

Review of Concepts for Lab 02

Naomi Wilcox
9/5/18

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

Linear Regression

The diagram illustrates the linear regression equation $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ with the following components labeled:

- Dependent Variable**: Points to Y_i .
- Population Y intercept**: Points to β_0 .
- Population Slope Coefficient**: Points to β_1 .
- Independent Variable**: Points to X_i .
- Random Error term**: Points to ϵ_i .

Below the equation, two components are grouped with curly braces:

- Linear component**: Groups $\beta_0 + \beta_1 X_i$.
- Random Error component**: Groups ϵ_i .

lm()

- Performs a linear regression

General syntax:

Multiple Linear Regression Example

```
fit <- lm(y ~ x1 + x2 + x3, data=mydata)
```

show results

```
summary(fit)
```

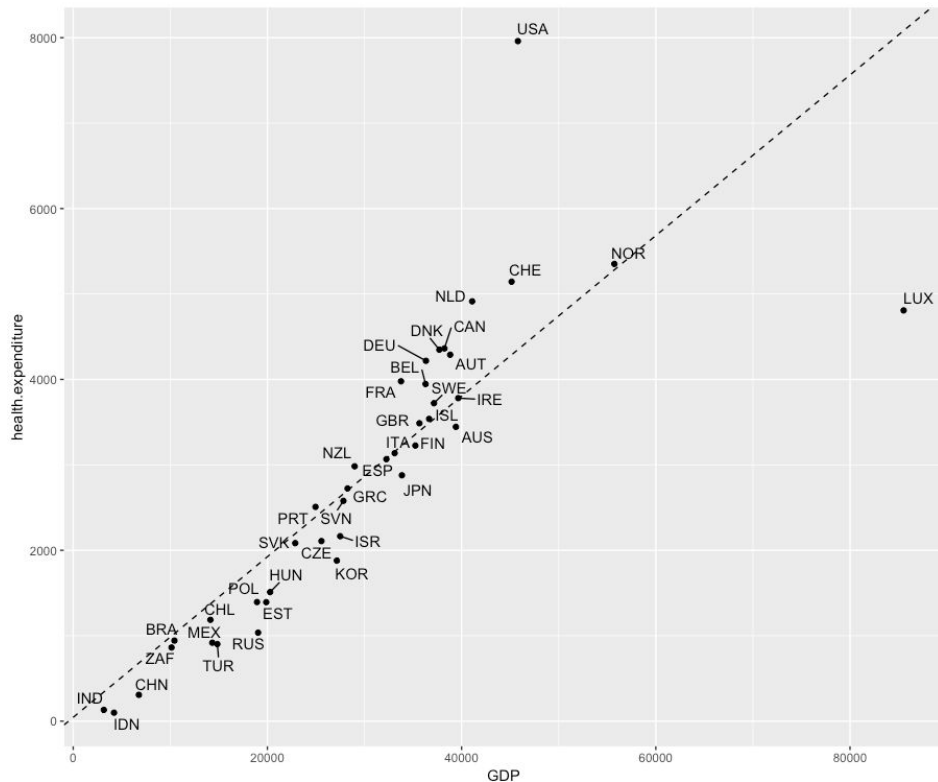
Note: Know how to interpret the output of a `summary(fit)` command

lm() example

First, plot the data. Make a scatterplot with `geom_point()` to assess whether the relationship between two variables is linear.

Then perform `lm()` and view the `summary()` of the fit to get exact intercept and coefficient (β) terms.

