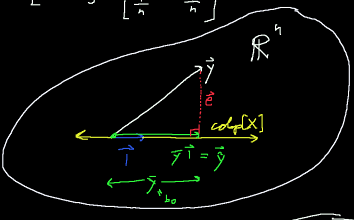
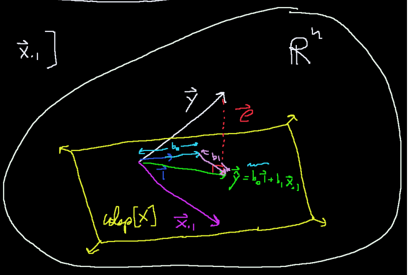
Let's examine the null model, p = 0 so that X = $\left[\vec{1}_n\right] \Rightarrow \vec{b} = b_0 = \bar{y}$

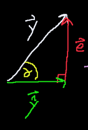
$$H = X\underbrace{(X^TX)^{-1}}_{\underbrace{n}_{\frac{1}{n}}}X^T = \frac{1}{n}\underset{n\times 1}{\vec{1}}\,\underset{1\times n}{\vec{1}^T} = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & & \vdots \\ \frac{1}{n} & & \frac{1}{n} \end{bmatrix}$$

$$\hat{\vec{y}} = H\vec{y} = \begin{bmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix} = \bar{y}\vec{1}_n$$



Consider p = 1 so that X = $\left[\vec{1} \;\; \vec{x}_{\cdot 1}\right]$

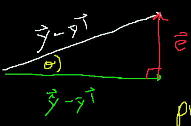




$\Longrightarrow$ (Pythag. Thm) $\|\hat{\vec{y}}\|^2 + \|\vec{e}\|^2 = \|\vec{y}\|^2$, $\cos^2(\theta) = \dfrac{\|\hat{\vec{y}}\|^2}{\|\vec{y}\|^2}$

Is the following illustration accurate? Yes

$\text{proj}_{\text{colsp}[X]}(\vec{1}) = \vec{1}$



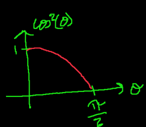
$$\vec{e} := \vec{y} - \hat{\vec{y}} = \vec{y} - \bar{y}\vec{1} + \bar{y}\vec{1} - \hat{\vec{y}} = (\vec{y} - \bar{y}\vec{1}) - (\hat{\vec{y}} - \bar{y}\vec{1})$$

$$\text{proj}_{\text{colsp}[X]}(\vec{y} - \bar{y}\vec{1}) = H(\vec{y} - \bar{y}\vec{1}) = H\vec{y} - \bar{y}H\vec{1}$$
$$= \hat{\vec{y}} - \bar{y}\vec{1} \checkmark$$

$\Downarrow$ Pythag. Thm.

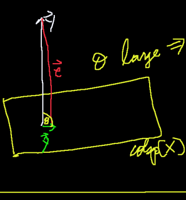
$$\|\vec{y} - \bar{y}\vec{1}\|^2 = \|\hat{\vec{y}} - \bar{y}\vec{1}\|^2 + \|\vec{e}\|^2, \quad R^2 = \frac{SST - SSE}{SST} = \frac{SSR}{SST} = \cos^2(\theta) \in [0, 1]$$

$$\underbrace{\sum(y_i - \bar{y})^2}_{SST} = \underbrace{\sum(\hat{y}_i - \bar{y})^2}_{SSR} + \underbrace{\sum e_i^2}_{SSE}$$



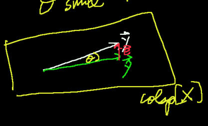
SST = SSR + SSE

sum of squares total, sum of squares regression, sum of squares error



θ large $\Rightarrow R^2 \approx 0$

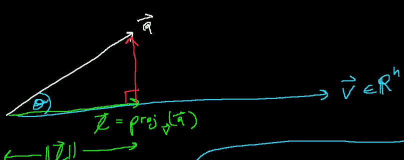
θ small $\Rightarrow R^2 \approx 1$



Back to linear algebra...

