some title

January 4, 2018

Regression

Correlation

The most crucial part in defining correlation geometrically is definig the dot product as it enables to compute the length of a vecotr:

$$|\vec{a}| = \sqrt{\langle \vec{a}, \vec{a} \rangle}$$

and the angle between any two vectors:

$$\cos(\vec{a}, \vec{b}) = \frac{\langle \vec{a}, \vec{a} \rangle}{|\vec{a}||\vec{b}|}$$

Now we define scalar product of two random vectors as covariation between them:

$$\langle X, Y \rangle = \operatorname{Cov}(X, Y)$$

The main characteristics of a random vector are its length and direction. So, we introduce the length

$$\sqrt{\operatorname{Cov} X, X} = \sqrt{\operatorname{Var}(X)} = \sigma_X$$

and the angle between two random vectors

$$\cos(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}} = \operatorname{Corr}(X, Y)$$

Note that from the definition of the angle it follows that correlation can range from -1 to 1.

The most widespread definition of correlation is

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}}$$

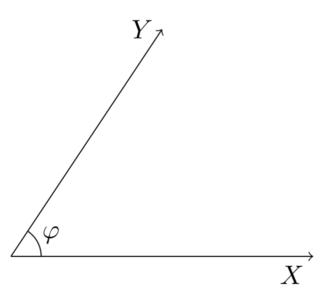


Figure 1: Random vector X of length σ_X and random vector Y of length σ_Y, $\cos\varphi$ is the angle between X and Y.

Another important geometrical tool is projection. Recall that for any two vectors the scalar product $\langle \vec{a}, \vec{b} \rangle$ can be interpreted as the length of projected $\vec{b}$ multuplied by the length of $\vec{a}$. The projection itself is $cos(\vec{a}, \vec{b})\vec{b}$. Same holds for random vectors. The projection of a random vector Y onto $\{cX | c \in \mathbb{R}\}$ is $\hat{Y} = \operatorname{Corr}(X, Y) \cdot Y$.

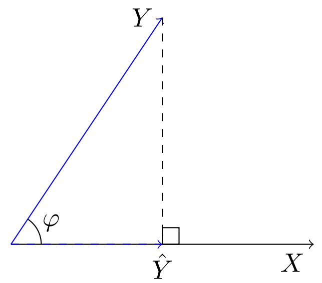


Figure 2: Random vector Y projected onto random vector X.

Looking at Figure , we can interpret the square of correlation coefficient. Using the fact that $cos^2\varphi$ is the squared ratio of the leg adjacent to φ to hypotenuse, we can write

$$\text{Corr}^2(X,Y) = \frac{\text{Var}(\hat{Y})}{\text{Var}(Y)}$$

as the variance of a random vecor is associated with the square of its length. Thus, the correlation coefficient squared shows the fraction of variance in Y which can be explained with the most similar random variable proportional to X.

Sample correlation coefficient in simple linear regression

Theorem 1. *A linear regression model with one explanatory variable and constant term has the property*

$$\text{sCorr}(y,\hat{y}) = sign(\hat{\beta}_2)\,\text{sCorr}(y,x)$$

Proof. Firstly, we consider the case when $\hat{\beta}_2 > 0$ so the main picuture is of the form depicted in Figure . It has been shown earlier that the correlation coefficient squared represents the angle betweem two random vectors.

Assuming the underlying relationship between x and y to be

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \quad i = 1,\ldots,n$$

where ε_i is an error term the following holds

$$\begin{aligned}
\text{sCorr}(y,\hat{y}) &= \frac{\text{sCov}(y)\,\text{sCov}(\hat{y})}{\sqrt{\text{sVar}(y)\,\text{sVar}(\hat{y})}} \\
&= \frac{\text{sCov}(y)\,\text{sCov}(\hat{\beta}_1 + \hat{\beta}_2 x)}{\sqrt{\text{sVar}(y)\,\text{sVar}(\hat{\beta}_1 + \hat{\beta}_2 x)}} \\
&= \frac{\text{sCov}(y)\,\text{sCov}(\hat{\beta}_2 x)}{\sqrt{\text{sVar}(y)\,\text{sVar}(\hat{\beta}_2 x)}} \\
&= \frac{\hat{\beta}_2\,\text{sCov}(y)\,\text{sCov}(x)}{|\hat{\beta}_2|\sqrt{\text{sVar}(y)\,\text{sVar}(x)}} \\
&= sign(\hat{\beta}_2)\frac{\text{sCov}(y)\,\text{sCov}(x)}{\sqrt{\text{sVar}(y)\,\text{sVar}(x)}}
\end{aligned}$$

