Introduction to linear regression

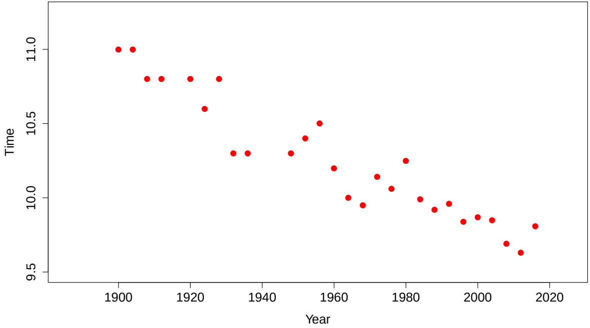
Rosario Barone*

*Email: rosario.barone@uniroma2.it

Outline of the lecture

- Introduction
- Least squares estimation
- Properties of the least squares estimators
- The coefficient of determination R^2
- Normal assumption and likelihood function
- Confidence intervals and hypothesis test
- Prediction

- Regression models analyze how one variable depends on others.
- Suppose to have two or more variables, some of which will be regarded as fixed, and others as random. The random quantities are known as **responses** and the fixed ones as **explanatory variables** or **covariates**.
- We shall suppose that only one variable is regarded as a response.
- In this lecture we outline the basic results for the simplest regression model, where a single response depends linearly on a single covariate



Winning Olympic 100-metres sprint times from 1900 to 2016

- The most obvious feature is that the winning time decreased by about 1 s. and 35 cs over that period
- A simple model is that of linear trend in the winning time (the response y) so in year j (the covariate) we have

$$y_j = \beta_1 + \beta_2 j + \epsilon_j$$

The straight-line regression model (or simple regression model) assumes that random variables Y_j satisfy

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j, \quad j = 1, \ldots, n$$

where

- $x_1 \ldots, x_n$ are known constants
- $\epsilon_1, \ldots, \epsilon_n$ are *i.i.d.* $N(0, \sigma^2)$ (homoskedasticity)
- β_1, β_2 and σ^2 are unknown parameters

Thus, the random variables Y_j are independent but not identically distributed and $Y_j \sim N(\beta_1 + \beta_2\, x_j, \sigma^2)$ for $j = 1, \ldots, n$

The data arise as pairs $(x_1, y_1) \ldots, (x_n, y_n)$, from which β_1, β_2 and σ^2 are to be estimated

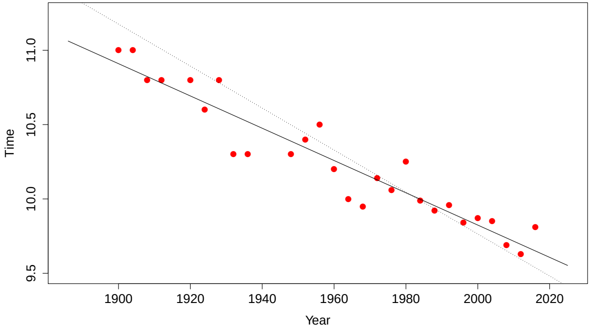
Least square estimates

To estimate β_1 and β_2 we can minimize the distance

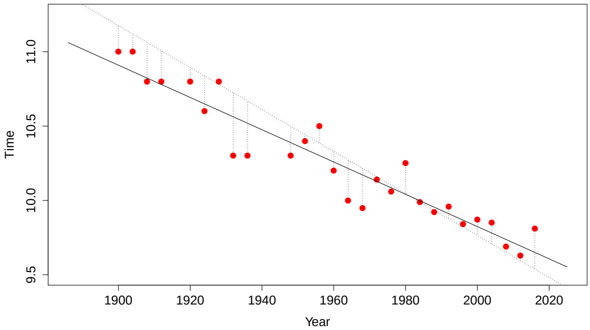
$$SS(\beta_1, \beta_2) = \sum_{j=1}^{n} (y_j - (\beta_1 + \beta_2 x_j))^2$$

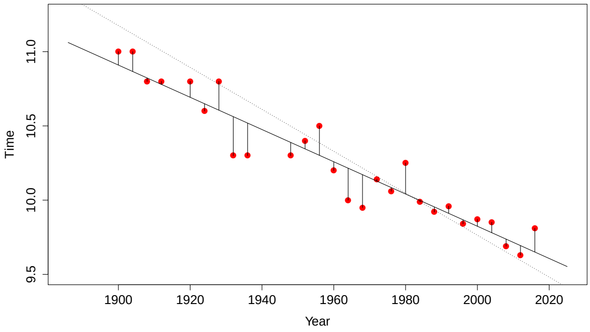
which is the sum of squared vertical deviations between the y_j and their means $\beta_1 + \beta_2\, x_j$ under the linear model.

