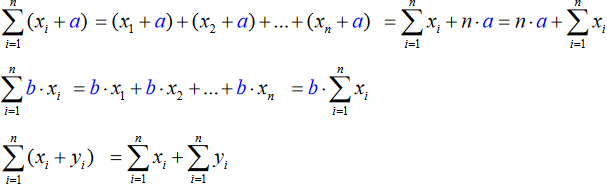
Formelsammlung

»Statistik I«

**Arten von Variablen:**

* Diskret/kategorial = endlich viele mögl. Ausprägungen (Kategorien)
* Kontinuierlich/stetig = beliebig feine Abstufungen in Ausprägungen mögl.
* Unabhängige Variable (UV): Variable, deren (kausaler) Einfluss auf die abhängige Variable in einer Hypothese behauptet wird. Die UV wird in einer Studie hergestellt (z.B. Computerspielen) oder erfasst (z.B. Interesse, Religionszugehörigkeit).
* Abhängige Variable (AV): Variable, auf die die UV einen Einfluss ausüben soll, in der sich ein Effekt zeigen soll. Die AV wird beobachtet (gemessen, erfasst). (UV🡪AV)
* Störvariable (SV): Variable, die mit der AV zusammenhängt, nicht aber systematisch als UV untersucht wird. Sie kann zu Scheineffekten (Artefakten) führen, wenn sie systematisch mit der UV zusammenhängt (Konfundierung). Handelt es sich um eine unsystematische SV, so wird es schwieriger, bestehende Effekte aufzudecken.(SV,UV🡪AV)
* Mediatorvariable(MedV): vermittelt Einfluss der UV auf die AV (UV🡪MedV🡪AV)
* Moderatorvariable(ModV): beeinflusst Art/Stärke der Wirkung/des Zusammenhangs der UV auf die AV)
* Zufallsvariable (siehe Abschnitt zu Wahrscheinlichkistheorie)

|  |  |
| --- | --- |
| **Notation/Generelles** |  |

* Rechnen mit Summenzeichen: 
* Urliste = Liste der Daten in der (beliebigen) Reihenfolge, in der sie angefallen sind
* Ausprägungen in Variablen werden zur Analyse numerisch codiert(Zahlen zugeordnet)
* Messen/Skalierung, wenn das Codierungsschema bedeutsame Relationen zwischen den Objekten widerspiegelt
* **Skalenniveau** = für eine Variable bedeutsame Relationen zwischen den Objekten (bestimmt zul. Transformationen und beschränkt Deutung auf die Relationen zwischen Objekten, die unter zul. Transf. invariant sind), bestimmt zulässige statistische Analysen, Vom Nominal- bis Verhältnisskalenniveau: erhöhte Messanforderungen, weniger zulässige Transf., Zahlen empirisch bedeutsamer, steigende Anzahl zul. Statistiken

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Skalenniveau | Voraussetzung | Zulässige Transf. | Deutbarkeit | Statistiken |
| Nominal | Eindeutige Zuordnung jedes Objekts zu einer Kategorie | Eineindeutig: Beibehalten von Verschiedenheit und Gleichheit von Objekten | Gleichheit oder Ungleichheit von Objekten in (Zahlenwert) der Variablenausprägung | Häufigkeitswerte,  Prozentwerte,  Modalwert |
| Ordinal | Objekte müssen hinsichtlich der Variablen in Rangfolge gebracht werden können | Ordnungserhaltende/ ordinale Transformationen/ streng monoton wachsende Abbildungen | Objekt in Variablenausprägung größer/kleiner/gleich eines anderen Objekts (keine Bedeutung von Rangdifferenzen) | + kumulierte Prozentränge,  + Median,  + Rangkorrelationen,  + Rangvarianzanalysen |
| Intervall | Differenzen in Ausprägungen bedeutsam | Lineare Transformationen, x’ = a + b\* x mit b>0 | Differenz in Variablenwerten zweier Objekte gleich oder um best. Faktor größer/kleiner als andere Differenz, Nullpunkt ohne Bedeutung | + Mittelwert,  + Varianz |
| Verhältnis | Verhältnisse bedeutsam, Existenz eines Nullpunktes | Ähnlichkeitstransformation, x‘ = b\*x mit b>0 | Variablenausprägung (um best. Faktor) größer/kleiner/gleich einer anderen, absoluter Wert bedeutungslos | + Variations- und  + Kongruenzkoeffizient |

**Grafische Darstellung**:

* Histogramm/Säulendiagramm: Säulenhöhe proportional zur Häufigkeit
* Balkendiagramm: s.o., um 90° gedreht
* Polygonzug: s.o. nur durch Linien verbundene Punkte statt Säulen
* Überlagerungen bei obigen Varianten gut möglich
* Kuchendiagramm/pie-chart: Winkel eines Tortenstücks proportional zur Häufigkeit
* Boxplots/Box-Whisker Plots: Variable auf Ordinate, IQA=box, Md=Trennlinie, Whiskers ab box bis Q1-1,5\*IQA und Q3+1,5\*IQA, Ausreißer innerhalb 3\*IQA = Sterne, extreme Ausreißer = Kreise
* Stem-and-Leaf Plots: erste Stelle(n) der Meswerte = linke Spalte/Stamm, letzte Stelle rechts 🡪 je zwei versch. Endungen pro Zeile, nicht auftretende Werte = leere Zeile hinter Stamm
* Streudiagramm: für bivariate Verteilungen zur Veranschaulichung von Zusammenhängen, Prädiktorvariable auf Ordinate, Kriterium auf Abszisse

**Ausreißer/outlier:**

* extrem hohe/niedrige/für Verteilung untypische Werte
* mögl. Ursachen: versehentliche oder mutwillige Fehler der Vp, falsches Verständnis der Instriktionen, Messgerätefehler, Fehler bei Dateneingabe, extreme Merkmalsausprägungen bei Vp
* Umgang: abhängig von Ursache, Korrektur zu richtigem Wert, Löschen(missing data), alle Werte einer Vp löschen, im Datensatz lassen (mit und ohne Ausreißerwerten analysieren und differenziert über Ergebnisunterschiede berichten), Verwendung robuster Statistiken (Md)

Sonstiges (Notation):

* *X* und *Y* Variablen mit den Ausprägungen *xi* bzw. *yi* mit *i* = 1, ..., *n* und *n* als Zahl der Untersuchungseinheiten (Personen). Ferner bezeichnet *x*(*i*) das *i*‐te Element bei aufsteigend sortierten Ausprägungen in *X*.
* *~~x~~* den arithmetischen Mittelwert, *s*2 die Varianz und *s* die Standardabweichung der Variablen *X*. Bei Uneindeutigkeit wird ggf. durch Indizierung deutlich gemacht, worauf sich die Statistiken beziehen, z.B. *sY* als Standardabweichung der Variablen *Y*.
* *r* bezeichnet die Produkt‐Moment Korrelation und *cov* die Varianz zweier Variablen. Ggf. wird durch Indizierung deutlich gemacht, zwischen welchen Variablen die Statistik bestimmt wurde, z.B. bei *rXY* zwischen den Variablen *X* und *Y*.

|  |  |
| --- | --- |
| **Univariate deskriptive(beschreibende) Statistik** |  |

**Häufigkeiten:**

**Diskrete Häufigkeitsverteilung**: Auflistung von Messwerten mit Auftretenshäufigkeit

|  |  |  |  |
| --- | --- | --- | --- |
| Bezeichnung | Form(-el) | Bedeutung | kumuliert |
| **absolute Häufigkeit** |  | Anzahl Messwerte für eine Kategorie |  |
| **relative Häufigkeit** |  | Anteil absoluter Häufigkeit an Stichprobenumfang |  |
| **prozentuale Häufigkeit** |  | Relative Häufigkeit als Prozentzahl ([0,100]) |  |

**Sekundäre Häufigkeitsverteilung**: bei kontinuierlichen Variablen/vielen Kategorien sinnvolle Verteilung der Messwerte auf Bereiche/Intervalle (möglichst gleich breit, breiter bei <n und >Wertebereich)

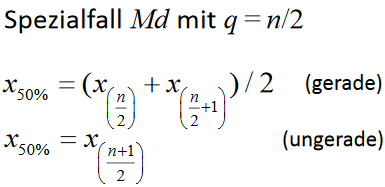
**Maße/Kennwerte/Statistiken der zentralen Tendenz** (Lage/Niveau der Verteilung):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bezeichnung | Formel | Skalierung | Sonstiges | Eigenschaften |
| **Arithmetischer Mittelwert** | Mit | Intervall | Daten gruppiert in c Kategorien: | * wenig robust ggü. Ausreißern * muss kein echter Messwert sein  * Summe aller Abweichungen = 0 * Summe quadr. Abw. = Minimum |
| **Geometrischer Mittelwert** | (alle *xi* > 0) |  |  |  |
| **Harmonischer Mittelwert** |  |  |  |  |
| **Modalwert** | *Mo* | Nominal | bei kontinuierlichen Variablen sekundäre Verteilung nutzen | * Uneindeutig bei bimodaler Verteilung |
| **Median** | n ungerade:  n gerade: | Ordinal | Aufsteigend sortierte Messwerte, 50% unter, 50% über Md | * Hängt nicht von allen Werten ab (🡪 robust ggü Ausreißern) * Summe abs. Abw. ist Minimum |

