^{(i)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(i)}$.
- This can be seen as a *multiplicative* model by exponentiating: $\hat{y}^{(i)} = e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 x_1^{(i)}}$.
- What is the interpretation of the estimate $\hat{\beta}_1$ now?
- Suppose there is a 1 unit increase in the predictor value x_1: $\frac{\hat{y}_{\text{new}}}{\hat{y}_{\text{old}}} = e^{\hat{\beta}_1} \approx 1 + \hat{\beta}_1$ for small $\hat{\beta}_1$.
- This means, $\hat{\beta}_1$ is the proportionate change in the response value for a unit increase in the predictor value.

Summary

Summary

- Transform data for capturing meanigful relationship using centering. standardizing, and logarithmic transformation.