Tests für Korrelationen

Frage-stellung

Besteht ein Zusammenhang zwischen zwei interval-, ordinal-, und/oder nominalskalierten Variablen?

Variablen

Jede der Variablen X und Y kann ein beliebiges Skalenniveau haben.
Wichtig: Wenn es einen gerichteten Zusammenhang zwischen den Variablen gibt, kann eine als UV (Prädiktor X) und die andere Variable als AV (Kriterium Y) bezeichnet werden.

Übersicht Korrelations-koeffizienten

Variable 2 / Variable 1	**Intervall-skaliert**	**Ordinal-skaliert**	**Künstlich dichtomisiert**	**Echt dichotom**
Intervallskaliert	Produkt-Moment Korrelation	Rang-korrelation - Kendall´s Tau - Spearman´s Rho	Biseriale Korrelation	Punkt-biseriale Korrelation
Ordinalskaliert			Biseriale Rang-korrelation	Biseriale Rang-korrelation
Künstlich dichtomisiert			Tetrachorische Korrelation	Phi-Koeffizient
Echt dichotom				Phi-Koeffizient

Inhaltliche Hypothesen

In einer Studie wird der Zusammenhang zwischen zwei Variablen untersucht. Im Einzelnen wird erwartet, dass es einen positiven/ negativen/ keinen oder einen Zusammenhang gibt.

Ableitung PV & Statistische Hypothesen

PH-gerichtet: Die beiden Variablen hängen positiv zusammen oder die beiden Variablen hängen negativ zusammen
PH-ungerichtet: Es besteht ein Zusammenhang zwischen den Variablen oder es besteht kein Zusammenhang zwischen den Variablen
⇨
PV: Vermuteter Zusammenhang zwischen den gemessenen Variablen.
⇨
Vorhersagekonforme Statistische Hypothesen:
Gerichtet: H_1: $\rho > 0$ H_0: $\rho \leq 0$
Ungerichtet: H_1: $\rho \neq 0$; H_0: $\rho = 0$
Wichtig: ρ (rho) steht hier für den jeweiligen Korrelationskoeffizienten

Testplanung per Hand

Determinanten: Alpha, Beta, Effektgröße Fisher Z, N

$$N = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{Z^2} + 3$$

Groß Z ist Fisher Z-transformierte Korrelation
$Z = 0{,}5 * In(\frac{1+\rho}{1-\rho})$ mit ρ (rho) als erwarteter Korrelation

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

<table>
<tr><td rowspan="3">Relevante Kennwerte bei intervallskal. Variablen</td><td>Mittelwert der Variablen X (Mittelwert Y entsprechend)</td><td>$M_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$</td><td>=MITTELWERT(Bereich)</td></tr>
<tr><td>Streuung der Variable</td><td>$S_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$</td><td>=STABW.N(Bereich)</td></tr>
<tr><td>Kovarianz</td><td>$S_{XY} = \frac{1}{n}\sum_{i}^{n}(x_i - \bar{x}) * (y_i - \bar{y})$</td><td>=KOVARIANZ.P(Daten Variable1; Daten Variable2)</td></tr>
<tr><td rowspan="3">Korrelationskoeffizienten</td><td>Produkt Moment Korrelation (X und Y intervallskaliert)</td><td>$r_{XY} = \frac{1}{n}\sum_{i}\frac{(x_i - \bar{x})}{S_X} * \frac{(y_i - \bar{y})}{S_Y} = \frac{1}{n}\sum_{i} z_{x_i} * z_{y_i}$
$= Produkt\ der\ z - standardisierten\ Werte$</td><td>=KORREL(Daten Variable1; Daten Variable2)</td></tr>
<tr><td>Punkt-biseriale Korrelation (X dichotom, Y intervallskaliert)</td><td>$r_{pbis} = \frac{M_1 - M_2}{S_Y} * \sqrt{\frac{n_1 * n_2}{(n_1 + n_2)^2}}$</td><td>=KORREL(Daten Variable1; Daten Variable2)</td></tr>
<tr><td>Phi-Korrelation
4-Felder Korrelation
(X dichotom, Y dichotom)</td><td><table><tr><td></td><td>Y=0</td><td>Y=1</td><td></td></tr><tr><td>X=0</td><td>a</td><td>b</td><td>a+b</td></tr><tr><td>X=1</td><td>c</td><td>d</td><td>c+d</td></tr><tr><td></td><td>a+c</td><td>b+d</td><td>N</td></tr></table></td><td>$\Phi = \frac{a * d - b * c}{\sqrt{(a + c) * (a + b) * (c + d) * (b + d)}}$</td></tr>
<tr><td>Statistischer Test</td><td colspan="2">z-test (Vergleich gegen 0)
$z_{emp} = Z * \sqrt{(N - 3)}$ mit Fisher Z transformierter Korrelation r $Z = 0{,}5 * ln(\frac{1+r}{1-r})$