By law of cosines,

$$\cos(\theta) = \frac{\vec{a}\cdot\vec{v}}{\|\vec{a}\|\,\|\vec{v}\|} = \frac{\|\vec{\ell}\|}{\|\vec{a}\|} \quad \text{(def. cosine)}$$



$$\Rightarrow \|\vec{\ell}\| = \frac{\vec{a}\cdot\vec{v}}{\|\vec{v}\|}$$

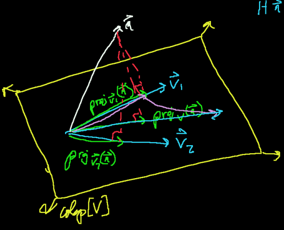
$$\Rightarrow \vec{\ell} = \|\vec{\ell}\|\cdot\frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{a}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{\vec{a}^T\vec{v}\,\vec{v}}{\|\vec{v}\|^2} = \frac{\vec{v}\vec{v}^T\vec{a}}{\|\vec{v}\|^2} = \underbrace{\frac{\vec{v}\vec{v}^T}{\|\vec{v}\|^2}}_{H}\vec{a} = H\vec{a}$$

$$H = \frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T = \left[\frac{v_1}{\|\vec{v}\|^2}\vec{v} \;\middle|\; \frac{v_2}{\|\vec{v}\|^2}\vec{v} \;\middle|\; \cdots \;\middle|\; \frac{v_n}{\|\vec{v}\|^2}\vec{v}\right], \quad \text{rank}[H] = 1$$

$$HH = \left(\frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T\right)\left(\frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T\right) = \frac{1}{\|\vec{v}\|^{\cancel{4}\,2}}\vec{v}\underbrace{\vec{v}^T\vec{v}}_{\cancel{\|\vec{v}\|^2}}\vec{v}^T = \frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T = H \checkmark$$

$$V = [\vec{v}_1 \mid \vec{v}_2] \quad \underbrace{\text{proj}_V(\vec{a})}_{H\vec{a}} \overset{?}{=} \underbrace{\overbrace{\text{proj}_{\vec{v}_1}(\vec{a})}^{H_1\vec{a}} + \overbrace{\text{proj}_{\vec{v}_2}(\vec{a})}^{H_2\vec{a}}}_{} = (H_1 + H_2)\vec{a}$$

Sometimes....



will always project onto colsp[V] but it may not be the correct length (it can over/under count). The correct length gives you the right angle:

$$\text{proj}_V(\vec{a})^T(\vec{a} - \text{proj}_V(\vec{a})) = 0$$

$\|\vec{u} + \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\|\vec{u}\|\|\vec{v}\|\cos(\theta)$ angle between $\vec{u}, \vec{v}$

$$\Rightarrow \text{proj}_V(\vec{a})^T\vec{a} - \text{proj}_V(\vec{a})^T\text{proj}_V(\vec{a})$$
$$= (H_1\vec{a} + H_2\vec{a})^T\vec{a} - (H_1\vec{a} + H_2\vec{a})^T(H_1\vec{a} + H_2\vec{a}) = (\vec{a}^TH_1 + \vec{a}^TH_2)\vec{a} - \|H_1\vec{a} + H_2\vec{a}\|^2$$
$$= \cancel{\vec{a}^TH_1\vec{a}} + \cancel{\vec{a}^TH_2\vec{a}} - \underbrace{\|H_1\vec{a}\|^2}_{\substack{(H_1\vec{a})^T(H_1\vec{a}) \\ \vec{a}^TH_1H_1\vec{a} \\ \vec{a}^TH_1\vec{a}}} - \underbrace{\|H_2\vec{a}\|^2}_{\substack{(H_2\vec{a})^T(H_2\vec{a}) \\ \vec{a}^TH_2H_2\vec{a} \\ \vec{a}^TH_2\vec{a}}} - 2\overbrace{\|H_1\vec{a}\|}^{>0}\overbrace{\|H_2\vec{a}\|}^{>0}\underbrace{\cos(\theta)}_{\in[0,1]}$$

angle between $\vec{v}_1, \vec{v}_2$

The only way to make this expression zero is if cos(θ) = 0 i.e. θ = a right angle. Thus, the full projection is a sum of the component projections if the components are orthogonal.

