# Regression

## Lien entre test t et régression

Il est connu que $b_1 = \frac{\sum\left[(X_i-\bar{X})(Y_i-\bar{Y})\right]/(n-1)}{S_X^2} = \frac{S_{XY}}{S_X^2} = r_{XY} \times \frac{S_Y}{S_X}$

Par ailleurs, il existe un lien entre la statistique $t$ et la statistique $r$:

$$t(n-2) = \frac{r_{XY}}{\sqrt{\frac{1-r^2}{n-2}}} \leftrightarrow |r_{XY}| = \left|\frac{t(n-2)}{\sqrt{n+t^2-2}}\right|$$

Donc, il existe un lien direct entre le coefficient $b_1$ et $t$:

$$b_1 = \frac{t(n-2)}{\sqrt{n+t^2-2}} \times \frac{S_Y}{S_X}$$

## Et dans le cas particulier où le prédicteur est catégoriel dichotomique?

La relation précédemment énoncée reste exacte.

Note: on sait aussi qu'à condition de définir les modalités du facteur par des codes contrastes, on peut simplifier le calcul de $b_1$: $b_1 = \frac{\sum \lambda_k \bar{Y_k}}{\sum \lambda_k^2}$ Mais cette information ne me semble pas pertinente à ce stade. Elle le sera peut-être plus tard, quand j'aurai fait le lien entre les stat de Student et de Welch.

## Réfléchir à ce que je vais pouvoir faire avec cette information.

Suite de la réflexion : - Exprimer t en fonction de beta? L'idée est de comprendre comment on arrive à la conclusion que beta diffère significativement de 0, précisément. Pe que si j'arrive à exprimer t en fonction de beta, ça m'éclairera? Y réfléchir