Analyse Zusammenhang zweier kategorialer Variablen **Chi-Quadrat-Tests bei zwei Variablen**	
Frage-stellung	Gibt es einen Zusammenhang zwischen den verschiedenen Ausprägungen der beiden Variablen?
Variab-len	Zwei kategoriale Variablen, die auf Nominalskalenniveau gemessen werden.
Inhaltliche Hypothesen	Es wird in einer Studie erhoben, wie welche Ausprägungen bei zwei kategorialen, nominalskalierten Variablen bei jeder der untersuchten Personen vorliegen. 1. Es wird vermutet, dass es einen Zusammenhang zwischen den Variablen gibt. Je nach Ausprägung der einen Variablen, liegt die andere Variable in bestimmten Ausprägung vor. 2. Es wird vermutet, dass es keinen Zusammenhang zwischen den Variablen gibt. Die Ausprägung der einen Variablen ist unabhängig von der Ausprägung der anderen Variablen.
Ableitung + Statistische Hypo-thesen	PH: Vermutete Abhängigkeit oder Unabhängigkeit der Ausprägungen der beiden Var. ⇨ PV: Vermutete Abhängigkeit oder Unabhängigkeit der Ausprägungen der operationalisierten Variablen ⇨ Vorhersagekonforme Statistische Hypothesen: H_1: $\Pi(Var1,Var2) \neq \Pi(Var1)^* \, \Pi(Var2)$ H_0: $\Pi(Var1,Var2) = \Pi(Var1)^* \, \Pi(Var2)$
Test-planung	Determinanten: Alpha, Beta, Effektgröße (w), Freiheitsgrade $\Rightarrow$ Mit G*Power zu berechnen, χ^2-Test, Goodness of fit test
Annahmen	• Unabhängigkeit der Messungen der beiden Variablen • Eindeutigkeit der Zuordnung zu einer Kategorie, d.h. einer Ausprägung der jeweiligen Variablen • Erwartete Häufigkeiten in 80% der Fälle >5, nie kleiner 1
Relevante Kennwerte	Beobachtete Häufigkeit in der Kategorienkombination i und j: $f_{b,ij}$ Erwartete Häufigkeit in der Kategorienkombination i und j: $f_{e,ij}$ Gesamtzahl an Beobachtungen/Personen: $N = \Sigma_i \Sigma_j f_{b,ij} = \Sigma_i \Sigma_j f_{e,ij}$ Berechnung der erwarteten Häufigkeiten bei H_0: $f_{e,ij}$ = Zeilensumme i * Spaltensumme j / N = $p_i{*}p_j{*}N$
Statistischer Test	Chi-Quadrat-Wert: $Chi^2_{emp} = \chi^2 = \Sigma_i \Sigma_j (f_{b,ij} - f_{e,ij})^2 / f_{e,ij}$ Vergleich mit $Chi^2_{krit}(\alpha,df)$= ? Chi^2_{krit} aus Excel: =CHIQU.INV.RE(alpha;df), df=(j-1)*(k-1) ODER Vergleich *p* = Wahrscheinlichkeit Chi^2_{emp} unter H_0 mit gesetztem Alpha *p* aus Excel: =CHIQU.VERT.RE(Chi^2_{emp};df)

Entscheidung über SH	Bei $Chi^2_{emp} \geq Chi^2_{krit}$ ODER $p \leq$ Alpha aus Testplanung ⇨ Annahme H_1 Bei $Chi^2_{emp} < Chi^2_{krit}$ ODER $p >$ Alpha aus Testplanung ⇨ Annahme H_0
Effektgröße/ Konventionen	Cohens w w = Wurzel(Chi^2_{emp}/N) w ≥ 0,1 ⇨ Kleiner Effekt w ≥ 0,3 ⇨ Mittlerer Effekt w ≥ 0,5 ⇨ Großer Effekt
Entscheidungen über PV und PH	Entscheidungen nach folgendem Schema: Bei abgeleiteter H_0 und eingetretener H_0 ⇨ PV eingetreten, PH bewährt Bei abgeleiteter H_0 und eingetretener H_1 ⇨ PV nicht eingetreten, PH nicht bewährt Bei abgeleiteter H_1 und eingetretener H_1 und w > w aus Testplanung ⇨ PV eingetreten, PH bewährt Bei abgeleiteter H_1 und eingetretener H_1 und w < w aus Testplanung ⇨ PV bedingt eingetreten, PH bedingt bewährt Bei abgeleiteter H_1 und eingetretener H_0 ⇨ PV nicht eingetreten, PH nicht bewährt Achtung: Bei abgeleiteter H_1 ist es notwendig, das Zusammenhangsmuster mit dem erwarteten Muster der Häufigkeiten optisch zu vergleichen. Die H_1 kann statistisch angenommen werden auch wenn das Muster nicht den Erwartungen entspricht.