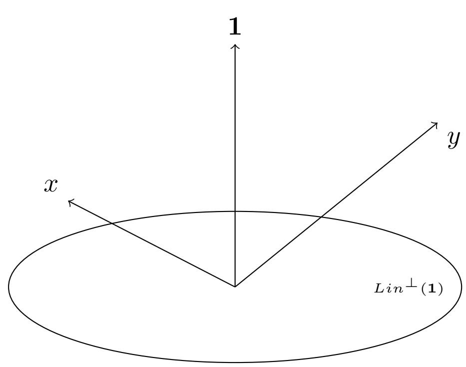
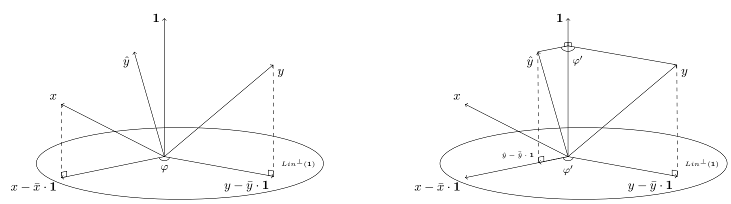


Figure 3: Vectors x, y and $\mathbf{1}$.

However, it seems to be difficult to compare the angles in the three dimensional space. That is why we start with projecting both x and y onto the space perpendicular to the vector of all ones $\mathbf{1}$ as shown in Figure 4(a). We denote this space as $Lin^{\perp}(\mathbf{1})$. The resulting vectors are $x - \bar{x} \cdot \mathbf{1}$ and $y - \bar{y} \cdot \mathbf{1}$ respectively since projection of any vector $\vec{a}$ on the line given by a vector of all ones yields the vector of averages $\vec{\bar{a}}$.

In order to get the angle between y and $\hat{y}$ we should start with regressing y on $Lin(x, \mathbf{1})$. Then the only thing thing left is to project $\hat{y}$ onto $Lin^{\perp}\mathbf{1}$ since the y vector has already been projected. The result of this step is shown in Figure 4(b).



(a) (b)

Figure 4: (a): 'Centred' x and y, i.e., projected onto $Lin^{\perp}(\mathbf{1})$; (b): 'Centred' $\hat{y}$, i.e., projected onto $Lin^{\perp}(\mathbf{1})$.

todo: whole picture, $\beta_2 < 0$

$\square$

Regression line and point of averages

Theorem 2. *The point of averages lies on the estimated regression line.*

Proof. For the geometrical proof it suffices to show that $\hat{y}$ is a linear combination of the regressors, which is true by construction, and that $\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \frac{1}{n}\sum_{i=1}^{n} y$. In order for the pictures to be more clear the proof will be presented for the case of two regressors.

The first step is regressing y on $Lin(\mathbf{1}, x)$. As shown in Figure 5(a), we obtain $\hat{y}$ as a linear combination of $\mathbf{1}$ and x. The next step is to regress both y and $\hat{y}$ on $\mathbf{1}$ which results in $\bar{y}$ and $\bar{\hat{y}}$ correspondingly. By the theorem of three perpendiculars, $\bar{y} = \bar{\hat{y}}$ which is shown in Figure 5(b).

