# *some title*

*February 4, 2018*

## *Geometric interpretation of random variables*

The fundamental part to start with is defining the geometric properties of random variables using some concepts of linear algebra.

Consider the vector space which consists of all random variables with finite mean and variance. We will regard each point in this space (or vector that correponds to that point in terms of linear algebra) a random variable. We define the scalar product of two random variables $X$ and $Y$ to be

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

It is of no difficulty to check that the definition satisfies the properties of scalar product assuming that $X$ and $Y$ are the same random variables if $\mathbb{P}(X = Y) = 1$.

Having defined the inner product, we are now able to introduce the squared length of a random variable $X$ which is

$$\|X\|^2 = \langle X, X \rangle = \mathrm{Cov}(X, X) = \mathrm{Var}(X),$$

so the length is simply the square root of this expression, i.e., the standard deviation of $X$ ($\sigma_X$).

Recall that for any non-random vectors $a$ and $b$ the angle between them is calculated with the formula

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

The same applies for the random variables and it is already clear that two random variables are uncorrelated if and only if their scalar product equals to $0$. Additionally, it means that these two random variables are orthogonal in the vector space.

The analogue for $cos(a, b)$ in the vector space of all the random variables is the correlation between two of them:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\langle X, Y \rangle}{\sqrt{\|X\|^2 \|Y\|^2}}.$$

From the equivalence of $\mathrm{Corr}(X, Y)$ to the $\cos(a, b)$ it automatically follows that the correlation coefficient can range from $-1$ to $1$.

A useful property of the geoemtry of random variables is that all the geometric theorems still hold. For instance, the Pythogorean theorem can be formulated as follows: if the ranadom variables $X$ and $Y$ are uncorrelated (which implies that they are orthogonal), then the variance of their sum equals the sum of their variances:

$$\mathrm{Var}(X + Y) = \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Translated to the non-random language, assumption of uncorrelatedness correspnds to the right triangle setting, the variance of the sum of two random variables stands for the hypotenuse squared and the sum of the variances is the sum of the legs squared.

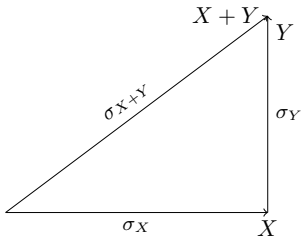



Figure 1: The Pythagorean theorem for random variables $X$ and $Y$.

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

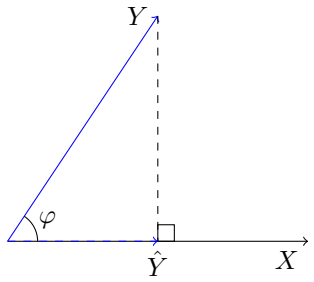



Figure 2: The projection of a random variable $Y$ onto the line spanned by a random varibale $X$.

Note that the squared lengths of the leg adjacent to $\varphi$ and the hypotenuse are $\mathrm{Var}(\hat{Y})$ and $\mathrm{Var}(Y)$. So, the Figure  gives a useful expression for the correlation coefficient squared:

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

### *The law of iterated expectations*

**Theorem 1**. *For any random variable $X$ and $Y$,*

$$\mathrm{E}(\mathrm{E}(X|Y)) = \mathrm{E}(Y).$$

*Proof.* Consider the vector space of all the random variables. The random variable which can be described as functions $h(X)$ of $X$ form a subspace of

Here is the proof for the case when $X$ and $Y$ are both discrete. Let $\mathrm{E}(Y|X) = g(X)$.

$$\begin{aligned}
\mathrm{E}(g(X)) &= \sum_x g(x)\mathbb{P}(X = x) \\
&= \sum_x \left( \sum_y y\mathbb{P}(Y = y|X = x) \right) \mathbb{P}(X = x) \\
&= \sum_x \sum_y y\mathbb{P}(X = x)\mathbb{P}(Y = y|X = x) \\
&= \sum_y y \sum_x \mathbb{P}(X = x, Y = y) \\
&= \sum_y y\mathbb{P}(Y = y) \\
&= \mathrm{E}(Y)
\end{aligned}$$

The proof in case of continous random variables is absolutely analogous.

that vector space, represented as a plane $\alpha$ in Figure . Another subspace is a subspace of constants, denoted as a vector $\mathbf{1} \in \alpha$.

In order to obtain $\mathrm{E}(Y|X)$, first, we need to project $Y$ onto the subspace corresponding to $X$. As a result of this step, we get $\mathrm{E}(Y|X)$ — the function of $X$ that predicts $Y$ the best. Next, projecting $\mathrm{E}(Y|X)$ onto the space of all constants, we end up with $E(Y)$.

Notice that the vector $Y - \mathrm{E}(Y|X)$ (which is also called the residual) is perpendicular to the plane $\alpha$. Moreover, the vecotr $\mathrm{E}(Y|X) - E(Y)$ is perpendicular to the vector of constants $\mathbf{1}$. Thus, we can apply the theorem of three perpendiculars and conclude that the vector $Y - \mathrm{E}(Y)$ is also perpendicular to the vector of constants $\mathbf{1}$.

So, we showed that the expectation of the random variable $Y$ can be obtained either in two steps or by its direct projection onto the subpace of constans.

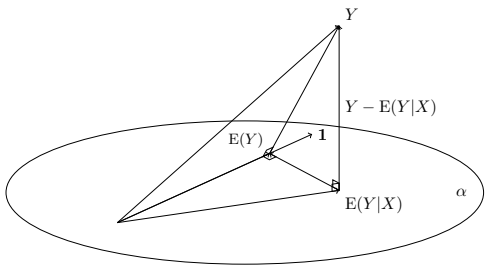



Figure 3: The law of iterated expectations. Equivalence of the two-step projecttion and direct projection of $Y$ onto $\mathbf{1}$.

□

*MSE decomposition*

**Theorem 2**. *The mean squared error of an estimator $\hat{\theta}$ with respect to an unknown parameter $\theta$ defined as $MSE(\hat{\theta}) = \mathrm{E}((\hat{\theta} - \theta)^2)$ can be decomposed into the sum of the variance of the estimator and its squared bias:*

$$MSE(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \mathrm{E}\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right]$$

*Proof.* We start with a random variable $\theta$ and its estimate $\hat{\theta}$ in the vector space. We know that an unbiased estimator's projection would be exactly the vector representing $\theta$. However, in general it does not have to and Figure  illustrates this case: the projection of the estimator falls onto the line spanned by the vector $\theta$.