**Streuungsmaße/Dispersions/Variabilitätsmaße** (quantifizieren Variabilität der Messwerte):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bezeichnung | Formel | Skala | Eigenschaften | Sonstiges |
| **Varianz** |  | Intervall | * Bei gleichen Werten = 0 * Invariant ggü. Addition einer Konstante * Wert\*b 🡪 s^2\*b^2 | Hängt von allen Werten ab, Ausreißer durch Quadrierung stark gewichtet, nicht in Einheit der Messwerte |
| **Standard-abweichung** |  | Intervall | * Konstante Add. Wie bei Varianz * Wert\*b 🡪 s\*b | Einheit der Messwerte |
| **Variations-/Spannbreite/Range/Streubereich** |  |  | Nur von Extremwerten abhängig 🡪 sensibel auf Ausreißer | Differenz zwischen größtem und kleinstem Wert |
| **Interquartilsabstand/-bereich** |  |  | * Robust ggü. Ausreißern * Q1 = 25. Quantil * Q3 = 75. Quantil | Gibt an, wie stark die mittleren 50% streuen |

**Quantile** (Messwert, der p Prozent der Verteilung abschneidet):

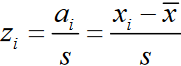
*p*‐tes Quantil mit [*q*] als nächstgrößere ganze Zahl, die auf *q* = *n* · *p* / 100 folgt:

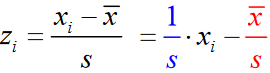
* falls *q* ganzzahlig 🡪 *xp*%=(*x*(*q*)+*x*(*q*+1))/ 2
* falls *q* nicht ganzzahlig 🡪 *xp*% = *x*([*q]*)

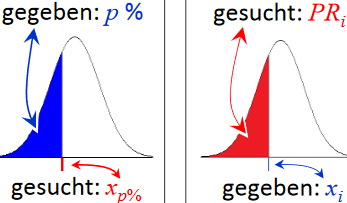
**Verteilungsformen:**

|  |  |  |  |
| --- | --- | --- | --- |
| Kriterium/Maß | Form(el) | Ausprägungen&Eigenschaften | Sonstiges |
| Asymmetrie/ Schiefe |  | * symmetrisch: = 0, außerdem:   (+unimodal (=eingipflig) 🡪=Mo)   * linkssteil >0, rechtssteil <0 | Schwerpunkt links(>0)/rechts(<0) vom Mittelpunkt |
| Gipfelform/ Exzess/ Kurtosis |  | Normalverteilung = 0  Platykurtisch/stumpf/breitgipflig >0  Leptokurtisch/schmal/steilgipflig <0 | Bei unimodaler Verteilung, Bezugsgröße ist Normalverteilung |

**z-Transformation/z-Standardisierung:**

Warum? Kriteriumsorientierte Bewertung eines Messwertes (Wert für bestimmtes Kriterium ausreichend?) lässt keine Bewertung hinsichtlich der Bezugsgruppe zu 🡪 normorientierte Bewertung (z.B. Abweichungswert (🡪Zentrierung) mit Normierung an Standardabweichung)

Berechnung: = 🡪lineare Transformation der Messwerte

Eigenschaften: Mittelwert = 0, Streuung = 1, Verteilungsform (Symmetrie/Schiefe s.o.) unverändert, dimensionslos

Bedeutung: wie viele Standardabweichungen ein Wert über Mittelwert liegt

**Prozentränge** (Wie viel Prozent der Verteilung weisen einen Wert x <= x(i) auf?):

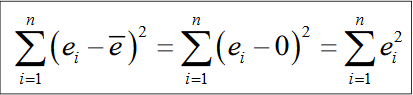
Bemerkung: (Skala) min. ordinal, keine lineare Transformation der Messwerte

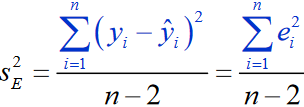
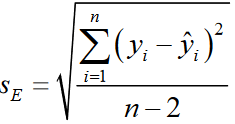
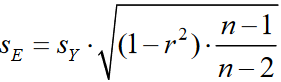
Berechnung: = kumulierte prozentuale Häufigkeit des Wertes in primärer Häufigkeitsverteilung

|  |  |
| --- | --- |
| **Bivariate deskriptive Statistik** |  |

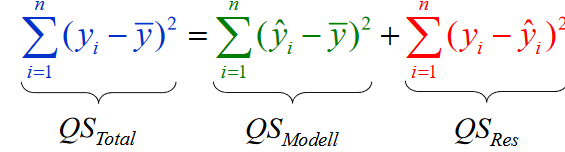
Wozu? Untersuchung von Zusammenhängen (linear(🡪*Y*(Kriterium)=a+b\**X*(Prädiktor)), monoton fallend, umgekehrt U-förmig,…), deren Richtung und Enge, darauf basierende Prädiktion

|  |  |  |  |
| --- | --- | --- | --- |
| Statistik | Formel | Eigenschaften | Voraussetzungen/Sonstiges |
| **Lineare Regression** (-sgerade): „von Y auf X“ | Regressionsgleichung:    Optimierungskriterium:  Berechnung: | * eindeutige Lösung * mathematisch einfach * Summe der Residuen = 0 * Kein Zsm.hang 🡪 = | * zur Vorhersage   Nur bei perfektem linearem Zusammenhang identisch für beide Vorhersagerichtungen,  sonst Schnittpunkt bei () |
| **Kovarianz** |  | * Größer, je mehr pos. Kreuzprodukte (Abw) * Kleiner, je mehr neg. Kreuzprodukte * Kein Zusammenhang, wenn Kreuzprodukte sich aufheben ( = 0), oder eine Variable keine Varianz zeigt | * Mindestens Intervallskala * Vorzeichen gibt Richtung an * **Maßstabsabhängig:**   invariant ggü. add. Konstanten , aber:     * Zusammenhang mit Regressionsgeradengleichung: |
| **Produkt-Moment-Kor-relation** | Pearson: | Wertebereich:  [-1 = perfekt negativ,  1 = perfekt positiv]  Invariant ggü. Linearen Transformationen | * Beide **Variablen intervallskaliert** * **Maßstabsunabhängig** * Kein Zsm.hang 🡪 = 0 * Enge und Richtung **linearer Zusammenhänge** * Zusammenhang mit Regressionsgeradengleichung: * Gleiche Vorzeichen von r und b |
| **Determinationsko-effizient** | = r^2 🡪 | Ungerichtet: Wert in[0,1]  = r bei 0 und1 | Interpretation: Anteil der Fehler-reduktion durch die Prädiktor-variable = um welchen Anteil wurden die Vorhersagefehler durch Regression verbessert? 1- |
| **Partial-korrela-tion** |  | Gibt den linearen Zusammenhang zweier Variablen *X, Y* an, aus dem der lineare Einfluss einer dritten Variable *Z* eliminiert wurde | * Aufdecken von Scheinzusammenhängen |
| **Rangkorreation nach Spearman** | Bei Reihen Rg(X), Rg(Y): rs= r [Rg(X), Rg(Y)]  Mit di =Rg(xi)-Rg(yi) und ohne Ties: | Quantifiziert Zusammenhang zwischen zwei min. ordinalskalierten Variablen X und Y | * Aufsteigend sortierten Messwerten Rangplätze Rg() zuweisen (bei ties Mittelwert) * Invariant ggü. streng mon. wachsenden und linearen Transf. |
| **Rangkorrelation nach Kendall (a)** | (*C* =Anzahl konkordante, *D* = Anzahl diskordante Paare): | C-D als Maß für Enge des Zusammenhangs (aber: abhängig von Stichprobengröße n), Normierung an der Gesamtzahl an Paaren | * Betrachtung aller Messwertpaare (i,j) * Anzahl aller mögl. Paare:   = C + D + TX + TY + TXY   * Nur konkordant (C = ) => +1 * Nur disordant (D = ) => -1 * C = D => 0 (kein Zusammenhang) * , wenn keine Ties exist. * Invariant ggü. streng mon. wachsenden und linearen Transf. * Unterschiedliche Gewichtung von Ties, stärkste mindernde Wirkung auf τ, weniger auf τb, nicht systematisch bei γ (kann als einziger Rangkorrelationskoeff. auch 1 sein bei Ties) |
| **Goodman und Kruskal** |  | Normierung an Summe der konkordanten und diskordanten Paare  + bei vielen Ties in Daten |
| **Rangkorrelation nach Kendall (b)** |  | Normierung and geom. Mittel aus Häufigkeiten nicht vorliegender Ties in Y und nicht vorliegender Ties in X |
| **Punktbiserale Korrelation** | rpb = r(X,Y)  =  und Mittelwerte(in X) der beiden 0-1 codierten Gruppen(Y) und n1, n0 Anzahl Personen in Gruppen | Intervall & natürl. dichotom |  |
| **Phi-Koef-fizient ϕ (oder φ)** |  | Natürl. dich. Variablen X, Y  4 mögl. Kombis, Randh. > 0 | * ϕ= r * -1 oder 1 wenn Randvert. Gleich (a+d oder c+b = 0) * Bei Umkodierung: Vorzeichenwechsel |
| **Cramérs V** | X mit k und Y mit m Ausprägungen: | Min eine polychotome (mehrkategoriale) Variable 🡪k x m Kontingenztafel  Zsm.hang[0=kein,1=perf]  (Richtung sinnlos) |  |
| **erwartete Häufig-keiten χ2** | erwartete abs. Häufigkeit: | min 1 natürl. nominal dich. Variable, Vierfeldertafel mit Häufigkeiten, gleiche Anteile (f) empirisch beobachteter Häufigkeiten wie bei Unabhängigkeit erwartet = kein Zsm.hang => 0 | * Rel. Häufigkeiten fij = hij/n * Für eij: Multiplikationstheorem * hi. und h.j Randhäufigkeiten * **χ2** größer je mehr Abw. zu eij |
| **tetrachorische Korrelation rtet** | Für b, c >0: | 2 künstlich dichotome Variablen (eigentlich kontinuierlich und normalverteilt) | * Je größer a\*d(Diagonale links unten-rechts oben) relativ zu b\*c, desto größer a\*d/(b\*c), desto größer der Nenner, desto mehr geht rtet gegen 1 |

