AUA CS108, Statistics, Fall 2020

Lecture 43

Michael Poghosyan

07 Dec 2020

Contents

Linear Regression

Intro to Linear Regression

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our aim was, given a Data, to get an information about the Distribution behind that data.

Intro to Linear Regression

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our aim was, given a Data, to get an information about the Distribution behind that data.

In particular, in our Parametric Statistics setting, we were assuming that our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

Intro to Linear Regression

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our aim was, given a Data, to get an information about the Distribution behind that data.

In particular, in our Parametric Statistics setting, we were assuming that our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

- to find a good Point Estimator and Estimate;

Intro to Linear Regression

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our aim was, given a Data, to get an information about the Distribution behind that data.

In particular, in our Parametric Statistics setting, we were assuming that our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

- to find a good Point Estimator and Estimate;
- to find a CI for θ of given CL;

Intro to Linear Regression

So far, except some exceptions $\ddot\smile$, we have considered inference about one r.v. and the Distribution behind: our aim was, given a Data, to get an information about the Distribution behind that data.

In particular, in our Parametric Statistics setting, we were assuming that our r.v. was $X \sim \mathcal{F}_\theta$, and our aim was, given a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta,$$

to get some information about θ, in particular,

- to find a good Point Estimator and Estimate;
- to find a CI for θ of given CL;
- to Test a Hypothesis about θ, say, is it likely that $\theta = 3.1415926$ or not.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Recall that, in the Descriptive Statistics part, we considered two Datasets, and defined the Sample Covariance and Correlation Coefficients, to measure the Linear Relationship between that Datasets.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Recall that, in the Descriptive Statistics part, we considered two Datasets, and defined the Sample Covariance and Correlation Coefficients, to measure the Linear Relationship between that Datasets. That was defined for two **Numerical Dataset**, without any assumptions behind the Process generating that Datasets.

Intro to Linear Regression

Now we want to talk about Modeling and inference for two (or more) Dependent Random variables, to talk about Modeling the Relationship between two or more Random Variables.

Recall that, in the Descriptive Statistics part, we considered two Datasets, and defined the Sample Covariance and Correlation Coefficients, to measure the Linear Relationship between that Datasets. That was defined for two **Numerical Dataset**, without any assumptions behind the Process generating that Datasets. Now, if we assume that that Datasets are coming from some Distribution, we are at the stage of doing a Statistical Inference, Statistical Analysis.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible. Even if possible, usually, interpolating curve is not the best descriptor of the relationship between r.v.s.

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible. Even if possible, usually, interpolating curve is not the best descriptor of the relationship between r.v.s.

Note: The coordinates of $\mathbf{X}$ are called **Explanatory, Predictor, Independent Variables or Features**,

Intro to Linear Regression

We assume that we have a pair of r.v.s $(\mathbf{X}, Y)$, and let $\mathcal{F}$ be their Joint Distribution:

$$(\mathbf{X}, Y) \sim \mathcal{F}.$$

Here $\mathbf{X}$ can be 1D r.v. or a d-D Random Vector. Our aim is, having some observations from that Distribution,

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n),$$

to get some information about $\mathcal{F}$. And, knowing about $\mathcal{F}$, we will know info about the relationship between $\mathbf{X}$ and Y.

Note: In fact, in the Probabilistic setting, we can have same $\mathbf{x}_k$-s with different y_k-s, so finding an interpolating curve will not be always possible. Even if possible, usually, interpolating curve is not the best descriptor of the relationship between r.v.s.

Note: The coordinates of $\mathbf{X}$ are called **Explanatory, Predictor, Independent Variables or Features**, and Y is called the **Dependent, Response Variable or the Label**.

Examples

Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS.

Examples

Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS. Our $\mathbf{X} = MWHSS$, is 1D, and $Y = STG$.