Vergleich zweier relativer Häufigkeiten z-Test

Frage-stellung	Weichen zwei beobachtete relative Häufigkeiten voneinander ab?
Variab-len	Zwei kategoriale Variablen, die auf Nominalskalenniveau gemessen werden
Inhaltliche Hypothese	Es wird vermutet, dass die relative Häufigkeit der Ausprägung Y=1 in der Gruppe 1 (X=1) größer/kleiner/ungleich der relativen Häufigkeit der Ausprägung Y=1 in der Gruppe 2 (X=2) ist.
Ableitung + Statistische Hypo-thesen	PH: s.o. ⇨ PV: Vermutung über relative Häufigkeiten der gemessenen Variablen ⇨ Vorhersagekonforme Statistische Hypothese: H_1: P_{b1} ≠/>/< P_{b2} H_0: P_{b1} =/<=/>= P_{b2}
Test-planung	Determinanten: Alpha, Beta, Effektgröße (w), Freiheitsgrade ⇒ Mit G*Power zu berechnen, z-Test, Proportions: Difference between two independent proportions
Annahmen	• Unabhängigkeit der Messungen • Eindeutigkeit der Zuordnung zu einer Kategorie, d.h. einer Ausprägung der Variablen • Großer Stichprobenumfang so dass Schätzung der relativen reliabel ist (n>=20 pro Gruppe)
Relevante Kennwerte	Relative Häufigkeiten in den Gruppen: $P_{b1} = f(Y=1\|X=1)$, $P_{b2} = f(Y=1\|X=2)$ Anteile in Gruppe 1 und Gruppe 2 Erwartete Wahrscheinlichkeit bei zwei relativen Häufigkeiten: $P_{erw} = \frac{P_{b1} * n_1 + P_{b2} * n_2}{n1 + n2}$
Statistischer Test	z-Wert: $z_{emp} = \frac{P1-P2}{\sqrt{P_{erw}*(1-P_{erw})*(\frac{1}{n1}+\frac{1}{n2})}}$ Vergleich mit $z_{krit}(\alpha)$= ? z_{krit} aus Excel: =NORM.S.INV(1-alpha) rechtsseitig ODER Vergleich *p* = Wahrscheinlichkeit z_{emp} unter H_0 mit gesetztem Alpha *p* aus Excel: =1-NORM.S.VERT(z_{emp};wahr)

Entschei-dung über SH	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	
Effektstärke und Konfidenzintervall	Risk difference RD = $P_1 - P_2$	SE(RD) = $SE_{RD} = \sqrt{\frac{a*b}{(a+b)^3} + \frac{c*d}{(c+d)^3}}$ a,b,c,d beziehen sich auf die Zellen in 4-Felder Tafel Untere Grenze 95%CI = RD – 1,96*SE(RD) Obere Grenze 95% CI = RD + 1,96*SE(RD)
	Odds ratio: OR = $(p_1/(1-p_1)) / (p_2/(1-p_2))$	
Konventionen	Risk Difference: keine allgemeinen Konventionen Odds ratio OR ≥ 1,44 ⇨ Kleiner Effekt OR ≥ 2,48 ⇨ Mittlerer Effekt OR ≥ 4,27 ⇨ Großer Effekt	
Entscheidungen über PV und PH	Entscheidungen nach folgendem Schema: Bei abgeleiteter H_0 und eingetretener H_0 ⇨ PV eingetreten, PH bewährt Bei abgeleiteter H_0 und eingetretener H_1 ⇨ PV nicht eingetreten, PH nicht bewährt Bei abgeleiteter H_1 und eingetretener H_1 und Power >= 0,8 ⇨ PV eingetreten, PH bewährt Bei abgeleiteter H_1 und eingetretener H_1 und Power >=0,6 ⇨ PV bedingt eingetreten, PH bedingt bewährt Bei abgeleiteter H_1 und eingetretener H_0 ⇨ PV nicht eingetreten, PH nicht bewährt	