Connecting vectors $\theta$ and $\hat{\theta}$, we obtain the right triangle which legs are $\hat{\theta} - \mathrm{E}(\hat{\theta})$, $\mathrm{E}(\hat{\theta}) - \theta$ and the hypotenuse $\hat{\theta} - \theta$. Applying the Pythagorean

$$\begin{aligned}
MSE(\hat{\theta}) &= \mathrm{E}((\hat{\theta} - \theta)^2) \\
&= \mathrm{E}\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta}) + \mathrm{E}(\hat{\theta}) - \theta\right)\right] \\
&= \mathrm{E}\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2 + 2\left((\hat{\theta} - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)\right)\right. \\
&\quad \left. + \left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] \\
&\quad + 2\,\mathrm{E}\left[\left((\hat{\theta} - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)\right)\right] \\
&\quad + \mathrm{E}\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] \\
&\quad + 2(\mathrm{E}(\hat{\theta} - \mathrm{E}(\hat{\theta}))\,\mathrm{E}(\hat{\theta} - \mathrm{E}(\hat{\theta})) \\
&\quad + \mathrm{E}\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{\theta} - \mathrm{E}(\hat{\theta})\right)^2\right] + \mathrm{E}\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right] \\
&= \mathrm{Var}(\hat{\theta}) + \mathrm{E}\left[\left(\mathrm{E}(\hat{\theta}) - \theta\right)^2\right]
\end{aligned}$$

theorem, we finish the proof:

$$\begin{aligned}
\|\hat{\theta} - \theta\|^2 &= \|\hat{\theta} - \mathrm{E}(\hat{\theta})\|^2 + \|\mathrm{E}(\hat{\theta}) - \theta\|^2 \\
\mathrm{E}((\hat{\theta} - \theta)^2) &= \mathrm{E}((\hat{\theta} - \mathrm{E}(\hat{\theta}))^2) + \mathrm{E}((\mathrm{E}(\hat{\theta}) - \theta)^2) \\
MSE(\hat{\theta}) &= \mathrm{Var}(\hat{\theta}) + \mathrm{E}((\mathrm{E}(\hat{\theta}) - \theta)^2)
\end{aligned}$$

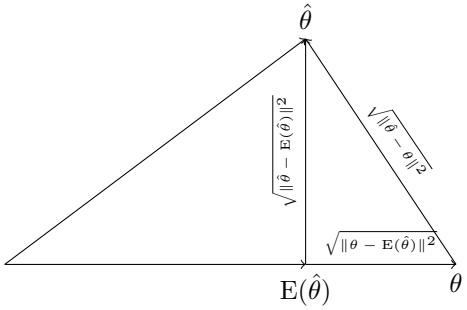



Figure 4: Decomposition of mean squred error into the variance and the bias squared.

□

## *Regression*

The concepts discussed in the following section could also be presented in random variables instead of sample ones. As the geometry of sample variables is almost of no difference comparing to the random ones, the logic of all the theorems is also the same.

### *Geometry of sample variables*

In the same manner which was performed in Section 1, we define the scalar product of two sample variables $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ as a sample covariation between them:

$$\langle x, y \rangle = \mathrm{sCov}(x, y).$$

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

$$\sqrt{\mathrm{sCov}(x, x)} = \sqrt{\mathrm{sVar}(x)} = \sigma_x$$

and the angle between two sample variables

$$\cos(x, y) = \frac{\mathrm{sCov}(x, y)}{\sqrt{\mathrm{sVar}(x)\,\mathrm{sVar}(y)}} = \mathrm{sCorr}(x, y).$$

$$\begin{aligned}
\mathrm{sCorr}(x, y) &= \frac{\mathrm{sCov}(x, y)}{\sqrt{\mathrm{sVar}(x)\,\mathrm{sVar}(y)}} \\
&= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})\frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})}}
\end{aligned}$$

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

Completely analogus to the case of random variables, the projection of such a sample variable $y$ onto $\{cx|c \in \mathbb{R}\}$ is $\hat{y} = \text{sCorr}(x, y) \cdot y$.

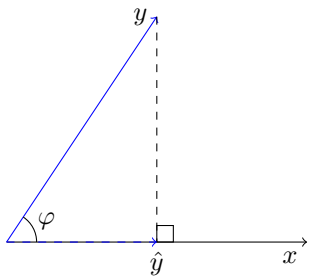



Figure 5: Vector $y$ projected onto vector $x$.

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

$$\text{sCorr}^2(x, y) = \frac{\text{sVar}(\hat{y})}{\text{sVar}(y)},$$

as the variance of a vector is associated with the square of its length. Thus, the sample correlation coefficient squared shows the fraction of variance in $y$ which can be explained with the most similar vector proportional to $x$.

### *Sample correlation when a constatnt vector added*

**Theorem 3**. *Adding a vector of constants does not affect the sample correlation coefficient:*

$$\text{sCorr}(x + \alpha\mathbf{1}, y) = \text{sCorr}(x, y)$$

*where* $\alpha \in \mathbb{R}$.

*Proof.* Firstly, we project vectors $x$ and $y$ onto $Lin^{\perp}(\mathbf{1})$ in order to get $x^c = x - \bar{x}$ and $y^c = y - \bar{y}$ ('c' stands for 'centred'). It can be shown that the matrix corresponding to projecting onto the line spanned by a vector of all ones has the following form

$$\frac{\mathbf{1}^T\mathbf{1}}{\mathbf{1}\mathbf{1}^T} = \frac{\begin{pmatrix} 1 & \dots & 1 \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}{\sum_{i=1}^{n} 1} = \begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Thus, projecting onto an orthogonal subspace is equivalent to substracting the projected vector, i.e., the vector of averages, from the original one.

$$\begin{aligned} \text{sCorr}(x + \alpha\mathbf{1}, y) &= \frac{\text{sCov}(x + \alpha\mathbf{1}, y)}{\text{sVar}(x + \alpha\mathbf{1})\,\text{sVar}(y)} \\ &= \frac{\text{sCov}(x, y)\,\text{sCov}(\alpha\mathbf{1}, y)}{\text{sVar}(x)\,\text{sVar}(y)} \\ &= \frac{\text{sCov}(x, y)}{\text{sVar}(x)\,\text{sVar}(y)} \\ &= \text{sCorr}(x, y) \end{aligned}$$

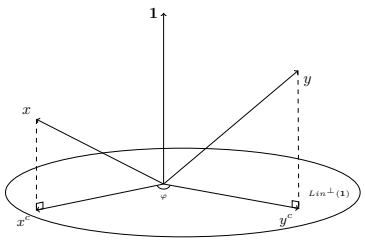



Figure 6: Centred vectors $x^c$ and $y^c$

Also note that the angle $\varphi$ between original and centred vectors remains the same. The results of this step is shown in Figure .

Then we need to derive a new vector $\tilde{x}$ with constants added to each component. Geometrically adding a vector of costants means adding a vector of all ones scaled by $\alpha \in \mathbb{R}$, i.e., $\alpha\mathbf{1}$. Then the new vector $\tilde{x}$ can be broken up into a sum of $\alpha\mathbf{1}$ and $\beta x$, $\alpha, \beta \in \mathbb{R}$, which can be seen in Figure . After that we will project this new vector $\tilde{x}$ onto $Lin^{\perp}(\mathbf{1})$. By the properties of projection it is of no difference whether to project the whole vector $\tilde{x}$ or project its parts $\alpha\mathbf{1}$ and $\beta x$ — the result is the same. So, while $\beta x$ is projected onto the span of $x^c$, the projection of $\alpha\mathbf{1}$ onto the orthohgonal space $Lin^{\perp}(\mathbf{1})$ yields zero as demonstrated in Figure . Moreover, it follows that the angle between $\tilde{x}$ and $y$ is still $\varphi$.