This is equivalent to find among all the possible straight lines $\beta_0 + \beta_1 x$ the one which minimizes the sum of the vertical distances between the points y_j and $\beta_0 + \beta_1 x_j$



11.0
10.5
10.0
9.5
Time
1900
1920
1940
1960
1980
2000
2020
Year


11.0
10.5
10.0
9.5
Time
1900
1920
1940
1960
1980
2000
2020
Year


11.0
10.5
10.0
9.5
Time
1900
1920
1940
1960
1980
2000
2020
Year

To find the least square estimates (ols) we can solve the system

$$\begin{cases} \frac{\partial SS(\beta_1,\beta_2)}{\partial \beta_1} = -2\sum_{j=1}^{n}(y_j - (\beta_1 + \beta_2 x_j)) = 0 \\ \frac{\partial SS(\beta_1,\beta_2)}{\partial \beta_2} = -2\sum_{j=1}^{n} x_j(y_j - (\beta_1 + \beta_2 x_j)) = 0 \end{cases}$$

which is equivalent to

$$\begin{cases} \sum_{j=1}^{n} y_j - n\beta_1 - \beta_2 \sum_{j=1}^{n} x_j = 0 \\ \sum_{j=1}^{n} x_j y_j - \beta_1 \sum_{i=1}^{n} x_j - \beta_2 \sum_{j=1}^{n} x_j^2 = 0 \end{cases}$$

From the first eqn we have $\beta_1 = \bar{y} - \beta_2 \bar{x}$ and the second becomes

$$\sum_{j=1}^{n} x_j y_j - \bar{y}_1 \sum_{i=1}^{n} x_j + \beta_2 \bar{x} \sum_{i=1}^{n} x_j - \beta_2 \sum_{j=1}^{n} x_j^2 = 0$$

Hence, the system solution is the point $(\hat{\beta}_1, \hat{\beta}_2)$ where

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum_{j=1}^{n} x_j y_j - \bar{y}\sum_{i=1}^{n} x_j}{\sum_{j=1}^{n} x_j^2 - \bar{x}\sum_{i=1}^{n} x_j} = \frac{n\sum_{j=1}^{n} x_j y_j - \sum_{j=1}^{n} y_j \sum_{i=1}^{n} x_j}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{i=1}^{n} x_j\right)^2} \\ &= \frac{\sum_{i=1}^{n}(y_j - \bar{y})(x_j - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} = \frac{s_{xy}}{s_x^2}\end{aligned}$$

and

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

Note that $\hat{\beta}_2$ cannot be calculated if all the x_j are equal.

The matrix of the second derivative of $SS(\beta_1, \beta_2)$ is positive definite so that $(\hat{\beta}_1, \hat{\beta}_2)$ minimizes $SS(\beta_1, \beta_2)$

The quantity $SS(\hat{\beta}_1, \hat{\beta}_2)$ known as *residual sum of squares*, is the smalles sum of square $SS(\beta_1, \beta_1)$ attainable by fitting the linear regression model to the data

The values $\hat{y}_j = \hat{\beta}_1 + \hat{\beta}_2 x_j$ for $j = 1, \ldots, n$ are called **fitted values** and the straight line $y = \hat{\beta}_1 + \hat{\beta}_2 x$ is the **least squares regression** line

Properties of the least squares estimators

We now show that $E(\hat{\beta}_2) = \beta_2$ and that $V(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2}$
In fact

$$\hat{\beta}_2 \quad = \quad \frac{\sum_{i=1}^n (x_j - \bar{x})(Y_j - \bar{Y})}{\sum_{j=1}^n (x_j - \bar{x})^2} = \sum_{j=1}^n w_j (Y_i - \bar{Y})$$

where $w_j = (x_j - \bar{x}) / \sum_{i=1}^n (x_i - \bar{x})^2$ so that $\sum_{j=1}^n w_j = 0$ and

$$\sum_{j=1}^n x_j w_j = \frac{\sum_{j=1}^n (x_j^2 - x_j \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{\sum_{j=1}^n x_j^2 - n\bar{x}^2}{\sum_{j=1}^n (x_j - \bar{x})^2} = 1$$