**Vorhersagegüte:**

* (Vorhersagefehler: )
* **Varianz der Residuen/Fehlervarianz** (nach Optimierungskrit.) =
  +  Gebräuchlicher: Wurzel daraus = **Standardschätzfehler** =
    - = 0 wenn Vorhersage perfekt
    - Bei Produkt-Moment-Korrelation: vereinfacht:

**Abw. Wert von Mittelwert** = Abw. Vorhersage vom Mittelwert + Abw. Wert von Vorhersage

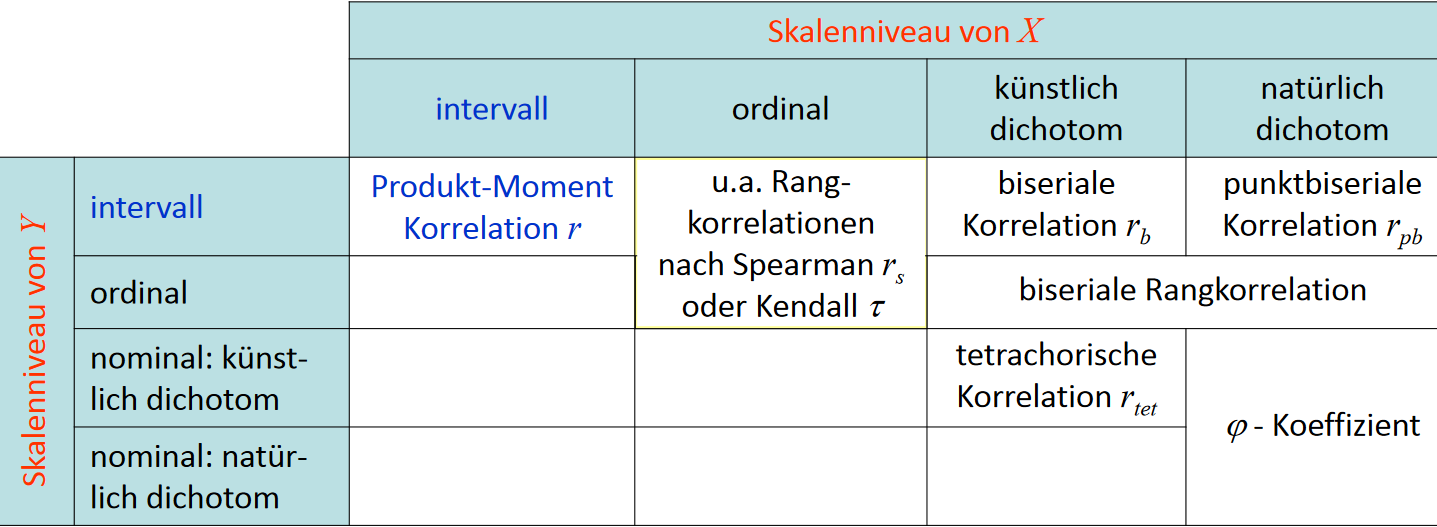
* Analog: **QuadratSummen** mit (**/n-1**) 

**Ursachen/Quellen von Verzerrungen in Korrelationsmaßen:**

* **Ausreißer**, v.a. wenn n klein
* Einschränkung in der „Varianz“ (**range**-restriction) von Prädiktor oder Kriterium
* Nichtlineare Zusammenhänge, z.B. U-förmig 🡪 nichtlineares Regressionsverfahren nehmen
* Heterogene Subgruppen

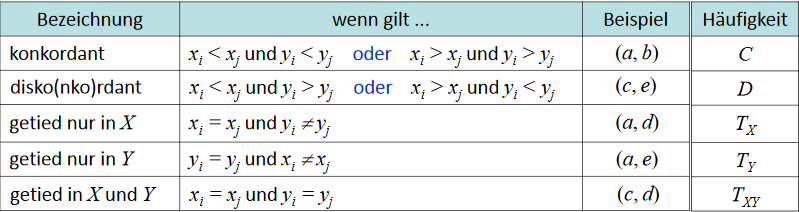
**!Korrelation gibt keine Auskunft über Kausalität des Zusammenhangs🡪**experimentelles Design

**Korrelationskoeffizienten nach Skalenniveau:**



* Wertebereich [-1,1], wobei-1 = perf.neg., 0 = kein und 1 = perf. positiver Zusammenhang
* alle Statistiken niedrigerer Skalenniveaus auch auf Variablen höherer Niveaus anwendbar

**Rangkorrelation:**

* Kendall und andere: Betrachtung aller Messwertpaare(i,j), Bestimmung der Häufigkeiten der verschiedenen Messwertpaar-Relationen:
* 

|  |  |  |
| --- | --- | --- |
|  | *y*1 = 0 | *y*2 = 1 |
| *x*1 = 0 | *a* = *h*11 | *b* = *h*12 |
| *x*2 = 1 | *c* = *h*21 | *d* = *h*22 |

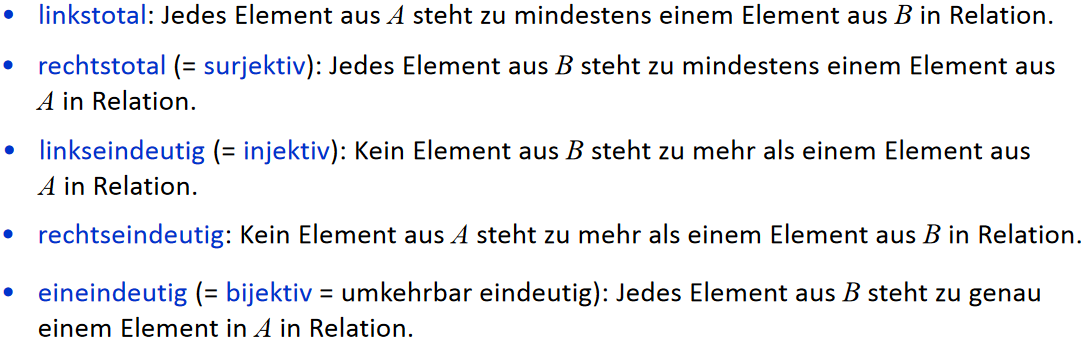
**Phi‐Koeffizient** (hier und im Weiteren mit folgender Notation für absolute Häufigkeiten):

**Deutung von Korrelationen:**

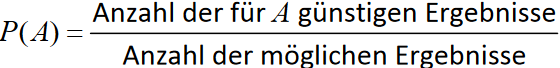
* Quadrierung ergibt Maß für erklärte Varianz
* Cohen: 0.1 = kleiner, 0.3 = mittlerer, 0.5 = starker Effekt
* Bewertung als hoch oder niedrig hängt stark vom Kontext ab

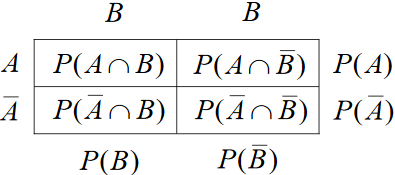
|  |  |
| --- | --- |
| **Wahrscheinlichkeitsrechnung** |  |

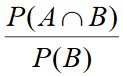
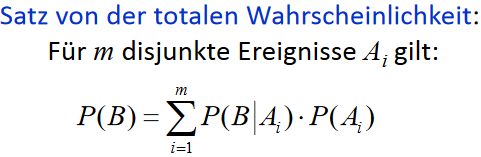
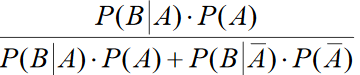
**Mengen:**

* enthalten voneinander *unterscheidbare Elemente*
* B *Teilmenge* von A, wenn alle Elemente aus B auch in A enthalten sind, echt wenn B != A
* Alle Teilmengen einer Menge A(auch leere Menge und A) = **Potenzmenge 2A**, mit leerer Menge als unmögliches und A als sicheres Ereignis
* **Komplement**:  ={x|x∉A}
* Durchschnitt: A∩B={x|x∈A ∧x∈B}, wenn leer 🡪A und B disjunkt
* Vereinigung: A∪B = {x|x∈A ∨x∈B}
* Kartesisches Produkt: A×B = {(a,b)|(a ∈A ∧ b∈B) ∨ (a ∈B ∧ b∈A)}, wobei (a,b) ≠(b,a)
  + N Mengen, dann A1×A2 ×...×An und jedes Element wird als n‐tupel (a1,a2, ...,an)bezeichnet
  + Teilmenge des kartes. Produktes: Relation R
    - Binär (2 Mengen): (a,b)∈R oder aRb
    - 
    - N Mengen: n stellige Relation
    - Relation R = Funktion/Abbildung/Transformation f, wenn linkstotal & rechtseindeutig
      * y = f(x), oder auch f: X🡪Y, x🡪f(x)

**Zufallsvorgang/-experiment:** wiederholbarer Prozess nach Vorschrift, dessen Ergebnis nicht 100% sicher vorhergesagt werden kann