Examples

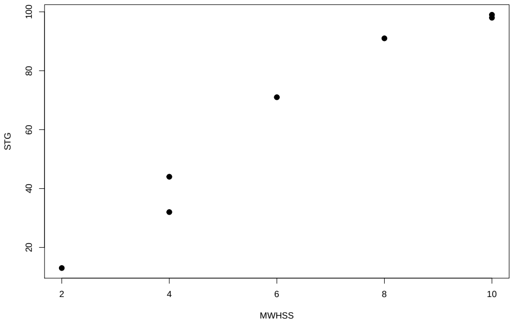
Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS. Our $\mathbf{X} = MWHSS$, is 1D, and $Y = STG$.To estimate the relationship, we need some data, say, we can have:

$$(10, 98);\ (6, 71);\ (10, 99);\ (4, 44);\ (2, 13);\ (8; 91);\ (4, 32).$$

Examples

Example: We want to find a relationship between the Stat Total Grade, STG, and the Mean Weekly Hours Spent on Statistics (during the Semester), MWHSS. Our $\mathbf{X} = MWHSS$, is 1D, and $Y = STG$.To estimate the relationship, we need some data, say, we can have:

$$(10, 98);\ (6, 71);\ (10, 99);\ (4, 44);\ (2, 13);\ (8; 91);\ (4, 32).$$



Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, “An Introduction to Statistical Learning”) Say, we want to get a relationship between Advertisement (AD) and Sales (S).

Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S). Here $\mathbf{AD} = (TV, Radio, Newspaper, SocialMedia)$ is 4D r.v. and S is 1D. And we have the following Data:

$$(10K\$, 5K\$, 0K\$, 30K\$, 1M\$); (7K\$, 15K\$, 2K\$, 0K\$, 31K\$), \ldots$$

Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S). Here $\mathbf{AD} = (TV, Radio, Newspaper, SocialMedia)$ is 4D r.v. and S is 1D. And we have the following Data:

$$(10K\$, 5K\$, 0K\$, 30K\$, 1M\$); (7K\$, 15K\$, 2K\$, 0K\$, 31K\$), \ldots$$

Example: (from Kaggle, Link) We want to predict the House Price by having some Features of the House (say, Floor, Zone, YearBuilt, RoofStyle etc). So we want to model the relationship between the House Price and all other variables.

Examples

Example: (similar to the example from the book of James, Witten, Hastie, Tibshirani, "An Introduction to Statistical Learning") Say, we want to get a relationship between Advertisement (AD) and Sales (S). Here $\mathbf{AD} = (TV, Radio, Newspaper, SocialMedia)$ is 4D r.v. and S is 1D. And we have the following Data:

$$(10K\$, 5K\$, 0K\$, 30K\$, 1M\$); (7K\$, 15K\$, 2K\$, 0K\$, 31K\$), ...$$

Example: (from Kaggle, Link) We want to predict the House Price by having some Features of the House (say, Floor, Zone, YearBuilt, RoofStyle etc). So we want to model the relationship between the House Price and all other variables. The Number of explanatory variables is 79. So here $\mathbf{X}$ is 79-D.

Intro to Linear Regression

Example: Say, given the height of a person, we want to predict the weight of a person.

Intro to Linear Regression

Example: Say, given the height of a person, we want to predict the weight of a person. Of course, here we are not seeking for an exact mathematical relationship, of the form

$$Weight = g(Height).$$

Intro to Linear Regression

Example: Say, given the height of a person, we want to predict the weight of a person. Of course, here we are not seeking for an exact mathematical relationship, of the form

$$Weight = g(Height).$$

Instead, we want to find a function g such that

$$Weight = g(Height) + random\ term.$$

Intro to Linear Regression

Example: Say, given the height of a person, we want to predict the weight of a person. Of course, here we are not seeking for an exact mathematical relationship, of the form

$$Weight = g(Height).$$

Instead, we want to find a function g such that

$$Weight = g(Height) + random\ term.$$

Say, see Wiki page for BMI.

Intro to Linear Regression

The simplest Regression Model is the Linear Model: in the Linear Regression Problem, we assume

$$Y = \beta_0 + \beta^T \cdot \mathbf{X} + \varepsilon,$$

where ε is a r.v., for each value of $\mathbf{X}$, with $\mathbb{E}(\varepsilon) = 0$.

Note: For each value of X, ε can be a different r.v.!