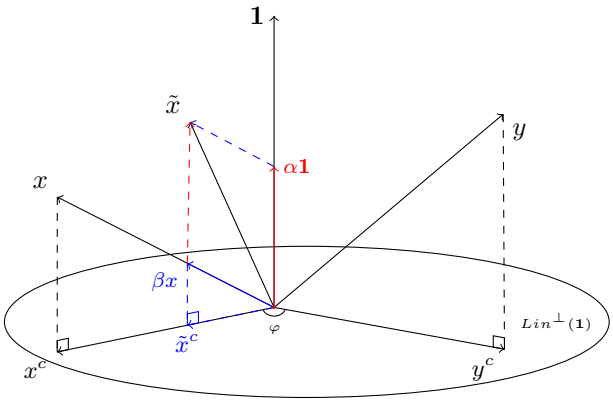



Figure 7: Decomposition and projection of $\tilde{x}$

Finally, putting everything together we finish the proof:

$$\text{sCorr}(x + \alpha\mathbf{1}, y) = \text{sCorr}(x, y)$$

□

*Sample correlation coefficient in simple linear regression*

**Theorem 4.** *A linear regression model with one explanatory variable and constant term*

$$y = \beta_1 + \beta_2 x + \varepsilon$$

Assuming the underlying relationship between $x$ and $y$ to be

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

where $\varepsilon_i$ is an error term the following holds

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

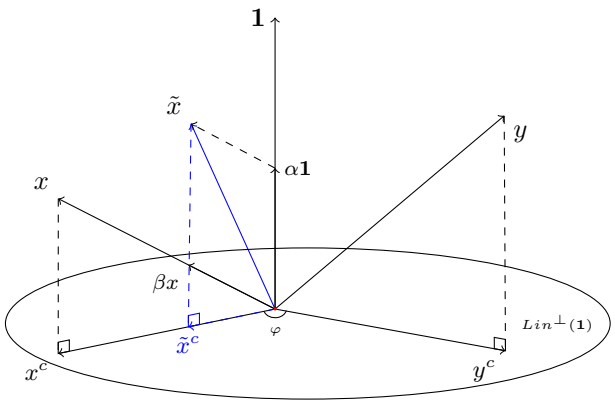



Figure 8: $\text{sCorr}(x + \alpha\mathbf{1}, y) = \text{sCorr}(x, y)$ as the corresponding angles are equal.

*has the property*

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

*Proof.* Firstly, we consider the case when $\hat{\beta}_2 > 0$ and introduce the base picuture of the proof as depicted in Figure . It has been shown earlier that the correlation coefficient represents the angle betweem two random vectors. So in order to complete the proof we need to find the appropriate angles and compare them.

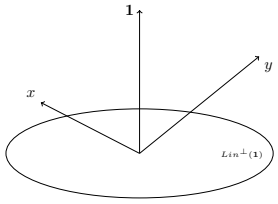
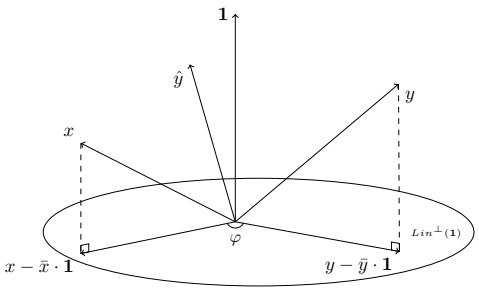



Figure 9: Vectors $x$, $y$ and $\mathbf{1}$.

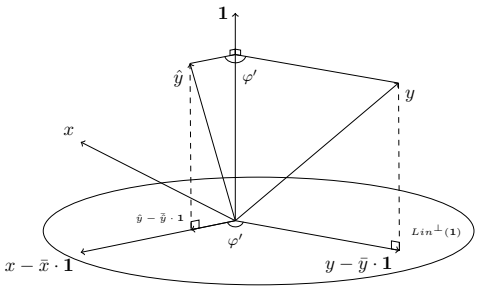
However, it seems to be difficult to compare the angles in the three dimensional space. That is why we start with projecting both $x$ and $y$ onto the space perpendicular to the vector of all ones $\mathbf{1}$ as shown in Figure 10(a). We denote this space as $Lin^{\perp}(\mathbf{1})$. The resulting 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. Note that the projected $\hat{y}$ falls onto tha span of vector $x - \bar{x} \cdot \mathbf{1}$ as it can be decomposed into a sum $ax + b\mathbf{1}$ where $a, b \in \mathbb{R}$. $ax$ is projected in the same way as $x$ and $b\mathbf{1}$ yields zero when projected onto the orthogonal space. The result of this step is shown in Figure 10(b).

Since the projection of $\hat{y}$ lies exactly on the span of vector $x - \bar{x} \cdot \mathbf{1}$, we can conclude that $cos\varphi = \cos\varphi'$ and to put it another way $\text{sCorr}(x, y) =$




(a)




(b)

Figure 10: (a): 'Centred' $x$ and $y$, i.e., projected onto $Lin^{\perp}(\mathbf{1})$; (b): 'Centred' $\hat{y}$, i.e., projected onto $Lin^{\perp}(\mathbf{1})$.

$\text{sCorr}(y, \hat{y})$.

Now consider the case when $\hat{\beta}_2 < 0$. Note that the sign of $\beta_1$ does not influence the correlation coefficient sign. The only difference is that now $\hat{y}$ is projected onto the span of $x - \bar{x} \cdot \mathbf{1}$ and not on this vector itself while the projections of $x$ and $y$ remain the same. Looking at Figure  we deduce that the angle betwween $y$ and $\hat{y}$ is compelement to the angle between $x$ and $y$. Using trigonometric properties, we simplify $\cos(180° - \varphi) = -\cos\varphi$ which in turn implies $\text{sCorr}(x, y) = -\text{sCorr}(y, \hat{y})$.

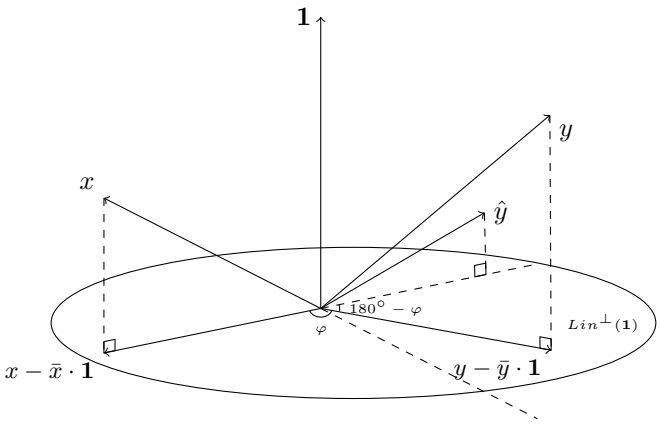



Figure 11: Case of $\beta_2 < 0$.

□

*RSS + ESS = TSS*

**Theorem 5**. *A linear regression model with $n$ observations and $k$ explanatory variables including a constant unit vector*

$$y = X\beta + \varepsilon$$

*has the following property*

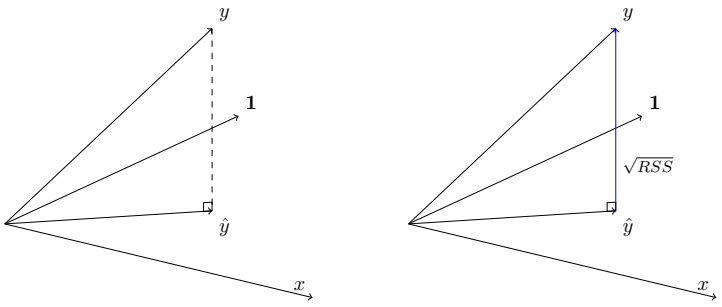
$$RSS + ESS = TSS$$

*where $RSS = \|y - \hat{y}\|_2^2$, $ESS = \|\hat{y} - \bar{y}\|_2^2$, $TSS = \|y - \bar{y}\|_2^2$.*

*Proof.* The proof will be presented for the case of two regressor $x$ and $\mathbf{1}$ in order for the picture to be clear. However, the same logic applies for the case of $k$ regressors.