Hence

$$\hat{\beta}_2 = \sum_{j=1}^{n} w_j(Y_j - \bar{Y}) = \sum_{j=1}^{n} w_j Y_i$$

and

$$E(\hat{\beta}_2) = E\left(\sum_{j=1}^{n} w_j Y_j\right) = \sum_{j=1}^{n} w_j E(Y_j) = \sum_{j=1}^{n} w_j(\beta_1 + \beta_2 x_j) = \beta_2$$

and

$$V(\hat{\beta}_2) = V\left(\sum_{j=1}^{n} w_j Y_j\right) = \sum_{j=1}^{n} w_j^2 V(Y_j) = \sum_{j=1}^{n} w_j^2 \sigma^2 = \frac{\sigma^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2}$$

It is also straightforward to prove that

- $E(\hat{\beta}_1) = \beta_1$
- $V(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \right)$
- $Cov(\hat{\beta}_1, \hat{\beta}_2) = -\bar{x} \frac{\sigma^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2}$

Note that all these properties (and also the least squares estimators) have been obtained without assuming the normality of the response variables but considering only their mean and variance and the independence of these random variables.

σ^2 estimator

Remember that the simple linear model assumes

$$y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

where $\epsilon_1, \ldots, \epsilon_n$ are *i.i.d* with $E(\epsilon_j) = 0$ and $V(\epsilon_j) = \sigma^2$.
Then

$$\epsilon_j = y_j - (\beta_1 + \beta_2 x_j) \quad j = 1, \ldots, n$$

and we can estimate σ^2 by calculating the variance of the **residuals**

$$e_j = y_j - (\hat{\beta}_1 + \hat{\beta}_2 x_j) \quad j = 1, \ldots, n$$

that is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} e_j^2$$

σ^2 estimator

Remember that the simple linear model assumes

$$y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

where $\epsilon_1, \ldots, \epsilon_n$ are *i.i.d* with $E(\epsilon_j) = 0$ and $V(\epsilon_j) = \sigma^2$.
Then

$$\epsilon_j = y_j - (\beta_1 + \beta_2 x_j) \quad j = 1, \ldots, n$$

and we can estimate σ^2 by calculating the variance of the **residuals**

$$e_j = y_j - (\hat{\beta}_1 + \hat{\beta}_2 x_j) \quad j = 1, \ldots, n$$

that is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} e_j^2$$

It is possible to prove that

$$E(\hat{\sigma}^2) = \frac{n-2}{n}\sigma^2$$

Hence an unbiased estimator for σ^2 is

$$S^2 = \frac{n}{n-2}\hat{\sigma}^2 = \frac{\sum_{j=1}^{n} e_j^2}{n-2}$$

Coefficient of determination

Once we have obtained the fitted value $\hat{y}_j$ it is important to evaluate how they fit the observed values y_j, that is we need to measure the goodness of fit of the regression model

Note that

$$\frac{1}{n}\sum_{j=1}^{n}\hat{y}_j = \frac{1}{n}\sum_{j=1}^{n}(\hat{\beta}_1 + \hat{\beta}_2 x_j) = \hat{\beta}_1 + \hat{\beta}_2\bar{x} = \bar{y} - \hat{\beta}_2\bar{x} + \hat{\beta}_2\bar{x} = \bar{y}.$$

Then, the **explained sum of squares (ESS)**, i.e. the sum of the squares of the deviations of the predicted values from their mean is

$$\begin{aligned} ESS &= \sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2 = \sum_{j=1}^{n}(\hat{\beta}_1 + \hat{\beta}_2 x_j - \bar{y})^2 = \\ &= \sum_{j=1}^{n}(\bar{y} - \hat{\beta}_2\bar{x} + \hat{\beta}_2 x_j - \bar{y})^2 = \hat{\beta}_2^2\sum_{j=1}^{n}(x_j - \bar{x})^2 \end{aligned}$$