Intro to Linear Regression

The simplest Regression Model is the Linear Model: in the Linear Regression Problem, we assume

$$Y = \beta_0 + \beta^T \cdot \mathbf{X} + \varepsilon,$$

where ε is a r.v., for each value of $\mathbf{X}$, with $\mathbb{E}(\varepsilon) = 0$.

Note: For each value of X, ε can be a different r.v.!

Here, if $\mathbf{X} = (X^1, ..., X^d)$ is d-Dim, then

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_d \end{bmatrix},$$

Intro to Linear Regression

The simplest Regression Model is the Linear Model: in the Linear Regression Problem, we assume

$$Y = \beta_0 + \beta^T \cdot \mathbf{X} + \varepsilon,$$

where ε is a r.v., for each value of $\mathbf{X}$, with $\mathbb{E}(\varepsilon) = 0$.

Note: For each value of X, ε can be a different r.v.!

Here, if $\mathbf{X} = (X^1, ..., X^d)$ is d-Dim, then

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_d \end{bmatrix},$$

so our Model, in the expanded way, is

$$Y = \beta_0 + \beta_1 \cdot X^1 + \beta_2 \cdot X^2 + ... + \beta_d \cdot X^d + \varepsilon.$$

Simple Linear Regression

Now we simplify further the story: we consider the 1D Explanatory Variable case, i.e., $\mathbf{X} = X$ is 1D.

Simple Linear Regression

Now we simplify further the story: we consider the 1D Explanatory Variable case, i.e., $\mathbf{X} = X$ is 1D. In that case, we will have our LR Model:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0.$$

Simple Linear Regression

Now we simplify further the story: we consider the 1D Explanatory Variable case, i.e., $\mathbf{X} = X$ is 1D. In that case, we will have our LR Model:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0.$$

This is called the **Simple Linear Regression Model**.

Example

Example:

Example

Example: (a Generative Model, how to think the Data is generated)

Example

Example: (a Generative Model, how to think the Data is generated)
Assume we have the actual Relationship between X and Y:

$$Y = -1 + 3 \cdot X + \varepsilon$$

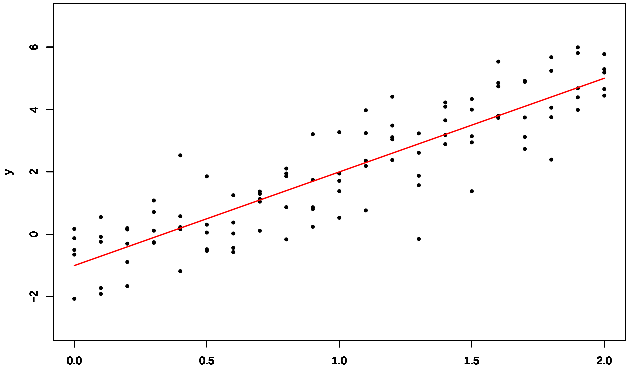
where $X \in [0, 2]$ and for each value of X, $\varepsilon \sim \mathcal{N}(0, 1)$, and for different values of X, ε-s are Independent.

Example

Example: (a Generative Model, how to think the Data is generated)
Assume we have the actual Relationship between X and Y:

$$Y = -1 + 3 \cdot X + \varepsilon$$

where $X \in [0, 2]$ and for each value of X, $\varepsilon \sim \mathcal{N}(0, 1)$, and for different values of X, ε-s are Independent.



Simple Linear Regression

Again, let us assume we have a Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0.$$

Simple Linear Regression

Again, let us assume we have a Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0.$$

The Problem is the Following: we have a Dataset of pairs:

$$(x_1, y_1), ..., (x_n, y_n),$$

and we want to Estimate β_0 and β_1.

Simple Linear Regression

Again, let us assume we have a Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0.$$

The Problem is the Following: we have a Dataset of pairs:

$$(x_1, y_1), ..., (x_n, y_n),$$

and we want to Estimate β_0 and β_1.