We start with depicting the vectors $y \in \mathbb{R}^{n-2}$ and $x, \mathbf{1} \in \mathbb{R}^2$. Then we project $y$ onto $Lin(x, \mathbf{1})$ and obtain $\hat{y}$ which is shown in Figure 12(a).

From this picture we can immediately derive $\sqrt{RSS}$ as by definition this is the squared difference between $y$ and $\hat{y}$.




(a) (b)

So as to visualize $ESS$ and $TSS$ we first need to visualize vector of averages $\bar{y}$. Geometrically this means projecting a vector onto a line spanned by vector $\mathbf{1}$.

Now we both project $y$ and $\hat{y}$ onto $\mathbf{1}$ and following the definition obtain $\sqrt{TSS}$ as the difference vector $y - \bar{y}$ and $\sqrt{ESS}$ as the vector $\hat{y} - \bar{y}$.

The final step is to put everything together. Note that since $y - \hat{y}$ is perpendicular to $Lin(x, \mathbf{1})$ it is also perpendicular to $\hat{y} - \bar{y}$ and $\mathbf{1}$ as these vectoros are in $Lin(x, \mathbf{1})$. Then, applying the theorem of three perpendiculars we conclude that the foot of vector $y - \bar{y}$ is the same point as the foot of the vector $\hat{y} - \bar{y}$. Thus, we obtain a right angle triangle and can apply the Pythagorean theorem for the catheti $\sqrt{RSS}$ and $\sqrt{ESS}$ and the hypotenuse $\sqrt{TSS}$:

$$(\sqrt{RSS})^2 + (\sqrt{ESS})^2 = (\sqrt{TSS})^2$$

Consider a regresion model with $n$ observations and $k$ explanatory variables including a constant unit vector

$$y = X\beta + \varepsilon$$

The OLS estimator for the vector of coefficients $\beta$ is

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

and the residual vector is

$$\begin{aligned} \hat{e} &= y - \hat{y} \\ &= y - X\hat{\beta} \\ &= y - X(X^T X)^{-1} X^T y \end{aligned}$$

Then we define residual sum of squares (RSS), explained sum of squares (ESS) and total sum of squares (TSS) as follows:

$$\begin{aligned} RSS &= \|y - \hat{y}\|_2^2 \\ ESS &= \|\hat{y} - \bar{y}\|_2^2 \\ TSS &= \|y - \bar{y}\| \end{aligned}$$

Figure 12: (a): Vectors $y \in \mathbb{R}^{n-2}$ and $\hat{y} \in Lin(x, \mathbf{1})$. (b): Residual sum of squares

Disclosing parentheses and using the fact that $y^T y = \hat{y}^T \hat{y}$

$$\begin{aligned} \hat{y}^T y &= \beta^T X^T y \\ &= y^T X (X^T X)^{-1} X^T y \\ \hat{y}^T \hat{y} &= \beta^T X^T X \beta \\ &= y^T X (X^T X)^{-1} X^T X (X^T X)^{-1} X^T y \\ &= y^T X (X^T X)^{-1} X^T y \end{aligned}$$

we obtain

$$\begin{aligned} RSS &= y^T y - \hat{y}^T \hat{y} \\ ESS &= \hat{y}^T \hat{y} - \hat{y}^T \bar{y} + \bar{y}^T \bar{y} \\ TSS &= y^T y - 2y^T \bar{y} + \bar{y}^T \bar{y} \end{aligned}$$

When putting everything together all the terms cancel out which proves

$$ESS + RSS = TSS$$

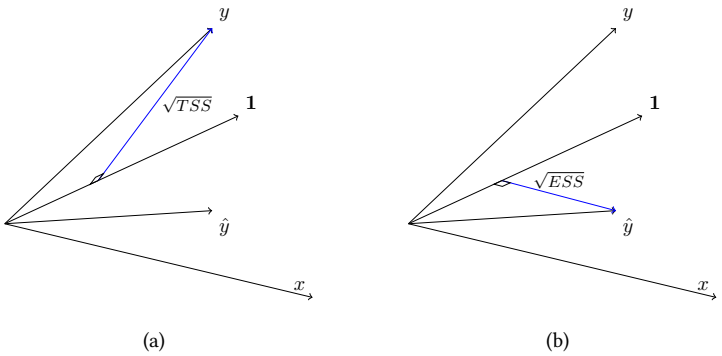



Figure 13: (a): Total sum of squares; (b): Explained sum of squares.

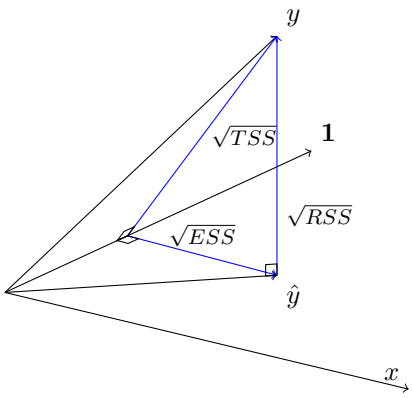



Figure 14: $(\sqrt{RSS})^2 + (\sqrt{ESS})^2 = (\sqrt{TSS})^2$

□

*Determination coefficient*

**Theorem 6.** *A linear regression model with $n$ observations and $k$ explanatory variables including a constant unit vector*

$$y = X\beta + \varepsilon$$

*has the following property*

$$R^2 = \text{sCorr}^2(y, \hat{y})$$

*Proof.* Proving this theorem geometrically means showing that the determination coefficient can be interpreted as some squared angle which happens to be eqaul to the squared angle betwen $y$ and $\hat{y}$.

Consider Figure  from the previous proof. It was shown there that the vectors $y - \bar{y}$, $y - \hat{y}$ and $\hat{y} - \bar{y}$ form a right triangle. Having defined the determination coefficient as

$$R^2 = \frac{ESS}{TSS}$$

$$\begin{aligned}
\text{sCorr}^2(y, \hat{y}) &= \left( \frac{\text{sCov}(y, \hat{y})}{\sqrt{\text{sVar}(y)\,\text{sVar}(\hat{y})}} \right)^2 \\
&= \frac{\text{sCov}(y, \hat{y})\,\text{sCov}(y, \hat{y})}{\text{sVar}(y)\,\text{sVar}(\hat{y})} \\
&= \frac{\text{sCov}(\hat{y} + e, \hat{y})\,\text{sCov}(\hat{y} + e, \hat{y})}{\text{sVar}(y)\,\text{sVar}(\hat{y})} \\
&= \frac{(\text{sCov}(\hat{y}, \hat{y}) + \text{sCov}(e, \hat{y}))(\text{sCov}(\hat{y}, \hat{y}) + \text{sCov}}{\text{sVar}(y)\,\text{sVar}(\hat{y})} \\
&= \frac{\text{sVar}(\hat{y})\,\text{sVar}(\hat{y})}{\text{sVar}(y)\,\text{sVar}(\hat{y})} \\
&= \frac{\text{sVar}(\hat{y})}{\text{sVar}(y)} \\
&= \frac{ESS}{TSS} \\
&= R^2
\end{aligned}$$

we conclude that its geometric interpretaion is

$$R^2 = \frac{ESS}{TSS} = \cos^2 \varphi$$

as shown in Figure .