$\square$

If the regression contains the intercept, the following equation holds:

$$\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$$
$$= X(X^TX)^{-1}X^TX\beta + X(X^TX)^{-1}X^T\varepsilon$$

Premultiplying both sides by X^T, we obtain:

$$X^T\hat{y} = X^TX(X^TX)^{-1}X^TX\beta$$
$$+ X^TX(X^TX)^{-1}X^T\varepsilon$$
$$= X^TX\beta + X^T\varepsilon$$

This is a system of equations. The first row of X^T is $\mathbf{1}$ vector, so we can write out the first equation:

$$\sum_{i=1}^{n}\hat{y}_i = \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}\beta_j$$

From the first equation in the system

$$X^T\hat{y} = X^Ty$$

we obtain

$$\sum_{i=1}^{n}\hat{y}_i = \sum_{i=1}^{n} y$$

And this finishes the proof:

$$\frac{1}{n}\sum_{i=1}^{n} y = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}\beta_j$$

Frisch–Waugh–Lovell theorem

Theorem 3. *Consider regression*

$$y = X_1\beta_1 + X_2\beta_2 + u \qquad (1)$$

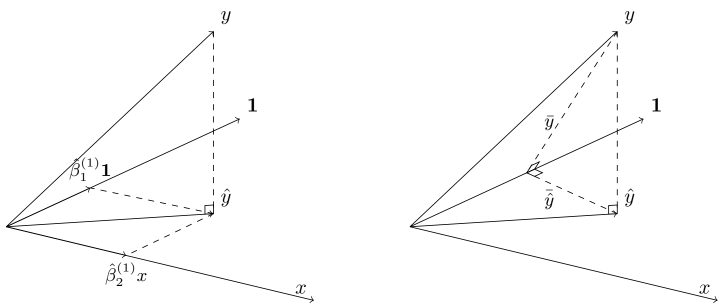
where $X_{n\times k} = [X_1X_2]$, i.e. X_1 consists of first k_1 columns of X and X_2 consists of remaining k_2 columns of X, β_1 and β_2 are comfortable, i.e. $k_1 \times 1$ and $k_2 \times 1$ vectors. Consider another regresion

$$M_1y = M_1X_2\beta_2 + M_1u \qquad (2)$$

From regresison 2 we get the following estimator:

$$\hat{\beta}_2 = ((M_1X_2)^TM_1X_2)^{-1}(M_1X_2)^TM_1y$$
$$= (X_2^TM_1^TM_1X_2)^{-1}X_2^TM_1^TM_1y$$
$$= (X_2^TM_1X_2)^{-1}X_2^TM_1y$$

As for regresison 1, let us note that due to



(a) (b)

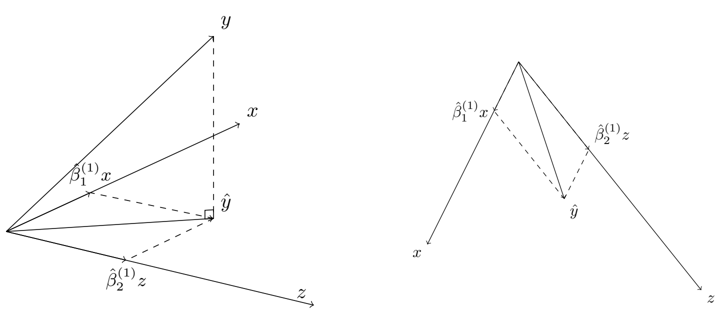
Figure 5: (a): Regression of y on $Lin(\mathbb{1}, x)$; (b): Regression of y and $\hat{y}$ on $\mathbb{1}$.

where $M_1 = I - P_1$ projects onto the orthogonal complement of the column space of X_1 and $P_1 = X_1(X_1^T X_1)^{-1} X_1^T$ is the projection onto the column space of X_1. Then the estimate of β_2 from regression 1 will be the same as the estimate from regression 2.

Proof. Geometrical proof will be presented for the following model:

$$y_i = \beta_1 x_i + \beta_2 z_i + u_i \tag{3}$$

We start with regression 'all-at-once' and will distinct its coefficients with index (1). The only step in obtaining $\beta_1^{(1)}$ is regressing y on $Lin(x, z)$ and then expanding $\hat{y}$ as a linear combination of basis vectors x and z, which is shown in Figure 6(a). Figure 6(b) depicts $Lin(x, z)$.



(a) (b)

Figure 6: (a): Regression of y on $Lin(x, z)$; (b): $Lin(x, z)$.