Of course, as we have done it many-many times, to generalize, we Model this by a Random Sample (Y_k-s are Random, but not X_k-s)

$$(X_1, Y_1), ..., (X_n, Y_n),$$

coming from a Model

$$Y_k = \beta_0 + \beta_1 \cdot X_k + \varepsilon_k, \qquad \mathbb{E}(\varepsilon_k) = 0$$

Simple Linear Regression

Again, let us assume we have a Simple Linear Regression Model:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon, \qquad \mathbb{E}(\varepsilon) = 0.$$

The Problem is the Following: we have a Dataset of pairs:

$$(x_1, y_1), ..., (x_n, y_n),$$

and we want to Estimate β_0 and β_1.

Of course, as we have done it many-many times, to generalize, we Model this by a Random Sample (Y_k-s are Random, but not X_k-s)

$$(X_1, Y_1), ..., (X_n, Y_n),$$

coming from a Model

$$Y_k = \beta_0 + \beta_1 \cdot X_k + \varepsilon_k, \qquad \mathbb{E}(\varepsilon_k) = 0$$

where ε_k-s are Independent, and our aim is to find good Estimators for β_0 and β_1.

Simple Linear Regression

Now, the idea of the Ordinary Least Squares Method for Estimating the Parameters β_0, β_1 is the following: Find

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{argmin} \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k \Big)^2 .$$

Simple Linear Regression

Now, the idea of the Ordinary Least Squares Method for Estimating the Parameters β_0, β_1 is the following: Find

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{argmin} \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k\Big)^2.$$

Here, the solution $(\hat{\beta}_0, \hat{\beta}_1)$ will be a pair of r.v.s, since Y_k-s are r.v.s. So we will obtain *Estimators* for β_0 and β_1.

Simple Linear Regression

Now, the idea of the Ordinary Least Squares Method for Estimating the Parameters β_0, β_1 is the following: Find

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{argmin} \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k \Big)^2.$$

Here, the solution $(\hat{\beta}_0, \hat{\beta}_1)$ will be a pair of r.v.s, since Y_k-s are r.v.s. So we will obtain *Estimators* for β_0 and β_1. If we will have an Observation (y_k, x_k), $k = 1, ..., n$, then we will solve

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{argmin} \sum_{k=1}^{n} \Big(y_k - \beta_0 - \beta_1 \cdot x_k \Big)^2,$$

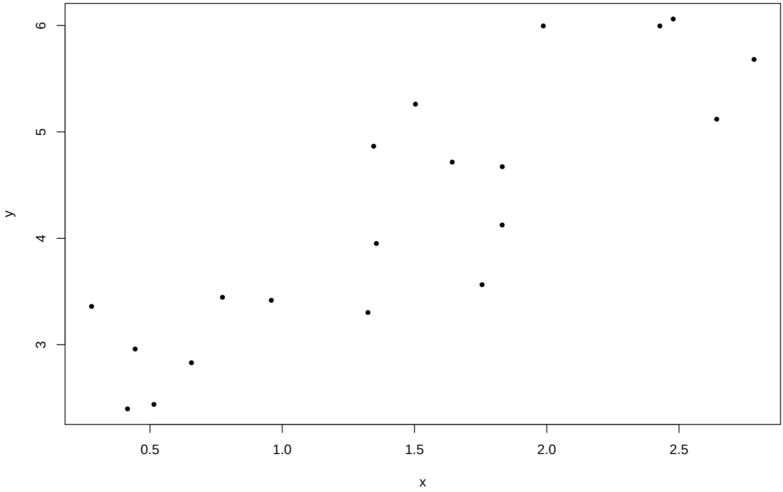
and we will find *Estimates* for β_0 and β_1.