Note also that the **residual sum of squares (RSS)** is

$$
\begin{aligned}
RSS &= \sum_{j=1}^{n} e_j^2 = \sum_{j=1}^{n}(y_j - \hat{\beta}_1 - \hat{\beta}_2 x_j)^2 \\
&= \sum_{j=1}^{n}\Big((y_j - \bar{y}) - \hat{\beta}_2(x_j - \bar{x})\Big)^2 \\
&= \sum_{j=1}^{n}(y_j - \bar{y})^2 + \hat{\beta}_2^2 \sum_{j=1}^{n}(x_j - \bar{x})^2 - 2\hat{\beta}_2 \sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y}) \\
&= \sum_{j=1}^{n}(y_j - \bar{y})^2 + \hat{\beta}_2^2 \sum_{j=1}^{n}(x_j - \bar{x})^2 - 2\hat{\beta}_2^2 \sum_{j=1}^{n}(x_j - \bar{x})^2 \\
&= \sum_{j=1}^{n}(y_j - \bar{y})^2 - \hat{\beta}_2^2 \sum_{j=1}^{n}(x_j - \bar{x})^2 = TSS - ESS
\end{aligned}
$$

where the **total sum of squares TSS** is $\sum_{j=1}^{n}(y_j - \bar{y})^2$

Thus we have the following identity

$$TSS = ESS + RSS$$

In general, the greater the ESS, the better the estimated model performs. In fact ESS represents the data variability explained by the regression model

The coefficient of determination

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

represents an index of goodness of fit for the simple regression model. It measures the fraction of data variability explained by the regression model. Note that $0 \leq R^2 \leq 1$ and values of R^2 approaching 1 represent a perfect fit. It is straighforward to prove that $R^2 = r^2$ where r is the correlation coefficient $s_{xy}/(s_x s_y)$

Thus we have the following identity

$$TSS = ESS + RSS$$

In general, the greater the ESS, the better the estimated model performs. In fact ESS represents the data variability explained by the regression model

The coefficient of determination

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

represents an index of goodness of fit for the simple regression model. It measures the fraction of data variability explained by the regression model. Note that $0 \leq R^2 \leq 1$ and values of R^2 approaching 1 represent a perfect fit. It is straighforward to prove that $R^2 = r^2$ where r is the correlation coefficient $s_{xy}/(s_x s_y)$

Normal assumption and likelihood function

Assuming that the variables Y_j are independent $N(\beta_1 + \beta_2 x_j, \sigma^2)$, the likelihood function based on $(x_1, y_1), \dots (x_n, y_n)$ is

$$L(\beta_1, \beta_2, \sigma^2) = \prod_{j=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(y_j - (\beta_1 + \beta_2 x_j))^2\right]$$

and the loglikelihood is

$$\ell(\beta_1, \beta_2, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - (\beta_1 + \beta_2 x_j))^2$$

For any $\forall\sigma^2$ maximizing over β_1, β_2 is equivalent to minimizing $SS(\beta_1, \beta_2) = \sum_{j=1}^{n}(y_j - (\beta_1 + \beta_2 x_j))^2$. Then, the maximum likelihood estimates (mle) for (β_1, β_1) are exactly the ols estimates

Normal assumption and likelihood function

Assuming that the variables Y_j are independent $N(\beta_1 + \beta_2 x_j, \sigma^2)$, the likelihood function based on $(x_1, y_1), \ldots (x_n, y_n)$ is

$$L(\beta_1, \beta_2, \sigma^2) = \prod_{j=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(y_j - (\beta_1 + \beta_2 x_j))^2\right]$$

and the loglikelihood is

$$\ell(\beta_1, \beta_2, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - (\beta_1 + \beta_2 x_j))^2$$

For any $\forall\sigma^2$ maximizing over β_1, β_2 is equivalent to minimizing $SS(\beta_1, \beta_2) = \sum_{j=1}^{n}(y_j - (\beta_1 + \beta_2 x_j))^2$. Then, the maximum likelihood estimates (mle) for (β_1, β_1) are exactly the ols estimates

The mle for σ^2 can be obtained by solving

$$\frac{\partial \ell(\beta_1, \beta_2, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{j=1}^{n} (y_j - (\hat{\beta}_1 + \hat{\beta}_2 x_j))^2 = 0$$

which leads to

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} (y_j - (\hat{\beta}_1 + \hat{\beta}_2 x_j))^2$$

Since $\hat{\beta}_1$ and $\hat{\beta}_2$ are linear combinations of normal random variables we have that

$$\hat{\beta}_2 \sim \mathcal{N}\left(\beta_2, \frac{\sigma^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2}\right) \quad \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2}\right)\right)$$

Moreover, it is possible to prove that

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2} \quad \text{i.e.} \quad \frac{S^2}{\sigma^2} \sim \frac{\chi^2_{n-2}}{n-2}$$

and that S^2 and $(\hat{\beta}_1, \hat{\beta}_2)$ are independent random variables

