Inferentie voor Correlaties, Albers (2015)

Introductie

De correlatie coëfficiënt, ρ, meet de kracht en richting van de lineaire relatie tussen twee variabelen. Wanneer er geen relatie is, ρ = 0, wanneer er een perfecte positieve relatie is, ρ =1, wanneer er een perfecte negatieve relatie is, ρ = -1. De formule voor de steekproef correlatie coëfficiënt is:

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \cdot \frac{y_i - \bar{y}}{s_y}\right) \qquad (72)$$

We zullen nu de basis conclusies van de correlatie coëfficiënten bespreken.

Hypothese toetsen voor

De steekproef correlatie coëfficiënt, *r*, vertelt ont iets over de sterkte en richting van de lineaire relatie tussen x en y. We moeten nu gaan kijken wat de steekproef correlatie coëfficiënt ons kan vertellen over de correlatie coëfficiënt, . De hypotheses zijn als volgt:

$H_0: \rho = 0$
$H_1: \rho \neq 0$

There is a strong relation between the correlation coefficient and the slope β_1 of the simple linear regression equation, y = $\beta_0 + \beta_1$x + ε. This gives the following formula:

$$\beta_1 = \rho \frac{\sigma_y}{\sigma_x} \leftrightarrow \rho = \beta_1 \frac{\sigma_x}{\sigma_y} \qquad (73)$$

So you can see that the correlation coefficient is 0 when β_1 is 0. So you can also name the hypotheses, $H_0: \beta_1 = 0$. The test is performed through the test statistic:

$$t = \frac{b_1}{SE_{b1}} \qquad (74)$$

Onder de nul hypothese, volgt de t een t-verdeling met n-2 vrijheidsgraden. Dus de alternatieve formule om t te berekenen is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (75)$$

Betrouwbaarheidsintervallen voor

Voor de meeste parameters word het betrouwbaarheidsinterval berekend met: (estimator) (critical value) x (standard error).

Deze berekening van het betrouwbaarheidsinterval heeft alleen wel een normale verdeling nodig. Maar de verdeling van *r* is niet normaal verdeeld om rho. Om dit probleem op te lossen en een betrouwbaarheidsinterval te kunnen berekenen,, gebruiken we de *Fisher z-transformation*. Deze transformatie zorgt ervoor dat alle waarden wel een normale verdeling volgen. De Fisher z-transformation transformeert een *r*- waarde in een nieuwe waarde, weergegeven door *r* ‘. (Of als z or r_z):

$$r' = 0.5 \log\left(\frac{1+r}{1-r}\right) \qquad (76)$$

Waarden van *r* dicht bij 1 zullen meer worden veranderd door de transformatie dan waarden dicht bij 0. Na de transformatie is de verdeling van de correlatie coëfficiënten normaal met het gemiddelde, ρ', en een variatie = $\frac{1}{n-3}$. Dit kan worden gebruikt om het betrouwbaarheidsinterval te berekenen:

De $(1-\alpha)$ % CI voor ρ_z is gegeven door:

De $(1 - \alpha)$ % CI voor ρ_z is gegeven door:

$$(r' - z^{*}_{\alpha(2)} \frac{1}{\sqrt{n-3}},\ r' + z^{*}_{\alpha(2)} \frac{1}{\sqrt{n-3}}) \quad (77)$$

Maar we zijn niet geïnteresseerd in ρ', we hebben dit alleen gebruikt om een interval van ρ te krijgen. We moeten het betrouwbaarheidsinterval van ρ ' veranderen in het betrouwbaarheidsinterval van ρ. De omgekeerde transformatie is:

$$r = \frac{e^{2r'} - 1}{e^{2r'} + 1} \quad (78)$$

Een toets om twee ρ 's te vergelijken.
Wanneer we twee groepen willen vergelijken hebben we de volgende hypothesen:
$H_1 : \rho_1 \neq \rho_2$ or $H_1 : \rho'_1 \neq \rho'_2$

Voor de laatste hypothese moeten we de Fischer transformatie uitvoeren. Voor de hypothese is dit the toetsingsgrootheid:

$$Z = \frac{r'_1 - r'_2}{\sqrt{\frac{1}{n1-3} + \frac{1}{n2-3}}} \quad (79)$$

Conclusie
Na het lezen van dit document moet je een hypothese toets, en betrouwbaarheidsintervallen voor de correlatie coëfficiënten kunnen maken. Je moet een hypothese kunnen maken om correlatie coëfficiënten te vergelijken, en begrijpen waarom de Fischer z-transformatie nodig is.