**Alberto Andrés Valdés González.**
**Degree:** Mathematical Engineer.
**Work position:** Data Scientist.
**Mail:** anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
**Location:** Santiago, Chile.

---

# Linear Regressions

Statistical and Optimization based approach

# Modeling

We can present many ways of modeling Linear Regressions and all these ways are equivalents. However, here we will introduce only two perspectives: The based in Statistics and the based in Optimization.

## Statistical based approach

We consider the random variable $Y$ and the data $X$ which satisfy the following relation:

$$Y = \beta \cdot X + \alpha + \epsilon$$

Where $\epsilon$ are a **normal random variable** which distributes $\epsilon \sim N(0, \sigma^2)$.

On the other hand $\beta, \alpha$ and $\sigma$ are **poblational parameters** we have to estimate.

The estimators for these parameters we will denote as $\hat{\beta}, \hat{\alpha}$ and $\hat{\sigma}$ respectively.

---

As we know in the real world, we don't see **random variables** but we see **observations** of random variables.

Let's consider the independent observations $y_1, ..., y_n, \epsilon_1, ..., \epsilon_n$ and the data $x_1, ..., x_n$. We have the following relations:

$$y_i = \beta x_i + \alpha + \epsilon_i \qquad \forall\ i \in \{1, ..., n\}$$

How $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ then $y_i \overset{iid}{\sim} N(\beta \cdot x_i + \alpha, \sigma^2)$.

We introduce the Likelihood function:

$$L(\beta, \alpha, \sigma) = \mathbb{P}(y_1, ..., y_n | x_1, ..., x_n, \beta, \alpha, \sigma) = \prod_{i=1}^{n} \mathbb{P}(y_i | x_i, \beta, \alpha, \sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp\left(-\frac{(y_i - [\beta \cdot x_i + \alpha])^2}{2\sigma^2}\right)$$

The values of $\hat{\beta}, \hat{\alpha}$ and $\hat{\sigma}$ are the arguments maximize $L(\beta, \alpha, \sigma)$.

A useful fact to maximize the likelihood function is that maximize a positive function is equivalent to maximize the logaritm of that function. Thus we introduce the log-likelihood function:

$$l(\beta, \alpha, \sigma) = ln(L(\beta, \alpha, \sigma)) = \sum_{i=1}^{n} \left(-\frac{1}{2} \cdot ln(2\pi) - ln(\sigma) - \frac{(y_i - [\beta \cdot x_i + \alpha])^2}{2\sigma^2}\right)$$

$$= -\frac{n \cdot ln(2\pi)}{2} - n \cdot ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - [\beta \cdot x_i + \alpha])^2$$

To obtain the values $\hat{\beta}, \hat{\alpha}$ we have to maximice $l(\beta, \alpha, \sigma)$ which es equivalent to maximize:

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - [\beta \cdot x_i + \alpha])^2$$

Which es equivalent to minimize:

$$\boxed{\sum_{i=1}^{n} (y_i - [\beta \cdot x_i + \alpha])^2} \quad (1)$$

## Optimization based approach

Let's consider the data $x_1, ..., x_n, y_1, ..., y_n$. We will asume that satisfy the relation:

$$y_i = \beta \cdot x_i + \alpha \qquad \forall\ i \in \{1, ..., n\}$$

However, actually this relation have a error considered and the following equations are fulfilled.

$$y_i = \beta \cdot x_i + \alpha + \epsilon_i \qquad \forall\ i \in \{1, ..., n\}$$

Where $\epsilon_i$ are the error component.

In this approach we want to minimice the mean square error (MSE), and the arguments that minimice the MSE are $\beta^*$ and $\alpha^*$.

$$MSE(\beta, \alpha) = \frac{1}{n}\sum_{i=1}^{n}(y_i - [\beta \cdot x_i + \alpha])^2$$

We can see that minimice MSE are equivalent to minimize:

$$\boxed{\sum_{i=1}^{n}(y_i - [\beta \cdot x_i + \alpha])^2} \quad (2)$$

---

As we can see by (1) and (2) the **both approaches are equivalents**.