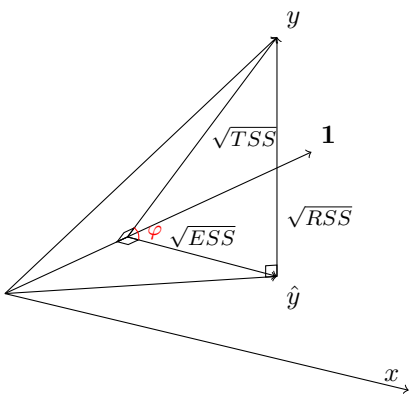



Figure 15: Determination coefficient as squared cos $\varphi$

Recall that the sample correlation coefficient two vectors was defined earlier as the angle between these two vectors. Thus, we conclude that $\text{sCorr}(y, \hat{y})$ is the angle between $y$ and $\hat{y}$ which is also eqaul to $\cos \varphi$. Finally, squaring both sides, we obtain

$$R^2 = \text{sCorr}^2(y, \hat{y})$$

□

*Regression line and point of averages*

**Theorem 7**. *In a linear regression model with one explanatory variable and constant term*

$$y = \beta_1 + \beta_2 x + \varepsilon$$

*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 16(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 16(b).

□

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

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

Premultiplying both sides by $X^T$, we obtain:

$$\begin{aligned}X^T\hat{y} &= X^TX(X^TX)^{-1}X^TX\beta \\ &\quad + X^TX(X^TX)^{-1}X^T\varepsilon \\ &= X^TX\beta + X^T\varepsilon\end{aligned}$$

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:

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

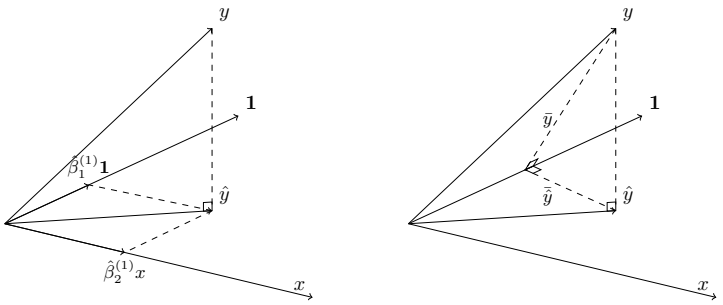



Figure 16: (a): Regression of $y$ on $Lin(\mathbf{1}, x)$; (b): Regression of $y$ and $\hat{y}$ on $\mathbf{1}$.

*Frisch–Waugh–Lovell theorem*

**Theorem 8**. *Consider regression*

$$y = X_1\beta_1 + X_2\beta_2 + u \tag{1}$$

*where $X_{n\times k} = [X_1 X_2]$, i.e. $X_1$ consists of first $k_1$ columns of $X$ and $X_2$ consists of remaining $k_2$ columns of $X$, $\beta_1$ and $\beta_2$ are comfortable, i.e. $k_1 \times 1$ and $k_2 \times 1$ vectors. Consider another regresison*

$$M_1 y = M_1 X_2 \beta_2 + M_1 u \tag{2}$$

*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 $\beta_2$ from regression 1 will be the same as the estimate from regression 2.*

There are two ways to visualize the proof of the Frisch-Waugh-Lovell theorem using geometric concepts. Both are presented below.

*Proof.* 1. Consider 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 17(a). Figure 17(b) depicts $Lin(x, z)$.

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' $y$.

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

From regresison 2 we get the following estimator:

$$\begin{aligned} \hat{\beta}_2 &= ((M_1 X_2)^T M_1 X_2)^{-1} (M_1 X_2)^T M_1 y \\ &= (X_2^T M_1^T M_1 X_2)^{-1} X_2^T M_1^T M_1 y \\ &= (X_2^T M_1 X_2)^{-1} X_2^T M_1 y \end{aligned}$$

As for regresison 1, let us note that due to $y = \hat{y} + \hat{u}$ $y$ can be decomposed as follows:

$$y = Py + My = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + My$$

Premultiplying both sides by $X_2^T M_1$, we obtain:

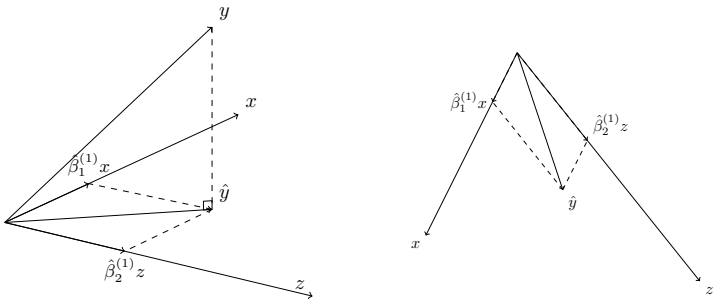
$$\begin{aligned} X_2^T M_1 y &= X_2^T M_1 X_1 \hat{\beta}_1 + X_2^T M_1 X_2 \hat{\beta}_2 + X_2^T M_1 M y \\ &= X_2^T M_1 X_2 \hat{\beta}_2 + X_2^T M_1 M y \\ &= X_2^T M_1 X_2 \hat{\beta}_2 \end{aligned}$$

On the last step we used the fact that

$$\begin{aligned} (X_2^T M_1 M y)^T &= y^T M^T M_1^T X_2 \\ = y^T M M_1 X_2 &= y^T M X_2 = 0^T \end{aligned}$$

Assuming $X_2^T M_1 X_2$ is invertible, we get the same estimator

$$\hat{\beta}_2 = (X_2^T M_1 X_2)^{-1} X_2^T M_1 y$$




(a) (b)

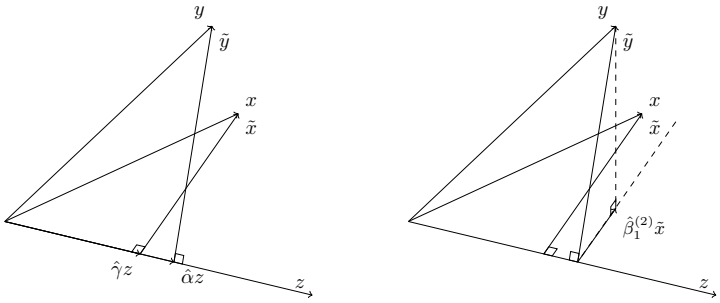
Figure 17: (a): Regression of $y$ on $Lin(x, z)$; (b): $Lin(x, z)$.

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 18(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 18(b)).




(a) (b)

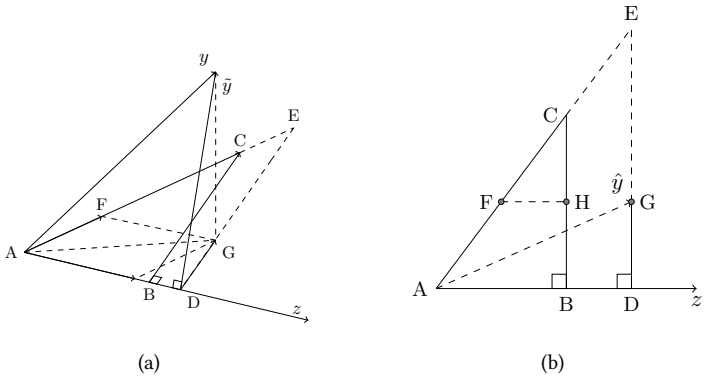
Figure 18: (a): Regression of $y$ on $z$ and of $x$ on $z$; (b): Translation of $\tilde{x}$.