Let $V = [\vec{v}_1 \mid \vec{v}_2 \mid \ldots \mid \vec{v}_d] \in \mathbb{R}^{n\times d}$, $\forall i,j \;\; \vec{v}_i \cdot \vec{v}_j = 0$

$$V = \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_d \end{bmatrix} \; n$$

$$\Rightarrow \text{proj}_{\text{colsp}[V]}(\vec{a}) = \text{proj}_{\vec{v}_1}(\vec{a}) + \ldots + \text{proj}_{\vec{v}_d}(\vec{a})$$
$$= \frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2}\vec{a} + \ldots + \frac{\vec{v}_d\vec{v}_d^T}{\|\vec{v}_d\|^2}\vec{a}$$
$$= \left(\frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2} + \ldots\ldots + \frac{\vec{v}_d\vec{v}_d^T}{\|\vec{v}_d\|^2}\right)\vec{a} = (\vec{v}_1\vec{v}_1^T + \ldots + \vec{v}_d\vec{v}_d^T)\vec{a} = \underbrace{\left(\sum_{i=1}^{d} v_iv_i^T\right)}_{H}\vec{a}$$

If $\|\vec{v}_1\| = \|\vec{v}_2\| = \ldots = \|\vec{v}_d\| = 1$, i.e. all unit length

$Q = [\vec{v}_1 \mid \ldots \mid \vec{v}_d]$, which is an "orthonormal matrix"

$$\underbrace{Q^TQ}_{d\times d} = \begin{bmatrix} \leftarrow \vec{v}_1^T \rightarrow \\ \vdots \\ \leftarrow \vec{v}_d^T \rightarrow \end{bmatrix}\begin{bmatrix} \uparrow & & \uparrow \\ \vec{v}_1 & \cdots & \vec{v}_d \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & & 1 \end{bmatrix} = I_d$$

$$\underbrace{QQ^T}_{n\times n} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_d \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} \leftarrow \vec{v}_1^T \rightarrow \\ \leftarrow \vec{v}_2^T \rightarrow \\ \vdots \\ \leftarrow \vec{v}_d^T \rightarrow \end{bmatrix} = \vec{v}_1\vec{v}_1^T + \vec{v}_2\vec{v}_2^T + \ldots + \vec{v}_d\vec{v}_d^T = H$$

$$= [A_1 A_2 \cdots A_d]\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_d \end{bmatrix} = A_1B_1 + A_2B_2 + \ldots + A_dB_d$$

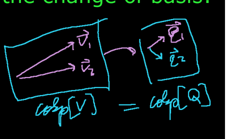
$$\Rightarrow QQ^T = V(V^TV)^{-1}V^T = H$$

where the columns of Q are the orthonormalized columns of V = [v_1 | ... | v_d]. Further colsq[Q] = colsp[V] since the column vectors in Q representes a change of basis of the column vectors of V.

$$\text{proj}_{\text{colsp}[Q]}(\vec{a}) = Q\underbrace{\underbrace{(Q^TQ)}_{I}{}^{-1}}_{I}Q^T = QQ^T$$

How can we convert matrix V to matrix Q? There is a computational algorithm called "Gram-Schmidt" and during the computation, you can collect a matrix that is the change of basis:

$$\underset{n\times d}{V} = \underset{n\times d}{Q}\,\underset{d\times d}{R} \Rightarrow VR^{-1} = Q$$



This is also called Q-R decomposition of a matrix. R will be upper triangular and full rank (and invertible).