* Ergebnismenge Ω= {ω1, ω2, ...,ωk} bei k möglichen Ergebnissen
* Ereignismenge A ⊂ Ω, wenn |A|=1, dann ist A Elementarereignis
* (**apriori**, unbedingte) (Auftretens) **Wahrscheinlichkeit** P(A): Zahl in [0,1] (= [unmöglich,sicher]) für Ereignis A
  + Oft unbekannt, Schätzung über relative Häuf.
    - Bernoulli-Theorem: Schätzung wird besser bei mehr Beobachtungen
  + Laplace: bei endlich vielen, gleich wahrscheinlichen Elementarereignissen =K
  + Potenzmenge bei Münzwurf mit K=Kopf, Z=Zahl:
  + **Kolmogorov Axiome** für Zahlen, die Wahrscheinlichkeiten repräsentieren:
    1. P(A) >= 0 für alle Ergebnisse A in der Potenzmenge
    2. P(Ω) = 1
    3. Für disjunkte Teilmengen A,B aus Ω, d.h. A∩B=∅: P(A∪B)=P(A) +P(B)
  + Wenn alle Ergebnisse gleich wahrscheinlich, dann P(Ergebnis)=1/K (siehe Tabelle)

 anpassen: apriori als Randwahrsch., Rest ergänzen

* + Additionstheorem: P(A∪B) = P(A) + P(B) + P(A∩B)
* **Bedingte Wahrscheinlichkeit P(A|B) = a posteriori Wahrscheinlichkeit** =
* Wenn P(A|B) = P(A), dann A und B **unabhängig** = P(A∩B)=P(A)\*P(B)
* P(A|B)=P(B|A)\*P(A)/P(B)=

da P(B) =

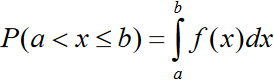
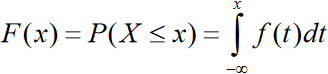
**Kombinatorik** ( K = Zahl möglicher Ergebnisse bei ***k* Ziehungen, *n* Einzelergebnissen**):

|  |  |  |
| --- | --- | --- |
|  | mit Zurücklegen | ohne Zurücklegen |
| Reihenfolge bedeutsam | *K nk* |  |
| Reihenfolge nicht bedeutsam |  |  |

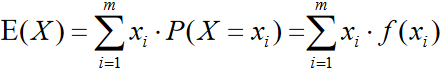
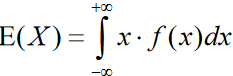
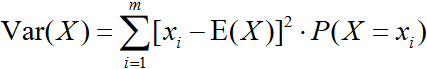
**Zufallsvariable** (Funktion X, die Ergebnissen reelle Zahlen(Messwerte) zuordnet): Ω→ R, ω🡪X(ω)

* diskret: kann endlich viele Werte annehmen (Kategorien, Intervall ganzer Zahlen)
  + oft als angenähert stetig betrachtet (Körpergröße, Intelligenz, …)
* stetig: Anzahl Ergebnisse überabzählbar unendlich

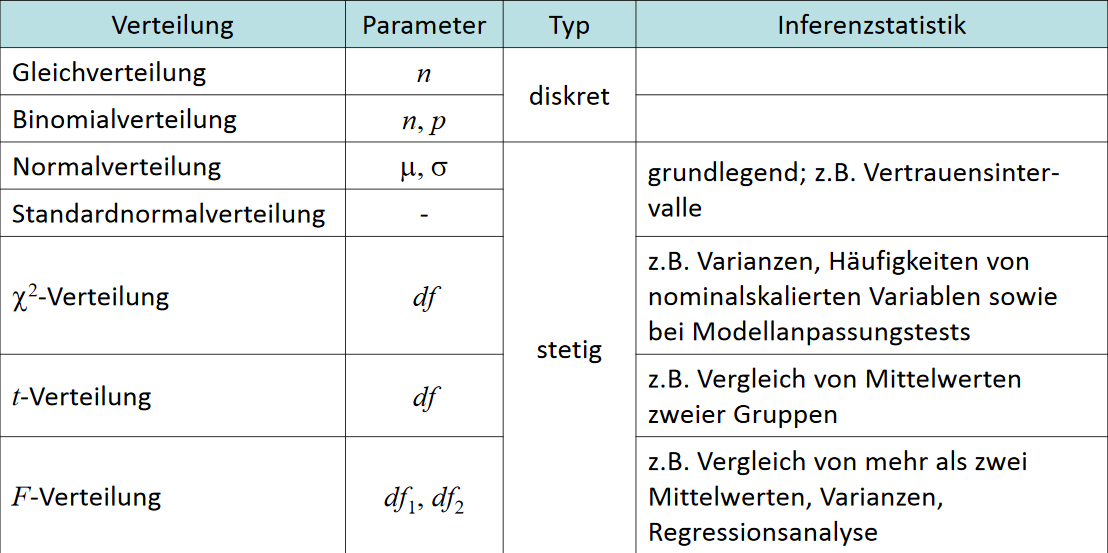
**Wahrscheinlichkeitsverteilungen:** Verteilung von Realisierungen (Funktionswerten) x von X:

* gibt an, mit welcher Wahrscheinlichkeit x-Werte vorkommen
* Wahrscheinlichkeitsfunktion (bei diskreten X): f(x) = P(X=x)
  + Verteilungsfunktion = kumulierte Wahrscheinlichkeitsfunktion =
* Dichte der Zufallsvariablen (bei stetigen X):
  + f(x)!= P(X=x), da P(X=x) = 0 🡪Integral statt Summe
  + Wahrscheinlichkeiten existieren nur für Intervalle von Dichten
  + Verteilungsfunktion (höchstens Wert von x=xp 🡪Integral bis xp):

Parameter von Wahrscheinlichkeitsverteilungen (X mit m Ausprägungen falls X diskret):

* **Erwartungswert** (nicht zwingend selbst Messwert):
  + Diskret:
  + Stetig:
* **Varianz** (**Variabilität** der Verteilung falls diskret):
* Rechenregeln (X,Y Zufallsvariablen, c,d reelle Konstanten):
  + E(c) = c
  + E(c\*X) = c \* E(X)
  + E(c\*X + d\*Y) = c\*E(X) + d\*E(Y)
  + Var(X) = E(X2) - E(X)2
  + Var(c\*X + d) = c2\*Var(X)
  + Var(c\*X + d\*Y) = c2\*Var(X) + d2\*Var(Y) + 2\*c\*d\*Cov(X,Y)
  + Var(c\*X + d\*Y) = c2\*Var(X) + d2\*Var(Y) falls X und Y unabhängig/unkorreliert

|  |  |
| --- | --- |
| **Wahrscheinlichkeitsverteilungen** |  |



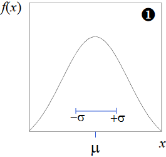
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Diskrete Verteilung** | Beschreibung | Funktion | Erwartungswert | Varianz |
| **Gleich-verteilung** (Bsp. Würfeln) | alle n möglichen Werte gleich wahrscheinlich |  |  |  |
| **Binomial-verteilung** (Bsp. Ziehen mit Zurücklegen)  **X ~ B(n,p)** | Bernoulli-Exp. (n-fache Wdh. eines Zufallsexperiments), ab n > 20 🡪 annähernd Normalv | P(A)=p, k= Auftre-tenshäufigkeit von A in n Durchführungen: | E(X)= n\*p | Var(X)= n\*p\*(1-p) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Stetige Verteilung** | Parameter/Eigenschaften | Dichte | Sonstiges |
| **Normalverteilung**  **X ~ N ( μ, σ)** | μ= Mittelwert = E(X)  σ >0 = Standardabw.  Var(X)= σ2   * symmetrisch um μ * gegen 0 für x 🡪±∝ |  | Umrechnung in standardnormal-verteilte Werte: |
| **Standard-normalverteilung** | μ= 0  σ = 1 |  | P(X<=x) = 1-P(X<=-x)  P(X>x) = 1-P(X<=x)  P(X<=-x) = 1-P(X<=x) |
| **χ2- Verteilung**  **X ~ χ2(df)** | Quadrierung und Aufsum-mierung unabhängiger standard-normalverteilter Zufallsvariablen(Z) Z1,…,Zk  df(Anzahl Freiheitsgrade)=k=E(X)  Var(X) = 2\*df |  | Je kleiner df, desto linkssteiler, je größer df, desto mehr annähernd an Normalverteilung |
| **t-Verteilung**  **X ~ t(df)** | Standardnormalverteilte Zufallsv geteilt durch Wurzel dazu unabhängiger χ2-verteilter Z/df  df = k  E(X) = 0, falls df(= k) > 1  Var(X) = df/(df-2), falls df(=k) >2 |  | Symmetrisch um 0, je größer df, desto mehr annähernd normalverteilt  Quadrat entspricht F-Verteilung mit df1= 1:  t2(df) ~ F(1,df) |
| **F-Verteilung**  **X ~ F (df1, df2)** | df1= Anzahl Freiheitsgrade Z1  df2= Anzahl Freiheitsgrade Z2  E(X) = df2/(df2−2), falls df2 > 2    falls df2 > 4 |  | Asymmetrisch bei kleinen df, symmetrischer bei wachsenden df,  0 < X <= +inf |

Suche nach Wahrscheinlichkeiten (p) oder Messwerten (x, gegeben: Perzentil) in Tabellenband

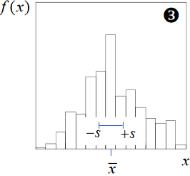
|  |  |
| --- | --- |
| **Inferenzstatistik:** (Testen von Aussagen/) Schlüsse auf die Population(-sparameter) |  |

Stichprobenfehler: Abweichung der Stichprobenparameter zwischen Stichprobe(n)&Population

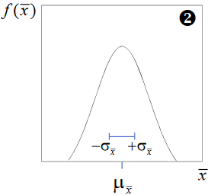
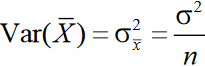
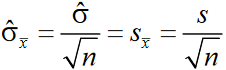
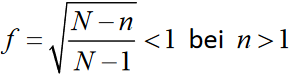
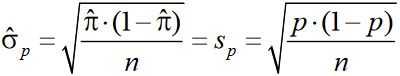
**Population** (Notation griech.): Ω = Menge aller Untersuchungseinheiten ω, |Ω| = N