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

In Figure 19(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:

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

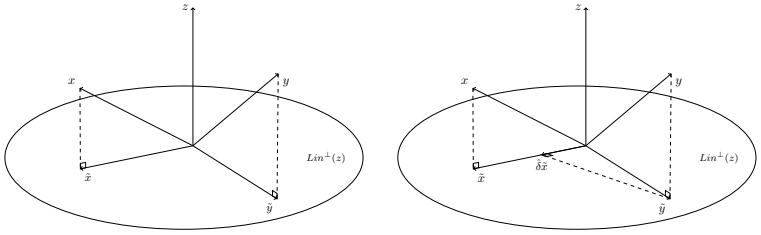
2. Alternatively, we could implement a concept close to the partial correlation. In the same model 3 we wiil treat $z$ vector fixed and again consecutively cleanse the $x$ and $y$ variables by projecting them onto the space




(a) (b)

Figure 19: (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): $Lin(x,z)$.

orthogonal to $z$, i.e., $Lin^{\perp}(z)$ as demonstrated in Figure 20(a). Then we perform a regression of the 'cleansed' $\tilde{y}$ on the 'cleansed' $\tilde{x}$ (see Figure20(b)).




(a) (b)

Figure 20: (a): 'Cleansed' variables $\tilde{x}$ and $\tilde{y}$; (b): 'Cleansed' $\tilde{y}$ regressed on 'cleansed' $\tilde{x}$.

Now we show that the latter regression produces $\hat{\beta}_1$ coefficient which is exatly the coefficient from the 'one-step' regression of original $y$ onto original $x$ and $z$. Recall that the vector $y$ can be split up into a sum of some multiple of $x$ and some multiple of $z$. Since the second term is the orthogonal component its projection yields zero. The multiple of $x$ is equal to $\hat{\beta}_1$ by construction.

Assume that the coefficient at $\tilde{x}$ is some unknown variable $\hat{\delta}$. Then consider the similar triangles in the $Lin(x,z)$. From the proportions we obtain:

$$\frac{CE}{CA} = \frac{CD}{CB} \Leftrightarrow \frac{\hat{\beta}_1 x}{x} = \frac{\hat{\delta}\tilde{x}}{\tilde{x}} \Rightarrow \hat{\beta}_1 = \hat{\delta}$$

□

## *Partial correlation*

### *Definition of partial correlation*

Partial correlation can be defined into two ways. We will provide both definitions and show their equivalence.

Partial correlation is the measure of degree of dependence between two random variables while controlling for the effect of other random variables:

$$\text{pCorr}(X,Y;Z) = \frac{\text{pCov}(X,Y;Z)}{\sqrt{\text{pVar}(X;Z)\,\text{pVar}(Y;Z)}}$$

where $\text{pVar}(X;Z) = \text{Var}(X-\alpha Z)$, $\alpha$ is such a constant that $\text{Cov}(X-\alpha Z, Z) = 0$, and $\text{pCov}(X,Y;Z) = \text{Cov}(X-\alpha Z, Y-\beta Z)$, $\alpha$, $\beta$ are such constants that $\text{Cov}(X-\alpha Z, Z) = \text{Cov}(Y-\beta Z, Z) = 0$

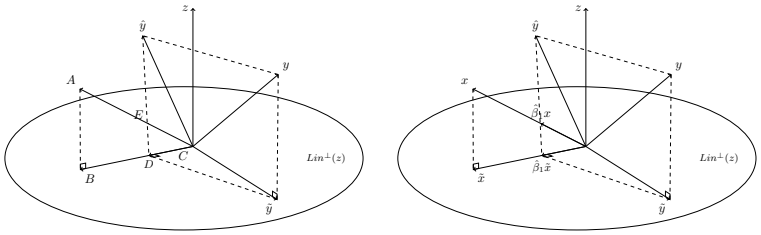



Figure 21: (a): Similar triangles: $\triangle ABC \sim \triangle EDC$; (b): Alternative proof for the Frisch-Waugh-Lovell theorem.

**Definition 1**. Partial correlation between random variables $X$ and $Y$ holding random variable $Z$ fixed is the correlation coefficient between the residuals in regression of $X$ onto $Z$ and the residuals in regression of $Y$ onto $Z$.

Firstly, we project random variable $X$ onto $Z$, which yields $\mathrm{E}(X)$. The residuals in this regression are $X - \mathrm{E}(X)$ — a vector in $Lin^{\perp}(Z)$. We will call this variable 'centered' and label it as $\widetilde{X}$. Repeating this step for $Y$ yields 'centerd' variable $\tilde{Y} = Y - \mathrm{E}(Y) \in Lin^{\perp}(Z)$. The angle between $\widetilde{X}$ and $\widetilde{Y}$ ($\varphi$ in Figure ) is the correaletion coefficient between these 'centred' random variables and the partial correlaiton between the original ones.

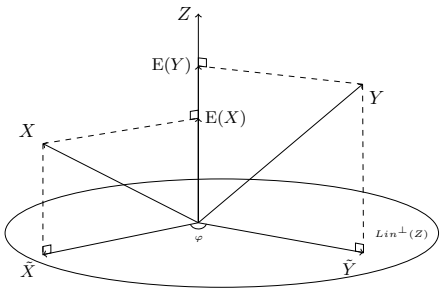



Figure 22: Partial correlation between $X$ and $Y$ while $Z$ is fixed.

**Definition 2**. Partial correlaiton between random variables $X$ and $Y$ holding random variable $Z$ fixed is the geometric mean between the coefficient $\hat{\beta}_{XY}$ in regression

$$\hat{X} = \hat{\beta}_{XY} Y + \hat{\beta}_{XZ} Z$$

and the coefficinent $\hat{\beta}_{YX}$ in regression

$$\hat{Y} = \hat{\beta}_{YX} X + \hat{\beta}_{YZ} Z$$

which has the same sign as the coefficients $\hat{\beta}_{XY}$ and $\hat{\beta}_{YX}$.

Following the definition, we need to start with regressing variable $X$ onto $Y$ and $Z$. Then, the vector we obtained $\hat{X}$ can be broken up into the sum of $\hat{\beta}_{XY} Y$ and $\hat{\beta}_{XZ} Z$. Projecting $\hat{\beta}_{XY} Y$ onto $Lin^{\perp}(Z)$ we results in a vector $\alpha \widetilde{Y}$ where $\widetilde{Y} = Y - \mathrm{E}(Y)$ is the projection of $Y$ onto $Lin^{\perp}(Z)$.