Confidence intervals and hypothesis test

Confidence intervals and hypothesis tests are based on the pivotal quantities[†] q_r

$$q_r = \frac{\hat{\beta}_r - \beta_r}{\sqrt{\hat{V}(\hat{\beta}_r)}} \quad r = 1, 2$$

where $\sqrt{\hat{V}(\hat{\beta}_r)}$ is the standard error of $\hat{\beta}_r$

Since $\hat{V}(\hat{\beta}_r) = S^2 V(\hat{\beta}_r)/\sigma^2$ we have that

$$q_r = \frac{\hat{\beta}_r - \beta_r}{\sqrt{\hat{V}(\hat{\beta}_r)}} = \frac{\hat{\beta}_r - \beta_r}{\sqrt{\frac{S^2}{\sigma^2} V(\hat{\beta}_r)}} = \frac{\frac{\hat{\beta}_r - \beta_r}{\sqrt{V(\hat{\beta}_r)}}}{\sqrt{\frac{S^2}{\sigma^2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-2}}{n-2}}} \sim t_{n-2}$$

where in last statement we have considered also the independence between $\hat{\beta}_r$ and S^2

[†]A function of observations and unobservable parameters such that the function's probability distribution does not depend on the unknown parameters.

Consider the following hypothesis test

$$\begin{cases} H_0 : \beta_r = \beta_r^{(0)} \\ H_1 : \beta_r \neq \beta_r^{(0)} \end{cases}$$

The test statistic

$$t_r = \frac{\hat{\beta}_r - \beta_r^{(0)}}{\sqrt{\hat{V}(\hat{\beta}_r)}}$$

under H_0 is a t_{n-2} distribution while under H_1 assumes large (positive or negative) values and the p-value is

$$\text{p-value} = P(|t_{n-2}| > |t_r^{oss}|) = 2P(t_{n-2} > |t_r^{oss}|)$$

$(1-\alpha)\%$ confidence intervals can be obtained by observing that

$$
\begin{aligned}
1-\alpha &= P(t_{n-2;\alpha/2} < q_r < t_{n-2;1-\alpha/2}) \\
&= P\left(-t_{n-2;1-\alpha/2} < \frac{\hat{\beta}_r - \beta_r}{\sqrt{\hat{V}(\hat{\beta}_r)}} < t_{n-2;1-\alpha/2}\right) \\
&= P\left(-t_{n-2;1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_r)} < \hat{\beta}_r - \beta_r < t_{n-2;1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_r)}\right) \\
&= P\left(\hat{\beta}_r - t_{n-2;1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_r)} < \beta_r < \hat{\beta}_r + t_{n-2;1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_r)}\right)
\end{aligned}
$$

Hence, the $(1-\alpha)\%$ confidence interval is

$$\hat{\beta}_r \pm t_{n-2;1-\alpha/2}\sqrt{\hat{V}(\hat{\beta}_r)}$$

Prediction

Let us consider now the unknown expected value

$$\mu_f = E(Y|x_f) = \beta_1 + \beta_2 x_f$$

A point estimate for μ_f is

$$\begin{aligned} \hat{y}_f &= \hat{\beta}_1 + \hat{\beta}_2 x_f \\ &= \bar{y} + (x_f - \bar{x})\hat{\beta}_2 \end{aligned}$$

Mean and variance of the estimator $\hat{Y}_f$ are

$$E(\hat{Y}_f) = E(\hat{\beta}_1 + \hat{\beta}_2 x_f) = \beta_1 + \beta_2 x_f = \mu_f$$

$$V(\hat{Y}_f) = V(\bar{Y} + (x_f - \bar{x})\hat{\beta}_2) = \frac{\sigma^2}{n} + \frac{\sigma^2 (x_f - \bar{x})^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2}$$

100 metres at the Olympics

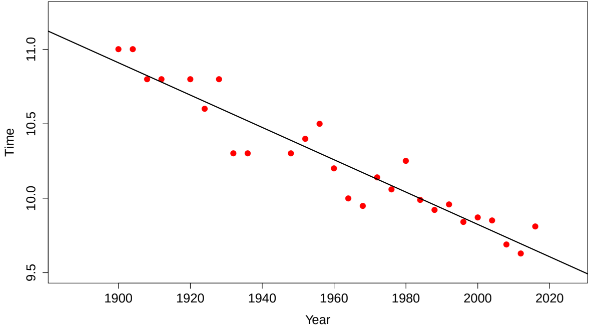
```
> olympics=read.table('olympics.txt',header=TRUE)
> m=lm(time~Year,data=olympics)
> summary(m)

Call:
lm(formula = time ~ Year, data = olympics)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.262434 -0.053855 -0.007824  0.079724  0.208744 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 31.5398334  1.4084088   22.39  < 2e-16
Year        -0.0108579  0.0007182  -15.12 4.39e-14

Residual standard error: 0.1314 on 25 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.9014,	Adjusted R-squared:  0.8975 
F-statistic: 228.6 on 1 and 25 DF,  p-value: 4.391e-14
```



