Linear Regression Models

Segment 1 – Simple Linear Regression Model

Topic 2 – Simple Linear Regression Model (SLRM) and its Assumptions

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

Topics

1. The Geometric Idea Behind Simple Linear Regression Model (SLRM)

2. Population & Sample - Revisited in the Context of SLRM

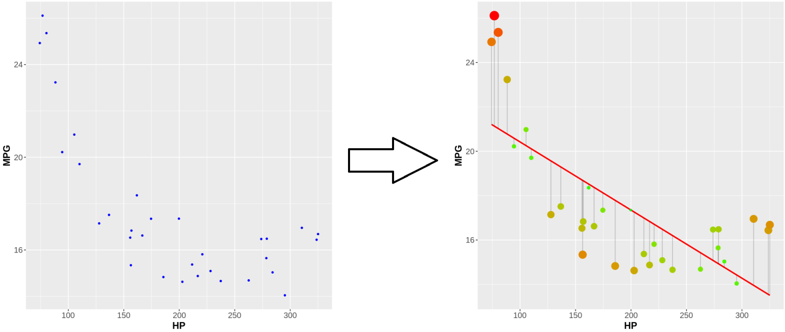
3. Ordinary Least Squares Estimation for an SLRM

4. Assumptions in SLRM

The Geometric Idea Behind Simple Linear Regression Model (SLRM)

The Geometric Idea Behind Simple Linear Regression Model (SLRM)

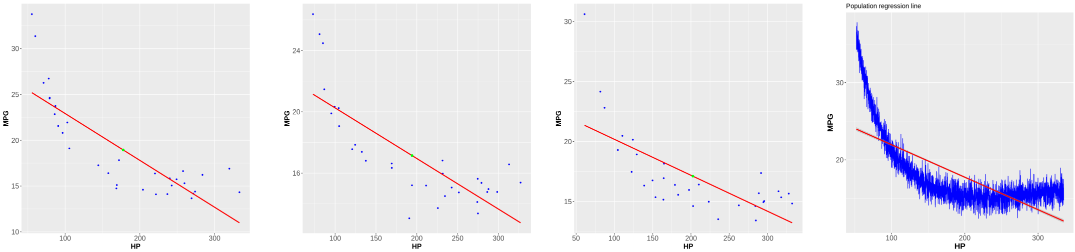
Given a dataset (*random samples from the population*), find a straight line that *fits* the data (*response variable and a single predictor*) well in an *average sense*:



Population & Sample - Revisited in the Context of SLRM

Population & Sample - Revisited in the Context of SLRM

Note that the straight line of best fit will depend on the dataset but there is <u>only one</u> unique straight line of best fit for the entire population data:



Ordinary Least Squares Estimation for an SLRM

Ordinary Least Squares Estimation for an SLRM

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.

Ordinary Least Squares Estimation for an SLRM

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

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.
- The true relationship (*real population model*) is $Y = f(X_1) + \varepsilon$,

Ordinary Least Squares Estimation for an SLRM

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.
- The true relationship (*real population model*) is $Y = f(X_1) + \varepsilon$, where f is an unknown and a possibly nonlinear function,

Ordinary Least Squares Estimation for an SLRM

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.
- The true relationship (*real population model*) is $Y = f(X_1) + \varepsilon$, where f is an unknown and a possibly nonlinear function, and ϵ is a random error term capturing the effects of noise inherent in the data, predictors that we may have missed, and other effects.

Ordinary Least Squares Estimation for an SLRM

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.
- The true relationship (*real population model*) is $Y = f(X_1) + \varepsilon$, where f is an unknown and a possibly nonlinear function, and ϵ is a random error term capturing the effects of noise inherent in the data, predictors that we may have missed, and other effects.
- In SLRM, we model the true population relationship using a linear function:

Ordinary Least Squares Estimation for an SLRM

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.
- The true relationship (*real population model*) is $Y = f(X_1) + \varepsilon$, where f is an unknown and a possibly nonlinear function, and ϵ is a random error term capturing the effects of noise inherent in the data, predictors that we may have missed, and other effects.
- In SLRM, we model the true population relationship using a linear function: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$.

Ordinary Least Squares Estimation for an SLRM

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

- Suppose Y is the response variable and we are interested in studying its relationship with a single predictor X_1.
- The true relationship (*real population model*) is $Y = f(X_1) + \varepsilon$, where f is an unknown and a possibly nonlinear function, and ϵ is a random error term capturing the effects of noise inherent in the data, predictors that we may have missed, and other effects.
- In SLRM, we model the true population relationship using a linear function: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$.
- Note that in the SLRM above, we use the same symbol ε for the random error term which now additionally includes the effect of missing out a possibly nonlinear relationship between Y and X_1.

Ordinary Least Squares Estimation for an SLRM - Continued

Ordinary Least Squares Estimation for an SLRM - Continued

- The SLRM model predicts Y as an approximation $\hat{Y} = \beta_0 + \beta_1 X_1$.

Ordinary Least Squares Estimation for an SLRM - Continued

- The SLRM model predicts Y as an approximation $\hat{Y} = \beta_0 + \beta_1 X_1$.
- The prediction error, also referred to as residual, is $R = Y - \hat{Y} = Y - (\beta_0 + \beta_1 X_1)$.

Ordinary Least Squares Estimation for an SLRM - Continued

- The SLRM model predicts Y as an approximation $\hat{Y} = \beta_0 + \beta_1 X_1$.
- The prediction error, also referred to as residual, is
  $R = Y - \hat{Y} = Y - (\beta_0 + \beta_1 X_1)$.
- The goal now is to compute *estimates* for the population parameters β_0 and β_1.