Simple Linear Regression

Geometrically, the Problem is to find the “best fit line”:

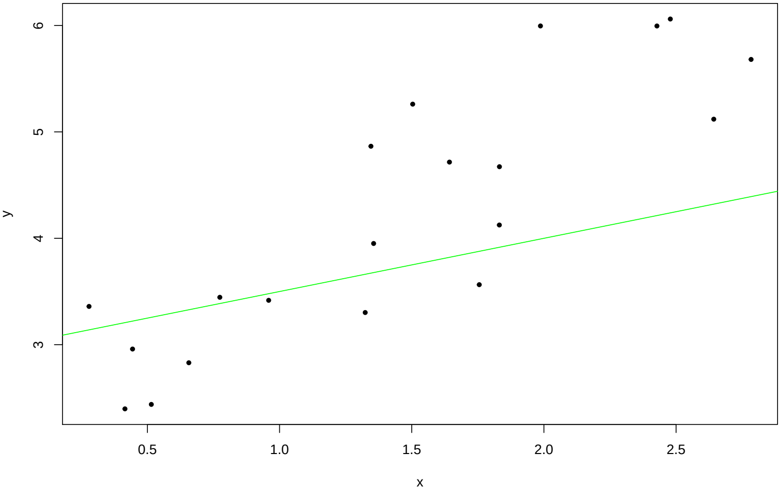
Simple Linear Regression

Geometrically, the Problem is to find the “best fit line”:



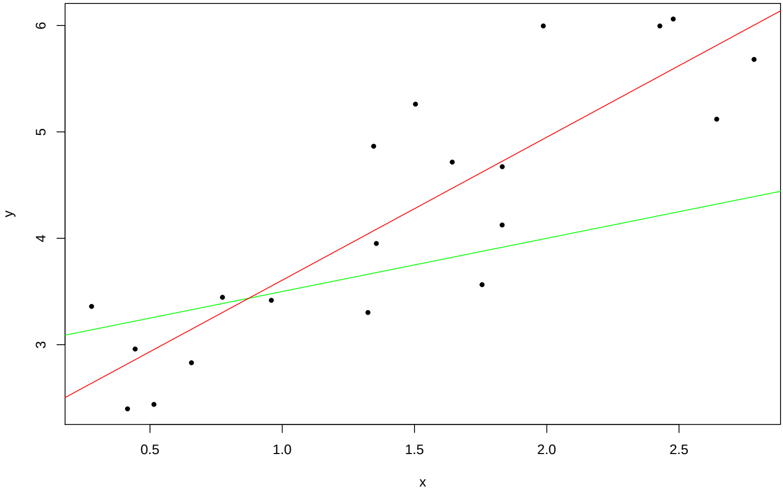
Simple Linear Regression

Geometrically, the Problem is to find the “best fit line”:



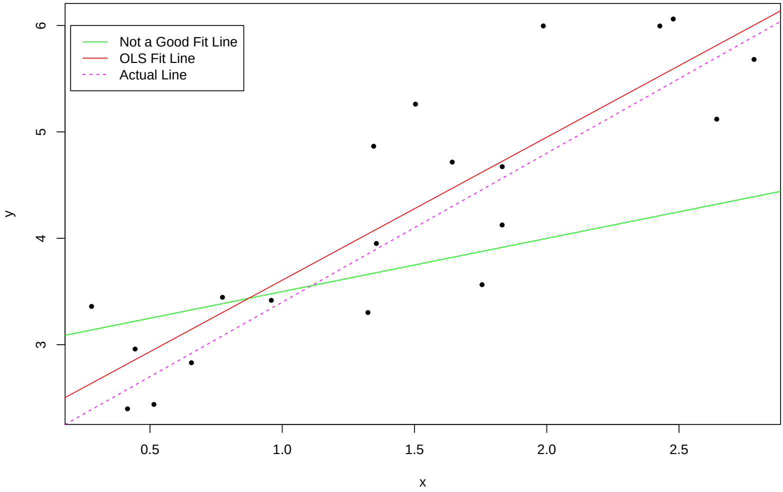
Simple Linear Regression

Geometrically, the Problem is to find the "best fit line":



Simple Linear Regression

Geometrically, the Problem is to find the “best fit line”:



Simple Linear Regression

Calculation of the best fit line is easy: we just define

$$\varphi(\beta_0, \beta_1) = \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k \Big)^2, \qquad (\beta_0, \beta_1) \in \mathbb{R}^2,$$

and using our Calc 3 knowledge, find the Minimum Point of φ by solving the System

$$\begin{cases} \dfrac{\partial \varphi}{\partial \beta_0} = 0 \\ \dfrac{\partial \varphi}{\partial \beta_1} = 0 \end{cases}$$