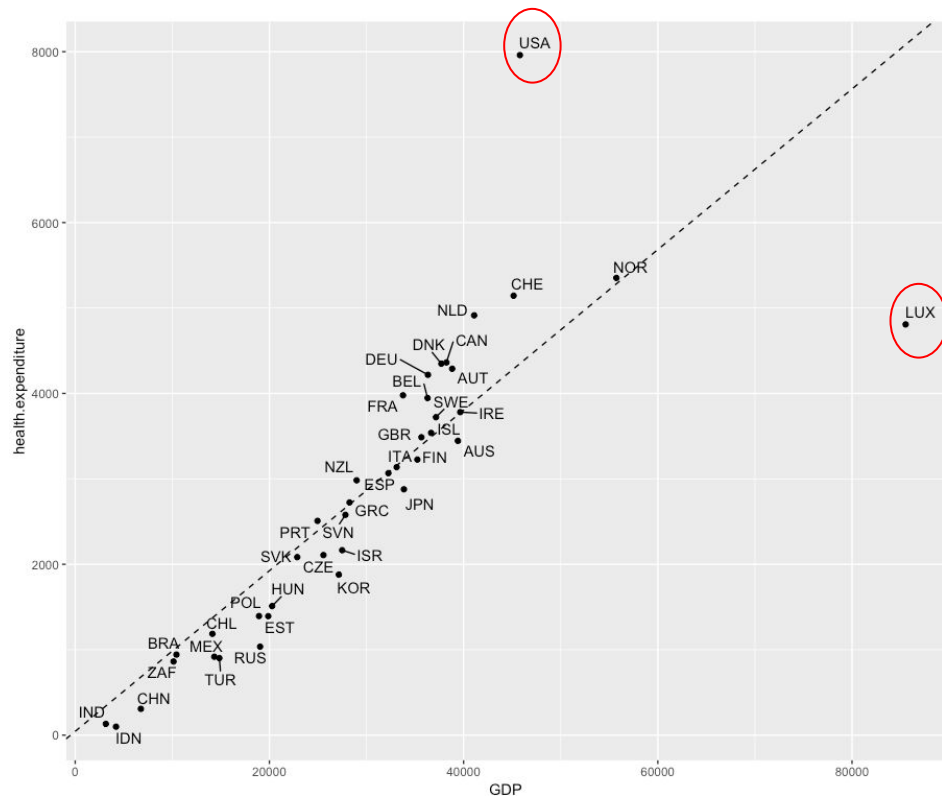
Finally, Interpret these values into words.



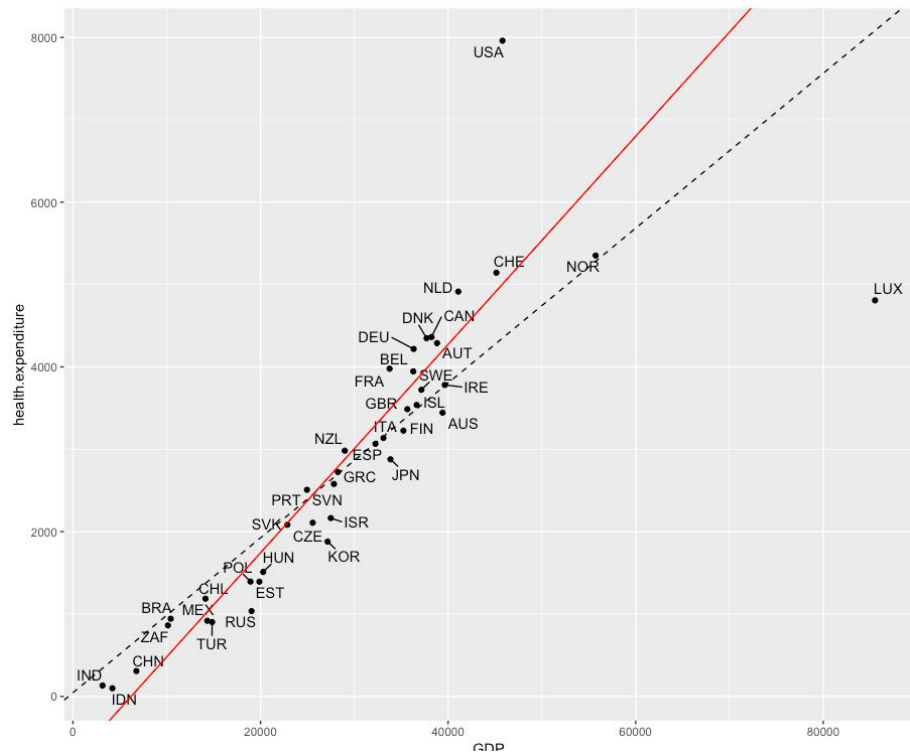
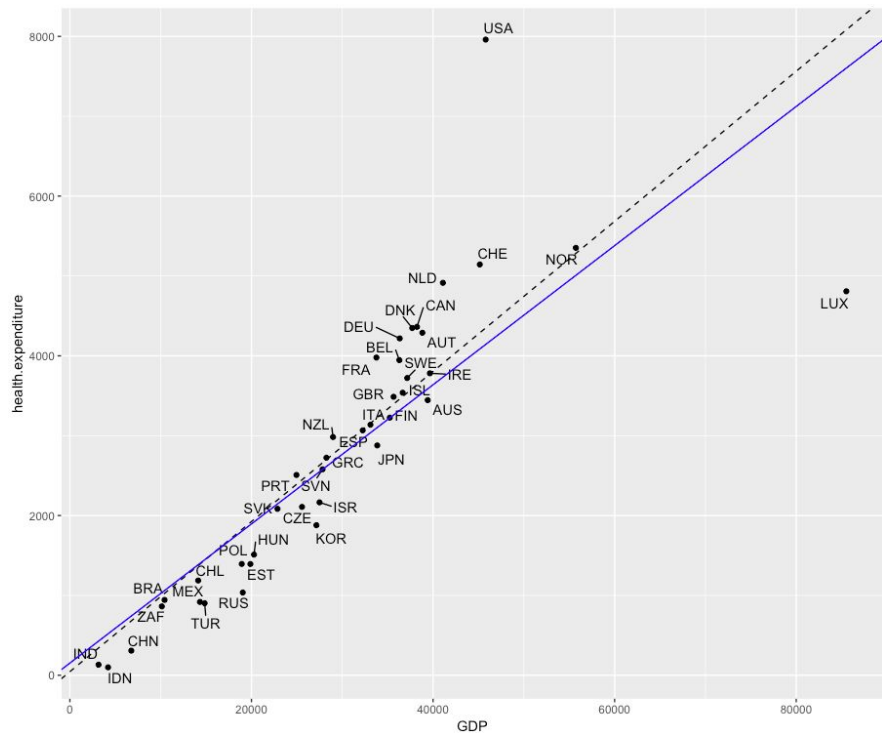
Example

