

Külön statisztikai próbával kell eldönteni, hogy ha $r \neq 0$, ez csak a véletlennek tulajdonítható, vagy annak, hogy X és Y nem függetlenek.

Nullhipotézis: $H_0: \rho = \rho_0$

Ellenhipotézis: $H_1: \rho \neq \rho_0$

Próbastatisztika: $t = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}}$ szabadsági fok: $n-2$

Kritikus tartomány: $K: \left\{ |t| > t_{1-\frac{\varepsilon}{2}} \right\}$

Ha $t \in K$ a nullhipotézis, azaz a függetlenség hipotézisét elvetjük, ellenkező esetben megtartjuk.

A t értékének kiszámítása megtakarítható, ha olyan táblázatot készítünk, amely adott ε tévedési valószínűséghez megadja azt a legnagyobb r értéket, amely független változók esetén pusztán a véletlen hatására is felléphet. (Függelék X. táblázat)

Megjegyzés: Ha a korrelációs együttható 0, akkor X és Y még nem feltétlenül függetlenek.

4.7.2. Kétváltozós lineáris regresszió

Feltételezésünk szerint X és Y között lineáris kapcsolat van, azaz

$$E(Y|X=x) = \alpha + \beta x$$

Az a feladat, hogy a tapasztalati adatok alapján α -ra és β -ra megfelelő becslést adjunk.

A becslésre több módszert dolgoztak ki, de leginkább elterjedt a "legkisebb négyzetek" módszere. E módszer alapján α becslt értékét "a"-t és β becslt értékét "b"-t úgy határozzuk meg, hogy az $y=a+bx$ egyenes és a megfigyelt értékeket jelentő pontok közötti függőleges távolságok négyzeteinek összege a lehető legkisebb legyen.