Simple Linear Regression

Calculation of the best fit line is easy: we just define

$$\varphi(\beta_0, \beta_1) = \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k \Big)^2, \qquad (\beta_0, \beta_1) \in \mathbb{R}^2,$$

and using our Calc 3 knowledge, find the Minimum Point of φ by solving the System

$$\begin{cases} \dfrac{\partial \varphi}{\partial \beta_0} = 0 \\ \dfrac{\partial \varphi}{\partial \beta_1} = 0 \end{cases}$$

The Solution is:

$$\hat{\beta}_1 = \frac{\sum_{k=1}^{n} (X_k - \overline{X})(Y_k - \overline{Y})}{\sum_{k=1}^{n} (X_k - \overline{X})^2}$$

Simple Linear Regression

Calculation of the best fit line is easy: we just define

$$\varphi(\beta_0, \beta_1) = \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k \Big)^2, \qquad (\beta_0, \beta_1) \in \mathbb{R}^2,$$

and using our Calc 3 knowledge, find the Minimum Point of φ by solving the System

$$\begin{cases} \dfrac{\partial \varphi}{\partial \beta_0} = 0 \\ \dfrac{\partial \varphi}{\partial \beta_1} = 0 \end{cases}$$

The Solution is:

$$\hat{\beta}_1 = \frac{\sum_{k=1}^{n} (X_k - \overline{X})(Y_k - \overline{Y})}{\sum_{k=1}^{n} (X_k - \overline{X})^2} = cor(X, Y) \cdot \frac{sd(Y)}{sd(X)} = \rho_{XY} \cdot \frac{S_Y}{S_X}$$

Simple Linear Regression

Calculation of the best fit line is easy: we just define

$$\varphi(\beta_0, \beta_1) = \sum_{k=1}^{n} \Big(Y_k - \beta_0 - \beta_1 \cdot X_k \Big)^2, \qquad (\beta_0, \beta_1) \in \mathbb{R}^2,$$

and using our Calc 3 knowledge, find the Minimum Point of φ by solving the System

$$\begin{cases} \dfrac{\partial \varphi}{\partial \beta_0} = 0 \\ \dfrac{\partial \varphi}{\partial \beta_1} = 0 \end{cases}$$

The Solution is:

$$\hat{\beta}_1 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{\sum_{k=1}^{n}(X_k - \overline{X})^2} = cor(X, Y) \cdot \frac{sd(Y)}{sd(X)} = \rho_{XY} \cdot \frac{S_Y}{S_X}$$

and

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \cdot \overline{X}.$$

Simple Linear Regression

The obtained Line

$$y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$$

is called the **Regression Line**.

Simple Linear Regression

The obtained Line

$$y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$$

is called the **Regression Line**.

Note: So the $cor(X, Y)$ is not the Slope of the Regression Line, but

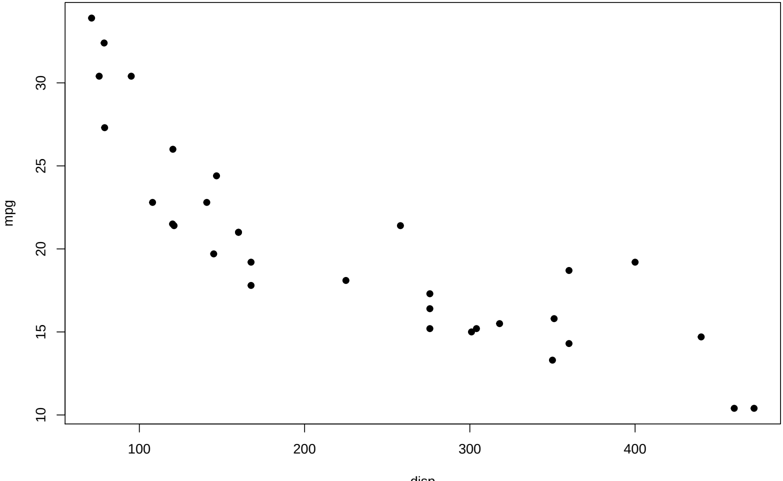
$$cor(X, Y) \cdot \frac{sd(Y)}{sd(X)}$$

is. Recall our Descriptive Statistics part!