Ordinary Least Squares Estimation for an SLRM - Continued

- The SLRM model predicts Y as an approximation $\hat{Y} = \beta_0 + \beta_1 X_1$.
- The prediction error, also referred to as residual, is $R = Y - \hat{Y} = Y - (\beta_0 + \beta_1 X_1)$.
- The goal now is to compute *estimates* for the population parameters β_0 and β_1.
- To that end, we use a randomly sampled dataset comprising n samples with the ith sample's predictor and response values denoted as $x_1^{(i)}$ and $y^{(i)}$, respectively.

Ordinary Least Squares Estimation for an SLRM - Continued

- The SLRM model predicts Y as an approximation $\hat{Y} = \beta_0 + \beta_1 X_1$.
- The prediction error, also referred to as residual, is $R = Y - \hat{Y} = Y - (\beta_0 + \beta_1 X_1)$.
- The goal now is to compute *estimates* for the population parameters β_0 and β_1.
- To that end, we use a randomly sampled dataset comprising n samples with the ith sample's predictor and response values denoted as $x_1^{(i)}$ and $y^{(i)}$, respectively.
- The ordinary least squares estimates of β_0 and β_1 is obtained by minimizing the sum of the squares of the residuals (RSS) for all samples in the dataset:

Ordinary Least Squares Estimation for an SLRM - Continued

- The SLRM model predicts Y as an approximation $\hat{Y} = \beta_0 + \beta_1 X_1$.
- The prediction error, also referred to as residual, is $R = Y - \hat{Y} = Y - (\beta_0 + \beta_1 X_1)$.
- The goal now is to compute *estimates* for the population parameters β_0 and β_1.
- To that end, we use a randomly sampled dataset comprising n samples with the ith sample's predictor and response values denoted as $x_1^{(i)}$ and $y^{(i)}$, respectively.
- The ordinary least squares estimates of β_0 and β_1 is obtained by minimizing the sum of the squares of the residuals (RSS) for all samples in the dataset:

$$\min \sum_{i=1}^{n} \left(r^{(i)}\right)^2 = \sum_{i=1}^{n} \left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)}\right)\right)^2.$$

Ordinary Least Squares Estimation for an SLRM - Continued

Ordinary Least Squares Estimation for an SLRM - Continued

- We minimize the RSS by calculating its partial derivative w.r.t. β_0 and β_1, and set them equal to zero:

$$\begin{cases} \dfrac{\partial(\text{RSS})}{\partial \beta_0} = 0 \Rightarrow -2\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)}\right)\right) = 0, \\ \dfrac{\partial(\text{RSS})}{\partial \beta_1} = 0 \Rightarrow -2\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)}\right)\right) x_1^{(i)} = 0. \end{cases}$$

Ordinary Least Squares Estimation for an SLRM - Continued

- We minimize the RSS by calculating its partial derivative w.r.t. β_0 and β_1, and set them equal to zero:

$$\begin{cases} \dfrac{\partial(\text{RSS})}{\partial\beta_0} & = 0 \Rightarrow -2\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)}\right)\right) = 0, \\ \dfrac{\partial(\text{RSS})}{\partial\beta_1} & = 0 \Rightarrow -2\sum_{i=1}^{n}\left(y^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)}\right)\right) x_1^{(i)} = 0. \end{cases}$$

- Solving this results in the estimates

$$\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n,$$
$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}},$$

Ordinary Least Squares Estimation for an SLRM - Continued

where

$$s_{xy} = \underbrace{\sum_{i=1}^{n} \left(x_1^{(i)} - \bar{x}_n\right)\left(y^{(i)} - \bar{y}_n\right)}_{\text{sample covariance-like measure}}$$

$$s_{xx} = \underbrace{\sum_{i=1}^{n} \left(x_1^{(i)} - \bar{x}_n\right)^2}_{\text{sample variance-like measure in the predictor}},$$

$$\bar{x}_n = \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_1^{(i)}}_{\text{sample mean of predictors}} \quad \text{and } \bar{y}_n = \underbrace{\frac{1}{n}\sum_{i=1}^{n} y^{(i)}}_{\text{sample mean of responses}}.$$

Assumptions in SLRM

Assumptions in SLRM

- For a random ith sample, note that in the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, the random error term $\varepsilon^{(i)}$ for the ith sample is the same as its residual $R^{(i)}$.

Assumptions in SLRM

- For a random ith sample, note that in the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, the random error term $\varepsilon^{(i)}$ for the ith sample is the same as its residual $R^{(i)}$.
- For deriving the ordinary least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, no assumptions about $\varepsilon^{(i)}$ are needed.

Assumptions in SLRM

- For a random ith sample, note that in the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, the random error term $\varepsilon^{(i)}$ for the ith sample is the same as its residual $R^{(i)}$.
- For deriving the ordinary least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, no assumptions about $\varepsilon^{(i)}$ are needed.
- For the purpose of deriving statistical inferences (mean, variance etc.) about the least square estimates, we will assume that $\varepsilon^{(i)}$ will have zero mean, constant variance, and uncorrelated across the samples that will be chosen from the population.

Assumptions in SLRM

- For a random ith sample, note that in the linear approximation $Y^{(i)} = \beta_0 + \beta_1 X_1^{(i)} + \varepsilon^{(i)}$, the random error term $\varepsilon^{(i)}$ for the ith sample is the same as its residual $R^{(i)}$.
- For deriving the ordinary least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, no assumptions about $\varepsilon^{(i)}$ are needed.
- For the purpose of deriving statistical inferences (mean, variance etc.) about the least square estimates, we will assume that $\varepsilon^{(i)}$ will have zero mean, constant variance, and uncorrelated across the samples that will be chosen from the population.
- Later, for the purpose of constructing hypotheses tests and confidence intervals for the least squares estimates, we will also assume that $\varepsilon^{(i)}$ is normally distributed.