By the properties of similar triangles

$$\frac{\beta_{XY}Y}{Y} = \frac{\alpha\widetilde{Y}}{\widetilde{Y}} \Leftrightarrow \beta_{XY} = \alpha$$

In the same way we perform a regression of $Y$ onto $X$ and $Z$ and repeat the same steps as for $X$. Finally, we get the whole picture:

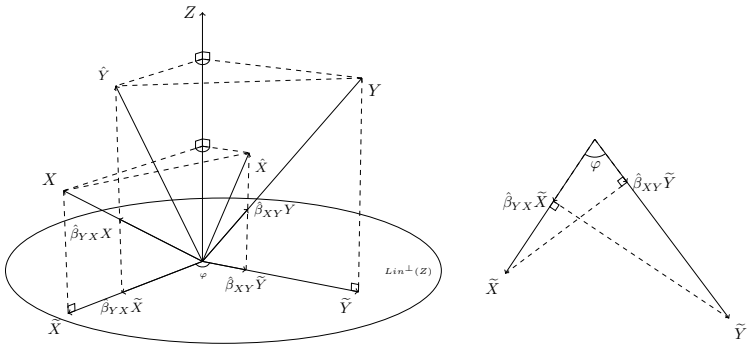



Figure 23: (a): Alternative definition of the partial correlation; (b): $Lin^{\perp}(Z)$.

Having plotted $Lin^{\perp}(Z)$ now we can express $\cos\varphi$ in terms of $\beta_{XY}$ and $\beta_{YX}$

$$\begin{aligned}
\cos\varphi &= \frac{|\beta_{XY}\widetilde{Y}|}{|\widetilde{X}|} = |\beta_{XY}| \\
\cos\varphi &= \frac{|\beta_{YX}\widetilde{X}|}{|\widetilde{Y}|} = |\beta_{YX}| \\
\cos^2\varphi &= |\beta_{XY}\beta_{YX}| \overset{sign(\beta_{XY})=sign(\beta_{YX})}{=} \beta_{XY}\beta_{YX}
\end{aligned} \tag{6}$$

Recall that the angle $\varphi$ can be interpreted as the partial correlation between $X$ and $Y$ holding $Z$ fixed, so it follows form equaitions (6)

$$\text{pCorr}^2(X, Y; Z) = \cos^2\varphi = \beta_{XY}\beta_{YX}$$

*Partial correlation as correlaiton between residuals*

???

**Theorem 9**. *Partial correlaiton between $X$ and $Y$ holding $Z$ fixed is the negative correlaiton coefficient betwenn the residuals $\hat{u}$ in the regeression model*

$$X = \alpha_1 Y + \alpha_2 Z + u$$

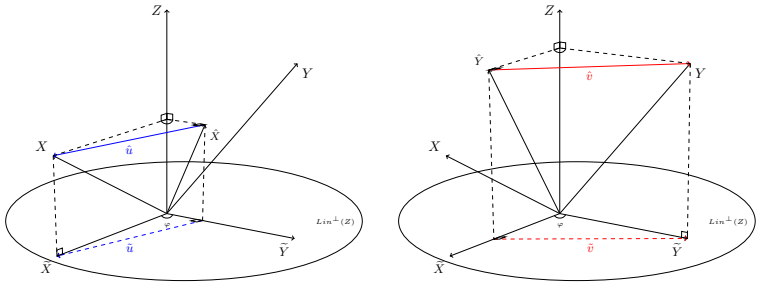
*and the residuals $\hat{v}$ in the model*

$$Y = \beta_1 X + \beta_2 Z + v$$

*Proof.* The first step is to find the residuals in the mentioned regresisions. For example, in order to get $\hat{u}$ we regress $X$ onto $Lin(Y, Z)$ which results

in $\hat{X}$. Then we take the difference $X - \hat{X} = \hat{u}$ and project it as well as $X$ itself onto $Lin^{\perp}(Z)$ as demonstrated in Figure 24(a).

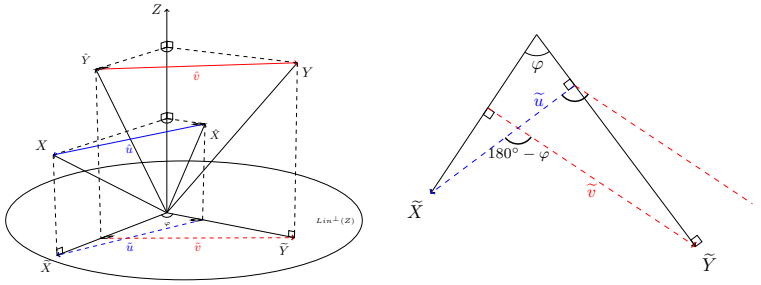
Figure 24(b) shows the same step for obtainig $\hat{v}$.




(a) (b)

Figure 24: (a): $\hat{u}$ form regression of x onto $Y$ and $Z$, $\hat{u}$ projected; (b): $\hat{v}$ from regression of $Y$ onto $X$ and $Z$, $\hat{v}$ projected.

After all we put these figure together and the goal is to find the angle between the red and blue lines which is the same as the angle between the dashed red and blue lines. However, before that we need to apply translating to them to get this angle as shown in Figure 25(b). Finally, using the properties of complementary angles, angles in triangles and parallelograms we conclude that the angle between the residuals $\tilde{u}$ and $\tilde{v}$ is eqaul to $\varphi$.




(a) (b)

Figure 25: (a): The residuals of both regressions; (b): $Lin^{\perp}(Z)$.

□

## *Probability distributions*

### *Normal*

### *Chi-squared*

**Theorem 10**. *Consider a random vector $z \in \mathbb{R}^n$ which components are independent and follow standard normal distribution, $z_i \sim \mathcal{N}(0,1)$. Consider also a fixed $k$-dimensional subspace $L$ in $\mathbb{R}^n$. Let the projection of vector $z$ onto the subspace $L$ be $\hat{z}$ and its length squared $Q$*

$$Q = \|\hat{z}\|^2 = \langle \hat{z}, \hat{z} \rangle = \hat{z}^T \hat{z}$$

Let $z_i \overset{iid}{\sim} \mathcal{N}(0,1)$. Then $Q$ follows the chi-squared distribution with $k$ degrees of freedom if it can be written as

$$Q = z_1^2 + z_2^2 + \ldots + z_k^2.$$

This definition is a particular case of the geometric one. Consider projecting a vector $z = (z_1, z_2, \ldots, z_n)$ form $\mathbb{R}^n$ onto the $k$-dimensional subspace $S$ of vectors which first $k$ coordinates are arbitrary and all the rest are zeros. As a result we would get

$$\hat{z} = (z_1, z_2, \ldots, z_k, 0, \ldots, 0).$$

Squaring the length of the projection, we obtain

$$Q = \|\hat{z}\| = z_1^2 + z_2^2 + \ldots + z_k^2.$$

*Then $Q$ follows the chi-squared distribution with $k$ degrees of freedom.*

*Proof.* First, it can be shown that the projected vector $\hat{z}$ is the original vector $z$ multiplied by the projection matrix $H = X(X^TX)^{-1}X^T$ where the columns of $X$ are fixed linearly independent vectors $x_1, \ldots, x_k$ in $L$ or equivalently $colX = Lin(x_1, \ldots, x_k)$. This matrix is also often referred to as 'hat-matrix'. Then the statement in the theorem can be rewritten as follows:

$$\hat{z}^T\hat{z} = (Hz)^THz = z^TH^THz = z^TH^2z = z^THz,$$

applying the idempotence property in the last step.

Another nice property of the hat-mattix is symmetry. Thus, it can be decomposed as

$$H = PDP^T,$$

where we choose the vectors of matrix $P$ to be unit and orthogonal, and $D = diag\lambda$ where $\lambda_i$ is an eigenvalue of $H$.