* Verteilung eines Merkmals X(wenn normalverteilt Mittelwert μ und Standardabweichung σ)
* Bestimmung von Erwartungswert und Standardfehler der Stichprobenkennwerteverteilung möglich (siehe Stichprobenkennwerteverteilung)
* Wenn Merkmal normalverteilt, dann auch Stichprobenkennw.verteilung:

**Stichprobe** (Notation lat.):

* Teilmenge S von Ω, and denen Messwerte erhoben wurden, |S|=n konstant,
* Wahl zufällig, Entnahmewahrscheinlichkeit gleichverteilt über der Population: P(ω∈S) = 1/N
* Wertunterschiede in (Kennwerten über) Zufallsvariable(n) ∈ Xn(Merkmale) abhängig von ω∈S
* Mittelwert , Standardabweichung s

Schätzen von **Populationsparametern** (Stichprobenkennwerte in lat. Buchstaben):

* (Stichprobenmittelwert)
*  (Stichprobenvarianz)
* (Korrelation)
* **Stichprobenkennwerteverteilung** (unendlich viele Stichprobenziehungen, n konstant):
  + Mittelwerte über X verschieden (sowie deren Abweichung vom Populationsmittel)
  + Theoretische Verteilung Stichprobenkennwerten über S1 , …, Sk (mit k🡪unendlich, (|S|∈ Sk) = n (=konstante Stichprobengröße))
  + Erwartungswert des Mittelwerts = Mittelwert in Ω:
  + Varianz des Mittelwerts (um n kleiner als in Ω):
  + Standardabweichung aka Standardfehler (d. Mittelwertes):
    - Annahme einer Zufallsziehung mit Zurücklegen bzw. N=unendlich
    - Wenn Populationsparameter bekannt:
    - Sonst: Schätzung:
    - Endlichkeitskorrektur falls n/N > 0.05: Standardfehler = Standardfehler\*
    - zeigt Präzision der Schätzung: je kleiner desto genauer (mit wachsendem n)
  + Standardfehler für relative Häufigkeit :
  + **Zentrales Grenzwerttheorem**: Annäherung der Verteilungsform der Stichprobenkennwerteverteilung des Mittelwerts ab n>30 an Normalverteilung

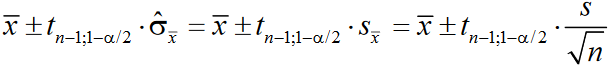
An der Stichprobe gewonnene Kennwerte (relative Häufikgeiten ***fk***, Mittelwert , Streuung ***s2***, Korrelation ***r***) können zur **Stichprobenbeschreibung** und **Schätzung** von Parametern der zugrundeliegenden Population( **πk**, **μ**, **σ2**, **ρ** )herangezogen werden.

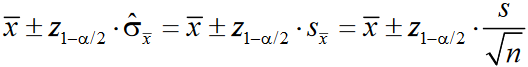
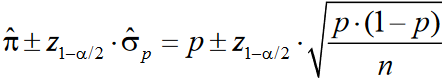
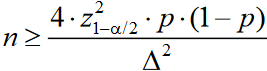
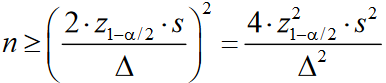
**Punktschätzung:**

* **Schätzung** von Populationsparameters **durch einzelnen Kennwert** aus Stichprobendaten
* **Gütekriterien** für (einzelne) Stichprobenkennwerte zur Parameterschätzung**:**
  + *Konsistenz:* Kennwert nähert sich bei wachsendem n immer mehr dem Parameter an, d.h. der Standardfehler des Schätzers geht für n gegen unendlich gegen 0
    - Stichprobenmittelwert, -standardabweichung und -varianz für μ, σ, σ2, sowie *fk*=*hk*/*n* für πk
  + *Erwartungstreue:* Erwartungswert der Stichprobenkennwerteverteilung des Schätzers entspricht dem des Parameters (systematische Abweichung = Bias)
    - Erwartungstreu: Stichprobenmittelwert für μ, Stichprobenvarianz(=MQA\*n/n-1) für σ2, Md ebenso für μ, wenn die Populationsverteilung symmetrisch ist
    - Nicht erwartungstreu: Mittlere quadratische Abweichung (MQA) für Variabilität
  + *Effizienz:* Präzision der Schätzung, bei Erwartungstreuen Schätzern bemessen an Standardfehler/Varianz der Stichprobenkennwertewerteilung
    - bei Normalverteilung in Population: Mittelwert und Median beide erwartungstreue und konsistente Schätzer für μ, Varianz der Stichprobenkennwerteverteilung des Medians ist aber um Faktor 1.56 (>1) größer als die des Mittelwertes 🡪 Mittelwert ist effizienterer Schätzer
  + *Suffizienz:* Verwendung aller in der Stichprobe vorliegenden Informationen bzgl. des Parameters
    - für μ, *s2* für σ2 (Einbezug aller n Werte in ihren Ausprägungen in die Berechnung)
    - Median nicht suffizient
  + P: Was ist mit Robustheit? Median wird immer als *robuster* als Mittelwert bezeichnet, steht aber in allen Gütekriterien dem Mittelwert nach (auch Effizienz/Präzision der Schätzung)

**Intervallschätzung:**

* Schätzung/Angabe eines Bereichs (Konfidenz-/**Vertrauensintervall**) **um** Kennwert bzw. (Punkt-) **Schätzer** aus den Stichprobendaten, in dem der entsprechende Populationsparameter mit bestimmter Sicherheit/Wahrscheinlichkeit(=Konfidenz) liegt
* Merkmal X ist normalverteilt mit μ und σ2, dann gilt:
  + Stichprobenkennwerteverteilung normalverteilt mit E[X] = μ, Standardfehler
  + Konfidenzintervall für μ mit Irrtumswahrscheinlichkeit α:
    - Wird breiter (= Vorhersage wird unpräziser), je kleiner α und n, bzw. je größer σ (Streuung von X) ist
* Merkmal X ist normalverteilt, aber σ unbekannt, dann:
  + Bestimmung des Konfidenzintervalls für μ mit *tn-1; 1-*α*/2* als t-Wert einer t-Verteilung mit *df = n-1* Freiheitsgraden und *1-*α*/2* = durch ihn bestimmter Abschnitt der Verteilung:



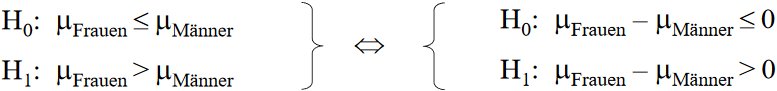
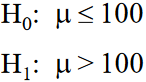
* Verteilung von Merkmal X unbekannt, bei n>30 mit Normalverteilungsannahme (σ unbekannt) nach zentralem Grenzwertsatz:
  + Bestimmung des Konfidenzintervall für μ mit:
  + Konfidenzintervall für Anteil **π** mit :
    - Breite abhängig von Kennwert ( ), größer, je größer p\*(1-p), maximal breit bei p=0.5 (Zufall)
* Bestimmung des Stichprobenumfangs n zur Erreichung einer Schätzung mit gegebener Präzision/Irrtumswahrscheinlichkeit α:
  + kleinster Stichprobenumfang n, sodass die Breite des Konfidenzintervalls für Mittelwert bei Schätzung von σ durch s maximal Δ ist ⬄
  + Analog bei Schätzung von πdurch Anteil *p*:

|  |  |
| --- | --- |
| **Hypothesentests** |  |

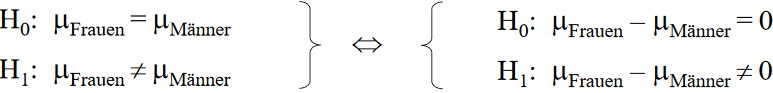
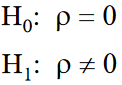
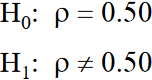
**Forschungshypothesen:**

* Allgemeingültige, theoretisch begründete, empirisch überprüfbare und prinzipiell widerlegbare (falsifizierbare) Behauptungen über Unterschiede oder Zusammenhänge von Merkmalen bzw. Merkmalsausprägungen in einer Population
* Gerichtet(=g) oder ungerichtet(=ung) mit Bezug auf Zsmhänge/Unterschiede in Population:
  + Unterschiedshypothese:
    - g: Frauen sind kreativer als Männer. (1)
    - ung: Anteil an Studierenden, die Studium in RS-Zeit absolvieren ist verschieden in Fächern BWL und Biologie.
  + Zusammenhangs-/korrelative Hypothese:
    - g: Je mehr Personen … tun, desto <aggressiver> sind sie.
    - ung: Es besteht ein Zusammenhang zwischen beruflicher Leistung und der Arbeitszufriedenheit. (2)
* Können auf Basis von Ergebnissen einer Stichprobenziehung nicht bewiesen oder widerlegt, sondern nur (mit gewisser (Un-)Sicherheit) beibehalten oder (zugunsten einer Alternativhypothese) abgelehnt werden (Wahrscheinlichkeitsaussagen möglich).