Example

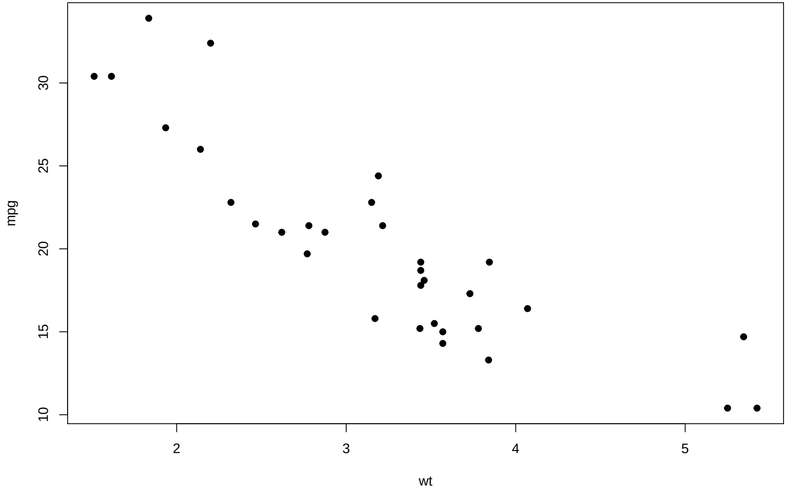
Example: We will use the `mtcars` Dataset from **R**:

```
plot(mpg ~ disp, data = mtcars, pch = 19)
```



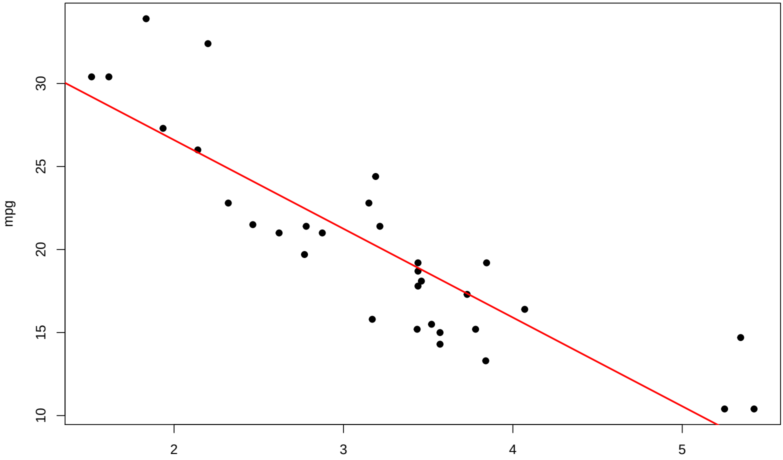
Example, Cont'd

```
plot(mpg ~ wt, data = mtcars, pch = 19)
```



Example, Cont'd

```
plot(mpg ~ wt, data = mtcars, pch = 19)
model <- lm(mpg ~ wt, data = mtcars)
abline(model, col = "red", lwd = 2)
```



Example, Cont'd

```r
model <- lm(mpg ~ wt, data = mtcars)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

Example, Cont'd

Now, we predict the value of `mpg` for a new values of a `wt` Variable:

```
pred <- predict(model, data.frame(wt=4.7))
pred
```

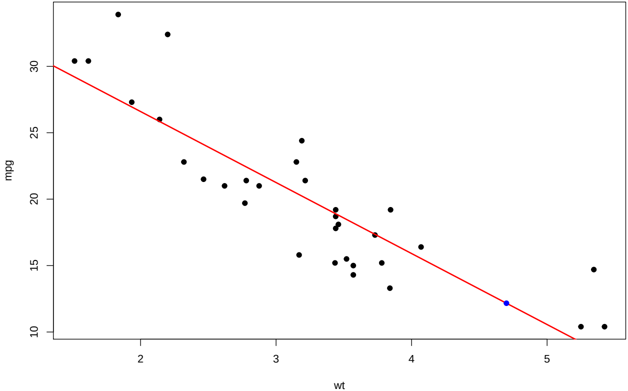
```
##        1
## 12.16611
```

Example, Cont'd

Now, we predict the value of `mpg` for a new values of a `wt` Variable:

```
pred <- predict(model, data.frame(wt=4.7))
pred
```

```
##        1
## 12.16611
```



Example

```r
x <- rnorm(100, mean = -1, sd = 1)
y <- runif(100, 2, 10)
z <- 2.7 - 1.7*x + 13.5*y + rnorm(100)
head(x)
```

```
## [1]  0.5271618 -2.0473435  0.4363159 -0.1708086 -0.7298618 -1
```

```r
head(y)
```

```
## [1] 3.977900 9.480366 2.867116 6.922705 7.797675 7.675757
```

```r
head(z)
```

```
## [1]  55.39792 133.41440  42.64337  95.71013 109.41316 108.627
```

Example

```
mod1 <- lm(z ~ x); summary(mod1)
```

```
##
## Call:
## lm(formula = z ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -58.755 -27.454   2.906  28.465  50.215
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   84.135      4.202  20.022   <2e-16 ***
## x             -3.130      2.917  -1.073    0.286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
##
## Residual standard error: 31.23 on 98 degrees of freedom
## Multiple R-squared:  0.01161,    Adjusted R-squared:  0.00152
## F-statistic: 1.152 on 1 and 98 DF,  p-value: 0.2859
```

Example, Cont'd

```r
mod2 <- lm(z ~ x + y); summary(mod2)
```

```
##
## Call:
## lm(formula = z ~ x + y)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.25835 -0.69422 -0.04329  0.72539  2.12851
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.92180    0.29018   10.07   <2e-16 ***
## x           -1.50675    0.09194  -16.39   <2e-16 ***
## y           13.46759    0.04283  314.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
##
## Residual standard error: 0.9827 on 97 degrees of freedom
## Multiple R-squared:  0.999,  Adjusted R-squared:  0.999
## F-statistic: 5.001e+04 on 2 and 97 DF,  p-value: < 2.2e-16
```

Properties of the Estimators: $\hat{\beta}_0$ and $\hat{\beta}_1$

Here we assume that $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Properties of the Estimators: $\hat{\beta}_0$ and $\hat{\beta}_1$

Here we assume that $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

Fact 1: Estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are UnBiased:

$$\mathbb{E}(\hat{\beta}_0) = \beta_0, \qquad Var(\hat{\beta}_0) = \frac{\sigma^2}{n} \cdot \frac{\sum_{k=1}^n X_k^2}{\sum_{k=1}^n (X_k - \overline{X})^2}$$

$$\mathbb{E}(\hat{\beta}_1) = \beta_1, \qquad Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{k=1}^n (X_k - \overline{X})^2}$$

Properties of the Estimators: $\widehat{\sigma^2}$

Fact 2: Assume σ^2 is unknown.

Properties of the Estimators: $\widehat{\sigma^2}$

Fact 2: Assume σ^2 is unknown. Then

$$s^2 = \frac{1}{n-2} \cdot \sum_{k=1}^{n} (\hat{\varepsilon}_k)^2$$

is an UnBiased Estimator for σ^2, and

$$\widehat{\sigma^2} = \frac{1}{n} \cdot \sum_{k=1}^{n} (\hat{\varepsilon}_k)^2$$

is the MLE for σ^2.

Properties of the Estimators: $\widehat{\sigma^2}$

Fact 2: Assume σ^2 is unknown. Then

$$s^2 = \frac{1}{n-2} \cdot \sum_{k=1}^{n} (\hat{\varepsilon}_k)^2$$

is an UnBiased Estimator for σ^2, and

$$\widehat{\sigma^2} = \frac{1}{n} \cdot \sum_{k=1}^{n} (\hat{\varepsilon}_k)^2$$

is the MLE for σ^2. Here

$$\hat{\varepsilon}_k = Y_k - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_k$$

is the k-th **residual**.

Fsyo!

Thank You!