---

# Probabilitiy Review

Let $(\Omega, \zeta, \mathbb{P})$ a probability space and $X$ a **random variable** over it.

a) If $X_+ \in L^1(\Omega, \zeta, \mathbb{P})$ or $X_- \in L^1(\Omega, \zeta, \mathbb{P})$ the expected value of $X$ its defined as follow:

$$\mathbb{E}(X) = \int X d\mathbb{P} = \int X dF_X$$

**Note:** $X = X_+ - X_-$ where $X_+ = \text{máx}\{X, 0\} = \frac{|X|+X}{2}$ and $X_- = \text{máx}\{-X, 0\} = \frac{|X|-X}{2}$.

b) For all $k \in \mathbb{N} : k \geq 1$ the $k$-th moment is given by $\mathbb{E}(X^k)$. We say $X$ has $k$-th moment if $X \in L^k(\Omega, \zeta, \mathbb{P})$.

c) If $X, Y \in L^2(\Omega, \zeta, \mathbb{P})$ we define its covariance as follows:

$$Cov(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

**Note:** $\mathbb{V}(X) = Cov(X, X)$, $\sigma_X = \sqrt{\mathbb{V}(X)}$ and $\rho_{X,Y} = \dfrac{Cov(X, Y)}{\sigma_X(X) \cdot \sigma(Y)}$.

---

# Estimators

Some **unbiased** estimators to considers are:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$$S_X = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{X,Y} = \frac{1}{(n-1)} \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

---

# Linear relation between random variables

If $X$ and $Y$ are random variables which satisfy the next relation:

$$Y = m \cdot X + n$$

then:

$$\rho_{X,Y} = \frac{Cov(X, m \cdot X + n)}{\sqrt{\mathbb{V}(X)} \cdot \sqrt{\mathbb{V}(m \cdot X + n)}} = \frac{m \cdot Cov(X, X)}{\sqrt{\mathbb{V}(X)} \cdot \sqrt{m^2 \cdot \mathbb{V}(X)}} = \frac{m \cdot \mathbb{V}(X)}{|m| \cdot \mathbb{V}(X)} = \frac{m}{|m|} = sign(m)$$

Thus, to see how strong the linearity is between two random variables, the important thing is to see the magnitude of the **absolute value of correlation**.

---

# Some facts about $\hat{\beta}$ and $\hat{\alpha}$

We can deduce from (1) and (2) that the values of $\hat{\beta}$ and $\hat{\alpha}$ satisfy the next:

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \quad (3)$$

$$\boxed{\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}} \quad (4)$$

**First Fact:** If instead of $x_i$ and $y_i$ we consider its normalizations, i.e:

$$u_i = \frac{x_i - \bar{x}}{\sqrt{S_X}}, \quad v_i = \frac{y_i - \bar{y}}{\sqrt{S_Y}}$$

then, we have:

$$\bar{u} = 0, \quad \bar{v} = 0$$

Therefore:

$$\hat{\beta} = \frac{\sum_{i=1}^n (u_i - \bar{u}) \cdot (v_i - \bar{v})}{\sum_{i=1}^n (u_i - \bar{u})^2} = \frac{\sum_{i=1}^n u_i \cdot v_i}{\sum_{i=1}^n u_i^2} = \frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sqrt{S_X}}\right) \cdot \left(\frac{y_i - \bar{y}}{\sqrt{S_Y}}\right)}{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sqrt{S_X}}\right)^2}$$

$$= \frac{\sqrt{S_X}^2}{\sqrt{S_X} \cdot \sqrt{S_Y}} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sqrt{S_X}^2}{\sqrt{S_X} \cdot \sqrt{S_Y}} \cdot \frac{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \frac{\sqrt{S_X}^2}{\sqrt{S_X} \cdot \sqrt{S_Y}} \cdot \frac{S_{X,Y}}{S_X} = \frac{S_{X,Y}}{\sqrt{S_X} \cdot \sqrt{S_Y}} = \hat{\rho}_{X,Y}$$

$\Rightarrow$

$$\boxed{\hat{\beta} = \hat{\rho_{X,Y}}}$$

On the other hand:

$$\hat{\alpha} = \bar{v} - \hat{\beta} \cdot \bar{u} = 0 - \hat{\rho}_{X,Y} \cdot 0 = 0$$

$\Rightarrow$

$$\boxed{\hat{\alpha} = 0}$$

---

**Second Fact:** If instead of $x_i$ and $y_i$ we consider its scaling by its desviations, i.e:

$$u_i = \frac{x_i}{\sqrt{S_X}}, \quad v_i = \frac{y_i}{\sqrt{S_Y}}$$

then, we have:

$$\bar{u} = \frac{\bar{x}}{\sqrt{S_X}}, \quad \bar{v} = \frac{\bar{y}}{\sqrt{S_Y}} \quad \Rightarrow \quad u_i - \bar{u} = \frac{(x_i - \bar{x})}{\sqrt{S_X}}, \quad v_i - \bar{v} = \frac{(y_i - \bar{y})}{\sqrt{S_Y}}$$