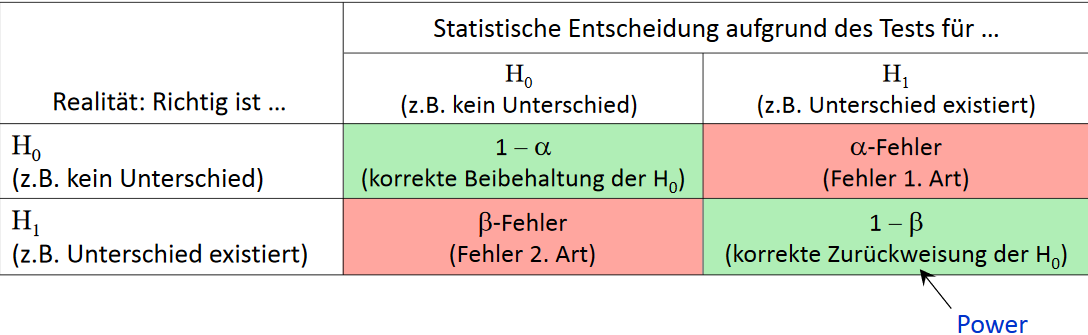
Prinzip/**Vorgehen bei Inferenzstatistik(„null hypothesis significance testing“, NHST):**

1. Überführung der Forschungshypothese in prüfbare **statistische Hypothesen**:
   * **Nullhypothese H0**: typischerweise zur Forschungshypothese gegenteilige Behauptung
   * **Alternativhypothese H1**: Gegenteil zu H0 (oft die formalisierte Forschungshypothese)
   * Bsp. einseitige Formulierung zu gerichteter Unterschiedshypothese(1):

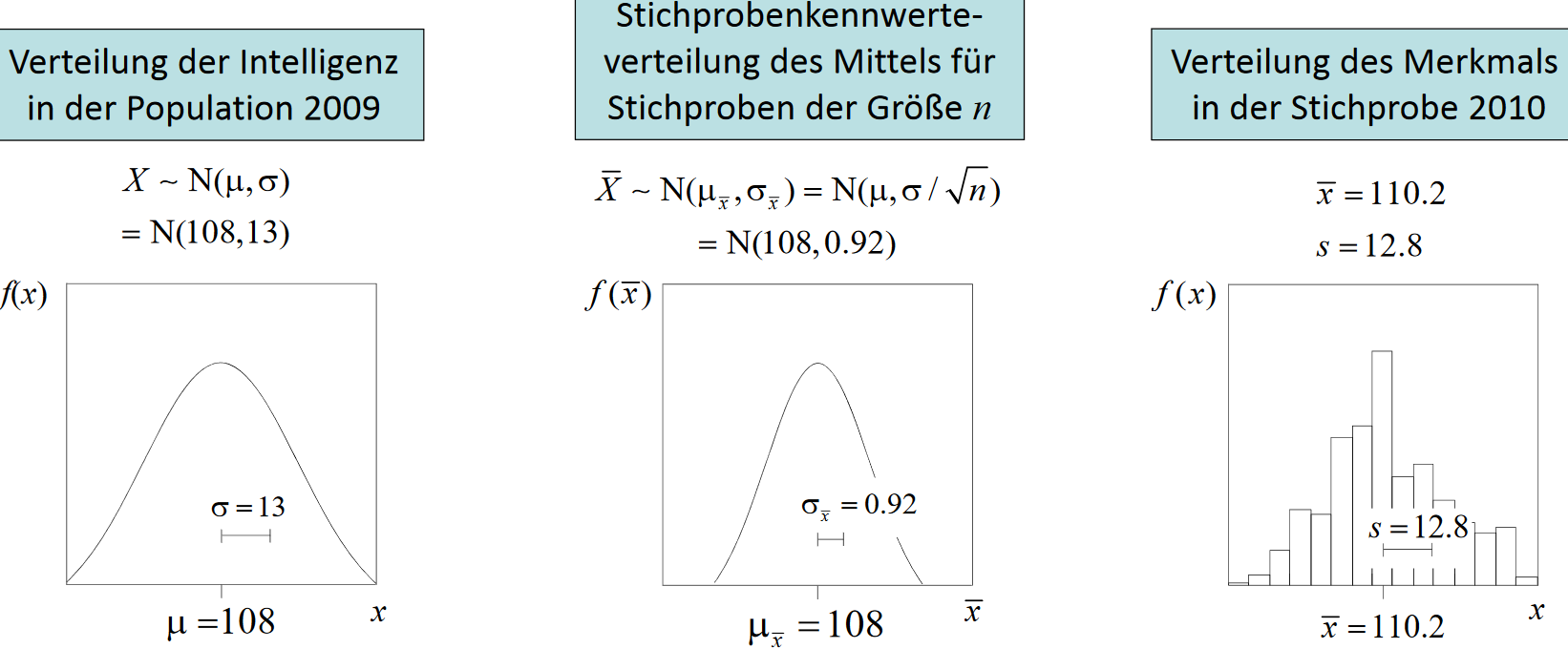
oder anderes Bsp. mit μ ≠0:

* + Bsp. zweiseitige Formulierung zu ungerichteter Unterschiedshypothese (1)(„Frauen und Männer unterscheiden sich in ihrer Kreativität.“):
  + Bsp. zweiseitige Formulierung zu ungerichteter korrelativer Hypothese: oder
  + **Festlegung des Signifikanzniveaus α** für statistischen Test

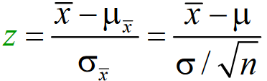
1. **Datenerhebung** Ziehung einer (Zufalls‐) Stichprobe aus der Population
2. **Statistische Testung** (Auswahl 🡪 Prüfung von Voraussetzungen 🡪 Durchführung)
3. Entscheidung über **Zurückweisung** von H0 (signifikantes Ergebnis🡪Alternative **H1 annehmen**) **oder Beibehaltung von H0** (nicht signifikantes Ergebnis🡪Alternative **H1 verwerfen**)

🡪mögliche Entscheidungen/Versuchsausgänge (mit α = P(Irrtum), 1-β = Power):

Wiederholung zu Verteilungen am Beispiel des Merkmals Intelligenz:



******Logik des Hypothesentestens**: Je weiter der ermittelte Stichprobenmittelwert von als theoretisch wahrscheinlichstem/zu erwartendem Mittelwert einer Stichprobe aus der Population abweicht, desto unwahrscheinlicher ist es, dass das gemessene Merkmal in der Population tatsächlich mit dem in der H0 behaupteten Mittelwert μ verteilt ist. Mit α wird festgelegt, wie unwahrscheinlich es mindestens sein soll, dass zu der aus der H0 gefolgerten/errechneten hypothetischen Stichprobenkennwerteverteilung mit μ= stammt. Ist der Mittelwert in der Stichprobe unwahrscheinlicher als durch α (oft 5%\* oder 1%\*\*). gefordert, kann die H0 zurückgewiesen werden, d.h. man geht davon aus, dass die Alternativhypothese stimmt (mit dem Restrisiko α, falls wir blöderweise eine wenig repräsentative Stichprobe gezogen haben).

**Beibehaltung von H0** (Prüfung mittels Konfidenzintervallgrenzen bzw. wenn |z| <= zcrit ):

1. Berechnung von z-Wert:
2. Test: (Entscheidung für ein- oder zweiseitige Testung vor Dateneinsicht sonst wird α höher!)

|  |  |  |  |
| --- | --- | --- | --- |
| **zcrit bei** | **Einseitige Testung** | **Zweiseitige Testung** | |
| **Möglich, wenn** | **H0 gerichtet** | **H0 gerichtet, H0 ungerichtet** | |
| **Prüfung von |z|<=** | z1-α | z1-α/2 | |
| **α = 0.01** | z0.99 = | z0.995 = 2.58 | |
| **α = 0.05** | z0.95 = 1.65 | z0.975 = 1.96 | |
| Annahmebereich | wie zweiseitig bloß nur + *oder* - | |  |
| **Konfidenzintervall** (H0 zurückweisen, wenn behaupteter Wert nicht enthalten) | wenn in H0 ‘<=’ 🡪 [ , inf[  wenn in H0 ‘>=’ 🡪 ]-inf, ] | |  |

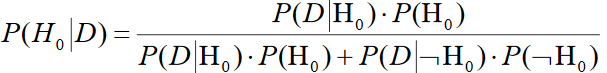
* + Argumente für/gegen zweiseitige Testung gerichteter Hypothesen:
    - + liegt Effekt in falscher Richtung, so muss H\_0 beibehalten werden (nur ungerichtete Schlüsse bei nicht signifikantem Ergebnis)
    - + Test spiegelt wider, dass theoretisch wenig Unterschied besteht, ob Effekt 0 ist oder in Gegenrichtung der Hypothese
    - + Richtung des Effektes schwer zu begründen 🡪Möglichkeiten offenhalten
    - - wenn Richtung begründet abgeleitet werden kann, dann sollte man diese Informationen auch nutzen
    - - Power bei einseitiger Testung höher (bei Effekt in vermuteter Richtung)

**p-Wert** akaP(Daten⏐H0)**:**

- Wahrscheinlichkeit, dass ein mindestens genauso unwahrscheinliches Ergebnis auftritt, wie das gemessene, wenn H0 wahr ist (Bestimmung anhand der Daten vs α als apriori gesetzte Wahrscheinlichkeit, die bestimmt, bis zu welchen p-Wert man Ergebnisse als signifikant ansieht)

- d.h. signifikantes Ergebnis, wenn p<α bei einseitiger, bzw. p<α/2 bei zweiseitiger Testung

- mathematisch formuliert: bedingte Wahrscheinlichkeit des Auftretens der Daten D (oder noch extremerer Daten), wenn die H0 richtig ist, also p = P(D⏐H0)

- zur Umrechnung von p (mittels folgender Formel) in P(H0|D) fehlt allerdings die noch P(H0):

**α - β Zusammenhang**:

* α (Irrtüml. Zurückweisung) klein ⬄ β (Irrtüml. Beibehaltung) hoch ⬄ 1-β (=Power) klein
* α hoch ⬄ β klein ⬄ 1-β hoch
* wenn H0  die Forschungshypothese beinhaltet: Wahl eines größeren α (10-20%) (🡪 geringe Chance (=β klein), irrtümlich die H0 (Forschungshypothese) beizubehalten)