z-Test (Vergleich zweier Korrelationen)
$z_{emp} = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$
mit Z_1 = Fisher Z von r_1 und Z_2 = Fisher Z von r_2</td><td>Fisher Z-Wert
=0,5*ln((1+r)/(1-r))

=FISHER(r)</td></tr>
<tr><td>Kritischer z-Wert; p-Wert</td><td>$z_{krit(\alpha)}$ = ?
p =?</td><td colspan="2">Kritischer z-Wert
=NORM.S.INV(1-Alpha) einseitig
=NORM.S.INV(1-Alpha/2) zweiseitig
Empirischer p-Wert Standardnormalverteilung
=1-NORM.S.VERT(z_{emp};wahr) einseitig positiver z_{emp}
=NORM.S.VERT(z_{emp};wahr) einseitig negativer z_{emp}
Empirischer p-Wert Standardnormalverteilung
=2*(1-NORM.S.VERT(z_{emp};wahr)) zweiseitig positiver z_{emp}
=2*NORM.S.VERT(z_{emp};wahr) zweiseitig negativer z_{emp}</td></tr>
</table>

<table>
<tr><td>Entscheidung über Hypothese</td><td colspan="2">Entscheidung nach folgendem Schema:
Bei $z_{emp} \geq z_{krit}$ ODER $p \leq$ Alpha aus Testplanung ⇨ Annahme H_1
Bei $z_{emp} < z_{krit}$ ODER $p >$ Alpha aus Testplanung ⇨ Annahme H_0</td></tr>
<tr><td>Effektstärke Konfidenzintervall</td><td>r ist die Effektstärke

Standardfehler Fisher Z
$SE_Z = \ ^1/_{\sqrt{N-3}}$ mit N = Anzahl Wertepaare

Konfidenzintervall von Z
• Untere Grenze: $Z - 1{,}96 * \ ^1/_{\sqrt{N-3}}$
• Obere Grenze: $Z + 1{,}96 * \ ^1/_{\sqrt{N-3}}$

Konfidenzintervall von r
• Untere Grenze: $\frac{e^{2*Z(untere\ Grenze)}-1}{e^{2*Z(untere\ Grenze)}+1}$

Obere Grenze: $\frac{e^{2*Z(obere\ Grenze)}-1}{e^{2*Z(obere\ Grenze)}+1}$</td><td>$e^x$ in Excel
=exp(x)

Berechnung Konfidenzintervall von r
=(exp(2*Z(Grenze)) -1) /
(exp(2*Z(Grenze))+1)
=FISHERINV(Grenze)</td></tr>
<tr><td>Effektstärke Konventionen</td><td colspan="2">r ≥ 0,1 ⇨ Kleiner Effekt
r ≥ 0,3 ⇨ Mittlerer Effekt
r ≥ 0,5 ⇨ Großer Effekt</td></tr>
<tr><td>Entscheidung über PV & PH</td><td colspan="2">Entscheidung nach folgendem Schema:
Bei abgeleiteter H_0 und angenommener H_0
⇨ PV eingetreten, PH bewährt
Bei abgeleiteter H_0 und angenommener H_1
⇨ PV nicht eingetreten, PH nicht bewährt
Bei abgeleiteter H_1 und angenommener H_1 + r > ρ(rho) aus Testplanung bzw.
Power ≥ 0.8
⇨ PV eingetreten, PH bewährt
Bei abgeleiteter H_1 und angenommener H_1 + r < ρ aus Testplanung bzw.
0,6 ≤ Power < 0,8
⇨ PV bedingt eingetreten, PH bedingt bewährt
Bei abgeleiteter H_1 und angenommener H_0
⇨ PV nicht eingetreten, PH nicht bewährt</td></tr>
</table>