As for the model 2, where several regressions are performed consecutively, we start with regressing y on z, resulting in $\tilde{y}$, which we will refer to as "cleansed" y.

$$\begin{aligned} y &= \alpha z + \varepsilon \\ \hat{\alpha} &= \frac{y^T z}{z^T z} \\ \tilde{y} &= \hat{\varepsilon} = y - \frac{y^T z}{z^T z} z \end{aligned} \tag{4}$$

Following that, x is regressed on z, resulting in $\tilde{x}$ — "cleansed" x.

$$\begin{aligned} x &= \gamma z + \nu \\ \hat{\gamma} &= \frac{x^T z}{z^T z} \\ \tilde{x} &= \hat{\nu} = x - \frac{x^T z}{z^T z} z \end{aligned} \tag{5}$$

Geometric results of these two steps are presented in 7(a).

Finally, 'cleansed' y must be regressed on 'cleansed' x. However, it cannot be performed immediately as $\tilde{y}$ and $\tilde{x}$ are skew lines. So at first, we fix this problem by translation and after taht obtain $\hat{\beta}_1^{(2)}\tilde{x}$ (see Figure 7(b)).

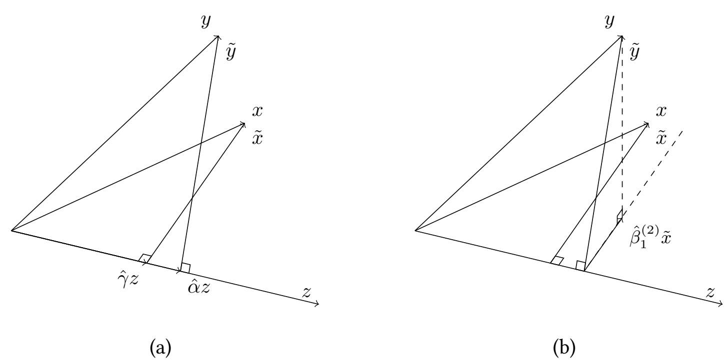


Figure 7: (a): Regression of y on z and of x on z; (b): Translation of $\tilde{x}$.

Now, let us picture all the results in one figure and mark some main points.

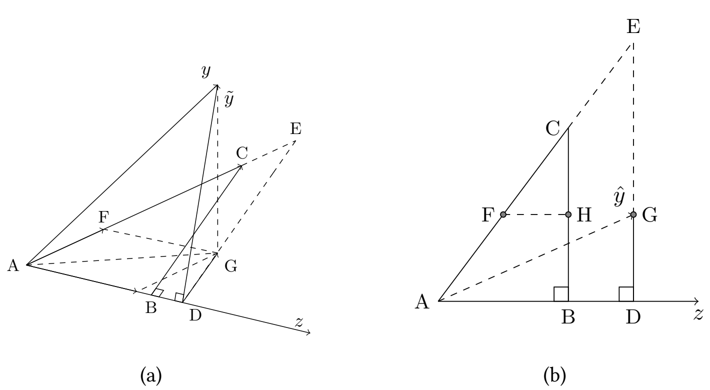


Figure 8: (a): Point A stands for the origin, B — $\hat{\gamma}z$, C — x, D — $\hat{\alpha}z$, E — intersection of vector x and line parallel to $\tilde{x}$, F — $\hat{\beta}_1^{(1)}x$, G — $\hat{\beta}_1^{(2)}\tilde{x}$; (b): $Lin(x, z)$.

In Figure 8(b) segments AF and BH = DG stand for $\hat{\beta}_1^{(1)}x$ and $\hat{\beta}_1^{(2)}\tilde{x}$ respectively, while segments AC and BC represent x and $\tilde{x}$. Having two congruent angles, triangles ABC and FHC are simillar. Then, it follows:

$$\frac{AF}{AC} = \frac{BH}{BC} \Leftrightarrow \frac{\hat{\beta}_1^{(1)}x}{x} = \frac{\hat{\beta}_1^{(2)}\tilde{x}}{\tilde{x}} \Leftrightarrow \hat{\beta}_1^{(1)} = \hat{\beta}_1^{(2)}$$

□