11.0
10.5
10.0
9.5
Time
1900
1920
1940
1960
1980
2000
2020
Year

Predictions for Tokyo 2020

```
> new <- data.frame(Year=2020)
> predict(m, new,interval="conf")

       fit      lwr      upr
1 9.606941 9.504984 9.708899

> predict(m, new,interval="pred")

       fit      lwr      upr
1 9.606941 9.317753 9.896129
```

Interpretation

Given the model

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

- β_1 is the intercept (often represented with α or β_0); it represents the value of Y_j when $x_j = 0$;
- β_2 is the slope of the regression line; i.e. if x increases (decreases) of one unit, Y increases (decreases) of β_2.

Interpretation

Given the model

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

- β_1 is the intercept (often represented with α or β_0); it represents the value of Y_j when $x_j = 0$;
- β_2 is the slope of the regression line; i.e. if x increases (decreases) of one unit, Y increases (decreases) of β_2.

Questions

We want to investigate the relationship between two variables Y and X;

- Correlation?

- By defining

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

  we assume that there is a **causal relationship**. One cannot "search" for causality with the regression, the regression can only be used if a causal relationship is assumed.

Questions

We want to investigate the relationship between two variables Y and X;

- Correlation?

- By defining

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

  we assume that there is a **causal relationship**. One cannot "search" for causality with the regression, the regression can only be used if a causal relationship is assumed.

Summary and...

1. model specification and assumptions:

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

   where $\epsilon_1, \ldots, \epsilon_n$ are $i.i.d.$ $N(0, \sigma^2)$.

2. point estimation:
   - $\hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$
   - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$
3. calculate standard errors
4. diagnostics:
   - $R^2 = \frac{ESS}{TSS}$
   - t-test
   - test for homoskedasticity! new entry
5. interpretation

Summary and...

1. model specification and assumptions:

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

   where $\epsilon_1, \ldots, \epsilon_n$ are $i.i.d.$ $N(0, \sigma^2)$.

2. point estimation:
   - $\hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$
   - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

3. calculate standard errors

4. diagnostics:
   - $R^2 = \frac{ESS}{TSS}$
   - t-test
   - test for homoskedasticity! new entry

5. interpretation

Summary and...

1. model specification and assumptions:

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

   where $\epsilon_1, \ldots, \epsilon_n$ are $i.i.d.$ $N(0, \sigma^2)$.

2. point estimation:
   - $\hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$
   - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

3. calculate standard errors

4. diagnostics:
   - $R^2 = \frac{ESS}{TSS}$
   - t-test
   - test for homoskedasticity! new entry

5. interpretation

Summary and...

1. model specification and assumptions:

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

   where $\epsilon_1, \ldots, \epsilon_n$ are $i.i.d.$ $N(0, \sigma^2)$.

2. point estimation:
   - $\hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$
   - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

3. calculate standard errors

4. diagnostics:
   - $R^2 = \frac{ESS}{TSS}$
   - t-test
   - test for homoskedasticity! new entry

5. interpretation

Summary and...

1. model specification and assumptions:

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

   where $\epsilon_1, \ldots, \epsilon_n$ are $i.i.d.$ $N(0, \sigma^2)$.

2. point estimation:
   - $\hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$
   - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

3. calculate standard errors

4. diagnostics:
   - $R^2 = \frac{ESS}{TSS}$
   - t-test
   - test for homoskedasticity! new entry

5. interpretation

Summary and...

1. model specification and assumptions:

$$Y_j = \beta_1 + \beta_2 x_j + \epsilon_j \quad j = 1, \ldots, n$$

   where $\epsilon_1, \ldots, \epsilon_n$ are $i.i.d.$ $N(0, \sigma^2)$.

2. point estimation:
   - $\hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$
   - $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

3. calculate standard errors

4. diagnostics:
   - $R^2 = \frac{ESS}{TSS}$
   - t-test
   - test for homoskedasticity! new entry

5. interpretation