Therefore:

$$\hat{\beta} = \frac{\sum_{i=1}^n (u_i - \bar{u}) \cdot (v_i - \bar{v})}{\sum_{i=1}^n (u_i - \bar{u})^2} = \frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sqrt{S_X}}\right) \cdot \left(\frac{y_i - \bar{y}}{\sqrt{S_Y}}\right)}{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sqrt{S_X}}\right)^2}$$

$$= \frac{\sqrt{S_X}^2}{\sqrt{S_X} \cdot \sqrt{S_Y}} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sqrt{S_X}^2}{\sqrt{S_X} \cdot \sqrt{S_Y}} \cdot \frac{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \frac{\sqrt{S_X}^2}{\sqrt{S_X} \cdot \sqrt{S_Y}} \cdot \frac{S_{X,Y}}{S_X} = \frac{S_{X,Y}}{\sqrt{S_X} \cdot \sqrt{S_Y}} = \hat{\rho}_{X,Y}$$

$\Rightarrow$

$$\boxed{\hat{\beta} = \hat{\rho_{X,Y}}}$$

On the other hand:

$$\hat{\alpha} = \bar{v} - \hat{\beta} \cdot \bar{u} = \frac{\bar{y}}{\sqrt{S_Y}} - \hat{\beta} \cdot \frac{\bar{x}}{\sqrt{S_X}} = 0 - \hat{\rho}_{X,Y} \cdot 0 = 0$$

$\Rightarrow$

$$\boxed{\hat{\alpha} = 0}$$

---

**Third Fact:** Using the (3) formula, we have that:

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{X,Y}}{S_X} = \frac{\sqrt{S_Y}}{\sqrt{S_Y}} \cdot \frac{S_{X,Y}}{\sqrt{S_X} \cdot \sqrt{S_X}}$$

$$= \frac{\sqrt{S_Y}}{\sqrt{S_X}} \cdot \frac{S_{X,Y}}{\sqrt{S_Y} \cdot \sqrt{S_X}} = \frac{\sqrt{S_Y}}{\sqrt{S_X}} \cdot \hat{\rho}_{X,Y}$$

$\Rightarrow$

$$\boxed{\hat{\beta} = \frac{\sqrt{S_Y}}{\sqrt{S_X}} \cdot \hat{\rho}_{X,Y}}$$

For this if two variables are strongly lineal relationated $\Rightarrow |\hat{\rho}_{X,Y}| \approx 1 \Rightarrow |\hat{\beta}| \approx \frac{\sqrt{S_Y}}{\sqrt{S_X}}$.

---

**Fourth Fact:** If we have the next linear regression:

$$\hat{y}_i = \hat{\beta} \cdot x_i + \hat{\alpha} \quad \forall i \in \{1, ..., n\} \quad (*)$$

And now we want to make other linear regression, but instead of consider $y_i, x_i$, we want to consider $v_i, u_i$, where:

$$v_i = c_1 \cdot y_i + c_2 \qquad u_i = d_1 \cdot x_i + d_2 \qquad \forall i \in \{1, ..., n\}$$

$$\hat{v}_i = \hat{\beta}_2 \cdot u_i + \hat{\alpha}_2 \qquad \forall i \in \{1, ..., n\}$$

Then the relation of $\hat{\beta}$ and $\hat{\alpha}$ with $\hat{\beta}_2$ and $\hat{\alpha}_2$ are the same if we only had replaced in $(*)$ with $x_i$ and $y_i$ in function of $u_i$ and $v_i$ respectively.