**Prüfung von Unterschieden in der zentralen Tendenz** in einer und/oder zwei Gruppen:

1. **Ein-Stichproben Tests** (z-Test, t-Test & co.) für Prüfung von Unterschiedshypothesen bzgl. Stichprobenmittel/andere Statistik und fixem Wert

|  |  |  |  |
| --- | --- | --- | --- |
| **Bezeichnung** | **Einsatzbereich** (Untersuchtes Maß der zentralen Tendenz, Hypothesen) | **Voraussetzungen** | **Vorgehen und Entscheidung** |
| **z-Test, ein-Stichproben Gauß-Test** | Hypothese, dass **Mittelwert** in Variable X bestimmten Wert aufweist:  zweiseitig:H0: μ=μ0, H1: μ≠μ0 oder entsprechend einseitig | X **metrisch und normalverteilt** in Population, **σ** **bekannt** | Bestimmung der Prüfgröße:    Zurückweisung von H0, falls  |z| > zcrit (Bestimmung s.Tabelle oben) |
| **t-Test für eine Stichprobe** | Mittelwertshypothesen, siehe z-Test | X **metrisch und normalverteilt** in Population | Bestimmung der Prüfgröße:  Zurückweisung von H0, falls |t|>tcrit mit tcrit = tn-1;1-α(/2 🡪wenn zweiseitig) |
| **Wilcoxon-Vorzeichen-Rangtest** | Median | X metrisch und symmetrisch verteilt |  |
| **Vorzeichentest** | Median | X ordinalskaliert |  |

1. **Parametrische Tests für Mittelwertsunterschiede zweier Gruppen** in X aka Zwei-Stichproben-Tests (t-Test für unabhängige und abhängige Gruppen, Welch Test)
   * Gruppeneinteilung durch unterschiedliche Versuchsbedingungen (Experimental-/Kontrollgruppe), Merkmale der Personen (Geschlecht) oder unterschiedliche Messzeitpunkte (vor/nach Therapie/Behandlung)
   * Unabhängige Gruppen durch:
     + 1. Voneinander unabhängige Ziehungen von Zufallsstichproben aus Teilpopulationen
       2. Ziehung einer Zufallsstichprobe mit randomisierter Zuweisung von VP auf Versuchsbedingungen (z.B. Instruktionen; Personenauswahl in einer Gruppe ohne Auswirkungen auf Personen in anderer Gruppe)
   * Abhängige Gruppen durch:
     + 1. Mehrfache Messungen in einer Stichprobe (repeated measures)

+ erhöhte interne Validität, da interindividuelle Unterschiede nicht mit der UV konfundiert sein können

+ erhöhte Power, da interindividuelle Unterschiede bei Analyse als Fehlerquelle

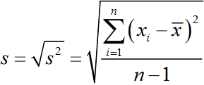
+ niedriger Vpn Bedarf

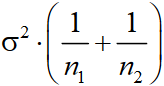
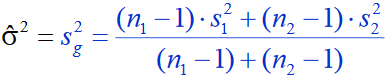
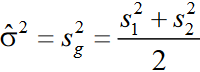
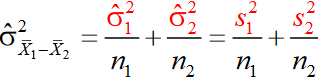
- Übungs-/Ermüdungseffekte, sodass Treatments nicht mehr unter gleichen Bedingungen untersucht werden

- drop-out möglich (Vpn zu späterem Zeitpunkt nicht mehr verfügbar)

* + - 1. Matching (Parallelisierung) der VP anhand min. einer Drittvariablen
      2. Verteilung natürlicher VP-Paare zufällig (z.B. Zwillinge) oder systematisch (z.B. Ehepartner) auf Gruppen

|  |  |  |  |
| --- | --- | --- | --- |
| **Bezeichnung** | **Einsatzbereich** (Untersuchtes Maß der zentralen Tendenz, Hypothesen) | **Voraussetzungen** | **Vorgehen und Entscheidung** |
| **t-Test - unabhängige Gruppen** | Mittelwertunterschiede in Merkmal X in 2 unabhängigen Gruppen;  zweiseitig:  H0: μ1=μ2, H1: μ1≠μ2 oder entsprechend einseitig | X **metrisch und normalverteilt** mit gleicher Varianz (**Homoskedastizität**) in beiden Teilpopulationen (Gruppen) | H0⬄μ1-μ2 = 0 ⬄ ~ 0  Bestimmung der Prüfgröße:    mit *df = n1 + n2-2*  Zurückweisung von H0, falls |t|>tcrit mit tcrit = t(n1+n2)-2;1-α(/2 🡪wenn zweiseitig) |
| **Welch-Test für unabhängige Gruppen** | Wie t-test oben mit Welch-Korrektur | X **metrisch und normalverteilt** in beiden Teilpopulationen (Gruppen) | wie t-Test oben mit  *df*=  (df zugunsten von H0  konservativ/klein wählen) |
| **z-Test für unabhängige Gruppen/Zwei-Stichproben Gauß-Test** | Mittelwerte | X metrisch, in beiden Gruppen normalverteilt und bekannte Varianzen σ2 |  |
| **t-Test für abhängige Gruppen** | Mittelwert, wie t-Test für unabhängige Gruppen | Normalverteilte Differenzwerte X1-X2 | Berechnung von Mittelwert und Standardabw. der (n) Differenzen (🡪d) von (n) Messwertpaaren zur Bestimmung der Prüfgröße (🡪t):  = und *sd*🡪  mit *df*=*n*-1  wie ein-Stichproben t-Test für Differenzwerte di mit 1<=i<=n |

**Standardfehler** der Stichprobenkennwerteverteilung der **Mittelwertsdifferenzen zweier Gruppen in Merkmal X (= Varianzfehler = quadrierter Standardfehler)**:

* X normalverteilt in beiden Gruppen:
  + Mittelwerte in Gruppen unabhängig: mit σ1 = σ2 = σ
    -  = ⬄ wenn n1 = n2 : sonst
  + Mittewerte in Gruppen unabhängig, Varianzen verschieden (ohne Homoskedastizität)

1. **Prüfung von Voraussetzungen** (Robustheit, Varianzhomogenität, Normalverteilung)

* Verletzung der **Annahme einer einfachen Zufallsstichprobe**: (Häufige) Verwendung von Gelegenheits- (anfallende, ad-hoc-) Stichproben oder „non-response“
  + systematische Verzerrungen (bias) möglich
* Verletzung von **Verteilungsvoraussetzungen**: z.B. linkssteile Reaktionszeitenverteilung bei Normalverteilungsannahme, Messfehler und resultierende Ausreißer
  + Prüfung mittels statistischer **Tests** (wobei widerum **α oder β‐Fehler** möglich), die unterschiedlich empfindlich bzw. robust gegenüber Voraussetzugsverletzungen sind (konkrete Tests dargestellt in nachfolgender Tabelle)
    - *Empfindliche* Tests: große **Veränderung** der **Stichprobenkennwerteverteilung** und deren **Parametern** bei Verletzungen; **faktisches und nominelles** **α** verschieden 🡪 faktische Wahrscheinlichkeit für **α-**Fehler höher als gesetztes (=nominelles) **α** -Niveau (= inflationiertes Risiko für **α-Fehler** 🡪 *progressive Entscheidungen zugunsten der* *H1*); Veränderung der Power der Tests (größer/kleiner)
    - *Konservative Entscheidungen zugunsten der H0*, wenn faktisches **α kleiner als nominell**
    - Prüfung der Robustheit z.B. mittels Computersimulationen
    - **Ein-Stichproben t-Test und t-Test für abhängige Gruppen**:
      * Verletzung der Normalverteilungsannahme: progressive Entscheidungen, umso mehr, je stärker die Abweichungen; Asymmetrie problematisch
      * Robustheit steigt mit wachsendem n (🡪Normalverteilung)
      * Robustheit höher bei zweiseitiger Prüfung mit α=5% als mit kleinerem α
    - **t-Test für unabhängige Gruppen:**
      * Robust gegenüber Verletzung der Homoskedastizität bei n1=n2 und normalverteilten Populationen
      * Empfindlich gegenüber Verletzung der Homoskedastizität bei n1≠n2
      * Robust gegenüber Verletzung der Normalverteilungsannahme bei n1=n2 (mäßige Asymmetrie / symmetrische nicht-Normalverteilungen unproblematisch)
      * Teilweise extrem empfindlich bei Verletzung der Homoskedastizität und Normalverteilungsannahme, insb. bei kleinem n (Stichprobenumfang)

**Tests zur Prüfung von Voraussetzungen:** Wunschhypothese oft H0 🡪 üblicherweise Wahl von α mit 10-20% zur Verringerung von β

