Inference on r

To develop a measure of association, or correlation, for two random variables X, Y, we start by standardizing them

$$X' = \frac{X - \mu_X}{\sigma_X}, \quad Y' = \frac{Y - \mu_Y}{\sigma_Y}$$

Each of X', Y' is free of units of measurement so their product is free of units of measurement. The expected value of their product, which is also the covariance, is a measure of association between X and Y called the population correlation coefficient.

$$\rho = E\left(\frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y} \right)$$

$\rho > 0$ when (X, Y) are both large or both small with high probability

$\rho < 0$ when X (or Y) is large and Y (X) is small

$$-1 \leq \rho \leq 1$$

$\rho = \pm 1$ perfect linear correlation

An estimator of ρ is $\boxed{r = \frac{1}{n-1} \sum_{i=1}^{n} X_i' Y_i'}$

$\Updownarrow$

r = sample correlation coefficient $\boxed{r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{S_X} \right) \cdot \left(\frac{Y_i - \bar{Y}}{S_Y} \right)}$

Fact:

$$t = (r - \rho) \cdot \sqrt{\frac{n-2}{(1-r^2) \cdot (1-\rho^2)}} \sim t(n-2)$$

Hypothesis Tests for Correlation

$H_0: \rho = \rho_0$

$H_a: \rho > \rho_0$ (a)

$\rho < \rho_0$ (b)

$\rho \neq \rho_0$ (c)

$$t^* = (r-\rho_0)\cdot\sqrt{\frac{n-2}{(1-r^2)(1-\rho^2)}} \overset{H_0}{\sim} t(n-2)$$

$$\text{p-value} = \begin{cases} P(T \geq t^*) & \text{(a)} \\ P(T \leq t^*) & \text{(b)} \\ 2\min\{P(T \geq t^*), P(T \leq t^*)\} & \end{cases}$$

OR equivalently fix a level of significance α and reject H_0 in favor of H_a

(a) if $t^* \geq t_{\alpha, n-2}$

(b) if $t^* \leq t_{1-\alpha, n-2} = -t_{\alpha, n-2}$

(c) if $t^* \geq t_{\frac{\alpha}{2}, n-2}$ or $t^* \leq -t_{\frac{\alpha}{2}, n-2}$

THE END!!