---

# Inference over $\hat{\beta}$

We know that $y_i \overset{iid}{\sim} N(\beta \cdot x_i + \alpha, \sigma^2)$ and $x_i$ are data. Also we know that:

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \quad (3)$$

We going to rewrite this formula using the next:

$$\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y}) = \sum_{i=1}^{n}(x_i \cdot y_i - \bar{x} \cdot y_i - x_i \cdot \bar{y} + \bar{x} \cdot \bar{y})$$

$$= \sum_{i=1}^{n} x_i \cdot y_i - \bar{x} \cdot \sum_{i=1}^{n} y_i - \bar{y} \cdot \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x} \cdot \bar{y}$$

$$= \sum_{i=1}^{n} x_i \cdot y_i - \bar{x} \cdot n \cdot \bar{y} - \bar{y} \cdot n \cdot \bar{x} + n \cdot \bar{x} \cdot \bar{y} = \sum_{i=1}^{n} x_i \cdot y_i - \bar{x} \cdot n \cdot \bar{y}$$

$$= \sum_{i=1}^{n} x_i \cdot y_i - \bar{x} \cdot \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} \bar{x} \cdot y_i$$

$$= \sum_{i=1}^{n}(x_i \cdot y_i - \bar{x} \cdot y_i) = \sum_{i=1}^{n}(x_i - \bar{x}) \cdot y_i$$

$\Rightarrow$

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \quad (5)$$

## Expected value of $\hat{\beta}$

$$\mathbb{E}[\hat{\beta}] \overset{(3)}{=} \mathbb{E}\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] = \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\mathbb{E}\left[\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})\right]$$

$$= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}\mathbb{E}\left[(x_i-\bar{x})\cdot(y_i-\bar{y})\right] = \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot\mathbb{E}\left[y_i-\bar{y}\right]$$

$$= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot\left(\mathbb{E}\left[y_i\right]-\mathbb{E}\left[\bar{y}\right]\right)$$

$\Rightarrow$

$$\boxed{\mathbb{E}[\hat{\beta}] = \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot\left(\mathbb{E}\left[y_i\right]-\mathbb{E}\left[\bar{y}\right]\right)} \qquad (6)$$

We know $\mathbb{E}\left[y_i\right] = \beta\cdot x_i+\alpha$, thus:

$$\mathbb{E}\left[\bar{y}\right] = \mathbb{E}\left[\frac{\sum_{j=1}^{n}y_j}{n}\right] = \frac{1}{n}\cdot\mathbb{E}\left[\sum_{j=1}^{n}y_j\right] = \frac{1}{n}\cdot\sum_{j=1}^{n}\mathbb{E}\left[y_j\right] = \frac{1}{n}\cdot\sum_{j=1}^{n}(\beta\cdot x_j+\alpha) = \beta\cdot\bar{x}+\alpha$$

$\Rightarrow$

$$\mathbb{E}\left[y_i\right]-\mathbb{E}\left[\bar{y}\right] = (\beta\cdot x_i+\alpha)-(\beta\cdot\bar{x}+\alpha) = \beta\cdot(x_i-\bar{x})$$

Substituting this in (6) we obtain:

$$\mathbb{E}[\hat{\beta}] = \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot\left(\mathbb{E}\left[y_i\right]-\mathbb{E}\left[\bar{y}\right]\right)$$

$$= \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot\beta\cdot(x_i-\bar{x}) = \frac{\beta}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2 = \beta$$

$\Rightarrow$

$$\boxed{\mathbb{E}[\hat{\beta}] = \beta} \qquad (7)$$

---

## Variance of $\hat{\beta}$

$$\mathbb{V}[\hat{\beta}] \overset{(5)}{=} \mathbb{V}\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] = \frac{1}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2}\cdot \mathbb{V}\left[\sum_{i=1}^{n}(x_i-\bar{x})\cdot y_i\right]$$

$$\overset{(iid)}{=} \frac{1}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2}\cdot\sum_{i=1}^{n}\mathbb{V}\left[(x_i-\bar{x})\cdot y_i\right] = \frac{1}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\mathbb{V}\left[y_i\right]$$

$$= \frac{1}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sigma^2 = \frac{\sigma^2}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$\Rightarrow$

$$\boxed{\mathbb{V}[\hat{\beta}] = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \tag{8}$$

---

## Statistical for Hypothesis Test

If we have to test the next hypotesis:

$$H_0 : \beta = \beta_0$$

We use the statistical:

$$T = \frac{\hat{\beta}-\beta_0}{\sqrt{\mathbb{V}(\hat{\beta})}} = \frac{\hat{\beta}-\beta_0}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} = \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\left(\frac{\hat{\beta}-\beta_0}{\sigma}\right)$$

Often we dont know the value of $\sigma$, for this reason we use $\hat{\sigma}$ and the statistical finally is:

$$T = \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\left(\frac{\hat{\beta}-\beta_0}{\hat{\sigma}}\right) \sim N(0,1)$$

---