STT 1000 - STATISTIQVES

ARTHVR CHARPENTIER

Correlation

file:///Users/acharpen/Downloads/correlation.pdf

FIGS/reg-galton.jpg

Galton, regression towards mediocrity in hereditary stature, 1886

Pearson's Correlation

$$\rho_{XY} = \text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$

$$\rho_{XY} = \frac{\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\mathbb{E}(X^2)-\mathbb{E}(X)^2}\cdot\sqrt{\mathbb{E}(Y^2)-\mathbb{E}(Y)^2}}$$

Note that $\rho_{XY} \in [-1,+1]$.

The empirical version is

$$r_{xy} \overset{\text{def}}{=} \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

Pearson's Correlation

- $\rho_{XY} = +1$ means that $X = aY + b$ with $a > 0$
- $\rho_{XY} = -1$ means that $X = aY + b$ with $a < 0$

In python

```
> from scipy.stats import pearsonr
> pearsonr(x, y)
```

and in R

```
> cor(x, y, method "pearson")
```

Régression Linéaire (courte introduction)