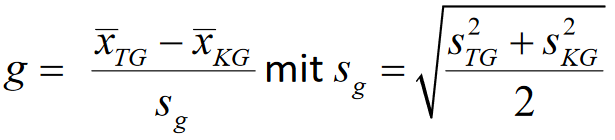
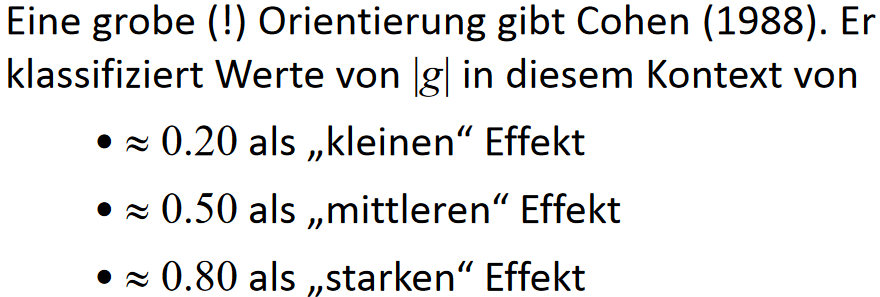
|  |  |  |
| --- | --- | --- |
| **Bezeichnung** | **Voraussetzungen u. Eigenschaften** | **Vorgehen/Entscheidung** |
| **Homoskedastizitätsprüfung:** auchTreatments können Varianz verändern, twl. auch zur Prüfung von Forschungshypothesen geeignet (z.B. Verfahren zur Testung zweier abhängiger Gruppen)  Test auf (Un-)**Gleichheit der Varianzen** zweier Gruppen in der Population: | | |
| **Levene-Test** | Normalverteilung in beiden Gruppen, robuster ggü. Verletzung dieser Annahme als F-Test (nicht vorgestellt),  auch verallgemeinerbar von j=2 auf 2<=j<=g Gruppen | Bestimmung abs. Abw. aller Gruppenwerte vom jew. Gruppenmittel, Mittelwerte davon in bzw. über Gruppe(n)  Prüfgröße:  F-verteilt mit  df1=1 u.  df2=n1+n2-2  Fcrit = F1;(n\_1)+(n\_2)-2;1-α  Zurückweisung von H0, falls W>Fcrit |
| **Brown-Forsythe-Test** | Wie Levene-Test, bloß  robuster (zunehmend mit n) als Levene-Test bei schiefen Verteilungen und Verteilungen mit Exzess ≠ 0, | wie Levene-Test, bloß Bestimmung der Abw. vom *Median* statt Mittelwert (sonst gleich) |
| **Anpassungstests zur Prüfung auf Normalverteilung** (oder andere Verteilungsform; goodness of fit tests):  H0: X ist in Population verteilt nach bestimmter Funktion Φ0(X) (ohne Abweichungen in einem Punkt) | | |
| **Kolmogorov-Smirnov (K-S) Test** | X stetig verteilt mit bekanntem/r Mittelwert μ und Standardabweichung σ (wenn beides unbekannt 🡪 zu konservative Entscheidungen zugunsten von H0) | Messwerte in X aufsteigend sortieren🡪für jeden Wert Anteil der Werte bestimmen, die kleiner sind (F(x(i)) = (i)/ n), mit F(x(n)) = 1) und entsprechenden Wert der angenommenen Verteilungsfkt. (z-Transf. des Messwertes, p für z-Wert nachschlagen)  🡪davon die max. Abweichung an oberer/unterer Grenze jedes Intervalls als Prüfgröße  0  Prüfgröße: |)  Zurückweisung von H0, wenn Dmax > Dcrit |
| **Lilliefors-Test** | X stetig verteilt, ist Korrektur des K-S Tests, wenn μ und σunbekannt | Wie K-S Test, bloß mit Punktschätzern aus Stichprobe für entsprechende Verteilungsparameter (bei NV also: Φ0(X) = NV(x#quer,s) statt NV (mue,sigma) |

1. **Nicht-parametrische (verteilungsfreie)** **Tests für Unterschiede zweier Gruppen** in der zentralen Tendenz (Mann-Whitney U-Test, Vorzeichen-Test, Wilcoxon-Test): Nutzung, falls Voraussetzungen nicht erfüllt bzw. auch möglich bei ordinalskalierten Variablen

|  |  |  |  |
| --- | --- | --- | --- |
| **Bezeichnung** | **Einsatzbereich/Hypothesen** | **Voraussetzungen** | **Vorgehen/Entscheidung** |
| **(Mann-Whitney) U-Test** | Ergebnis wie Wilcoxon-Rangsummen-Test, Test auf Verschiedenheit von Verteilungen/zentralen Tendenzen (Mediane) in X in 2 Gruppen   1. 2) 2. .5   Problem: bei Ties können mittlere Rangplätze vergeben werden 🡪 zu konservative Entscheidungen (zugunsten von H0)   * Korrekturformel:     Mit ti = Zahl Personen, die sich Rangplatz i teilen,  k = Zahl der Tie-Blöcke | 2 unab-hängige Gruppen, X mindes-tens ordinal-skaliert | Messwerte der Gruppen in gemeinsame Rangordnung bringen (mit n1+n2 Rängen insg.), Berechnung der Summe der Ränge der Messwerte jeder Gruppe, bestimmen:  U2  U1+U2=n1·\* n2(ohne Ties)  Je ungleicher die Anteile, desto verschiedener die Gruppen  Prüfgröße: **U = Min(U1,U2)**  Ucrit bei kleinen n aus Tabellen entnehmen,  **Zurückweisung von H0, wenn U < Ucrit**  sonst annähernd normalverteilt:      Zurückweisung falls |z|>zcrit zweiseitig: zcrit = z1-α/2  Einseitig: zuerst prüfen, ob Effekt in gew. Richtung geht (🡪Rangplatzmittelwerte), dann U-Wert nachschlagen bzw. zcrit= z1-α |
| **Vorzeichen-Test** | Hypothese, dass sich die Verteilungen/zentralen Tendenzen (Mediane) der Variable X in 2 Gruppen unterscheiden   * H0: Φ(X1) = Φ(X2) und H1: Φ(X1)≠Φ(X2) * H0: π(X1<X2) = 0.5 und H1: π(X1<X2)≠0.5 und entsprechend einseitig * H0: ηD=0 und H1: ηD≠0 (ηD = Median der Differenzen der Messwerte) und entsprechend einseitig * Bei symmetrischen Verteilungen:   H0: η1=η2 und H1: η1≠η2 und entsprechend einseitig  **Stetigkeitskorrektur** für z bei Approximation der Binomialverteilung durch N:  🡪 konservativere  Entscheidungen  Ties (Null-Differenzen di=0) 🡪 oft ignoriert | 2 abhängige Gruppen, X mindes--tens ordinal-skaliert | V- = Anzahl negative di = xi1 – xi2  V+ = Anzahl positive di = xi1 – xi2  Mit n = V- + V+, größere Abw. von V-Werten 🡪 Gruppenunterschiede  V ~ B(n,p=0.5) liefert Wahrscheinlichkeit, dass 0 <= k <= Min(V+,V-) mal die minimale (positive oder negative) Differenz auftritt (Summe über P(X=k) bei B(n, p=0.5))  Bei zweiseitiger Testung u. α= 0.05 müsste diese Wahrscheinlichkeit < 0.025 sein  Um H0 abzulehnen  Mit Normalverteilungsapprox.:    Zurückweisung von H0, falls |z|>zcrit |
| **Wilcoxon (Vorzeichen Rang-) Test** | Hypothesen wie Vorzeichentest  Korrektur dieser bei Ties in Rangsummen  ti = Zahl der di, die sich Rangplatz i teilen, k = Zahl der Tie-Blöcke    Bzgl. **Ties** (di=0) wie oben oft ignoriert 🡪 **progressive** Entscheidungen  Oft höhere Power als Vorzeichen-Test | X ordinal-skaliert, di Messwert-differenzen ordinal-skaliert (z.B. durch Gleichabstän-digkeit der Messwerte) | Rangreihe Rg von |di=xi1–xi2| bilden und getrennt für pos/neg di die Rangplätze summieren:  W=Min(W+, W-)  H0 zurückweisen, wenn W<Wcrit  Bzw. |z| > zcrit  Mit Normalverteilungsapproximation: |

* Vorzeichentest, wenn nur Informationen über Richtung des Unterschieds existieren

**Kritik am NHST:**

* Viel wichtiger als die Signifikanz eines Ergebnisses ist, wie stark der Effekt ist: Bei sehr großem n können auch triviale, minimale Effekte signifikant werden.
  + Berechtigt, Signifikanztest prüft nur, ob sich die Mittelwerte unterscheiden. Daher ist es sinnvoll, zusätzlich zur statistischen Signifikanz Effektstärkemaße zu bestimmen. Anwendung für Mittelwertsunterschiede zwischen zwei unabhängigen Gruppen:
    - Hedges g:
* Voraussetzungen häufig nicht geprüft/erfüllt
  + Unbestritten, aber
    - parametrische Tests robust unter bestimmten Bedingungen
    - voraussetzungsarme oder -lose(nonparametrische) Verfahren vorhanden
    - Strategien im Falle von Voraussetzungsverletzungen:
      * Transformation der Daten (z.B. Normalisierung linkssteiler Verteilung durch log-Transformation)
      * Berechnung robuster Statistiken bei Ausreißern, mit denen dann getestet wird (z.B. getrimmte/winsorisierte Mittelwerte)
      * Bootstrapping (Computersimulation) zur Erzeugung der Stichprobenverteilung statt theoretischer Ableitung
* In den Nullhypothesen wird oft vollständiges Fehlen irgendeines Effektes postuliert. Dies ist unrealistisch, da kleine Effekte immer zu erwarten sind.
  + Möglich, einen von 0 abweichenden Effekt zu postulieren, z.B. bei der Produkt-Moment-Korrelation (H0: ρ= .30 statt H0: ρ= 0).
* Viele Studien haben zu geringe Stichprobengrößen, es wird zu häufig der β‐Fehler gemacht und die Power ist zu gering.
  + Power-Analyse im Vorfeld erlaubt Abschätzung erforderlicher Stichprobengröße zur Bestimmung eines Effekts mit angenommener Stärke bei gewünschter Power
* Zufallsstichproben i.d.R. nicht gewährleistet
* Publikationsverzerrung: überproportionale Repräsentation von Studien mit statistisch signifikanten Befunden 🡪 Verzerrung, Effekte werden überschätzt, Rate der Fehler 1.Art ist höher als durch alpha suggeriert (faktisches und nominelles alpha auseinanderklaffend)
* NHST als Kombi von zwei verschiedenen twl. unvereinbaren statistischen Vorgehensweisen abgelehnt
* Nicht nur Prüfung von Theorien (für/gegen Theorie Entschiedungen), sondern auch Replikationen sollten mehr im Fokus stehen