What are these points called?

Ans: Outliers and influential points



Example



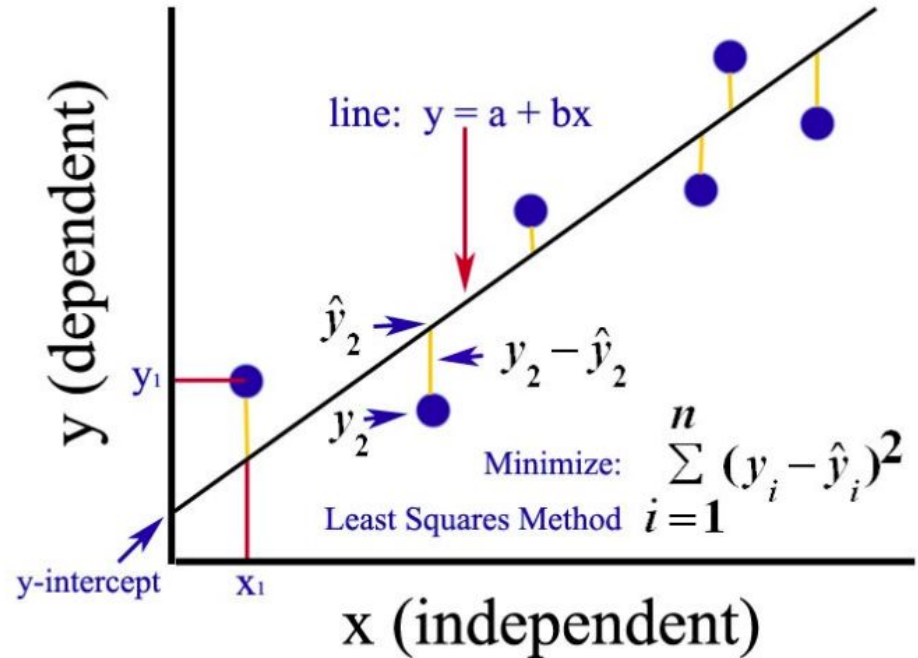
Terms

Correlation - r is the slope of the least squares regression line when we measure both x and y in standardized units.

r^2 - square of the correlation - the fraction of the variation in one variable that is explained by least-squares regression on the other variable.

Influential observation - an individual point that substantially changes the correlation or the regression line. Outliers in the x direction are often influential for the regression line.

Extrapolation - the use of a regression line for prediction for values of the explanatory variable well outside the range of the data from which the line was calculated



log()

log

From [base v3.5.1](#)
by [R-core](#) [R-core@R-project.org](#)

16th
Percentile

Logarithms And Exponentials

`log` computes logarithms, by default natural logarithms, `log10` computes common (i.e., base 10) logarithms, and `log2` computes binary (i.e., base 2) logarithms. The general form `log(x, base)` computes logarithms with base `base`.

`log1p(x)` computes $\log(1 + x)$ accurately also for $|x| \ll 1$.

`exp` computes the exponential function.

`expm1(x)` computes $\exp(x) - 1$ accurately also for $|x| \ll 1$.

Keywords [math](#)