Since $H^2 = H$ the eigenvalues are either 0 or 1. Recall that $H$ projects a vector onto $colX$. Then for any $x_i, i = 1, \ldots, k$, $Hx_i = x_i \cdot 1$ since any $x_i$ is already in $colX$. This implies that $\lambda_1 = \ldots = \lambda_k = 1$. There are also $n - k$ vectors in the subspace orthogonal to $colX$. So for any $x_i$, $i = k+1, \ldots, n$, the orthogonal projection yields zero. We conclude that $\lambda_{k+1} = \ldots = \lambda_n = 0$.

Rewritting the theroem statement further, we obtain

$$z^THz = z^TPDP^Tz = (P^Tz)^TD(P^Tz) = \tilde{z}^TD\tilde{z} = \tilde{z}_1^2 + \ldots + \tilde{z}_k^2.$$

Now we explore $\tilde{z}$ given $z \sim \mathcal{N}(0, I)$:

$$\begin{aligned}
&\tilde{z} = P^Tz \\
&\mathrm{E}(\tilde{z}) = \mathrm{E}(P^Tz) = P^T\,\mathrm{E}(z) = 0 \\
&\mathrm{Var}(\tilde{z}) = \mathrm{Var}(P^Tz) = P^T\,\mathrm{Var}(z)(P^T)^T = P^TP = I
\end{aligned}$$

So we conclude that $\tilde{z}_1^2 + \ldots + \tilde{z}_k^2 \sim \chi_k^2$.

□

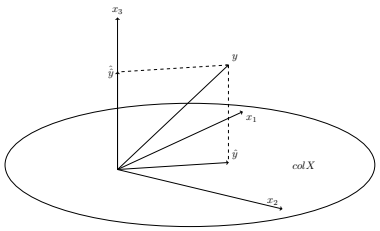



Figure 26: Consider a 3-dimensional example, $colX = Lin(x_1, x_2)$ and $col^\perp X = Lin(x_3)$. $Hx_1 = x_1$ and $Hx_2 = x_2$ since they are in $colX$. However, $Hx_3 = 0$ as $x_3 \perp colX$. Projecting an arbitrary vector onto $colX$ yileds $Hy = \hat{y} \in Lin(x_1, x_2)$ while projecting onto $col^\perp X$ results in $(I - H)y = \hat{\hat{y}} \in Lin(x_3)$.

*Student's*

*t-test*

In a simple linear regression model

$$y = \beta_1 + \beta_2 x + \varepsilon$$

the adjusted t-value $\frac{t}{\sqrt{n-2}}$ when $H_0 : \beta_2 = 0$ is tested can be expreesed in terms of the angle between $y$ and $\hat{y}$ $\varphi$ and is equal to $\mathrm{ctg}\,\varphi$.

Recall that the t-statistic is defined in the following way:

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})}$$

Adjusting this formula for the null hypothesis $H_0 : \beta_2 = 0$, we obtain

$$t = \frac{\hat{\beta}_2}{s.e.(\hat{\beta}_2)} \tag{7}$$

Then, we need to express $s.e.(\hat{\beta}_2)$ in terms of vectors which can be plotted. From standard OLS procedure it follows that

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{8}$$

Since actual $\sigma$ is unknown the estimator will be used instead:

$$\hat{\sigma}^2 = \frac{RSS}{n-2} \tag{9}$$

Substituting (8) and (9) into (7) divided by $\sqrt{n-2}$, we obtain

$$\begin{aligned}
\frac{t}{\sqrt{n-2}} &= \frac{\hat{\beta}_2}{\sqrt{n-2}\widehat{s.e.}(\hat{\beta}_2)} \\
&= \frac{\hat{\beta}_2}{\sqrt{n-2}\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \\
&= \frac{\hat{\beta}_2\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}{\sqrt{n-2}\frac{\sqrt{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}}{\sqrt{n-2}}} \\
&= \frac{\hat{\beta}_2|x^c|}{\sqrt{RSS}}
\end{aligned}$$

where $|x^c| = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ is the length of the centred vector $x$.

Now the result can be demonstrated visually. Again we will project $x$ and $y$ vectors onto the $Lin^{\perp}(\mathbf{1})$ so as to get their centred versions $x^c$ and $y^c$. Then, we perform regression of $y$ onto $Lin(x, \mathbf{1})$ which results in $\hat{y}$. Following that, we project $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ onto $Lin^{\perp}(\mathbf{1})$ which yields $\hat{\beta}_2 x^c$. After all, we translate $\sqrt{RSS}$ onto $Lin^{\perp}(\mathbf{1})$. These steps are demonstrated in Figure 27(a).

Looking at Figure 27(b) which depicts the $Lin^{\perp}(\mathbf{1})$, we derive

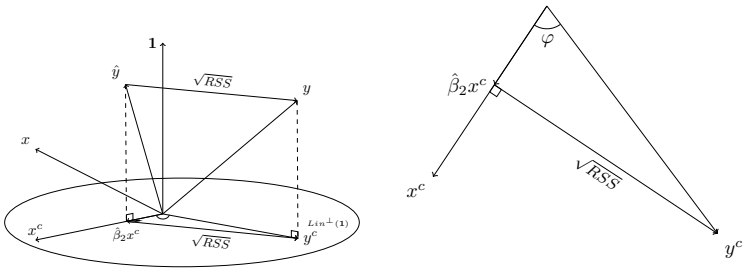
$$\text{ctg}\,\varphi = \frac{\hat{\beta}_2|x^c|}{\sqrt{RSS}} = t$$

### *F-distribution*

### *F-test*

The significance of several coefficients at once can be tested with the F-test. The F-statistic has the following form

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n - k_{UR})}$$




(a) (b)

Figure 27: (a): Regression of $y$ onto $Lin(x, \mathbf{1})$ and appropriate projections; (b): $Lin^{\perp}(\mathbf{1})$.

where indices $R$ and $UR$ stand for the restricted and unrestricted models respectively, $n$ — number of observations, $k$ — number of regressors, $q$ — number of equtions used in the null hypothesis.

Due to plotting limitations, we consider the unrestricted model to be

$$y = \beta_1 + \beta_2 x + u$$

and the restricted model to be

$$y = \alpha_1 + v$$

Note that there was a choice in the restricted models.

We perform both regressions in order to get the ressiduals and plot them in Figure . Adjusted to the degrees of freedom, the ratio can be expressed in terms of the angle between two vectors, $\varphi$, as demonstrated in Figure

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n - k_{UR})} = \text{ctg}^2\,\varphi$$

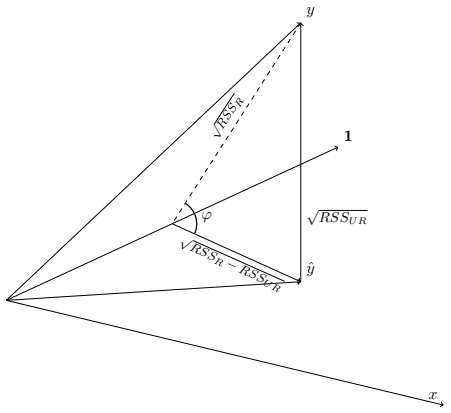



Figure 28: F-statistic as the cotangens squared of $\varphi$